ISSN 2752-7417

UNDERGRADUATE LINGUISTICS
ASSOCIATION OF BRITAIN

PROCEEDINGS OF ULAB 2021

UNIVERSITY OF ABERDEEN
16-18 APRIL 2021

EDITED BY
LYDIA WIERNIK

Undergraduate Linguistics Association of Britain

ULAB XI Proceedings

University of Aberdeen

16-18 April 2021

Editor: Lydia Wiernik
Copyeditors: Rachel Shannon, Essi Harbord, Zsófia Varga, Eleonora Kacl
Cover design: Lydia Wiernik

ISSN 2752-7417

Contents

National Committee 2020-21

Chair	Cliodhna Hughes, University of Edinburgh
Secretary	Amira Ramdani, University of Oxford
Treasurer	Beatrix Livesey-Stephens, University of Aberdeen
Webmaster	Louis Van Steene, University of Cambridge
Social Media Coordinator	Anne Eschenbrücher, University of Wolverhampton
Archivist	Tom Williamson, University of Cambridge
Institutional Representative Coordinator	Riley Crouch, University of Edinburgh
Institutional Representatives	Michael Gössler, University of Edinburgh Núria Bosch Masip, University of Cambridge Amira Ramdani, University of Oxford Anastasia Putilova, University College London Ondřej Boček, Lancaster University Julieta Gil Loizaga, Queen Mary University of London Arabella Graves, University of York Hafren Vaughan, University of Aberdeen Laura Mills, University of Huddersfield Grace Cotton, Ulster University Emily Copeland, University of Kent Laura Collins, Anglia Ruskin University Natalie Lawrence, Stellenbosch University Leo Adams, University of Leeds María Teresa Borneo, National University of Córdoba Jaidan McLean, University of Oregon Justin Malčić, University of Maryland Brandon Papineau, Stanford University

Local Committee 2020-21

Chair	Beatrix Livesey-Stephens, University of Aberdeen
Secretary	Emily Lea Heggernes, Univeristy of Aberdeen
Treasurer	Tina Wolff, University of Aberdeen
Social Media Coordinator	Sharon Methven, University of Aberdeen
Tech Coordinator	Emma Johnson, University of Aberdeen
EasyChair Manager	Hafren Vaughan, University of Aberdeen
Liaison Officer	Stephanie McWilliams, University of Aberdeen
Ordinary Member	Emme Strickland, University of Aberdeen

With special thanks to BSL interpreters Lesley Crerar and Margaret Kinsman.

Foreword from the National Chair, 2020-21

Cliodhna Hughes, University of Edinburgh

Despite everything, 2020-2021 has been a hugely successful period for ULAB: we set up a linguistics magazine, began organising regular online academic and social events for our members, founded an academic journal for undergraduate linguists to publish their research, and found and advertised linguistics-related opportunities for students to undertake in their spare time. We also worked on expanding the reach of the Association, recruiting 18 Institutional Representatives across several different countries, and we have been working on international collaborations with other national student organisations including Junge Sprachwissenschaft in Germany, and FNAEL in France.

On top of all of this, we organised our first ever online conference, which was a great success! Thanks to our Archivist, Lydia, you'll be able to read here about some of the work presented there, and the slides from some of the presentations will also be archived on our website. There were some truly amazing presentations this year and we hope that others will enjoy reading about the work that undergraduate linguists carry out.

I am incredibly proud of everyone on the 2020-2021 Committee, and thankful for all the time and energy they put into the Association. I have the pleasure of continuing to work with ULAB for the 2021-2022 period, and I look forward to seeing what we can accomplish in the next academic year. Going forward, we hope to keep the momentum going, and we remain committed to our goals of providing opportunities for undergraduates to showcase their research in linguistics, providing information and support for those looking to pursue postgraduate studies or careers in linguistics, and facilitating connections between linguistics students across the UK and beyond.

Foreword from the Local Chair, 2020-21

Beatrix Livesey-Stephens, University of Aberdeen

ULAB 2021 was our first ever virtual conference, an exciting (and mandatory) new initiative. The virtual medium actually ended up hugely contributing to the success of the conference. We were able to invite plenary speakers, panel speakers, and attendees from as far as the United States and Japan, including many universities that had never been present at ULAB before.

Our new Discord server was put to great use as we filled it with conference-specific channels for ULABers to talk between the presentations over MS Teams. We were even able to invite secondary schools this year, which is something we hope to expand on in the future. We also made leaps and bounds regarding the accessibility of ULAB this year. Attendees were able to attend ULAB from the comfort of their own home, and we scheduled access breaks between each talk.

We were also successful in applying for funding from the University of Aberdeen Development Trust in order to have BSL interpretation at the conference. I am so proud of the atmosphere we fostered over MS Teams, Discord, and the #ULAB2021 Twitter hashtag. ULAB 2022 will no doubt be a triumphant return to the in-person conference in Edinburgh.

Proceedings of ULAB XI: Introduction

Lydia Wiernik, University of Edinburgh

Despite the turbulence of a pandemic-ridden, online academic year, the ULAB 2021 conference proceeded flawlessly. Undergraduate linguists spanning universities across the globe came together to share their research and connect with one another. It is my honor to showcase the incredible work of these undergraduates and welcome you to the 2021 Proceedings of ULAB.

This year's Proceedings are critical, not just to highlight undergraduate research but as a testament to the might of the previous committee, who were the sole reason the ULAB conference was able to carry on this spring after the pandemic halted it in 2020.

ULAB 2021 was hosted on Microsoft Teams. Though wholly digital, the conference had a unique humanity. Spotty Internet connections, audio breaks, and other difficulties arose from time to time, but these circumstances fostered compassion in the commonality of our situations despite our physical locations.

Concessions had to be made in order to accommodate an online platform; not all attendees could enjoy each presentation as there were simultaneous presentations occuring in different "rooms." However, a digital conference proved to promote and cultivate accessibility. The previous committee capitalized on Teams' features to make the conference the best it could be. Teams' automatic closed captioning feature paired with committee-provided BSL interpreters allowed for minimized hindrances to enjoying and engaging with the presentations.

The 2021 conference was phenomenal. I am absolutely delighted to record its success as the new ULAB Archivist. These Proceedings have been the first of many items on my agenda during my tenure as Archivist. My goal as ULAB Archivist is to collect as much ULAB materials and data as possible, going back to the first conference in 2011. I strive to paint a broader picture of ULAB as an organization and to keep a record of those who have participated in the conference and helped it grow. These materials include but are not limited to presentation slides, data on past committee members, conference programs, past constitutions, and past issues of ULAB literature U-Lingua and JoULAB. Nothing is ancillary in an archive.

The Proceedings are their own archive: this years' Proceedings feature the works of 15 students who presented at the conference. Many papers are theses or dissertations, though equally many are not. At ULAB we prioritize accessibility and inclusivity, and the Proceedings are certainly not exempt! Following the guidelines set by Jamie Bailey, ULAB's first Archivist, papers are eligible for publishing in the Proceedings if they have received a Class I or II.1 mark from their university. Papers are then categorized as Section A, full length, or Section B, shorter write-ups and squibs, therefore opening publishing opportunities to all students, regardless of paper length or year of study.

The 2021 Proceedings also mark an exciting advance in ULAB's prowess - we have been granted an ISSN by the British Library, uniquely identifying the Proceedings as an official recorded publication.

Compiling the Proceedings has been a long process, and not without help from many. I'd like to thank Rachel Shannon, Essi Harbord, Zsófia Varga, and Eleonora Kacl for copyediting the Proceedings, Tom Williamson for making my transition to the role of Archivist so smooth, and the entire 2021 ULAB Committee, without whom such a fantastic conference in the most harrowing of circumstances could not have happened. But my greatest thanks go to the students who presented their research at ULAB 2021 and contributed it to the proceedings. I am honored to publish your insightful, innovative work.

Please enjoy the 2021 Proceedings of ULAB.

SECTION A

Assessed Contributions

The following contributions have been assessed by academic institutions and awarded a Class I, Class II.1, or international equivalent.

Factors Affecting Chinese-L2 Learners' Use of Classifiers

Jiahuan Zhang
University of Cambridge

Abstract. This study explores a number of factors influencing Chinese L2 learners' classifier acquisition: L1, task type, and classifier type. I developed a picture-based description test, including composition, free cloze, and multiple-choice cloze questions to elicit the use of classifiers. Participants were 50 Chinese L2 learners from Arabic, English, and Japanese L1 backgrounds. An analysis of potential predictors suggested that, although L1 is not a significant predictor of test performance, Japanese L1 participants performed numerically better. Certain tasks (composition) are conducive to use of test-taking strategies and reveal a higher classifier accuracy to more constrained tasks (multiple-choice). Type of task was found to interact with L1. Results also showed a universal path of acquisition of classifier types, with shape classifiers receiving higher scores.

Keywords: Chinese classifiers, second language acquisition, L1 transfer, task effect, hierarchical development

---

1 Introduction

A numeral classifier is an obligatory grammatical unit within a numeral noun phrase in classifier languages (Allan, 1977). Although ill-formed grammar may not hinder speech comprehension, semantics becomes vague if the compulsory classifier is eliminated. For example, "one person" is *yi* (one) *ge* (CL) *ren* (people) in Chinese, where classifier *ge* is a compulsory element. The expression of *yi* (one) *ren* (people) is not clear given an absence of the classifier *ge* even with the numeral "one" present. This is because, conventionally, the word "person" can be matched to multiple classifiers/quantifiers such as *yi kou ren* "one person" (only used in census); *yi qun ren* "a bunch of people"; *yi dui ren* "many people"; *yi che ren* "a car of people". Notably, 40% of frequently used nouns can be matched to the generic classifier *ge*, and because of this, young Chinese first language (L1) acquirers and adult Chinese as a second language (L2) learners often overuse *ge* in natural contexts (Erbaugh, 1986; Liang, 2009; Polio, 1994; Zhang, 2007).

The past two decades has witnessed a growing interest in the investigation of Chinese classifier acquisition in the L2, with an increasing amount of research being focused on L1 effects. The research on crosslinguistic influence aims to determine how previous language learning affects the acquisition of an additional L2, which is a useful approach for predicting errors in L2 production (e.g., Gass & Mackey, 2013; Odlin, 1989). The L1 is argued to impact on all linguistic levels, including phonology, lexis, syntax, semantics, and discourse (e.g., Chan & Leung, 2020; Cheong et al., 2019; De Vincenzi & Lombardo, 2000; Jarvis & Odlin, 2000; Jarvis & Pavlenko, 2008; Odlin, 1989; Pienemann, 2005; Quesada & Lozano, 2020). An alternative consideration is that difficulties in L2 learning can be predicted by a hierarchical development of learning whereby L1 effects are minimised (e.g., Krashen, 1983; Pienemann, 1998; 2005). From a longitudinal perspective, learners are predicted to acquire linguistic knowledge involving morphemes and syntactic structures following a universal route (e.g., Dyson, 2009).

A relative lack of research studying L1 effects on L2 acquisition of Chinese classifiers means that the impact of L1 on L2 classifier acquisition remains unclear. Methodologically, existing studies mostly focus on comparisons of a single group of participants whose L1 is a non-classifier language to that of classifier languages, which limits the possibility of comparison between speakers of different non-classifier languages (cf. Liang, 2009; Polio, 1994; Zhang & Lu, 2013). Existing studies indicate quite mixed results, which could be explained by the types of tasks employed; with most using a single task, the effect of task is poorly understood (cf. Polio, 1994; Zhang & Lu, 2013). Therefore, it remains unclear whether speakers of different non-classifier languages would vary in their acquisition of classifiers, and whether using different task types would affect the interpretation of classifier proficiency. Additionally, the potential order of L2 classifier acquisition remains underexplored (cf. Zhang & Lu, 2013).

Therefore, the aim of the present study is to contribute to and expand on existing research on the L1 effects, task, and hierarchical development of acquisition associated with the L2 use of Chinese classifiers. I use a specially designed test with three written tasks completed by three L1 groups to perform a fine-grained analysis of classifier production. As a consequence, I obtain a fuller description and understanding of the use of classifiers than that achieved in previous studies.

2 Literature review

2.1 Classifiers

Classifiers are absent or marginal in the majority of European languages, and these languages are representative non-classifier languages (Allan, 1977). In non-classifier languages such as English, only mass nouns require a unit of quantification, and there are no classifiers as a distinct part of speech (Saalbach & Imai, 2012; Shi, 2014). Instead, one of the articles *as, or the* is utilized ahead of the noun when an individual object is counted. Yet, there is "…an open class of words that are functionally similar to classifiers" (Lehrer, 1986, p. 109). If needed, a unique format "a+N1+of+N2" is used to express collective meanings when an uncountable object is counted (e.g., 'a cup of water').

On the other hand, many Asian languages are typical classifier languages, including Japanese and Chinese (see detailed description of classifier systems in Downing, 2002; Kuo & Sera, 2009; Zhang, 2007). Classifiers often categorize nouns based on their characteristics, with categorization being different across languages. There are over 900 classifiers in Chinese; and a Chinese numeral classifier is structurally obligatory within a noun phrase when the head noun is quantified, as demonstrated in (1) (Zhang, 2007). The sortal Chinese classifier *tiao* (which refers to long, thin, flexible objects) must be inserted to convey the meaning of "two ropes".

(1) liang *tiao* shengzi
two CL rope
'Two ropes'

The Chinese classifier system contains a fine-grained categorization of having shape-, animate-, and inanimate-based types of classifiers (cf. Gao & Malt, 2009). Aside from these concrete uses of classifiers, concept-typed classifiers are equally important in Chinese, which refer to the extended use of classifiers and their metaphorical function in discourse (Aikhenvald, 2003; Littlemore, 2009).

2.2 L1 transfer

Positive transfer occurs when features of the L1 and the L2 match, such that L2 acquisition is facilitated. Negative transfer (also called 'L1 interference'), on the other hand, hinders the L2 acquisition because the L1 and L2 differ in particular respects. Accordingly, classifier language speakers are anticipated to outperform non-classifier language speakers in learning Chinese classifiers. This is because non-classifier language speakers are unfamiliar with the use of classifiers owing to the lack of a corresponding category in their L1, while classifier language speakers are already familiar with the use of classifiers before learning Chinese. Evidence can be found in Liang's (2009) study, which reported that Korean L1 speakers outperformed English L1 speakers in using Chinese classifiers. The prior knowledge of using classifiers facilitates the Korean speakers' classifier learning. Likewise, in my pilot study (Zhang, 2019), the Thai L1 participant showcased a superior mastery of Chinese classifiers compared to the English L1 counterpart in a dynamic assessment writing task, although these two participants were both intermediate learners of Chinese.

On the other hand, L1-L2 similarities can impede L2 learning on the premise that similarities in languages create confusion (Andersen, 1984). Beginning L2 learners often rely on "one meaning-one-form mapping" between L1 and L2, which would easily result in L2 errors (Andersen, 1984). Tang (2005) examined L2 Chinese learners' classifier production by comparing multiple written tasks (exams, homework, and papers) across L1. It was noticed that Korean learners made some common mistakes in using classifiers, such as overuse of *ge* and incorrectly using classifiers with the same pronunciation or similar meaning to Korean classifiers. One potential explanation for this is that there are similarities and differences between the two classifier systems and Korean learners' reliance on their L1 knowledge may result in L2 mistakes where there are differences. In other words, learners are accustomed to mapping the functional or semantic similarities of L1 items onto the L2, particularly at the very early stages, which would often incur unwanted ramifications (cf. Jarvis & Odlin, 2000; Ringbom & Jarvis, 2009).

2.3 Hierarchical development of morphosyntax

Although L1 effects related accounts illustrate the reasons for some common types of errors in L2, they may depict an incomplete picture of how learners acquire an L2 from a longitudinal perspective. In this regard, one robust explanation is that the universal developmental sequence plays an essential role in both L1 and L2 learning, hence L1 transfer interacts with this developmental sequence but does not override it (Ortega, 2014). The natural order hypothesis was one of the first accounts to theorize language developmental patterns. It hypothesizes that L1 is acquired in a predictable order that is not determined by formal simplicity or influenced by instruction (Krashen, 1983). The observation of the development of individual morphemes such as pronouns and syntactic structures such as negation presents strong support for the existence of the developmental sequence (Ellis, 1994). Erbaugh's (1986) is one of the most representative L1 acquisition studies in this strand. It is a longitudinal study that recorded three young Chinese children's acquisition of classifiers in a naturalistic conversational context. Findings indicated that children might follow a trajectory of learning classifiers, especially those classifiers related to shape were earliest mastered and frequently produced from an early age; and the generic classifier *ge* can be a substitute for virtually any other classifier without causing misunderstanding in daily conversations (see also Jiang, 2017). The overgeneralization of *ge* has since been widely observed in the L2 empirical study of classifier

learning (e.g., Hu, 1994; Zhang & Lu, 2013). Ellis (1994) believes that the L1 acquisition sequence in general offers a reference point for predicting L2 acquisition sequences, while an important issue of whether L1 and L2 acquisition are exactly the same or whether there are differences remains.

The universal developmental sequence can also account for certain errors occurred in the process of L2 acquisition (Ellis, 1994). 'Processability theory', proposed and refined by Pienemann (1998), explicates the hierarchical development of morphosyntax. It posits that learners pass through six distinct developmental stages when acquiring English L1 morphosyntax, and that the same is true for both L1 and L2 learning of other languages (Pienemann, 2005). A representative illustration is Hyltenstam's (1977) investigation of L2 acquisition of Swedish negation. Results showed that the acquisition of grammatical structure is a dynamic and productive process, and learners with different language backgrounds follow the same path of acquisition as well as error production. Hence, it is argued that L2 learners potentially conform to the universal route of acquisition, in which L1 background is overshaded.

With regard to classifier acquisition, the 'numeral classifiers accessibility hierarchy' (NCAH) hypothesizes an ordered learning for different types of classifiers (Craig, 1986). That is, classifiers of 'animate human', 'animate non-human', 'shape', and 'function' type are expected to be acquired in succession. The least marked distinction (i.e., animate human) is expected to appear earliest and be retained longest in learners' acquisition, whereas the most marked distinction (i.e., function) is expected to appear last and be easiest to lose after the onset of attrition. Even though the NCAH has been formulated on the basis of a comparison of limited classifier types, a body of empirical studies substantiated that animacy classifiers are acquired first, followed by classifiers denoting inanimacy, and then concrete objects with salient features such as shape in L1 acquisition of Kilivila, Hokkien, and Cantonese (e.g., Hu, 1994; Luke & Harrison, 1986; Senft, 1996). Furthermore, Hansen and Chen (2001, p. 84) noticed that English L1 speakers' acquisition of Japanese and Mandarin Chinese, in particular, conformed to the order predicted by the NCAH. While slightly different to the NCAH order, Aikhenvald (2003) states that shape-based classifiers are acquired relatively early by speakers of Mandarin Chinese, and that classifiers referring to non-extended objects are acquired earlier than classifiers that refer to extended objects (including classifiers referring to concepts). Altogether, these contrasting findings suggest a sequential acquisition of Chinese classifiers.

2.4 Empirical studies on Chinese L2 classifier acquisition

As discussed above, it remains unclear which factor (i.e., L1 or acquisition order) plays a more important role in L2 acquisition. The investigation of Chinese classifiers in L2 may help to clarify this, and it has been of particular interest for language acquisition in general during the past decades. Many empirical studies have examined L2 classifier acquisition and have offered mixed findings as to the effect of L1 (e.g., Liang, 2009; Paul & Gruter, 2016; Polio, 1994; Zhang & Lu, 2013). Polio (1994), for example, found no significant L1 effects. Polio's (1994) study investigated L2 use of Chinese classifiers from 21 English-L1 speakers and 21 Japanese.

L1 speakers. Both groups of participants were intermediate Chinese-L2 learners, required to narrate a story for Chinese native speakers after watching a film. Findings suggested that: 1) no clear L1 effects were found in either group in terms of classifier acquisition; 2) L2 speakers did not avoid using classifiers in most cases; 3) both groups tended to overuse the generic classifier *ge*. This study offered valuable insights into classifier acquisition at a specific stage and underlying difficulties of learning classifiers from a crosslinguistic perspective. Yet, the data come from one single implicit task, and the task effects remain

unclear (cf. Quesada & Lozano, 2020). Participants could avoid difficult classifiers and just use those they were very confident with in their story-telling task (cf. Ellis, 1994).

Much contrary to Polio (1994), Liang (2009) found an L1 effect with Korean participants outperforming English participants in their use of Chinese classifiers overall. Methodologically, Liang (2009) improved on Polio's (1994) study by elaborating on learner proficiency and making a distinction between novice, intermediate and advanced levels, which helped to elucidate the effect of proficiency. It was not surprising that participants with a higher proficiency level outperformed the lower proficiency group within each L1 cohort. Equally importantly, Liang (2009) compared the results from the comprehension and production tasks, which demonstrated the potential task effects and may be the reason for the differing L1 effect finding. To be specific, intermediate English participants outperformed their Korean counterparts in the comprehension task, but not in the production task.

The above-mentioned studies offered a snapshot of classifier acquisition. Zhang and Lu (2013) administered a research study centring on adult Chinese-L2 learners' developmental acquisition of classifiers. Data were collected from a corpus of 657 essays written by L2 speakers (the majority of which were English-L1 speakers) in the same Chinese language classes throughout two consecutive academic semesters. One essential finding is the change in diversity of classifiers. Specifically, token (the number of classifiers that were used) frequency declined while type frequency increased. This supports the probability of a sequential mastery of different classifier types, although the authors had different research priorities and did not elaborate the detailed development of classifier types in their research. Compared to most other empirical studies, the data collection method (i.e., a composition test) had the merit of being situated in a naturalistic setting, as the participants were not told that the research focused on classifier acquisition (cf. Polio, 1994; Quesada & Lozano, 2020). Additionally, the study focused on written production instead of spoken communication. However, purely relying on classifier analysis from a composition task may problematize the interpretation of proficiency. Similarly, to the story narrative conducted by Polio (1994), participants could choose to use classifiers or not as per their individual preference when no instructions on the use of classifiers were provided.

Taken together, existing empirical studies find quite mixed effects of L1 transfer in Chinese L2 classifier use. On top of that, none of these studies examined speakers of more than one non-classifier language in a single study. It remains unclear whether different groups of non-classifier language speakers would perform differently in applying classifiers under the same conditions. At the methodological level, the majority of previous studies employed a single task (i.e., written composition, story narrative, or daily record) for data collection, which may incur inaccurate and incomplete interpretations (Ellis, 1994; Ellis & Barkhuizen, 2005; Johnstone, 2000). Additionally, limited studies have ascertained the existence of a developmental sequence for L2 classifier learning. Thus, this study aims to address the following research questions:

(1) Does L1 affect classifier acquisition?
(2) Does task type affect test performance in relation to classifiers?
(3) Is there a predictable hierarchical development of different types of classifiers?

3 Method

3.1 Participants

Participants (n = 50, female = 34) came from two language backgrounds in relation to classifiers: classifier L1 (Japanese) and non-classifier L1 (Arabic and English). All participants were taking Chinese language courses at their home universities in Australia, Egypt, and Japan: 17 Egyptian Arabic-L1 speakers (age 18-26, mean = 21.18, SD = 3.09), 15 Australian English L1 speakers (age 20-30, mean = 23.67; SD = 4.69), and 18 Japanese-L1 speakers (age 20-22; mean = 20.11, SD = 0.58). To ensure their comparability in terms of Chinese proficiency, I only recruited students enrolled in intermediate Chinese language classes, using Chinese textbooks pertaining to intermediate level, equivalent to HSK3 ('Haiyu shuiping kaoshi', Chinese Proficiency Test 3); and all of them rated themselves as intermediate learners in the demographic questionnaire (cf. Chan & Leung, 2020; Suzuki & Sunada, 2020).

3.2 Materials

There were 3 written tasks based on a description of the same picture: a composition, a free cloze test, and a multiple-choice cloze test, all intended to elicit classifiers (see Appendix). To provide a prompt for the three writing tasks, a picture (Figure 1) with many objects was designed by the first author in her pilot study. Before data collection with L2 speakers, I ran a norming study with four native Chinese speakers (aged 22-24, university students). They were invited to write a composition based on the same picture; these examples were used to develop the cloze tests. Table 1 lists all of the classifiers along with their corresponding head nouns that appear in the intermediate-level textbook 'New Practical Chinese Reader' (Liu, 2011) and that set as prompts in the test. These classifiers represent different types, including shape, animate, inanimate, and concept. Some classifiers not found in the textbooks could be elicited by the picture, such as the classifier *shan* for windows. Such classifiers were not included in the prompts to tasks 2 and 3, but any correct use of classifiers in task 1 could get a score.

Table 1: *Classifiers and types.*

Question	Classifier-object	Type
1	ge-people	animacy
2	dui-couple	concept
3	zhang-sofa	shape
4	zhi-cat	animacy
5	bei-wine	shape
6	tiao-dog	animacy
7	ge-cupboard	inanimacy
8	ke-tree	inanimacy
9	fu-calligraphy	inanimacy
10	zhang-table	shape
11	ge-vase	inanimacy
12	ge-goldfish bowl	inanimacy
13	tiao-goldfish	animacy
14	ge-home	concept
15	ge-moment	concept

Figure 1: *Picture used in the test.*

The three tasks were constructed following 'funnel' principle to narrow down the test scope. Task 1 was a short descriptive composition, instructed a compulsory word-length of 150-200 characters. It was aimed at capturing learners' naturalistic use of classifiers. Task 2 was a free cloze task intended to examine noticing of the compulsory application of classifiers in specific context. Note that task 2 accepted both formal and informal use of classifiers, which may result in a 'one-to-more mapping' in each question. The multiple-choice cloze task, task 3, aimed to ascertain accuracy in selecting classifiers, where only one formal use of the classifier was considered as correct in each question. On the whole, the research design allows data triangulation from multiple sources and thus can offer a more accurate interpretation of participant accuracy (cf. Revesz et al., 2019 for triangulation of data analysis).

3.3 Data collection

All the participants took part in one individual session in a language teaching classroom at their home university. The test was a traditional paper-and-pen test. The term 'classifier' was not used in any of the task instructions, so that participants remained unaware of the language feature that was being examined. This was to ensure a natural production of classifiers and to avoid confirmation bias. The test was not timed, but the majority finished it in around 45 minutes. Participants were prohibited to refer back to previous tasks or to consult a dictionary at any time. They could not progress to the next task until finishing the previous one either. The procedure of data collection was reviewed and approved by an (anonymised) Human Research Ethics Committee (protocol 2019/167).

3.4 Data coding and statistical analysis

Accuracy in the use of classifiers was scored as 1 or 0 for each test item. Multiple answers could be marked as correct for each head noun in tasks 1 and 2, depending on context; on the other hand, only one correct answer was accepted for each question in task 3. Considering the possible unbalanced use of classifiers (some participants wrote longer compositions), each participant was given an overall percentage score of correct use of classifiers for each task, instead of a raw correct score.

To determine the effects of participant L1 and task on the accuracy of classifier use, a linear mixed-effects regression model was fitted to the data. An interaction between task and participant L1 was used as a predictor, and participant was included as a random effect; the formula is: lmer(score ~ task*L1+(1|participant). To further examine whether there is a developmental acquisition order, a logistic mixed-effects model with L1 and classifier type as fixed effects was used to predict the scores; formula: glmer(score ~ L1*type+(1|participant) +(1|classifer). Here the random effects included individual participants and classifiers. For this model, I only consider the score from task 3 as the response variable, because there is only one correct answer assigned for each test item. The major difference between linear and logistic mixed-effects models lies in the dependent variable, with the former using score accuracy percentage and the latter employing binary accuracy response. Both models were computed using the lmerTest package in R (Kuznetsova et al., 2017; R Core Team, 2020). As the sample size was relatively small, I performed bootstrapping to validate both of the models (see Levshina, 2015 for bootstrapping in linguistic studies). The bootstrapping method suggested a stable 95% percentile type of confidence interval (0.112 - 0.351) for R^2 in the linear model and an optimism slope of 0.020 for the logistic model, both of which validated the statistical modelling.

4 Results

4.1 Classifier accuracy across L1s and tasks

Table 2 presents ranges, means and SDs of test accuracy rates in all nine (3*3) conditions. Japanese L1 participants scored highest overall with a mean of 71.94 (SD =12.47, range 42.33 - 88.67), followed by English L1 participants (mean 66.51, SD =14.65, range 33.33-89.00), and Arabic L1 participants (mean 59.35, SD = 20.05, range 30.00-93.33). In addition, each L1 group exhibited a decreasing trend from task 1 to task 3, which means that each cohort of speakers scored highest in task 1 and lowest in task 3. Descriptive analysis suggests that there are both L1 and task effects. Statistical analysis can confirm whether the observed numerical differences are in fact significant.

Table 2: *Descriptive statistics for classifier accuracy.*

L1	Task	Range	Mean	SD
Arabic	1	0.00-100.00	70.47	28.28
(N = 17)	2	20.00-87.00	54.76	21.31
	3	20.00-93.00	53.82	21.52
English	1	33.00-100.00	72.87	16.71
(N = 15)	2	27.00-93.00	69.27	17.91
	3	27.00-80.00	57.40	13.73
Japanese	1	50.00-100.00	78.56	14.91
(N = 18)	2	27.00-93.00	71.39	17.26
	3	33.00-93.00	65.89	16.09

The linear model is summarized in Table 3. To elaborate, the ‘estimate’ and the ‘standard error’ columns show the predicted score and standard error for a level, respectively. For the base level, i.e., intercept (Arabic L1 speakers performing task 1), the predicted score is 70.471. To calculate the predicted score for a different level, the respective value in the ‘estimate’ column is added or subtracted. For instance, compared to task 1, the Arabic L1 participants received a score of 16.706 lower in task 2 and 16.647 lower in task 3. These differences were significantly different ($p < 0.001$ and $p < 0.001$) as indicated in the ‘Significance’ column. This means that the scores of the Arabic participants were significantly different between tasks.

Table 3: *Summary for the linear model (Arabic L1 and task 1 as intercept).*

	Estimate	Standard error	df	t-value	Pr(>|t|)	Significance
(Intercept)	70.471	4.647	87.778	15.166	0.000	***
task2	-16.706	4.405	94.000	-3.792	0.000	***
task3	-16.647	4.405	94.000	-3.779	0.000	***
L1English	2.396	6.787	87.778	0.353	0.725	
L1Japanese	8.085	6.480	87.778	1.248	0.215	
task2: L1 English	13.106	6.434	94.000	2.037	0.045	*
task3: L1 English	1.180	6.434	94.000	0.183	0.855	
task2: L1 Japanese	9.539	6.143	94.000	1.553	0.124	
task3: L1 Japanese	3.980	6.143	94.000	0.648	0.519	

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

At the same time, no significant differences were found between the Arabic, English, and Japanese participants in task 1, which means that all L1 groups performed comparably in task 1. However, there was a significant interaction between task and L1 ($p < 0.05$), such that the drop from task 1 to task 2 was less by 13.106 for English participants in comparison to Arabic ones. The interaction between task and L1 is plotted in Figure 2.

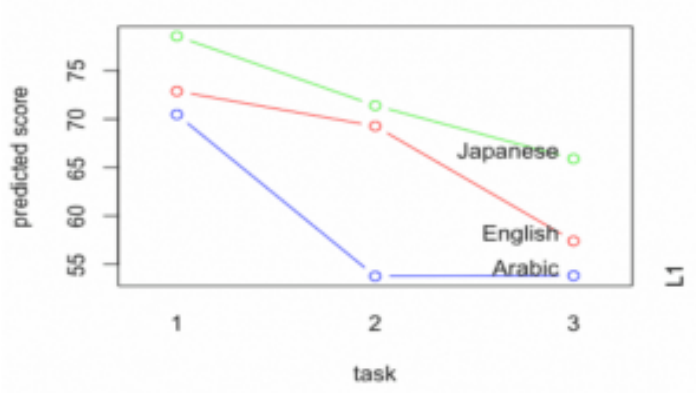


Figure 2: *Model prediction for scores of the three L1 groups in the three tasks.*

In an attempt to explain the difference between the two non-classifier language groups, I calculated the individual number of *ge* tokens in task 2, with results demonstrating that the Arabic, English, and Japanese L1 groups yielded an individual token number of 4.53, 6.13 and 4.88 in task 2 respectively, suggesting a higher usage of *ge* by the English participants. To ascertain whether the difference is significant, a Mann Whitney U Test was run. Results showed a significant difference ($p < 0.05$) between the Arabic and English participants in using *ge*.

4.2 Hierarchical development of classifier use

Table 5 shows the descriptive statistics for classifier types by L1 in task 3. Shape-type classifiers attract highest accuracy rate, followed by animacy, inanimacy, and concept. For all classifier types but shape, Japanese participants had highest accuracy score, followed by English and Arabic participants. For the shape type, however, Arabic participants were the most accurate. I run a logistic model to ascertain the significance of these effects.

Table 4: *Descriptive statistics for classifier types in task 3.*

Type	L1	Accuracy	Total
Animacy	Arabic	61.76	69.50
	English	71.67	
	Japanese	75.00	
Concept	Arabic	33.33	45.33
	English	44.44	
	Japanese	57.41	
Inanimacy	Arabic	43.53	51.60
	English	52.00	
	Japanese	58.89	
Shape	Arabic	80.39	72.67
	English	60.00	
	Japanese	75.93	

Our model suggests that neither the interaction nor L1 was a significant predictor, so I pruned the model to having type as the only fixed effect. Table 6 represents the final model. The type of concept was chosen as the reference level (the intercept). The 'estimate' column in the table represents the log odds of the dependent variable being one factor rather than the other. Positive values in the column mean a higher chance of scoring under a particular condition, while negative values mean a lower chance of scoring. The estimate for the shape type is positive at 1.617 and is significantly different from the estimate for the concept type ($p < 0.05$), as indicated in the 'significance' column. This means that classifiers of the shape type were significantly more likely to get correct scores in comparison to the concept type. There was a trend for animacy type to receive a higher score as well, but this difference did not reach significance at the traditional α level ($p = 0.063$).

Table 5: *Summary for the logistic model (concept type as intercept).*

	Estimate	Standard error	z-value	Pr(>|z|)	Significance
(Intercept)	-0.184	0.529	-0.347	0.729	
type.shape	1.617	0.747	2.164	0.030	*
type.animacy	1.278	0.686	1.862	0.063	.
type.inanimacy	0.635	0.687	0.924	0.356	

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

An effect plot (Figure 3) was made to visualise the predicted probabilities for type, with 95% error bars presented to indicate the uncertainty of the estimates. The differences in accuracy across classifier types suggest a potential acquisition order from a developmental perspective, according to which shape type classifiers are more likely to be acquired first, followed by animacy, inanimacy, and concept type classifiers.

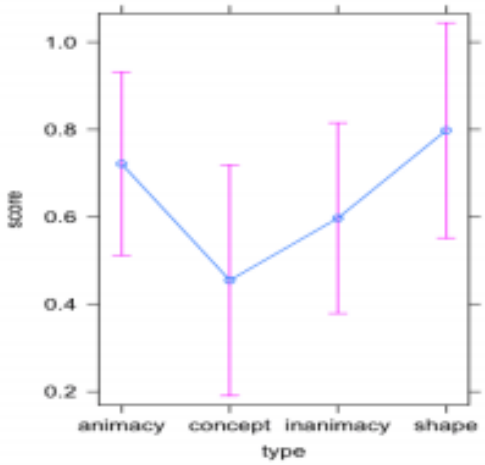


Figure 3: *Model prediction for classifier types.*

5 Discussion

5.1 L1 transfer

The descriptive statistics of the L1 effect suggest that similarity could facilitate learners' classifier acquisition. The Japanese participants scored highest numerically within each task and overall, in comparison to the English and Arabic participants. As Japanese is a typical classifier language, it exhibits frequent use of classifiers (Hansen & Chen, 2001). In contrast, Arabic and English are two non-classifier languages, so theoretically the Arabic and English participants are more likely to encounter difficulty in learning classifiers. This is in line with Liang's (2009) study of Chinese classifier acquisition, where Korean participants were found to outperform their English counterparts in producing classifiers. It also aligns well with me pilot study where the Thai participant scored much higher in using classifiers than the New Zealand participant in a short composition task.

However, statistical analysis of the data suggests that these differences did not always reach significance. While Arabic participants scored lower than Japanese in task 2, the difference did not reach significance in tasks 1 or 3 or between English and Japanese in any task. Similar conclusions can be found in another SLA research. For instance, Polio (1994) reported no clear L1 effect in the oral production of classifiers by comparing the data from the English and Japanese speakers. Liang (2009) observed a minimal difference between English and Japanese speakers' use of Chinese classifiers. That is, both groups showed a higher and correct use of shape-typed classifiers in the production task, while this varied in the other tasks

employed in his study. Hence, the current results are inconclusive in relation to L1 transfer effect and suggest that L1 effects may be modulated by task effects.

Taken together with previous research, my results suggest that L1 is not a significant predictor of classifier accuracy by itself, and classifier-L1 speakers do not outperform non-classifier-L1 speakers, at least at the intermediate level of proficiency. It may be that the L1 advantage is realised differently, perhaps, through the effort required to learn the feature or at beginner levels of proficiency.

5.2 Task effects

Task was found to be a significant predictor of test performance. Numerically, task 1 yielded the highest scores, followed by task 2 and task 3. Differences in the scores between tasks are reflections not of the different types of questions asked, but of the inherent difficulty of the tasks. From the perspective of language testing, it was not surprising that participants performed best in task 1, where there was no obligation for the use of classifiers and the score was calculated by accuracy percentage (cf. Douglas, 2014; Zhang & Lu, 2013). Participants could use the classifiers they were confident with and avoid using those they were less confident with, regardless of L1, resulting in no significant difference between the L1 groups. On the other hand, the use of classifiers became obligatory in task 2, so less space was left for the use of test-taking strategies (Downing, 2002). This task revealed a significant difference for the Arabic group. However, it becomes another story for task 3, which was intentionally designed to ascertain the correct and formal use of classifiers and there was only one fixed correct answer for each question (cf. Badger & Yan, 2012). This potentially required a superior command of classifiers in comparison to task 2 since there was no chance to choose *ge* as a placeholder without the option of *ge* being provided in the corresponding questions. The reason that participants still scored higher in task 2 than in task 3 could be the overuse of *ge* in task 2 (cf. Polio, 1994; Zhang & Lu, 2013).

In addition to the effect of task type *per se*, certain L1 groups may perform better in a specific task type than others, and test-taking strategies could affect the test performance in specific contexts (Douglas, 2014). Statistically, the linear model indicates a significant interaction between task and L1 with Arabic participants showing a substantial drop in accuracy from task 1 to task 2 and English participants from task 2 to 3. These findings could be explained in view of the combination of task nature, that the 'one-to-many mapping' allows the leeway for scoring; and test-taking strategies, namely, the overproduction of the generic classifier *ge* (Douglas, 2014). As the results demonstrated, the English participants were observed to use *ge* more frequently as a placeholder in the questions in task 2. Future research can further explore the effect of task-taking strategies.

5.3 Hierarchical development of classifiers

Results of the logistic model indicate no interaction between L1 and type, which implies a universally sequential mastery of different types, which aligns with no significant effect of L1 on accuracy. The descriptive statistics suggest a successive mastery of shape, animacy, inanimacy, and concept classifiers; and the model confirmed significant differences between shape and concept types. Accordingly, participants were more likely to get a correct score for the different types of classifiers as per the sequence predicted, and to score worse in the case of classifiers acquired later, which further implies that longitudinally the sequence could be a representation of a developmental path of L2 classifier learning. The results are in line with a longitudinal observation of L2 Chinese classifier acquisition by Zhang and Lu

(2013), who observed that the number of classifier types acquired increased over time. That is, L2 learners generally acquire different types of classifiers in succession. Although this study did not specify the sequence of classifier types, it can still, through the lens of increasing token and type frequency, support the assumption that the mastery of classifier types is developmental in nature. Moreover, Aikhenvald's (2003) study elucidates that Chinese-L1 speakers acquired shape-based classifiers earlier than the non-extended use of classifiers, which is congruent with the findings captured.

However, the findings of the current study are incongruent with the L1 acquisition of other classifier languages such as Kilivila, Hokkien or Cantonese, in which animacy classifiers are acquired first, followed by inanimacy classifiers, and then shape classifiers (e.g., Hu, 1994; 1994; Luke & Harrison, 1986; Senft, 1996). The findings also partially contradict the NCAH, which hypothesizes that classifiers of the animate type are to be acquired earlier than the shape typed ones (see also Craig, 1986). Comparison of all the classifier acquisition sequences above shows areas of agreement as well as disagreement. Debate mainly exists with respect to whether shape or animacy classifiers are acquired earlier. A consistent conclusion is that animacy classifiers are acquired relatively early, whereas concept classifiers (i.e., those that refer to abstract notions or extended meaning) are the last to be acquired.

5.4 Limitations and future research

Apart from discussing the results of the study, it is also necessary to recognise the limitations of the research. The most important one concerns the inhomogeneous distribution of the three participant groups. They were studying in three different countries, where language exposure, teaching style, or even cultural factors may affect their learning rate. In this study, this was necessary in order to recruit enough participants from different L1 groups. Future researchers are encouraged to collect data from one location to minimise the effect of such extraneous factors. Next, a gender bias clearly exists in the participant pool as the number of female participants (n = 34) was more than double that of the male participants (n = 16). Even though some studies have demonstrated that gender does not technically affect language learning outcomes (e.g., Gafni et al., 2017), future studies are recommended to recruit participants more evenly in terms of gender and other demographic variables. Moreover, the study did not take individual differences such as memory capacity and motivation to learn classifiers into account. Much research has indicated that memory capacity and motivation can lead to important influence on the rate of learning given the same learning situation and L1 (e.g., Marini, Eliseeva & Fabbro, 2016; Schuetze, 2015). However, due to the practicality and the limited scope of the current study, such data was not collected. Still, the three factors could be further explored in later research.

5.5 Conclusions

The current study focused on Chinese-L2 learners' production of classifiers by a specially designed language test and shed new light on how the performance was constrained by different factors (L1, task, and classifier type) that were not considered together in previous research. This study takes the position that L1 does not significantly affect the use of classifiers. Different types of tasks influence the classifier production because of the inherent difficulties of task setting and participants' test-taking strategies applied. Finally, the hierarchical development of classifiers is supported, with no L1 effects observed. The methodological implications of the study suggest that future research should carefully consider the pros and cons of using various types of tasks as an elicitation method and, perhaps, use several for triangulation

purposes. In the pedagogical practice, Chinese L2 language teachers may choose to introduce different types of classifiers in the order suggested here, in lieu of paying much attention to learners' L1, at least at the intermediate level.

6 References

Aikhenvald, A. (2003). Classifiers: A typology of noun categorization devices. Oxford: Oxford University Press.

Allan, K. (1977). Classifiers. Language, 53(2), 285-311.

Andersen, R. (Ed.). (1984). Second languages a cross-linguistic perspective. Rowley/London/Tokyo: Newbury House Publishers.

Badger, R., & Yan, X. (2012). The use of tactics and strategies by Chinese students in the Listening component of IELTS. Cambridge: Cambridge University Press.

Barkhuizen, G. (2005). *Analysing learner language.* Oxford: Oxford University Press.

Chan, R., & Leung, J. (2020). Why are lexical tones difficult to learn? Studies in Second Language Acquisition, 42(1), 1-27.

Cheong, C. M., Zhu, X., Li, G. Y., & Wen, H. (2019). Effects of intertextual processing on L2 integrated writing. Journal of Second Language Writing, 44, 63-75.

Craig, C. (Ed.). (1986). Noun classes and categorization. Amsterdam/Philadelphia: John Benjamins Publishing Company.

De Vincenzi, M., & Lombardo, V. (2000). Cross-linguistic perspectives on language processing. Boston: Kluwer Academic.

Douglas, D. (2014). *Understanding language testing*. New York: Routledge.

Downing, S. M. (2002). Threats to the validity of locally developed multiple-choice tests in medical education: Construct-irrelevant variance and construct underrepresentation. *Advances in Health Sciences Education*, *7*(3), 235-241.

Dyson, B. (2009). Processability theory and the role of morphology in English as a second language development: A longitudinal study. *Second Language Research*, *25*(3), 355-376.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R., & Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. Language and Cognitive Processes, 24(7-8), 1124-1179.

Erbaugh, M. S. (1986). Taking stock: The development of Chinese noun classifiers historically and in young children. In C. Craig (Ed.), Noun classes and categorization (pp. 399-436). Amsterdam/Philadelphia: John Benjamins Publishing.

Gafni, R., Achituv, D. B., & Rahmani, G. (2017). Learning foreign languages using mobile applications. *Journal of Information Technology Education Research*, 16, 301-317.

Gass, S. M., & Mackey, A. (Eds.). (2013). The Routledge handbook of second language acquisition. London: Routledge.

Hansen, L., & Chen, Y. L. (2001). What counts in the acquisition and attrition of numeral classifiers. *Jalt Journal, 23*(1), 83-100.

Hu, Q. (1994). Overextension of animacy in Chinese classifier acquisition. In E. V. Clark (Ed.), *Proceedings of the twenty-fifth annual child language research forum* (pp. 127-136). Stanford, USA: Stanford University Press.

Hyltenstam, K. (1977). Implicational patterns in interlanguage syntax variation. *Language Learning*, *27*(2), 383- 410.

Jarvis, S., & Odlin, T. (2000). Morphological type, spatial reference, and language transfer. *Studies in Second Language Acquisition*, *22*(4), 535-556.

Jarvis, S., & Pavlenko, A. (2008). Crosslinguistic influence in language and cognition. London: Routledge. Jiang, S. (2017). The Semantics of Chinese Classifiers and Linguistic Relativity. Abingdon: Routledge.

Johnstone, B. (2000). Qualitative methods in sociolinguistics. Oxford: Oxford University Press. Krashen, S. (1983). Principles and practice in second language acquisition. TESOL Quarterly, 17(2), 300-305.

Kuo, J. Y. C. & Sera, M. D. (2009). Classifier effects on human categorization: The role of shape classifiers in Mandarin Chinese. Journal of East Asian Linguistics, 18(1), 1-19.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins Publishing.

Lehrer, A. (1986). English classifier constructions. Lingua, 68(2), 109-148. doi:10.1016/0024-3841(86)90001-X.

Liang, S. Y. (2009). The Acquisition of Chinese Nominal Classifier Systems by L2 Adult Learners. (Unpublished Doctoral dissertation). University of Texas Arlington, USA.

Liu, X. (2011). *New practical Chinese reader.* Beijing: Beijing Language and Cultural University Press.

Littlemore, J. (2009). *Applying cognitive linguistics to second language learning and teaching*. London: Palgrave Macmillan.

Luke, K. K., & Harrison, G. (1986). 'Young Children's Use of Chinese (Cantonese and Mandarin) Sortal Classifiers'. In H. S. Kao & R. Hoosain (Eds.), *Linguistics, Psychology and the Chinese Language* (pp. 125-147). Hong Kong: University of Hong Kong.

Marini, A., Eliseeva, N., & Fabbro, F. (2016). Impact of early second-language acquisition on the development of first language and verbal short-term and working memory. *International Journal of Bilingual Education and Bilingualism, 22*(2), 1-12. doi:10.1080/13670050.2016.1238865

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Ortega, L. (2014). *Understanding second language acquisition*. London: Routledge.

Paul, Jing Z, & Grüter, Theres. (2016). Blocking Effects in the Learning of Chinese Classifiers. *Language Learning*, *66*(4), 972-999.

Pienemann, M. (1998). *Language processing and second language development: Processability theory (Vol. 15).* Amsterdam: John Benjamins Publishing.

Pienemann, M. (Ed.). (2005). *Cross-linguistic aspects of Processability Theory (Vol. 30).* Amsterdam: John Benjamins Publishing.

Polio, C. (1994). Non-native speakers' use of nominal classifiers in Mandarin Chinese. *Journal of the Chinese Language Teachers Association, 29*(3), 51-66.

Quesada, T., & Lozano, C. (2020). Which factors determine the choice of referential expressions in L2 discourse? Studies in Second Language Acquisition, 42(3), 1-28. doi:10.1017/S0272263120000224.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Australia. Retrieved from https://www.r-project.org

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors: A mixed methods study. Studies in Second Language Acquisition, 41(3), 605-631.

Ringbom, H., & Jarvis, S. (2009). The importance of crosslinguistic similarity in foreign language learning. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 106-118). Hoboken: Blackwell Publishing.

Saalbach, H., & Imai, M. (2012). The relation between linguistic categories and cognition: The case of numeral classifiers. *Language and Cognitive Processes*, *27*(3), 381-428.

Schuetze, U. (2015). Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long term memory. Language Teaching Research, 19(1), 28-42. doi:10.1177/1362168814541726

Senft, G. (1996). Classificatory Particles in Kilivila. New York: Oxford University Press.

Shi, Y. (2014). Comparison of individual classifiers and collective classifiers between Chinese and English. Theory and Practice in Language Studies, 4(9), 1961-1965. doi:10.4304/tpls.4.9.

Suzuki, Y., & Sunada, M. (2020). Dynamic interplay between practice type and practice schedule in a second language: The potential and limits of skill transfer and practice schedule. *Studies in Second Language Acquisition, 42*(1), 169-197. doi:10.1017/S0272263119000470.

Tang, C. J. (2005). Nouns or Classifiers: A Non-Movement Analysis of Classifiers in Chinese. *Language and Linguistics (Taipei)*, *6*(3), 431.

Zhang, J., & Lu, X. (2013). Variability in Chinese as a foreign language learners' development of the Chinese numeral classifier system. *The Modern Language Journal*, *97*(S1), 46-60.

Zhang, J. (2019). Crosslinguistic influence on Chinese-L2 learners' acquisition of classifiers. *ANU Undergraduate Research Journal*, *9*, 156-171. Retrieved from http://studentjournals.anu.edu.au/index.php/aurj/article/view/138

Zhang, H. (2007). Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics, 16*(1), 43-59. doi:10.1007/s10831-006-9006-9.

7 Appendix

Materials: Test of Classifiers

Task 1: short composition
Please describe what you see in the picture with as many details as possible (150-200 characters).

Task 2: gap-filling
Please fill the blanks based on the picture (Note: only one character is acceptable for each blank).

照片里有两 1._人，他**们应该**是一 2._夫妻。他**们**坐在一 3._**黄**色的沙**发**上，女生抱着 一 4._橘**黄**色的小猫，男生手里**举**着一 5._**红**酒。他**们**看起来很幸福。旁**边**有一 6._白 色的狗，也笑得很开心。窗外的**风**景很好，看得**见蓝**天，白云和几**栋**大楼。窗台下面

有一 7._柜子，上面摆了一些绿色的小盆栽。柜子的旁边还有一 8._树。 后面的墙壁 上有一 9._大大的倒过来的“福” 字，红红的，很喜庆，渲染着过年的气氛。靠着墙壁 还放了一 10._桌子，上面有一 11._金鱼缸和一 12._花瓶。花瓶里插着几枝梅花，十 分优雅。金鱼缸里面有两 13._可爱的小金鱼，它们在水里快乐地游来游去。这应该是
一 14._很温暖的家，这家人正在享受着一 15._幸福的时刻。

Task 3: multiple-choice
Please choose the correct answer for each blank based on the picture.

照片里有两 1._人，他们应该是一 2._夫妻。他们坐在一 3._黄色的沙发上，女生抱着 一 4._橘黄色的小猫，男生手里举着一 5._红酒。他们看起来很幸福。旁边有一 6._白 色的大狗，也笑得很开心。窗外的风景很好，看得见蓝天，白云和几栋大楼。窗台下 面有一 7._柜子，上面摆了一些绿色的小盆栽。柜子的旁边还有一 8. 树。 后面的墙
壁上有一 9._大大的倒过来的“福” 字，红红的，很喜庆，渲染着过年的气氛。靠着墙 壁还放了一 10._桌子，上面有一 11._金鱼缸和一 12._花瓶。花瓶里插着几枝梅花， 十分优雅。金鱼缸里面有两 13._可爱的小金鱼，它们在水里快乐地游来游去。这应该 是一 14._很温暖的家，这家人正在享受着一 15._幸福的时刻。

1. A. 组 B. 只 C. 个 D. 位
2. A. 个 B. 位 C. 对 D. 双
3. A. 条 B. 张 C. 段 D. 面
4. A. 匹 B. 个 C. 头 D. 只
5. A. 壶 B. 杯 C. 个 D. 瓶
6. A. 个 B. 条 C. 匹 D. 头
7. A. 台 B. 面 C. 条 D. 个
8. A. 根 B. 棵 C. 行 D. 个
9. A. 页 B. 幅 C. 面 D. 本
10. A. 台 B. 面 C. 张 D. 条
11. A. 只 B. 个 C. 盆 D. 杯
12. A. 个 B. 盆 C. 支 D. 瓶
13. A. 个 B. 匹 C. 头 D. 条
14. A. 所 B. 间 C. 个 D. 座

15. A. 分 B. 点 C. 个 D. 门

Beyond Anglo-Norman: the Lexical Influence of Old French Dialects on Middle English

Beth Beattie
University of Glasgow

Abstract. This project examines lexical borrowings between the Old French (OF) of Champagne and Picardy and Middle English, identified through spelling variations, and what these borrowings reveal about cultural links between English and France. The methodology consisted of using the Middle English Dictionary (MED) and the Linguistic Atlas of Early Middle English (LAEME) to establish the frequency of variant spellings of open-class Old French words, thus indicating the strength of the potential borrowing of an Old French word or feature. An examination of the context of texts in which a high concentration of dialectal forms were found was used to determine the cultural background that facilitated such transmission. The different source materials for the MED and LAEME resulted in a marked culling of forms to those found in both datasets, but the variant spellings given for those that remained proved ample. The analysis of the texts containing these variant spellings produced fewer results than expected; The majority of the manuscripts featured few words of note, with the exception of Arundel 57, which provided a glimpse into OF usage that was decidedly more varied than solely Anglo-Norman or Central French. This lack of final data, in combination with the wide distribution of the lexical items across manuscripts about which not much is known, made it impossible to focus on a particular lexical source and examine the reasons behind lexical transmission. The possibility to undertake such research remains, with further extended examination required.

Keywords: Old French; Middle English; loanwords; spelling; historical dialectology

---

1 Introduction

Due to rigorous examination and centuries of study, the Dark Ages can no longer be accurately named ‘dark’. The term is now very seldom used outside of popular culture, yet the Middle Ages in Western Europe still hold a certain fascination, especially to historical linguists. The role of Old French (OF) in the development of Middle English (ME) – especially the Anglo-Norman (AN) and Central French (CF) dialects – has been the recipient of an inordinate amount of attention in uncovering the history of English, yet they were not the only OF dialects to exist during this period. Spanning the width of the langue d’oïl was a continuum of around twelve major dialects, each with their own distinct features that differentiated them from the incipient CF standard. Given the proximity and cultural ties between England and France between the years of 1100 and 1400, it seems likely that other dialects played a role in contributing to ME vocabulary, even if a fleeting one. Such discoveries would be instrumental in detailing the picture of cross-Channel links, both on a linguistic and a societal level.

1.1 Research questions

The research questions this dissertation aims to answer are:

(1) Is it possible to identify loanwords into Middle English from Old French dialects other than Anglo-Norman and Central French?
(2) Can the origins of these loanwords be determined through spellings?
(3) What can these borrowings reveal about cultural links between England and France?

1.2 Method

First, differentiating features of Old French dialects that result in spelling variation will be established, providing examples of these features in literature to confirm usage. The MED shall then be used to search for open-class words known to be of Old Northern French origin, and examine the variant forms given in each entry to identify possible features found in a specific Old French dialect. These potentially relevant forms will form a basis for a LAEME search to establish the frequency of these forms, thus indicating the strength of the potential borrowing of an Old French word or feature. If particular texts feature a high volume of Old French dialectal forms, those texts will then be examined more in-depth to identify the reason for such concentrations, most likely due to scribal influences or – more interestingly for my research – cultural links with a particular region of France.

1.3 Old French Dialect Features

The most suitable elaboration of the distinguishing features between the various OF dialects can be found in Einhorn's 1974 handbook on Old French, and it is their table and description of these features that forms the basis for each stage of analysis.

Table 1: *Old French dialectal features, based on Einhorn, 1974, p. 137*[1].

Characteristics	S	SW	W	N	AN	P	Wn	Ch	L	FC	B
1. [o:] > *ou, o*	xx	xx	xx	xx			xx	x	xx	xx	xx
2. *ei* stays, or > *e*	xx	xx	xx	xx	xx						
3. *ie* > *e*	x	xx	xx	x	xx						
4. Cl.1 impf. in *-oue, -oe*	x	x	x	x	x						
5. Tonic *el, eu* > *al, au*		xx	x								

[1] References to these features will be labelled numerically according to this chart. For example, '16' refers to '16. Use of *w*'

6. 1st p.pl. *-om(s)*, *-on*		xx	xx	xx	x						
7. [ɛ] (+ ˜ *n*) stays [ɛ̃]			xx	xx	xx	xx	xx				
8. Graphy *ai* for *ei*			xx								
9. 1st p.pl. *-um(s), -un(s)*				x	xx						
10. [k], [g] for *ch, j*				x	x	xx					
11. *c* (+*e, i*) > *ch*				x	x	xx	x	x			
12. *eau* > *iau*				x		xx	x				
13. [eː] > *ei*				x			xx	x	xx	x	
14. [oː], [o] > *u*					xx						
15. Final *z* > *s* 12th c.					x	xx					
16. Use of *w*					x	xx	xx	x	xx		
17. *la* > *le*						xx	xx				
18. Final *t* remains						xx	xx	x	x		
19. Cl.1.impf. in – *(i)eve*							x		x	x	x
20. *a* > *ai*								x	xx	xx	x
21. Initial *e* > *a*									xx	x	
22. *lo, lou* = *le*									xx	x	x
23. Tonic [ɛ] > *a*										xx	x
24. *al, able* > *aul, auble*										x	xx
25. *ei* (+nasal) > *oi*											xx

The dialects and the areas in which they were spoken are:

- South (S): Barbonnais, Nivernais, Berry, Orléans
- South-West (SW): Angoumois, Saintonge, Aunis, Poitou
- West (W): Touraine, Anjou, Maine, Brittany
- Norman (N): Normandy
- Anglo-Norman (AN): England
- Picard (P): Picardy, Artois
- Walloon (Wn): North East (Belgium)
- Lorraine (L)
- Franche-Compté (FC)
- Burgundy (B)
- Champagne (Ch)

Table 1 lists the identifying features and the dialects in which they appear, illustrating a strong presence with two crosses and a weaker presence with one cross. Einhorn elaborates on the condensed versions of the features and gives examples of each:[2]

1. Tonic (or most stressed) syllables containing [oː] became [ø], spelled <eu>, in Parisian French and Picard, but [u] in other dialects, spelled <u> in Anglo-Norman and <ou> or <o> elsewhere.

- *nevou (neveu)*
- *seignor (seigneur)*

2. The diphthong /ei/, instead of becoming <oi> during the twelfth century, was lowered through [ɛi] to [ɛ], spelled <ei> or <e>.

- *saveir (savoir)*
- *le rei (le roi)*
- *aveit (avoit)*
- *la metié (la moitié)*
- *la vee (la voie)*

3. Tonic syllables containing [iɛ] became [ɛː], spelled <e>.

- *chevaler (chevalier)*
- *ben (bien)*
- *cel (ciel)*
- *manere (maniere)*
- *la pere (la pierre)*

4. In the thirteenth century, Class 1 verbs (where the infinitive ends in <-er>) used <-oue>, and later <-oe>, imperfect conjugations alongside the standard endings.

Standard OF Imperfect Conjugations	**Dialectal Imperfect Conjugations A**

[2] Einhorn, 1974, pp. 135–140; own italics); The forms in brackets are the equivalent Parisian French forms of the dialectal examples. The IPA symbols have also been updated to correspond with current usage.

-eie	-o(u)e
-eies	-o(u)es
-eit	-o(u)t
-iiens	-iiens
-iiez	-iiez
-eient	-o(u)ent

5. Tonic syllables containing <el> and <eu> appeared as <al> and <au>.

- *quaus (quels)*
- *tau (tel)*
- *corporau (corporel)*

6. The endings <-om(s)> and <-on> were used instead of <-ons> in the 1st person plural.

- *aloms ! (allons !)*
- *cum nos disiom (comme nous disons)*
- *nous voulon (nous voulons)*

7. [ɛ] followed by <n> remained [˜ ɛ] and was not lowered to [˜ ɑ̃].

- *prent* does not rhyme with *avant*

8. The spellings <ei> and <e> (see the sound change in feature 2) were at times spelled <ai>.

- *trais (trois)*
- *monaie (monoie)*
- *saient (soient)*
- *otraierent (otroierent)*

9. The endings <-um(s)> and <-un(s)> were used instead of <-ons> in the 1st person plural.

- *donum (donons)*
- *volums (voulons)*
- *nus volun et comanduns (nous voulons et comandons)*

10. [k], spelled <c>, <k> and [g], spelled <g>, replaced <ch> and <j> respectively.

- *camp (champ)*
- *castel (chastel)*
- *cose (chose)*
- *escaper (eschaper)*
- *gardin (jardin)*
- *goie (joie)*

11. <c> followed by <e> or <i> became <ch>.

- *chité (cite)*
- *grache (grace)*
- *merchi (merci)*

12. The triphthong <eau> became <iau>.

- *biaus sire (beaus sire)*
- *les oisiaux (les oiseaus)*

13. Tonic syllables containing [eː] became the dipthong [ei].

- *teil (tel)*
- *sa bontey (sa bonté)*
- *doneir (doner)*
- *l'assembleie (l'assemblée)*
- *son peire (son pere)*

14. Tonic syllables containing [oː] and [o], and initial [o], became [u], spelled <u> and later <ou>.

- *duner (doner)*
- *pur sue amur (por soe amor)*
- *sun seignur (son seignor)*

15. Final [ts], spelled <z>, soon became [s] (In Picard, the grapheme <z> was rare, with 2nd person plural using <-(i)és>).

- *assés (assez)*
- *se vos volés (se vos volez)*
- *vaillans (vaillanz)*
- *vos disiés (vos disez)*

16. Germanic initial <w> was retained instead of becoming <g> or <gu>; <w> could replace initial <v> or <vu> and was sometimes used as an intervocalic glide.

- *warder (garder)*
- *ju wel (je vuel)*
- *wages (gages)*
- *il lowent (il loent)*
- *Willaume (Guillaume)*
- *awoust (aoust)*

17. The feminine article and pronoun "la" became "le", which was not contracted after "a" or "de. "Li" could replace it in the nominative.

- *le contesse (la contesse)*
- *li vostre amie (la vostre amie)*
- *a lequele (a laquele/auquele)*

18. Final <t> was retained. The ending <-eit> was common.

- *volentet (volenté)*
- *tenut (tenu)*
- *portet (porté)*
- *la veriteit (la verité)*

19. <-er> verbs used <-eve> and <-ieve> endings in the imperfect.

Standard OF Imperfect Conjugations	Dialectal Imperfect Conjugations B
-eie	-(i)eve
-eies	-(i)eves
-eit	-(i)eveθ
-iiens	-iiens
-iiez	-iiez
-eient	-(i)event

20. [a] became [ɛ], especially before [ʧ], [ʤ], and [θ], and was written <ai>, sometimes <ei> and <e>.
- *saiche (sache)* - *usaige (usage)*

21. Initial syllables containing <e> becomes <a>.

- *mairdi (mardi)* - *il saront (il seront)*
- *nos davons (nos devons)*

22. “Lo” and “lou” replaced “le” as the masculine definite article.

- *Crucifie lou ! (crucifie le)*

23. Tonic syllables containing [ɛ] became [a].

- *farme (ferme)* - *je promat (je promet)*

24. <-al> and <-able> became <-aul> and <-auble>.

- *especiaul (especial)* - *honorauble (honorable)*

25. <ei> became <oi>, even before a nasal.

- *soignor (seignor)* - *poine (peine)*

1.4 Choice of Dialects

The Old French dialects selected as a basis of comparison are Picard and Champenois. As previously discussed, the north of France had a high level of prestige during the Middle Ages, both culturally and linguistically, which indicated that Picardy would have been an important regional centre. Furthermore, Einhorn notes that there are multiple differences in distinguishing features between Picard and Anglo-Norman (see Table 1) which has the potential to help with separating out Picard forms when there is etymological doubt (a common occurrence when examining Old French loanwords). Such importance and distinctiveness make it likely that Picard left some sort of mark on the words used just over the Channel.

Focusing on Champagne, the region played an integral role in trade, politics, and culture, which had a strong impact on its language. Ayres-Bennett observes that 'while Champenois is essentially a central dialect, which therefore has much in common with Francien, it also shares features with its neighbouring dialects to the north-east (Walloon), east (Lorraine), and south-east (Burgundian)' (1996, p. 69). Such sharing of features is typical of dialects in close proximity; however, the prestige of Champagne maintains the distinctiveness of the dialect and makes it an interesting dialect to analyse.

1.5 Benefits

This research aims to move beyond the giants of Anglo-Norman and Parisian French to offer a wider linguistic analysis of the influence of Old French lexis on Middle English. It is very easy to fall into the trap that contacts between languages during the medieval period was between two homogenous masses, yet this could not be further from the truth. Dialectal diversity is known to be significant during this period with French not having the single dominant standard that it does today, and the history books do not seem to take this into account. A more nuanced picture of language contact between Old French and Middle English is required, which must by definition take into account the variation within each language.

2 Research Context

2.1 Old French Overview

Before delving into the development of medieval French it must first be established exactly what is meant by the term 'Old French'. Huchon helpfully gives a grammatical reasoning behind his labelling:

> 'La denomination d'ancien français englobe souvent l'état de la langue du IX^{e}siècle au XVI^{e} siècle. On préférera toutefois parler pour le XIV^{e}siècle et le XV^{e}siècle de moyen français, conservant le terme d'ancien français pour l'époque où le français est une langue à déclinaisons à deux cas' (Huchon, 2003, p. 53).
>
> [Its southern counterpart had a surprising uniformity across the region: "The denomination 'Old French' often encompasses the state of the language from the ninth to the fourteenth century. Nevertheless, for the fourteenth and fifteenth centuries, it is better to use the term 'Middle French', reserving the term 'Old French' for the period where French distinguished between two grammatical cases] (own translation)

Such grammatical variances are of little importance for our purposes, but the period that Huchon defines through such differentiation is a sufficient base for further elaboration and exploration.

Much like the popular notion of the origins of the French nation, the French language is traditionally believed to originate from the early 800s with the writing of the Oaths of Strasbourg in 842 (Huchon, 2003, p. 27). However, works written in the vernacular were not commonplace until much later, and even then, the first increase in vernacular writings is found at the end of the eleventh century in the langue d'oc in the south of France, rather than in the langue d'oïl in the north (Lodge, 1993, p. 110). It is during the twelfth century that vernacular literary works appear with any frequency in the langue d'oïl, and "only in the thirteenth century is French used in written prose and in non-literary as well as literary texts" (Lodge, 1993,

p. 107). The reason behind this spread in vernacular writing from south to north is due to the growing literary interests of the aristocracy: "They found in the courtly literature which originated in the langue d'oc an expression in the vernacular language of the earthly values of chivalric fame and physical pleasure they were beginning to adopt" (Lodge, 1993, p. 109). This literature that the aristocracy was so fond of was designed for oral performance, however, with specific stylistic features that Latin and other languages could not accurately capture- in other words, "a vernacular performance required a vernacular script" (Lodge, 1993, pp. 108-9).

One of the most marked differences in the use of the vernacular in the langue d'oïl, compared to in the langue d'oc, is that there was a high degree of regional variation from the outset[3] (Lodge, 1993, p. 113).

> 'Du point de vue linguistique, ce qui frappe dans l'occitan des troubadours, c'est qu'il présente, dès ses premières manifestations, c'est-à-dire dès le XI^{e}siècle, une assez grande unité : les différences dialectales y sont en effet minimes et sans aucun rapport en général avec la provenance dialectale du troubadour : l'idiome est sensiblement le même du limousin jusqu'à la Méditerranée' (Bec, 1967, p. 69)
>
> [From the linguistic point of view what is striking in the Occitan of the troubadours is that it presents a reasonable degree of unity from its earliest manifestations onwards, that is from the eleventh century; dialectal differences are in fact minimal and bear no relationship with the dialectal origin of the troubadour; the language is perceptibly the same from the Limousin to the Mediterranean] (trans. Lodge, 1993, p. 111)

Despite the influence of the southern troubadours on the development of the courtly literature in the north that acted as a means of disseminating vernacular writings, this lack of a uniform variety across the region allows us to better examine the specific regional varieties in the langue d'oïl.

In discussing the state of Old French around 1200, Einhorn observes that the dialect of Paris was "only one of many competing dialects in northern France for in the passage from Latin to Old French local differences had developed, slight or more marked, merging or overlapping into neighbouring regions" (1974, p. 135). These varying features are more on the level of phonetics as opposed to lexis or syntax, as it has been observed that a lot of the dialects had a shared vocabulary (Huchon, 2003, p. 60). What is particularly noteworthy is that a number of these regional varieties gained quite high levels of prestige as the wealth and influence of their population centres grew (Lodge, 1993, p. 98). With Paris having not yet grown to the position of cultural and economic dominance that it has today, the status of places like Picardy and Champagne was considerably higher than in more modern times, and this is reflected in textual evidence:

> 'The earliest twelfth-century texts contain numerous Norman or western features (reflecting the literary influence of the Plantagenet court). The famous romances of Chrétien de Troyes (c.1180) contain Champenois features. Many literary texts composed in the thirteenth century reflect the linguistic usage of the great Picard towns. Interestingly, it is only in the thirteenth century that we begin to find vernacular texts written in the Paris region' (Lodge, 1993).

Economics was the reason for the rise in status of the northern towns around Picardy, and it was also the cause of their decline during the fourteenth century. Furthermore, the annexation of Normandy and Anjou at the beginning of the thirteenth century led to their regional features falling out of favour (Lodge, 1993, p. 132; Nezirovic, 1980, pp. 183-5). Paris had been growing in both economic and political influence since the end of the eleventh century, due to its prime location surrounded by excellent agricultural land and the easy access to the city afforded by the Seine (Lodge, 1993). There were other factors beyond its position that contributed to Paris' rise in status: the expansion of royal power, the centralisation of administrative power, the creation of its reputation as an intellectual hub (Huchon, 2003, p. 61). All of these developments helped to establish Paris as the crucible where a more standard French began its construction.

2.2 French Influence on Middle English

The history of French influence on English does not begin in 1066, but well before, due to contact between Norman and Anglo-Saxon rulers; Ethelred II even sought protection in Normandy to escape Viking raiders (Blake, 1992, p. 423). After 1066 is when things step up a gear in terms of greater contact between the two languages. However, post-Conquest, the evidence of English usage is very little, and mostly anecdotal (Blake, 1992, p. 424). An important consideration regarding the Conquest itself is that William's invading force was not a monolithic horde of Norman speakers; other languages and dialects such as Breton and Picard would have been well-represented (Rothwell, 1998, pp. 149-50). Despite the fact that subsequent settlers were mostly Norman, it is important to remember that overgeneralisation of dialectal variation during this period can result in an inaccurate linguistic overview.

A major shift in the usage of French in England comes relatively quickly after the Conquest, as French gradually fell out of usage as a vernacular in favour of English. The decline of Anglo-Norman began soon after the Conquest in certain areas and was only accelerated when Normandy was lost to France during the reign of King John (Blake, 1992, p. 427). However, as the prestige of Anglo-Norman was on the wane, the rise in prestige of Paris and Parisian French grew. It was seen as the 'langue du jour' and was the key to social advancement, as evidenced by the development of grammars and word lists to assist in the acquisition of the fashionable continental dialect (Blake, 1992, p. 423; 427).

Before the influence of Paris raised the prestige of its dialect, Continental French gained a foothold in England through Henry II's marriage to Eleanor of Aquitaine:

> 'First, the king's marriage to Eleanor of Aquitaine brought a new influx of courtiers just at the time when the descendants of the first Norman French invasion were becoming totally assimilated. Furthermore, her patronage of French literary production and the prestige of the new literary forms evolving in France, the romance and lyric poetry, encourages the use of French in the literate segments of society' (Kibbee, 1991, p. 14).

Such patronage resulted in an explosion of literary compositions that began with Henry II and did not fade until the death of Henry III. Anglo-Norman culture had a thriving native literary tradition, which provided a ready market for the Continental French works. Despite the separate development of the two traditions, both Anglo-Norman and Continental French works shared similar themes and language (Kibbee, 1991, p. 18). Therefore, the incursion of Continental French into Anglo-Norman literature was aided by cultural similarities, resulting in the phasing out of the former leviathan.

Contrasted with the new sophisticated French of the court, the speakers of Anglo-Norman began to feel that their language was inferior, which no doubt contributed to its early decline (Kibbee, 1991, p. 24; Burnley, 2000, pp. 66-7). However, as English began to reclaim its lost territory during the fourteenth century, the effects of all manner of French varieties would be left on the language.

2.3 Loanwords

Early Middle English was a time of great expansion in the lexicon. Of the 60,000 lexemes recorded in the MED between 25% and 30% of them are loanwords, taken from Latin, Old French, and Old Norse for a variety of purposes, such as filling lexical gaps due to the period's technological and societal upheaval (Brinton & Bergs, 2017, p. 173). An important factor to consider when looking at information concerning loans from French and Latin is that it is often challenging (or even impossible) to differentiate Latin and French loans; at times, the Latin and French spellings are identical, whereas at others, the Latin spellings were changed due to Middle English conventions. Brinton and Bergs cite 'allegory' (ME 'allegorie' < Latin *allegoria*) and 'desk' (ME 'deske' < Medieval Latin *deska*) as examples of this (2017, p. 173).

What we do know about French loanwords in English begins prior to the Conquest, with words pertinent to the nobility: 'prūd' (valiant), 'castel' (castle), 'gingifer' (ginger), and 'capun' (capon) (Blake, 1992, p. 429). Subsequent borrowings are traditionally divided into two waves: the first arriving with the Normans, and the second coming from Continental or, more specifically, Parisian French (Blake, 1992, p. 426). More recently, Durkin acknowledges that this view of the period is too narrow, and the English/French/Latin trilingualism of medieval England complicates this rather neat perspective (2014, p. 229). However, earlier borrowings from Norman are noticeably different from later ones, mostly due to spelling differences (Blake, 1992, p. 430), so this distinctiveness of Norman is useful to keep in mind when examining the circumstances of loanwords.

The division between earlier and later loans is also exemplified by an increase in borrowing from 1250 into the fourteenth century, during which time the language of writing in England gradually changed to English. The prestige of the semantic categories of these loans is similar in places to those of the earlier Norman words, but more have appeared over a broader scope:

> 'These later loans occur in many additional semantic areas where French was prestigious, such as administration, fashion, social life, food, medicine, and learning, but also in core areas of everyday life… French loans are, however, rare in shipping and seafaring, as well as in farming, which possible reflects the lower prestige of these fields' (Brinton & Bergs, 2017, p. 174).

When discussing lexical borrowing it is easy to gloss over the entire country assume that everyone everywhere experienced the same depth of borrowing. Brinton and Bergs remind us that 'not all Middle English dialects nor their speakers experienced language contact to the same extent' (2017, p. 166). This idea links in with the domains in which French loanwords are found; you are more likely to encounter French if you are wealthy and educated, and less likely if you are a member of the peasantry.

Relating to the prestigious use of Continental French, Brinton and Bergs chose to distinguish Law French as a separate variety existing from the thirteenth century dedicated solely to being used in court and in records (2017, pp. 186-7). The strength of their argument for differentiating this usage as a separate

variety is not entirely convincing, but it is important to remember the diversity of roles that French played and the effect that can have on the language.

2.4 Anglo-French Relations

The relationship between England and France underwent many fluctuations throughout the medieval period, swinging from close allies to bitter enemies on multiple occasions. The very nature of this relationship, however, came to be redefined during the centuries in the lead up to the thirteenth century:

> 'One feudal and political, the antagonism between [England and France] now became primarily a commercial one, in which the stakes were the wine trade of Gascony and the woollen manufacturers of Flanders, which were dependent of English raw wool. A conditional factor was sea-power or maritime supremacy in the Bay of Biscay and the English Channel, then often called the 'Narrow Sea'. A glance at the map will show that the bond of connection between England and Flanders and between England and Gascony was the sea' (Thompson, 1960, p. 55)

There are multiple facets to appreciate, with the most important being the growing mercantile echelon of society acting as a stabilising factor in the intense squabbling between powerful domains. Thompson goes into great detail of the precise nature of these interactions and their impact on the language: the role of English wool was integral in forging trade routes from Yorkshire to London and Dover, before continuing on to Bruges to fuel the renowned Flemish clothmakers. This vital relationship with northern regions of France resulted in a greater presence of Picard forms in treatises from this period. Furthermore, the nobility's love of wine maintained the importance of the English lands in Gascony, thus the same usage of Gascon forms has been noted in trade documents (Thompson, 1960, pp. 61-80). Gascon is less relevant for our purposes, but the role of Picard in England is extremely promising in looking for any fingerprint it leaves on Middle English.

Trade was not the only reason for contact between England and France during the Middle Ages. Diplomatic missions, the ability to study at French institutions, and also military expeditions increased the likelihood of people picking up words from the continent to bring back home with them (Rothwell, 1998, p. 144). Such 'cross-Channel traffic' embodies the image of how contact between countries results in lexical borrowing.

2.5 The Importance of Scribes

When examining medieval texts, especially manuscripts, there are numerous considerations to be aware of. The most relevant for our purposes is the importance of the scribe, as "le personage important… est celui qui a tenu la plume" [the most important person is he who held the pen] (Monfrin, 1968, p. 33). With copies of each text only being possible by writing them out by hand, the scribe is in a unique place to add their own flavour to a text by using their own language and spellings. There are three possible transmission outcomes from exemplar to copy:

> 'A: [The scribe] may leave the language more or less unchanged. This appears to happen only somewhat rarely.

> B: He may convert it into his own kind of language, making innumerable modifications to the orthography, the morphology and the vocabulary. This happens commonly.
> C: He may do something somewhere between A and B. This also happens commonly.'
> (Benskin and Laing, 1981, p. 56; cf. McIntosh, 1973, p. 61)

For translated texts, the most relevant outcome is (A) since that would best preserve any original dialect forms from the source text. However, not all of the texts are translations, so this preference is not universally applicable.

The possible variation in transcription is well-understood by those who study manuscripts, but it is important to note that 'writers and scribes in England did not live in total isolation, fixed in one spot from cradle to grave, cut off from all contacts and influences outside their place of residence' (Rothwell, 1998, p. 155). Where they are from is not necessarily where they learned to write, and who they learned to write from was potentially from a different place altogether, as well as any other possible contacts and influences that must be considered when examining spellings or other variances in textual transmission.

The effect that this has on loanwords is staggering, especially during the Middle English period. Due to the lack of a standard language form, one word can have any number of spellings for any number of reasons. And such variability is cumulative, resulting in the fact that 'loanwords do not inevitably or quite so obviously bear the stamp of their originary situation on them' (Dance, 2014, p. 171). Filtering through these layers can be impossible at times, but they each play a role in charting a word's history.

2.6 Things to Consider

Whilst it would be ideal to take every possible angle of lexical borrowing into account, this research must limit its focus to specific aspects. However, there are some considerations that must be acknowledged. First, there is the role of the MED and the Anglo-Norman Dictionary (AND) in widening the accessibility of the study of Middle English:

> 'Over the last forty years the steady publication of the voluminous MED has profoundly altered the whole landscape of English etymology, whilst the appearance of the first edition of the AND has made possible at least a preliminary understanding of the presence in medieval England of a French that often differs considerable both in form and, more importantly, in semantic content from what were until recently accepted as the norms of standard Central French' (Rothwell, 1998, p. 145).

Despite the wide-ranging benefits of these resources for enhancing our ability to study such historical languages, the very nature of historical dictionaries means that they are restricted to using literary texts as sources, thus resulting in an underrepresentation of non-literary texts (Brinton & Bergs, 2017). The language represented in these dictionaries could be seen to lack the more down-to earth and less florid language of treatises and trade documents, for example, but literary texts are often easier to find and thus make more readily-available sources for these dictionaries. Nonetheless, the level of language represented in the MED must be remembered when using it as a source of data.

Another element detailed by Rothwell whose importance in lexical borrowing cannot be overstated is semantics. Differing meanings between Anglo-Norman and Continental French words is one thing, but

when these words are borrowing into English, there is often a discrepancy between the spelling and sense, with one having come from a different side of the Channel from the other (Rothwell, 1998). As integral as it may be to examining Middle English loanwords, the semantics of the loans is too much of a wild card to do proper justice to in this research and is worthy of further consideration in relation to spelling in additional work.

Even the very definition of a loanword must be acknowledged in order to clarify the focus of this research. Brinton and Bergs state that 'there is no unequivocal way of deciding when a lexical item from one language that is used during discourse in another language – whether by a single speaker, or repeatedly in a community – should be considered a loanword', in addition to distinguishing between loans that fill a semantic gap in the target language from those that become productive (Brinton & Bergs, 2017, p. 111). All of these notions are worth considering but are not entirely relevant for this research. The importance is not placed on the productivity of the forms selected, nor making a significant distinction between whether a form is used by one person, a town, or the entire country. If the relevant form is cited and used in a deliberate way, then it is worth studying, no matter if it only cited once; any and all uses are significant and worth consideration.

3 Stage 1: MED

3.1 About the MED

The MED is one of three legs that form the trivium of the Middle English Compendium, with the other two being the MED Bibliography and the Corpus of Middle English Verse and Prose. The Bibliography lists all of the source texts used in the MED, greatly expanding on the citations given in the dictionary itself. The Corpus is not an actual corpus, but a collection of searchable texts, useful for examining collocations and the context of words and phrases.

The formulation of the MED began in 1925, drawing together around three million quotations from primary sources dating from between 1175 and 1500. It is widely regarded as the most comprehensive evaluation of Middle English vocabulary thanks to its inclusion of all types of evidence for a form or lexeme. The online edition – first published in 2000 and since revised – allowed the easiest access to the dictionary and its data and is the source of the initial data in this research.

3.2 Purpose

This first portion of this research is dedicated to gathering the base set of lexemes known to derive from some form of Old French and their variant spellings that could possibly have originated in Picardy or Champagne. To identify these relevant variant spellings, the descriptions of the dialects' distinguishing features as given by Einhorn are used to pinpoint spellings that could show evidence of those dialectal features.

3.3 Method and Analysis

The data from the MED was gathered in two steps: the first, by manually recording each entry that fit the search parameters, and the second by searching through this initial data wave for any of the features

belonging to Picard and Champenois as indicated in Table 1. This method was effective for reducing the thousands of Old French-derived lexemes to the fifty-two discussed in this chapter.

The MED includes multiple possible filters through which to examine its entries, giving you the option of limiting your search by subject, part of speech, and source language. Specifying a period was not available nor required due to the specialised nature of the MED (as indicated in the name). It did not seem relevant to restrict the search by subject since semantics had not been selected as a factor, but since open-class words are the most frequently borrowed, focusing on nouns, verbs, and adjective was logical. Furthermore, such constraints provided a broad spectrum of possible loans whilst eliminating closed-class words that were less likely to contain interesting features (Haspelmath & Tadmore, 2009).

The source language filter was more complicated, however. The MED lists 12,063 words as being derived from Old French, which is too many to analyse for this research. By searching for forms derived at least in part from Old Northern French (ONF) and Continental French (CF), the number of solely Anglo-Norman words was reduced, as was influence from Latin, Occitan, and Germanic languages beyond English. As a result of these filters, the total number of potential lexemes became a much more manageable 176.

The second stage of MED analysis consisted of searching the variant spellings of the chosen lexemes to find any evidence of the dialectal features associated with Champenois and Picard. Any dialectal remnants would have been evidenced in spelling, so it was the multitude of variant spellings for each lexeme that were examined, with reference to the etymologies given in the MED for the 'standard' Old French equivalents. All potential features were considered, both those identified as weak (one x) and strong (two x). Therefore, the selected features were:

7. [ɛ̃] (+*n*) stays [ɛ̃]
10. [k], [g] for *ch, j*
11. *c* (+*e, i*) > *ch*
12. *eau* > *iau*
13. [eː] > *ei*
15. Final *z* > *s* 12th c.
16. Use of *w*
18. Final *t* remains
20. *a* > *ai*

Despite '17. *la* > *le*' being a strong feature in Picard, because it refers to the definite article it was excluded from the list.

If the lexeme and/or its variant spellings and etymology indicated the potential presence of one of the above features, they were recorded in a spreadsheet along with the date of first citation as noted in the MED and the possible feature or features evidenced. To facilitate easier application to LAEME search parameters, the present-day English equivalent (or most modern version) of the lexeme was also recorded, using the links to the Oxford English Dictionary provided in each lexeme's MED page.

An example of this process can be illustrated with 'sc$_{\breve{o}}$rnen'. The MED gives multiple possible etymologies for this word: OF 'escharnir', Anglo-French (AF) 'charnir', AF/ONF 'escarnir', and ONF 'eskarnir'. These potential sources do not agree on whether the first consonant is [k] (spelled <c/k>) or [ʃ] (typically spelled <ch>), which is itself evidence for feature 10. This same discrepancy is noted in the variant spellings listed in the MED: 'scorn(e)', 'scoren', 'skorn(e(n))', 'schorne(n)', 'scorni(e)', 'skorni',

'scarne(n)', 'skarnen', and 'scoarnen' are just the possible ways of spelling the infinitive. Based on the feature description given by Einhorn and the etymology given in the MED, it can be determined that the spellings featuring <c/k> were evidence for this feature being borrowed into English and these forms warranted further investigation.

3.4 Data

The effect of both stages of analysis resulted in whittling down the entire MED to a total of fifty-six lexemes, whose variant forms clearly exemplified just three OF dialectal features: numbers 10, 11, and 16[4]. Forms featuring <w> where more standard OF forms use <g> (16) was the most common feature with 39 instances, with orthographic representations of [k] and [g] instead of <ch> and <j> (10) making up the majority of the remainders with 17 occurrences. There are just two instances in this data set of <c> becoming <ch> when followed by <e> or <i> (11), but those familiar with modern French dialects and Dany Boon films would recognise this feature as still being present in the 'ch'ti' of Picardy, so it is important to include. Two of the lexemes featured evidence for two features: 'wiket' and all its variant forms are examples of both features 10 and 16, as is 'wāğŏur'.

I also chose to organise the earliest possible dates of first citation of each lexeme in the data set in chronological order. For a large number of manuscript sources, dating can be more of an approximation, hence the earliest possible date of composition is used. The results displayed in figure 1 indicate that the majority of loanwords in this data set were borrowed between 1200 and 1400, with only a few words appearing outwith these boundaries. The absolute peak of this borrowing was during the fourteenth century, which confirms existing research which concludes that that the majority of OF loanwords came into English during the switch in bureaucratic language from French and Latin to English.

[4] The full list of lexemes can be found in Appendix A.

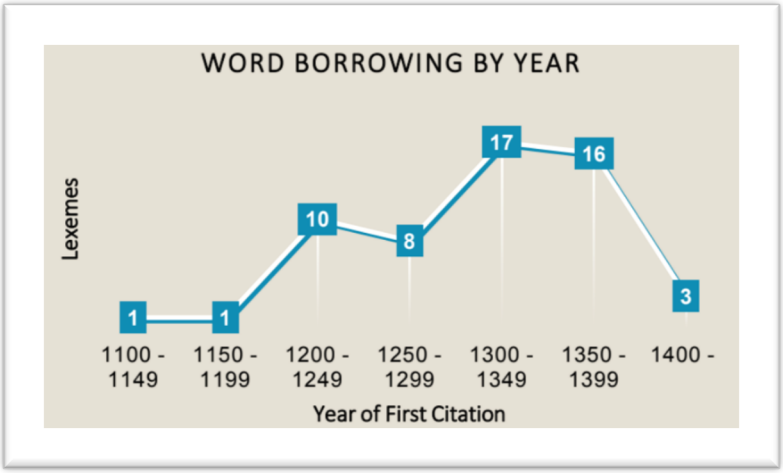


Figure 1: *Number of OF loanwords in this dataset borrowed into English between 1100 and 1400.*

3.5 Evaluation

This is a challenging period in which to examine etymologies due to the sheer amount of doubt and confusion when it comes to etymologies. 'Strācūr' is proposed to have come from an unknown AN or ONF form of the OF 'estrace', but because the MED only has one citation of the word from one source text there is very little to go on when trying to uncover a more solid etymology. Other words, like 'ward(e)' and 'waider' have been determined to be quite resolutely Anglo-Norman in origin, but that does not mean that other forms did not derive from elsewhere. Furthermore, both features 10 and 16 are more well-attested in Picard than in Norman and Anglo-Norman, so it is reasonable to look at Picard origins in these features before examining Norman paths.

When identifying possible OF dialectal features in the variant spellings, the lack of a standard spelling system was both a blessing and a curse. Consonantal features were easy to pinpoint, but due to the variability of vowels, as well as sound changes that have occurred over the years, it was significantly more challenging to find evidence for vowel-centric features. Furthermore, English has no definite way of marking nasal vowels in nativised words, so feature 7 relating to [ɛ̃] was impossible to identify given the constraints of this projects. Given more time and a larger data pool, greater focus on vowels could have been achieved, but the evidence for the consonantal features previously mentioned was solid enough to be sufficient.

Also, whilst it is reassuring that the data reflects previous research relating to when lexical borrowing was at its peak during Middle English, one must also take the linguistic situation in France into account. The decline of the northern towns and rise in status of Paris occurred during the fourteenth century, so in order to increase the likelihood of minimal Parisian influence on the OF dialects earlier loanwords are preferred. However, that does not mean to say that the words borrowed during the 1300s should be discounted entirely- the growing influence of Paris and the desire to conform to an incipient standard should be taken into account for later loanwords.

4 Stage 2: LAEME Forms and Frequencies

4.1 About LAEME

LAEME is described by its creators Margaret Laing and Roger Lass as the 'daughter atlas' of 'A Linguistic Atlas of Late Middle English' (LALME) but is quite different in scope and purpose. It is comprised of a corpus of searchable tagged texts and a searchable database of information about the manuscripts and texts used, in addition to numerous explanatory documents. The corpus contains 650,000 words, where each lexical item and derivation and inflectional morpheme is tagged in incredible detail.

The production of LAEME began after the publication of its predecessor 'A Linguistic Atlas of Late Middle English' (LALME) in 1987. The initial plan of using a questionnaire-based method of manuscript analysis was abandoned in favour of lexico-grammatical tagging, facilitating the creation of the corpus. The data gathered has facilitated greater research into early ME scribal practises and writing systems, as well as providing a large repository of early ME texts that is a lot easier to work with than searching for each manuscript manually.

4.2 Purpose

This second phase takes the lexemes and variant spellings extracted from the MED and applies them to LAEME in order to gather more detail about those forms. More specifically, they can be used to examine more in-depth the uses and distribution of variant spellings in specific texts. Furthermore, the data subsequently gathered on the frequency of these forms across texts provides a basis for identifying the manuscripts of greatest interest for further examination in the next stage of analysis.

4.3 Method and Analysis

The method used for this stage of data-gathering consisted of searching the 'Tag Dictionary' part of LAEME for the present-day English equivalent of each MED lexeme; where possible, the MED gives a link to the Oxford English Dictionary (OED)'s entry on the word. This is not always possible, however, such as with 'scapelen' and 'waider', so these words had to be omitted from the LAEME search. To further complicate matters, the OED had entries for a number of MED lexemes, such as 'warisoun', but LAEME would not accept this as a lexeme — 'garrison' would have to be inputted and forms beginning with <w> were extracted from the variant forms manually.

LAEME offers a wide scope of criteria for searching the tag dictionary, such as using suffixes, grammatical words, numbers, and other types of tag as the search string, in addition to restricting the scope of the search to the beginning, middle, end, or entirety of the 'lexel'. In this instance, using a lexical string and broadening the scope to the entirety of the lexel produced the best results.

Each variant spelling for the lexemes was recorded in a spreadsheet along with their frequency and the number of texts the forms appear in. LAEME is incredibly thorough when it comes to grammatically tagging each form, distinguishing between nouns and nouns that function as direct objects, even going so far as to note which forms appear in a rhyming position. Such a high level of detail can be useful when focusing on specific spellings, but for a wider overview, the grammatical functions were reduced to their broader categories as recorded from the MED. This resulted in some forms being repeated within the wider

word classes, so these duplicates were removed in order to simplify the resulting data.

4.4 Data

A surprisingly large number of words taken from the MED were not recorded in LAEME - just over half of the fifty-six words extracted. The result was a grand total of twenty lexemes identifiable in LAEME, of which four were different grammatical classes of two words: ‘warrant’ exists in this data set as a noun and a verb, and ‘waste’ as a noun and an adjective. Without levelling the duplicate spellings resulting from the broadening of grammatical categories, those twenty lexemes have one hundred and thirty-four forms between them; with levelling, that number reduces to ninety-five. The words with the highest number of variant spellings are ‘wait’ and ‘war’, which is not surprising given the period in question. To limit confusion, the total number of forms without duplicates is used for all further discussion.

Table 2: *Highest frequency spelling variations of lexemes in this dataset across texts in LAEME.*

PDE	Form	Freq.	Texts
war	peorre	14	11
warden	pardeins	11	3
waste	pesste	11	3
snake	snaken	8	7
war	werre	8	4
wain	waʒʒn	5	1
wait	wayteþ	5	4
warden	pardains	5	2
waste	peste	5	4
waste	peste	5	4
wile	piles	5	4
warnish	warniste	4	3
war	wer	4	4
war	weorre	4	3
war	perre	4	3

In terms of frequency of forms, the vast majority of spelling forms identified are only found once in one text, according to LAEME. Fifty-eight forms have just one citation in one text, with a further seventeen appearing twice across one or two texts. The fifteen highest frequencies are listed in Table 2, with the entire list given in Appendix B.

4.5 Evaluation

First of all, the discrepancy between the number of lexemes identified in the MED and in LAEME is quite dramatic but easily explained. The MED aims to be comprehensive in its study of ME, which by nature involves recording words which possibly have only one citation in one text. By contrast, LAEME's focus is on the texts themselves rather than the breadth of words covered in the corpus, so it is logical that there are words not found in their manuscripts that have evaded tagging. The early ME period is notorious for being hard to quantitively study due to the availability of material, so the discrepancy between the MED and LAEME's sources is to be expected.

The number of forms with just one citation is also to be expected due to the nature of ME. With the lack of a standard spelling system, countless spelling variations are possible. However, not all forms are made equal. In the steady creep toward standardisation, some spellings are more commonplace than others, as indicated by the high frequency of 'peorre' as a form of 'war' across texts shown in Table 2. Just as some spellings are preferred, others exist but are more peripheral for any number of reasons. It is reasonable to assume that 'uuerre' is rare because of its orthography, and the quirky 'peorrre' because of its potentially erroneous triple <r>.

Furthermore, the fact that LAEME distinguishes forms that appear in rhyme position from those that do not could be thought to affect the number of one-time spellings. The manipulation of spelling in order to fit a certain rhyme scheme is not unheard of during the ME period, but it does not seem that the role of rhyme is significant in this instance. Eighteen forms are described as occurring in rhyme position, yet they are relatively evenly distributed between the highest and lowest frequency forms. This indicates that the significance of rhyme in affecting spelling is relatively low for our purposes and can therefore be discounted as a major factor to consider. If the purpose of this research were to be analysing the variant forms identified in the MED and LAEME in and of themselves, these orphan spellings would be discounted. Nevertheless, by grouping these forms by text, they can play a more important part in identifying larger patterns within texts as a whole.

5 Stage 3: LAEME Texts

5.1 Purpose

The final and most lengthy portion of the analysis puts the word and feature lists from the previous chapters into context and examine their citation texts more fully in order to determine the likelihood of these variant spellings deriving from French dialects that are not AN or CF. Through using Einhorn's dialectal feature table, in addition to taking each texts' metadata into account, the aim is to identify the most prominent OF dialect in each manuscript (if any) and examine the possible reasons behind this, whether more general, such as relating to culture or historical events, or more text-specific, such as the idiosyncrasies or personal connections of the scribe or author.

5.2 Method

The initial identification of each form's text(s) involved creating an item list through LAEME, which listed the identification number of each text the feature in question appeared in. By putting all of these numbers for all of the features in numerical order, the texts with the highest total of relevant features were easy to determine. However, each text does not necessarily correspond to an entire manuscript- where a manuscript

is known to have been written in multiple hands, the manuscript is often split up with each text entry in LAEME consisting of the portions of that manuscript written in a specific hand. In order to gain the fullest picture, where a manuscript was divided into multiple searchable texts, each text was examined, regardless of whether that specific text featured one relevant spelling or eight. Grouping together related texts made it significantly clearer how many manuscripts were to be examined and provided the broadest perspective of a longer manuscript within the confines of this research.

As with the previous stages of analysis, only the most relevant texts were selected for further examination; manuscripts (both those consisting of multiple tagged texts and those not) containing two or more interesting spellings made the cut, as did a couple – such as Egerton 613 – whose metadata hinted at noteworthy ties to France. After discounting the surplus texts, forty tagged texts remained, deriving from nineteen manuscripts.

Each tagged text was subject to more direct searching for further indications of OF dialectal spellings using Einhorn's feature list as a guide. Examples of these indicative spellings include:

- Initial <gu>, to rule out the use of <w>
- Final <ez>, to rule out feature 15
- <che> and <chi>, to test for feature 11
- <eau> and <iau>, to test for feature 12

By examining the number of such spellings and features within each text as a whole, it is possible to attest more solidly whether these spellings indeed derive from Picard or Champenois or another dialect entirely.

5.3 Data

Of the forty texts examined, the vast majority had very little in the way of relevant forms or even French-derived vocabulary. However, five texts featured notable forms and other details that make them worthy of further elaboration.

The first text comes from MS Laud Misc. 108, which was written circa 1300 in Oxfordshire and contains rubrics written in AN and Latin in addition to the ME passages tagged in LAEME. This text is notable because of the clear French influences in the spelling; 'beau-frere' features the standard spelling of 'beau', indicating a lack of <iau> forms found in Picardy (see feature 12), as well as the clear presence of final <z> further discounting Picard – and possibly AN – as an influential dialect (see feature 15). These terminal <z> spellings are particularly interesting because a number of them are found in the French first-person plural ending <-ez> but applied to English contexts, such as in 'we ne findez nouȝt' and 'mani men þinchez'. The first sentence shows the correct person being used for this particular conjugation, however the second shows <-ez> being used with the third-person plural. This could be interpreted as an overextension of <-ez> usage in an attempt to make the language of this text seem more ornate and 'French', but such a hypothesis would require further examination. What is clearer, however, is the definite OF influence in this text, despite the inability to pinpoint precisely which dialect was most significant.

The first LAEME text from the MS of *Cursor Mundi* housed at the Royal College of Physicians is an excellent example of a mixed palette of dialectal features that make concrete conclusions challenging. The text is peppered with overtly French words like 'danais' and 'delices', and the preference for French forms of nationalities and countries is notable. The <ch> in 'da(n)emarche' could be read as [k] in other contexts, but given the preference for French forms elsewhere, this spelling should be taken as evidence

against feature 10, thus side-lining Picard and AN. However, the spelling of 'curtaisi' muddies the waters; it features the classic AN <u> (see feature 14) but the <ai> is most likely a Walloon spelling (see feature 8). Such a mix of dialects is in itself interesting, however, and definitely indicates a strong French influence in this text.

A further example of a healthy variety of OF dialectal features is the copy of 'Ancrene Riwle' found in Cotton Titus D xviii. The only uses of <gu> spellings are in the Latin portions of the texts, suggesting that <w> forms predominated (see feature 16). There is also no evidence for features 11 and 12, which are strong indicators of Picard or AN forms, as is feature 10, which is hinted against thanks to spellings like 'pa{-}t(r)iarches'. The verb 'auez' is present, however, exemplifying feature 15, which further counts against Picard and AN but does indicate a French influence. The word 'culuert' could provide a clue, but the MED gives the etymology as just deriving from OF with no further details. The spelling looks to indicate the presence of <u> forms (see feature 14), which would contradict the lack of AN spellings elsewhere in the passage. All in all, it is a rather confusing picture that is painted, but OF does affect at least some of the forms used.

Digby 86 is worth an honourable mention, as it is a manuscript in which half the material is in French. With other manuscripts containing other languages, LAEME's transcriptions skip over the French (but the Latin is often included), but in this instance some of the French has been transcribed and was searchable as a part of the English portions of the manuscript. There was no evidence in either OF or ME of the <che> and <chi> forms indicative of feature 11, and the presence of both terminal <z> and <eau> forms in the phrase 'Les diz de seint bernard / comencent Ici tresbeaus' suggest that the OF of the manuscript is not Picard. Unfortunately, very few of the interesting features of the French carry over to the English passages, but the ability to perform a direct comparison in situ without having to resort to the manuscript facsimile was noteworthy.

5.4 Arundel 57

Despite all these crumbs hinting at the presence of OF dialects, the strongest evidence for their presence comes from Arundel 57, otherwise known as 'Ayenbite of Inwyt'. The manuscript was written at St Augustine's in Canterbury and is known to have been completed on 27th October 1340 (Gradon, 1979, p. 1). The text is the only known ME translation of the OF 'Somme le Roi' – which was originally composed around 1280 for the children of Philip III of France ('Laurent D'Orléans, 'La Somme le Roi,'' n.d.) – and was written (and possibly translated) by Dan Michel, originally a secular clerk who became a priest at St Augustine's in 1296. The abbey is known to have had two copies of 'Somme le Roi', of which one was owned by Michel himself and was his most likely exemplar. This copy has not survived to the present day, although the other is found in Cotton Cleopatra A (Gradon, 1979, pp. 53–4). It has been noted that the language of 'Ayenbite' is quite archaic for its period, explained by the fact that Michel was quite an old man when writing it (Gradon, 1979, p. 12). The LAEME transcription of the manuscript also includes many annotations concerning mistranslations, suggesting that OF features would not be direct copies from the exemplar, but in fact subtler influences if not from Michel's own system.

The 'Frenchness' of Michel's spelling has been well-recorded and discussed, most notably with regard to his vowels. His use of <ou> to represent [u] follows the French system (Gradon, 1979, p. 14), and Wallenberg notes that his <i/y> spellings could represent an analogy with French, in addition to representing [je:] (Wallenberg, 1923, p. 121). However, his spellings do not correspond with AN usage. With regard to Michel's use of <ie>, Gradon notes that 'if Dan Michel was familiar, not only with the AN

pronunciation in words such as *chef*, but also with the CF pronunciation, he could just as well have used <ie> as a graph for [ie] as for [e:]˛˛' (Gradon, 1979, p. 32). In addition, Gradon observes that the variation in Michel's use of <o> cannot be explained by aligning these spellings with AN sound change. Instead, CF nasalisation of [a] must be taken into account in order to make sense of some of the uses (Gradon, 1979, p. 40). The result of such confusion is more the ruling out of AN as a source and bringing CF into the fore than providing clear-cut solutions.

The text has clear French influences that move beyond just spelling. An annotation on a use of 'þet' reads 'sic – 'that' for 'than' is a calque on French usage', which implies subtle French grammatical uses. Furthermore, the 'che' in 'che manere' appears to be a variant of *ce*, the French for 'this/that', building on this basic impact of French.

The lack of importance of AN in 'Ayenbite' is corroborated by the lack of the <u> described in feature 14, shown in 'glotounye'. Furthermore, the <ou> spelling corroborates feature 1, which is common to all dialects except AN and Picard. While there may not be a Picard flavour in that particular form, the only example of the <iau> spelling (feature 12) indicative of Picard is found in this text in 'hysiaus'. With further evidence for feature 10 in 'askapie', it cannot be ignored that Picard plays as big a role in the language of this text, if not more.

5.5 Evaluation

Across all of the texts analysed, there was a surprising amount of variation in dialectal forms used; no text favoured solely one dialect. Even with 'Ayenbite', whilst it is possible to conclude that AN was not the majority dialect preferred by Michel, it is not feasible to discern one that was. As evidenced by Gradon's analysis, taking CF into account makes sense of a lot of the vowels, as well as numerous Picard forms being found through this research. Therefore, without more in-depth research, it must be said that the overall use of OF spellings consists of a lot more variation than the historical record suggests.

6 Final Analysis

6.1 Is it possible to identify loanwords into Middle English from Old French dialects other than Anglo-Norman and Central French?

Through the course of this research, it has become clear that the first two questions examined are symbiotic in nature. It is indeed possible to identify variant forms of a lexeme that have been borrowed into English from OF, but such identification is only possible through spelling. Furthermore, these forms appear to be few and far between and when they do exist, they are incredibly difficult to find due to their scattered distribution.

It is impossible to hypothesise a spelling variation's origins from only examining dictionary material, yet it is entirely feasible to take a series of manuscripts or other texts and search for dialectal features without having previously referred to dictionaries. The benefit of having used both the MED and LAEME was that it narrowed down the Brobdingnagian collection of OF loanwords and provided a focus on the forty tagged texts explored. Further research with a previously determined set of texts would not require the initial lexicographical stages.

What is required for identifying loanwords is a reference point of OF dialects, which is where Einhorn's guidelines proved invaluable. It is also useful to have a working knowledge of ME, as certain indicative spellings occur numerous times natively in ME. For example, <chi> was an important feature to search for but occurred most often in 'child' and 'children'. A similar issue was faced with <w>, which is why it became easier to search for instances of <gu> in order to indicate the likelihood of <w> forms. Ultimately, it was the textual context of each potential form that indicates whether it was likely from an OF dialect or a victim of ME's spelling idiosyncrasies.

In terms of identifying a specific dialect of a loanword, however, this was not possible due to the lack of data in the tagged texts. As previously discussed, there was a large amount of variation in features used within texts, with some forms often contradicting each other when it came to suggesting a common dialectal origin. There was greater evidence for Picard forms than those of Champagne, however, most likely due to the greater geographical proximity to Paris resulting in its features being more familiar and widely accepted. Significant further research would need to be undertaken in order to clearly identify such forms' dialect of origin.

6.2 Can the origins of these loanwords be determined through spellings?

This is a challenging period to research because of the general spelling variability on both sides of the Channel. The difference in phonetic inventories between OF and ME means that vowels are particularly susceptible to spelling changes as they make the journey. Variation between scribes, as well as the inherent variability within ME, meant that no overarching conclusions about particular vowel-centric features could be made. It is an area which could provide a great deal of insight with further examination.

As previously mentioned, there were a great deal of spelling features that occurred in ME naturally, but filtering these out was possible, if time-consuming. What posed a greater challenge was the overlap in features, in particular between AN and Picard. While there were features specific to both dialects – such as <iau> for Picard and <u> for AN – they were not always present in the same text. Therefore, being able to distinguish between other northern dialects and AN was not always clear, but reasonable suggestions could be made, particularly in the case of Arundel 57.

6.3 What can these borrowings reveal about cultural links between England and France?

Due to the lack of data resulting from the analyses, there was insufficient evidence to even hint at the reasons behind the forms identified. Those that were present were almost randomly scattered across texts, which combined with a lack of background about the manuscripts and their scribes themselves meant that they revealed nothing about their reason for appearing in their texts. The greater evidence for more Picard forms could be explained by geography, or also the prestige of the region and its proximity to other regional centres and trading hubs, as well as England itself. This is pure speculation, however, with little substantiation to back it up.

Even with 'Ayenbite', where a great deal is known about both the manuscript and its scribe, there were very few conclusions to be drawn. The fact that the origin text 'Somme le Roi' was written for the offspring of the French king makes it unlikely to have ever been written in AN (despite what the British

Library says), yet without the exemplar that Dan Michel used to translate his version, it seems hasty to suggest that his Picard-leaning forms came from this OF manuscript.

Despite the lack of conclusions about why these dialectal forms came into ME, there is still a great deal of potential for further research with other documents whose origins are clearer and more useful.

7 Conclusion

The aim of this research was to bring together multiple resources in order to examine the relationship between medieval France and England on the level of lexical borrowing. The variability of scribal practices during the Middle Ages was the ideal medium through which to identify obscure lexical forms that have otherwise escaped notice. These forms, once unearthed, aimed to provide a gateway into a more nuanced understanding of the nature of links between England and regions of France beyond Normandy and Paris.

The MED proved to be an invaluable resource in narrowing the field to focus on open-class words from specifically the northern portion of France. Its lists of each spelling variation for each ME lexical item and known etymologies were instrumental in creating the initial dataset. Combined with the data from Einhorn concerning the features of each dialect, this dataset was tailored to indicate which features were most likely to be relevant in future stages of analysis.

Once familiarised with, LAEME provided an almost cumbersome amount of detail with regards to frequency of forms and their distribution across texts, building on the MED data. The different sources of the two resources resulted in a marked culling of forms, but the variant spellings given for those that remained proved ample. By organising these forms by text, the analysis of each individual text began, with fewer results than expected. Most manuscripts featured very few words of note, with the exception of Arundel 57, which provided a glimpse into OF usage that was decidedly more varied than solely AN or CF. Still, a large variability in forms remained, which resulted in the inability to draw hard and-fast conclusions.

The ultimate combination of the uncovered variant forms with a cultural reason for their existence was not ultimately possible beyond pure conjecture. A lack of final data, in combination with their wide distribution and the lack of knowledge about manuscripts' origins, made it impossible to focus on a particular lexical source and examine the reasons behind such lexical transmission. The possibility to undertake such research remains, with further extended examination required. The use of a more specific manuscript source – ideally one where the history of the scribe and text itself has survived – or examining non-literary texts would lead to clearer outcomes and be able to further enlighten the ties between England and more peripheral areas of France.

8 References

Anon (n.d.) Laurent D'Orléans, *'La Somme le Roi'*. Available from https://www.bl.uk/collection items/illustration-of-prowess-idleness-david-labour-from-laurent-dorleanss-la-somme-le roi [Accessed 30th March 2020].

Ayres-Bennett, W. (1996) *A History of the French Language through Texts*. London and New York, Routledge.

Bec, P. (1967) *La langue occitane*. Paris, PUF.

Benskin, M. and Laing, M. (1981) Translations and Mischsprachen in Middle English manuscripts. In: Benskin, M. and Samuels, M.L. eds. *So meny people longages and tonges: philological essays in*

Scots and medieval English presented to Angus McIntosh. Edinburgh, Middle English Dialect Project, pp. 55–106.

BL Cotton Titus D xviii, fols. 14r-105r: Ancrene Riwle. In Laing, 2013.

Blake, N. (1992) *The Cambridge History of the English Language: 1066-1476*. Cambridge, Cambridge University Press.

Brinton, L. and Bergs, A. eds. (2017) *The History of English: Middle English*. Berlin, De Gruyter Mouton.

Burnley, D. (1992) *Lexis and Semantics*. Cambridge, Cambridge University Press.

Burnley, D. (2000) *The History of the English Language: A Source Book*. 2nd ed. Harlow, Pearson Education.

Dance, R. (2014) Getting a word in: Contact, etymology and English vocabulary in the twelfth century. *Journal of the British Academy*, 2, pp. 153–211.

Durkin, P. (2014) *Borrowed Words: a history of loanwords in English*. Oxford, Oxford University Press.

Edinburgh, Royal College of Physicians, MS of Cursor Mundi, entry 1. In Laing, 2013.

Einhorn, E. (1974) *Old French: a concise handbook*. London, Cambridge University Press.

Gradon, P. (1979) *Dan Michel's Ayenbite of Inwyt*. Early English Book Society. Oxford, Oxford University Press.

Haspelmath, M. and Tadmore, U. eds. (2009) *Loanwords in the world's languages: a comparative handbook*. Berlin, De Gruter Mouton.

Huchon, M. (2003) *Histoire de la langue française*. Paris, Librairie Générale Française.

Kibbee, D. (1991) *For to speke Frenche trewely: the French language in England, 1000-1600: its status, description and instruction*. Amsterdam, Benjamins.

Laing, M. (2013). *A Linguistic Atlas of Early Middle English, 1150-1325, Version 3.2.* Edinburgh: The University of Edinburgh. http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html (28 March 2020)

Lodge, R. (1993) *French: from dialect to standard*. London, Routledge.

London, British Library, Arundel 57, fols. 2r-4r; 13r-96v: Ayenbite of Inwyt, Pater Noster etc. In Laing, 2013.

McIntosh, A. (1973) Word geography in the lexicography of medieval English. *Annals of the New York Academy of Sciences*, 211, pp. 55–66.

McSparran, F., et al. 2000-2018. *Middle English Dictionary*. University of Michigan Library. https://quod.lib.umich.edu/m/middle-english-dictionary/dictionary (28 March 2020).

Monfrin, J. (1968) Le mode de tradition des actes écrites et les études de dialectology. *Revue de linguistique romane*, 32 (125–126), pp. 17–47.

Nezirovic, M. (1980) Le vocabulaire dans deux versions du Roman de Thèbes. Clermont-Ferrand, *Association des Publications de la Faculté des Lettres de Clermont-Ferrand*.

OED Online. Oxford University Press. https://oed.com (28 March 2020)

Oxford, Bodleian Library, Digby 86, all the texts that are suitable for mapping. In Laing, 2013.

Oxford, Bodleian Library Laud Misc 108, part 1, fols. 1r-200v, hand A: Life of Christ, Infancy of Christ, South English Legendary. In Laing, 2013.

Pope, M. (1966) From Latin to modern French with especial consideration of Anglo-Norman phonology and morphology. 2nd ed. Manchester, Manchester University Press.

Rothwell, W. (1998) Arrivals and Departures: The Adoption of French Terminology into Middle English. *English Studies: A Journal of English Language and Literature*, 79 (2), pp. 144–165.

Thompson, J. (1960) *Economic and social history of Europe in the later Middle Ages (1300-1530*. New York, Ungar.
Wallenberg, J.K. (1923) *The Vocabulary of Dan Miche's Ayenbite of Inwyt*. Uppsala.

9 Appendices

9.1 Appendix One: Full MED Lexemes

Earliest Possible Citation	PDE Equivalent	Lexeme	Part of Speech	Feature
1131	wile	wīle	N	16. Use of w
1160	war	wer(re	N	16. Use of w
1200	scar	scar	N	10. [k], [g] for *ch, j*
1200	warrant	warant	N	16. Use of w
1200	ward	ward(e	N	16. Use of w
1200	warden	wardein	N	16. Use of w
1200	waste	wāst(e	N	16. Use of w
1200	wimple	wimple	N	16. Use of w
1225	ward	warden	V	16. Use of w
1335	wait	waiten	V	16. Use of w
1225	wicket	wiket	N	16. Use of w
1230	scorn	scōrnen	V	10. [k], [g] for *ch, j*
1255	wimble	wimble	N	16. Use of w
1273	-	waider	N	16. Use of w
1275	scald	scalden	V	10. [k], [g] for *ch, j*
1275	ware	wāren	V	16. Use of w
1278	skew	skeu	N	10. [k], [g] for *ch, j*
1287	-	strācūr	N	10. [k], [g] for *ch, j*
1295	scoop	scọ̄pe	N	10. [k], [g] for *ch, j*
1299	snake	snak	N	10. [k], [g] for *ch, j*
1300	scarce	scārs(e	N	10. [k], [g] for *ch, j*
1300	partridge	partrich(e	Aj	10. [k], [g] for ch, j

1300	archer	archēr, -iēr	N	11. c (+e,i) > ch
1300	warray	werreien	V	16. Use of w
1300	warison	warisŏun	N	16. Use of w
1300	warnement	warnement	N	16. Use of w
1300	waste	wās̆t(e	Aj	16. Use of w
1303	wardecorps	warde-cors	N	16. Use of w
1318	warnestore	warnestōr(e	N	16. Use of w
1328	-	waspail	N	16. Use of w
1330	warnish	warnishen	V	16. Use of w
1330	warish	warishen	V	16. Use of w
1330	waynpain	wain-pain	N	16. Use of w
1333	wain	wain	N	16. Use of w
1335	were	wēr̥(e	N	16. Use of w
1338	warnison	warnisŏun	N	16. Use of w
1338	-	warnisŏur	N	16. Use of w
1350	kennet	kenet	N	10. [k], [g] for ch, j
1350	-	werpishen	V	16. Use of w
1350	wager	wāğŏur	N	16. Use of w
1350	waynoun	wainŏun	N	16. Use of w
1359	wode	waid(e	N	16. Use of w
1364	-	scapelen	V	10. [k], [g] for *ch, j*
1373	warence	warance	N	16. Use of w
1373	wasp	wasp	N	16. Use of w
1375	wallop	walop	N	16. Use of w
1376	wernard	wernard	N	16. Use of w
1378	wardrobe	warde-rōbe	N	16. Use of w
1382	wage	wāğe	N	16. Use of w
1387	scarcity	scās̆rsetē	N	10. [k], [g] for *ch, j*
1387	botch	bocche	N	11. *c* (+*e,i*) > *ch*
1390	warrant	waranten	V	16. Use of w
1391	strick	strik(e	N	10. [k], [g] for *ch, j*
1400	tuck	tukken	V	10. [k], [g] for *ch, j*
1400	warrok	warroken	V	16. Use of w

1404	skellat	skellet	N	10. [k], [g] for *ch, j*

9.2 Appendix Two: Full LAEME Lexemes

No.	PDE	Form	Freq.	Texts
1	war	þeorre	14	11
2	warden	pardeins	11	3
3	waste	þesste	11	3
4	snake	snaken	8	7
5	war	werre	8	4
6	wain	waȝȝn	5	1
7	wait	wayteþ	5	4
8	warden	pardains	5	2
9	waste	peste	5	4
10	waste	þeste	5	4
11	wile	piles	5	4
12	warnish	warniste	4	3
13	war	wer	4	4
14	war	weorre	4	3
15	war	perre	4	3
16	wain	wayne	3	1
17	wait	paiteden	3	2
18	warden	pardein	3	3
19	waste	west	3	3
20	war	porre	3	3
21	scarce	scarse	2	2
22	scorn	scornunge	2	1
23	scorn	scarned	2	2
24	snake	snakes	2	2
25	wain	wayn	2	2
26	wain	waine	2	1
27	wait	waite	2	2
28	wait	waites	2	2
29	warrant	waraunt	2	2

30	warrant	parant	2	2
31	waste	weste	2	2
32	were	pere	2	1
33	war	worre	2	2
34	war	were	2	2
35	war	werres	2	1
36	wimple	wimpel	2	2
37	wimple	pimpel	2	2
38	archer	archer	1	1
39	partridge	pertrich	1	1
40	scald	scalden	1	1
41	scald	scoldeÞ	1	1
42	scald	scaldant	1	1
43	scorn	scornige	1	1
44	scorn	scærninge	1	1
45	scorn	schornigis	1	1
46	scorn	schorningis	1	1
47	scorn	scorneÞ	1	1
48	scorn	scornes	1	1
49	scorn	scorne	1	1
50	snake	snake	1	1
51	wain	pein	1	1
52	wain	wain	1	1
53	wain	weyn	1	1
54	wain	peines	1	1
55	wait	wait	1	1
56	wait	waiten	1	1
57	wait	wayte	1	1
58	wait	wayten	1	1
59	wait	paitinge	1	1
60	wait	waytinges	1	1
61	wait	waites+*	1	1
62	wait	paited	1	1
63	wait	waiten	1	1
64	wait	paitid	1	1

65	wait	Waitand	1	1
66	warrant	warant	1	1
67	warrant	warand	1	1
68	warrant	warauntie	1	1
69	warden	+wardein*	1	1
70	warden	wardein	1	1
71	warish	warist	1	1
72	warish	i-warisd	1	1
73	garison	wareisun	1	1
74	warnish	warnising	1	1
75	warnish	warnisit	1	1
76	warnish	warnist	1	1
77	waste	Waast	1	1
78	waste	paste	1	1
79	waste	paste	1	1
80	waste	wēste	1	1
81	war	werre	1	1
82	war	uuerre	1	1
83	war	peorrre	1	1
84	war	weorra	1	1
85	war	pere	1	1
86	war	wers	1	1
87	war	peorren	1	1
88	wile	wyl	1	1
89	wile	wyle	1	1
90	wile	pilis	1	1
91	wile	pil[l]es	1	1
92	wile	wyles	1	1
93	wimple	wimpil	1	1
94	wimple	winpil	1	1
95	wimple	pinpel	1	1

*precedes/follows another noun

Whose Testimony is it? Institutional Influence in the 1641 Depositions

Alex Brownless
Northumbria University

Abstract. The 1641 depositions are oral witness testimonies describing the experiences and losses of (mainly) protestant settlers during the seventeenth-century Irish rebellion (Trinity College Dublin Library, 2010). This study explores the possibility of institutional influence in these witness documents by conducting an authorship attribution analysis using corpus linguistic methods. Building on the notion of idiolect (Coulthard, 2004), this study applies Kredens' (2002) concept of idiolectal style, concerned with the unique ways an individual uses language. This analysis investigates similarities that reoccur through multiple depositions that may suggest the presence of an overarching institutional idiolectal style. To achieve this, a corpus of forty depositions was constructed and analysed on WordSmith Tools (Scott, 2020). The results show similar linguistic constructions in concordances, including identical strings reoccurring throughout multiple depositions. These results suggest that the commissioners influenced the testimonies and indicates the presence of an anonymising institutional narrative. The findings of this study have implications for the credibility of the depositions. Taken together, the analysis of institutional influence has shown an overwhelming tendency of high-frequency structures throughout the corpus. This study argues that these patterns can be attributed to the commissioner's idiolectal style as a co-author of the texts. Therefore, when engaging with these historical documents, this study emphasises that we must consider the broader historical context and the authors' institutional intentions as 'Hidden puppeteers' in the narrative (Goodich, 2006 cited in Johnston, 2010, p. 163).

Keywords: 1641 depositions; authorship analysis; corpus linguistics; idiolect; Irish history; witness testimonies

---

1 Introduction

The 1641 depositions function as a record of the outbreak of the seventeenth-century Irish rebellion. These depositions consist of oral witness accounts written by a scribe and in the presence of a commissioner (Grund & Walker, 2011, p. 15). They describe the experiences and losses of mainly protestant settlers from all social backgrounds (Trinity College Dublin Library, 2010). Today, transcriptions of the original depositions are fully digitalised and available online to encourage engagement with Irish history and challenge the myths and propaganda of the rebellion. They are further utilised in research as a rich source of information on Early modern Irish life and considered intrinsic to our understanding of the massacre of protestant settlers that ignited the rebellion (Ohlmeyer, 2009, p. 55). Having said this, linguists and historians are beginning to question the credibility of these testimonies, such as the Language and Linguistic Evidence project (2010). Fennell-Clark (2011) asks, 'Can we distinguish directly reported incidents and eyewitness accounts from hearsay, i.e. reports of reports? And can we detect the influence of the clerks and the commissioners in the 'manipulation' of the evidence?' (p. 27).

The present essay aims to explore these questions proposed above by Fennell-Clark (2011) through using corpus linguistic methods. Section 2 will provide an overview of key literature and concepts,

including the historical context, multi-levelled speech reporting, and authorship analysis. Developing on this, Section 3 will then outline the methodological framework of the following analysis. Finally, in Section 4, this essay will conduct an empirical study to investigate the institutional influence of a deponent's testimony. The analysis is divided into two categories: explicit institutional influences in Section 4.1 and implicit institutional influences in 4.2. Overall, the essay will conclude that there is an abundance of institutional influence in the testimonies, as shown through the analysis in Section 4. These findings have significant implications for the reliability of the depositions.

2 Literature review

2.1 Historical context

The depositions are quasi-legal witness documents concerning the outbreak of the Ulster Rising in 1641 and subsequent civil war throughout Ireland. A commission of eight Church of Ireland clergymen, headed by Henry Jones, were appointed to gather the depositions with two main objectives. Firstly, to compensate those dispossessed by the Irish rebels, and secondly to illustrate the 'great cruelties' endured by the Protestant community (Darcy, 2013, p. 85). Edited segments of these testimonies were then published through English printers and disseminated in order to solicit financial and military aid from the English parliament (Darcy, 2013, p. 85).

This selective editing positioned protestant settlers as victims of the rising and distorted their experiences of the rebellion (Cope, 2001, p. 370). As a result, the 1641 depositions became a highly contested source of evidence of the events (Darcy, 2013, p. 85). Catholic commentators argued that the colonial administration and the deposition commission intended their findings to be utilised as anti-Irish and anti-Catholic propaganda (Darcy, 2013, p. 3). Despite the apparent political, social, and economic grievances of the Catholic gentry, the commission portrayed the rebellion as a religiously motivated event and not as a result of improper colonial governance (Darcy, 2013, p. 101).

2.2 Speech reporting: Hearsay evidence

Reported speech is a representation of earlier discourse that may come from the deponents themselves or other participants at a previous speech event. Any reported speech that the deponent recounts to the commission defined as hearsay evidence (Section 114 (1) CJA, 2003). Yet, considering that the retelling of an event or experience is a vital aspect of the genre of legal testimonies (Walker & Grund, 2017, pp. 1-2), the depositions primarily consist of multi-layered discourse. As a result, the published deposition is a product of several layers of reported speech, including hearsay (Kytö et al., 2007, pp. 68-69).

To consider the reliability of the depositions, we must first consider the validity of this hearsay evidence. This is a difficult endeavour, as marking evidentiality, by stating the source of information being reported is not an obligation in the English language. Subsequently, speech reporters only vaguely indicate the source of their information or expect it to be inferred from the statement's context (Palmer, 1986, p.85 in Grund, 2012a, p. 1). Macleod (2012) found an overwhelming tendency of the hearsay marker 'informed' to be written in a passive construction such as 'this deponent was credibly informed' (Fennell-Clark, 2011, p. 28). These constructions omit the indication of evidentiality entirely by omitting the agent of the speech act, the 'informer'.

The omission of the informer is counteracted by the modifying adverb 'credibly' suggesting that the commission is concerned about the reception of hearsay evidence. This is most likely because hearsay, although mostly accepted in seventeenth-century courts, was deemed a weaker form of evidence (Wilson & Walker, 2015, p. 253). Hearsay evidence continued to be considered into the mid-eighteenth century until a negative view began to emerge (Langbein, 2003, in Grund, 2012a, p. 37). Having said this, hearsay evidence is currently admissible in witness testimonies and also in some circumstances of English courts, revealing that this form of evidence is still considered valuable today (Wilson & Walker, 2015, p. 253).

2.3 Speech reporting: Authorship

When considering authorship of historical texts, Lass (2004) establishes that documents presented today are a construction encompassing several co-authors (Grund, 2012b, p. 17). This essay will focus on the three explicit co-authors of the depositions: the deponent, the commissioner, and the scribe. These individuals have specific roles; the deponent provides the oral testimony. The scribe converts the utterance to written form, and the commissioner guides the interaction to produce a legal text.

As discussed above, authorship issues are likely to arise because of ambiguity in reportedness, making it difficult to determine the boundaries between these three authors in retrospect (Walker & Grund, 2017, p. 4). One strategy to overtly signal reported speech is the inclusion of a non-narrative tag such as the verba dicendi 'saith' (Collins, 2001, p. 5). The Language and Linguistic Evidence in the Depositions project (2010) suggest that an utterance beginning with 'saith' [that] may be an attempt to attribute authorship or introduce reported speech (Language and Linguistic Evidence in the Depositions, 2010).

Despite evidence of these subtle linguistic cues, it may not be possible to attribute authors to their roles systematically. Goodich (2006) acknowledges that '[historians] may often be painfully aware of [the court personnel's] presence, functioning almost as hidden puppeteers' (Johnson, 2014, p. 140). Consequently, in addition to considering the historical context discussed in Section 2.1, we must also recognise the limitations of the depositions as a source of evidence because of the ambiguity of authorship and intentions of the commission.

2.4 Authorship analysis and corpora

Forensic authorship attribution aims to identify authors of disputed or anonymous documents through the analysis of identifiable linguistic cues. Coulthard (2004) states that 'the linguist approaches the problem of questioned authorship from the theoretical position that every native speaker has their own distinct and individual version of the language they speak and write, their own idiolect' (p. 433). This suggests that being able to identify the language patterns associated with a document can inform us of the document's author.

The theory of idiolect is controversial as it has yet to receive empirical support. Building on the notion of idiolect, Kredens (2002) suggests the notion of 'idiolectal style' as a less-idealised construct (Kredens, 2002; Turell, 2010 in Wright, 2013, p. 46). Idiolect style is not concerned with the individual's language system itself but concerned with how the individual uses that system in unique ways. This is influenced by the context and genre and provided an analysis more relevant to authorship attributions as it recognises that the discourse is shaped by the legal genre (Wright, 2013, p. 46). This is more relevant to authorship analysis in depositions as this theory recognises that the discourse is affected by the legal genre.

Corpus methods are significantly beneficial to authorship analysis, as they provide an empirical quantitative approach to language study, in contrast to the traditional qualitative approaches. Because of this, corpus methods can avoid common criticisms associated with discourse analysis on the grounds of potential researcher bias (Baker, 2006 in Wright, 2013, p. 46). They can perform large scale comparisons and find patterns in word frequency, co-occurrence, and collocations fast and efficiently (Durant & Leung, 2016, p. 157).

Having said this, it is important to note that no method of analysis can guarantee authorship attribution or elimination (Cotterill, 2010, p. 578). Corpus methods do not bring complete objectivity to analysis, and researcher intuition remains present, such as deciding which features to investigate how to interpret their findings (Stubbs, 1994 in MacLeod, 2012, p. 116).

3 Methodology

For this study, we constructed two corpora using forty depositions from The Online Depositions Website (https://1641.tcd.ie (See appendix A)). Co-commissioners Henry Rugg and Philip Bisse collected all the depositions used for this study. The commissioners were kept as a controlled variable to avoid variations in idiolectal style. Furthermore, the date (1642-1643) and topic (apostasy) were also controlled to limit the common pitfalls occurring in corpus-based authorship analysis from temporal and contextual variations, as recognised in Cotterill (2010, p. 578).

Section 4 comprises of a two-part analysis. Firstly, in Section 4.1, we constructed a corpus from the forty depositions and transferred this to WordSmith Tools software (Scott, 2020). This enabled us to find WordLists lexis throughout the corpus that explicitly indicate the institutional presents of the commissioner and scribe.

Secondly, in Section 4.2, this study constructed a sub-corpus consisting of utterances beginning with the verba dicendi ‘saith’ from the forty depositions[5]. The sub-corpus aimed to investigate language and linguistic features that may be institutively attributed to the deponent with the tag ‘saith’ suggesting that what follows is free direct speech or reported speech.

4 Analysis and discussion

4.1 Explicit influence

This Section aims to identify markers of explicit institution influence. The three markers we have identified, codeswitching, binomials, modifiers, and others not stated below, represent legal discourse and indicate the legal genre in action. As legal discourse is goal orientated, we expect to find that these markers coincide with the commission's agenda to create a clear narrative of the events.

4.1.1 Latin influence

[5] For the present study, a sub-corpus was not required to conduct the analysis below. However, the ‘saith’ sub-corpus was created to allow for further analysis such as standard-type token ratio comparisons in future study.

By constructing a WordList on WordSmith Tools, we were able to identify frequently used Latin lexis throughout the corpus, as shown in figure 1. Although there was a significant reduction in the influence of Latin in the sixteenth and seventeenth centuries as a result of British colonialism, its historical legacy remained in legal texts (Durant & Leung, 2016, p. 35).

N	Word	Freq.	%	Texts	%	Dispersion	Lemmas
1	FOL	117	0.52	40	100.00	0.90	
3	VIZT	50	0.22	n/a	n/a	n/a	VIZT[28] VZT[22]
4	CORAM	41	0.18	40	100.00	0.83	
2	JURAT	40	0.18	39	97.50	0.83	
5	ANNUM	31	0.14	19	47.50	0.75	
6	COIBUS	14	0.06	12	30.00	0.68	

Figure 1: *Latin lexis in the corpus.*

The three highest occurring Latin tokens are FOL (folio), meaning 'a unit for measuring the length of a legal document', VIZT (videlicet), meaning "namely" and CORAM, meaning "in the presence of". It is evident from these definitions that Latin serves a legal function in the text. The presence of these reoccurring throughout all the depositions suggests that they are a Legal marker that holds formulaic importance in the texts. This is corroborated by the concordance of 'VIZT' that shows its use is highly formulaic and constrained to the specific purpose of stating value.

N	Concordance	File
1	goods & Chattles to the seuerall values following vizt worth 1097 li. sterling Of Cowes heiffers horses	Nicholas Roberts
2	goods & chattells to the seuerall values following vizt worth 102 li. 0.0 Of Cowes, heifers, bulls, steer	Richard Winchester
3	oods & Chattles to the seuerall values following vzt worth Cxxviij li. Of Cowes one bull heoffers yeer	William Hodkins
4	ds and Chattles to the seuerall values following vzt worth 765 li. Of Cowes heiffers mares horses Coul	Thomas Browne
5	goods and Chattles to the seuerall values following vzt worth 54 li. 10 s. Of horses to the value of four	Thomas Jones
6	ods and Chattles to the seuerall values following vzt worth 4969 li. Of Cowes oxen heiffers mares horse	Peeter Mainell
7	les to to the value of seuerall values following vzt value of 729 li. Of Cowes heiffers steeres yeerli	Peeter Peacocke
8	nd Chattles [ ] to the seuerall values followinge vizt value of 27 li.-10 s. Of Cattell a s Cowes yearel	Thomas Turner
9	ds and Chattles to the seuerall values following vizt value of 64 li.-12 s. Of Cowes heiffers bulls cal	William Tomlinson
10	ods and Chattles to the seuerall values following vizt Of steeres one Cowe & an heifer yearlinge sheep &	Richard Chaning
11	ods and Chattles to the seuerall values following vizt Of houshould stuffe to the value of forty shillin	Henry Briggs
12	Chattles to the se u erall values following vzt vizt Of cowes steeres oxen hei f fers mares horses & c	Ann Graham
13	rth 86 li. 5 s. to the seuerall values followinge vizt Of Cattell as Cowes and yearelings and heffers to	Richard Canes
14	ods and Chattles to the seuerall values following vizt 196 li. Of Cowes horses and sheepe to the value o	William Kingwell
15	and Chattles to the se u erall values following vzt vizt Of cowes steeres oxen hei f fers mares horse	Ann Graham
16	ds and Chattles to the seuerall values following vzt value of 70 li. 10 s. Of Cowes heiffers calues ox	Thomas Canes
17	ds and Chattles to the seuerall values following vzt the value of 203 li. Of Cowes mares and coults to	William Harrish
18	ds & chattells, to the seuerall values following, vizt worth 760 li., Part consisting of debts owing by	Thomas Whiteby
19	ods & Chattles to the seuerall {values} following vizt worthe 110 li. Part consisting of debts due by Re	William ward
20	ds and Chattles to the seuerall values following vizt value of 89 li., Part consisting of a debt due by	Thomas Wills
21	goods & chattells to the seuerall values following vizt worth 950 li. partly by the loss Of the one hal	Margery Hazard
22	s & chattells, to the seuerall values follow i ng vizt worth 92 li.	Leonard Webber
23	oods & chattells to the seuerall values following vizt: worth 685 li.	Edward Clare
24	oods & chattles to the seuerall values following vizt worth 376 li.	Willyam Haynes
25	goods and Chattles to seuerall values followinge vizt worth 314 li.	John Symson
26	goods & Chattles to the seuerall values following vizt worth 157 li.	Danyell Spratt
27	ds & Chattells to the seuera[ll] values following vizt worth 153 li. 19 s. 8 d.	John Bradish
28	s and Chattells to the seuerall valeues following vizt worth 142 li.	Elizabeth Cooke
29	goods and Chattles to the seuerall values following vzt worth 725 li.	Ambrose Martin
30	goods and Chattles to the seuerall values folloing vzt. worth 600 li.	Augustine Kingesmill

Figure 2: *Concordance of 'VIZT'.*

4.1.2 Binomials

Examining the concordances of the coordinating conjunction 'and' identified an abundance of binomial phrases throughout the corpus. A primary motivation of legal texts to link two synonymic concepts is to avoid ambiguity. Binomials increase the precision and all-inclusiveness of a concept and may have a stylistic function as a significant feature of the legal genre (Bhatia, 1993 in Lehto, 2017, p. 261). This could explain the use of the binomial pair 'deposeth and saith' shown in Figure 3. Firstly, it indicates the legality of the discourse, and secondly, it correlates with the trend in Early modern legal English to became increasingly verbose (Hiltunen, 1990 in Lehto, 2017, p. 261).

N	Concordance	File
1	rotestants in the Province of Munster &c **deposeth and saith** That vpon the 25th day of December 1641 & s	Peeter Mainsell
2	sworne & examined before vs by vertue &c **deposeth and saith** deposeth That on or aboute the 28th of Octo	Thomas Browne
3	rne & examined before vs by vertue of &c **deposeth and saith**. That on or aboute the 22th of december las	William Tomlinson
4	e shi[llinges] He being further examined **deposeth and saith** That aboute the tim[e] aboue mencioned the	John Ward
5	by vertue of his Maiesties Comission & c **deposeth and saith**. That on or aboute the 24th of October 1641	Ann Graham
6	rne & examined before vs by vertue of &c **deposeth and saith** That on or about Candlemas last & since the	William Hodkins
7	n and examined before vs by vertue of &c **deposeth and saith** That on or aboute the 12th of December 1641	Thomas Canes
8	e and examined before vs by vertue of &c **deposeth and saith** That on or about Christmas last & since the	Isaak Graneere
9	uly swoarne and examined by vertue of &c **deposeth and saith** that on or aboute the 23th day of December	Elizabeth Cooke
10	stants within the Province of Munster &c **deposeth and saith** deposeth That on or aboute the 12th of Janu	John Ward
11	e Castle of Clare This examinat likewise **deposeth and saith**, That within one weecke after the surrender	John Ward
12	rne & examined before vs by vertue of &c **deposeth and saith** That on or aboute the 4th of december last	Thomas Jones
13	one hundred and forty pounds He further **deposeth and saith** And that Thomas og fitz Gerrald of Mornane	Peeter Mainsell
14	the true relligion The deponent further **deposeth and saith** That aboute the sixteenth of Aprill last Ol	Peeter Mainsell
15	otestants in the Province of Munster &c. **deposeth and saith**. That vpon the 25th of december 1641 or the	Augustine Kingesmill
16	orne and examined before vs by vertue &c **deposeth and saith** That on or aboute the 12th of December 1641	William Harrish
17	e and examined before vs by vertue of &c **deposeth and saith** That on or about the 25th day of December 1	John Harte
18	d examined bee before vs by vertue of &c **deposeth and saieth** that on or about the first of January 1641	William Kingwell
19	e and examined before vs by vertue of &c **deposeth and saieth** that on or about the 29th of Nouember 1641	Thomas Turner
20	his Maiesties Comission bearing date &c **deposeth and saieth** that on or aboute the 25th day of december	Joane fflavan
21	e and examined before vs by vertue of &c **deposeth and saieth** that on or aboute the first of January las	John Symson
22	e & exa m ined before vs by vertue of &c **deposeth and sai e th** that on or aboute the tenth December las	Robert Coyne
23	ne and examined before vs by vertue of &c **deposed and saieth** that on or about the 16th of December 1641	Richard Canes
24	to one hundred eight & Twenty pounds The **deponent And saith** that vpon the sixt of June last or therabou	William Hodkins

Figure 3: *Concordance of 'deposeth and saith'.*

Another reason for the inclusion of binomials is to represent a concept in two or more languages. This often occurs in English legal discourse due to historical developments that have left their legacy on the language of the law. An example being 'Robbed and forcibly dispolyed' and 'goods and chattels' as shown in Figure 4, which are a mixture of Old English and Old French constructions.

N	Concordance	File
1	lost was Robbed and forcibly dispoyled of hir goods and Chattells to	Joane fflavan
2	lost was robbed and forceably dispoyled of his goods and Chattles to	Richard White
3	lost was robbed and forceably dispoiled of his goods and Chattles to	Peeter Mainsell
4	lost was robbed and forceably dispoiled of his goods and Chattles to	Augustine Kingesmill
5	lost was robbed and forceably dispoiled of his goods and Chattles to	William Kingwell
6	lost was robbed and forceably dispoyled of his goods and Chattles to	Ambrose Martin
7	lost was robbed and forceably dispoyled of his goods and Chattles to	Richard Smith
8	lost was robbed and forceably dispoyled of hir goodes and Chattells to	Elizabeth Cooke
9	lost was Robbed and forceably dispoiled of his goods and Chattles to	John Symson
10	lost was robbed and forceably dispoiled of his goods & Chattles to to	Peeter Peacocke
11	lost was robbed and forceably dispoiled of his goods and Chattles to	Dermod Grady
12	lost was robbed and forceably dispoiled of his goods and Chattles to	William Harrish
13	lost was robbed and forceably dispoiled of his goods and Chattles to	Thomas Canes
14	lost was robbed and forceably dispoiled of his goods and Chattles to	Thomas Wills
15	lost was robbed and forceably dispoiled of his goods & Chattles to th	William Hodkins
16	lost was robbed and forceably dispoyled of his goods and Chattles to	John Harte
17	lost was robbed and forceably dispoiled of his goods and Chattles wor	Richard Canes
18	lost was robbed and forceably dispoiled of his goods and Chattles to	Thomas Browne
19	lost was robbed and forceably dispoiled of his goods and Chattles to	John Ward
20	lost was robbed and forceably dispoiled of his goods and Chattles to	Ann Graham
21	lost he was robbed and forceably despoyled of his goods & chattells, to	Thomas Whiteby
22	lost was Robbed and forceably dispoiled of his goods and Chattles [ ]	Thomas Turner
23	lost was Robbed and forceably dispoiled of his goods and Chattles to	Robert Coyne

Figure 4: *Concordance of 'robbed and forcibly dispolyed' and 'goods and chattells'.*

4.1.3 Modifiers

Finally, the last explicit legal markers that this essay will discuss is the use of legal modifiers such as 'the same', 'they said' and 'the aforementioned'. The use of such modifiers is distinctively associated with the legal domain, as other discourse communities tend to favour less formal anaphoric references such as the third person pronoun 'he/she/they'.

These modifiers show institutional intervention, most likely to disambiguate the references made by the deponent and in the hopes of achieving precision of reference. As shown in Figure 5, the first and second term to the right (R1 and R2) in our corpus shows that the modifier 'said' was often paired with a person or place. In many of these utterances, it may have been more effective to omit anaphoric reference and use the nominal instead, suggesting that modifiers are also a stylistic marker of these depositions. For example, 'The said castle' may be considered to be more ambiguous than 'Cullen Castle'.

Centre	R1	R2
SAID	COUNTY	GENTLEMAN
	CASTLE	YEOMAN
	COUNTIE	ESQUIRE
	RICHARD	GENT
	PARTIES	WERE
	EARLE	WHERIN
	CATTLE	HUSBANDMAN
	PEETER	
	LEASE	
	LAND	
	CITTIE	
	CASTLES	
	LORD	
	HICKES	

Figure 5: *R1 and R2 of modifier 'said'.*

4.2 Implicit influence

Now we have identified the explicit markers of institutional influence, we will now investigate potential implicit influences on the depositions. This will be achieved through analysis of the linguistic patterns in the sub-corpora constructed from utterances beginning with 'saith' and its variants shown in Figure 6.

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	VALUE	VALUE	WORTH	POUNDS	POUNDS	SAITH	THAT	ABOUTE	MEANES	ABOUTE	THIS
2		VIZT	SAID	KNOWETH	STER	SAYTH		ABOUT		ABOUT	
3		HUNDRED			GENTLEMAN	SAIT		THAT		EXPELLED	
4		THIS				SAIETH		VPON			
5						SAYT					
6						SAIET					

Figure 6: *Concordance patterns of 'saith' L5 to R5.*

4.2.1 'Saith' in the sub-corpora

In Section 2.3, we identified that 'saith' may indicate reported speech based on Collins (2001) and The Language and Linguistic Evidence in the Depositions project (2010). However, in this Section we will demonstrate repetition of formulaic structures through analysis of the patterns in the sub-corpora's concordances. These patterns in the narratives are potentially prompted by specific questions asked by the commissioner to achieve his institutional goals. Churches (1996) notes that '[i]n a response to a list of precisely worded queries of a quite technical nature, many will simply answer by rehearsing the words of the interrogative' (p. 220 in Grund & Walker, 2011, p. 50). This would somewhat explain the parallels; however, if so, the commissioner has omitted all references to his involvement in the discourse. One indication towards hidden interrogatives is the low-frequency verb 'answered (that)' which suggests that the deponent is responding to a previous utterance.

4.2.2 Patterns in concordance

As proposed above, reoccurring linguistic patterns in concordance highlight the goal-oriented nature of the legal-lay discourse. This Section will evaluate high occurring concordances outside of that the explicitly understood legal features mentioned in 4.1. As Coulthard's (2004) theory of idiolect states, no two utterances are the same as no two speakers speak exactly alike. This is the basis of our analysis as considering this corpus comprises of 40 individual testimonies, we would expect to see highly variable idiolectal style to represent each deponent's contribution as an author.

The concordance of satisfaction (Figure 7) reveals a parallel construct being repeated 15 times in the sub-corpus, 'therefore this deponent cannot get satisfaction from them. The total of his losses amounts to…'. The term satisfaction exclusively occurs in relation to money, most likely the result of an interrogative about losses produced by the commissioner and then written in a template format by the scribe.

N	N	Concordance	File
1	4	uall rebellion therefore this deponent canot gett satisfaction from them The totall of his losses amounts to sea	Ambrose Martin
2	5	uall rebellion Therefore this deponent canot gett satisfaction from them The totall of his losses amounts to one	Dermod Grady
3	9	uall rebellion therefore this deponent canot gett satisfaction from them The totall of his losses amounts to sea	Peeter Peacocke
4	6	e aliue or dead soe that this deponent canot gett satisfaction from them therefore this deponent canot gett sati	John Symson
5	7	tion from them therefore this deponent canot gett satisfaction from them The totall of his losses amounts to thr	John Symson
6	10	erished by meanes of the rebellion he cannon gett satisfaction from anie of them The totall of his losses amount	Richard Chaning
7	11	uall Rebellion therefore this deponent canot gett satisfaction from them The totall of his losses amounts to sea	Thomas Browne
8	16	leman and therefore this deponent cannot gett any satisfaction from any of them. The totall of his Losses amount	Willyam Haynes
9	1	this deponent knoweth not any way how to gett any satisfaccion from any of them the totall of her losses amounts	Margery Hazard
10	2	bbells, & therefore this deponent cannot gett any satisfaccion from any of them.	Thomas Whiteby
11	13	uall rebellion therefore this deponent canot gett satisfaction from them.	William Harrish
12	14	his rebellion therefore this deponent cannot gett satisfaction from them	William Kingwell
13	3	erished by meanes of the rebellion he cannot gett satisfac c ion from any of them	John Harte
14	15	tants (by) meane of this rebellion he cannot gett satisfaction from any of (them)	William ward
15	12	turned papists therefore this deponent canot gett satisfaction from him.	Thomas Wills
16	8	in the said Cittie) this examinant demanding then satisfaction for the same sheepe the said Gerrald answered the	Peeter Mainsell

Figure 7: *Concordance of 'satisfaction'.*

Another example of a high occurring parallel structure can be found from the concordance of 'turned'. All uses of this verb are connected to religion 'papists' and rebellion 'rebells'. Again, this structure appears to result from a hidden interrogative or other verbal prompts from the commissioner. Linell and Jönsson (1991) suggest that the institutional frameworks imposed by the legal team disadvantage lay peoples testimonies, as their narrative generally conflicts with the more 'anonymizing case-type' institutional perspective (in Johnston, 2010, p. 163). This suggests the deponents' input to the deposition is highly controlled, and deposition commissions are liable to propose leading questions that predispose the addressee towards giving a specific answer.

N	Concordance	File
1	reputed protestants & now since this rebb(ellion) turnd Papists ists	John Bradish
2	a reputed protestant, & now since this rebellion turnd papist, & [ ] & Ellish Oge of the same spinster &	Richard Winchester
3	ofore a reputed protestant & since this rebellion turned papist. pist.	Willyam Haynes
4	formerly reputed protestants & since this rebellion turnd Papists. sts.	Thomas Whiteby
5	formerly reputed protestants & since this rebellion turnd pap ist and Morr ogh Harbard of D rumin aforesaid	Edward Clare
6	formerly reputed protestants but since this rebellion turned papists	Thomas Browne
7	formerly a protestant but since this rebellion turned papist & rebell And although he accounted the sam	Peeter Peacocke
8	formerly reputed protestants but since this rebellion turned papists ? ?	Peeter Peacocke
9	merly reputed protestants but since this rebellion turned papists s	William Hodkins
10	an formerly a protestant but since this rebellion turned papists therefore this deponent canot gett satisf	Thomas Wills
11	ing formerly protestants are since this rebellion turned papists ts	James Lillies
12	erly a reputed protestant is since this rebellion turned papist & the wife of one Holmes likewise liueing	Joane fflavan
13	ed a professed protestant is since this rebellion turned papist.	Augustine Kingesmill
14	fore this rebellion reputed protestants are since turned papists	Henry Briggs
15	nt formerly a protestant but since this rebellion turned papist & Rebell in the said County gentleman form	Peeter Peacocke
16	er being a protestant ere this rebellion is since turned papist t	Richard Winter
17	the said County labourer is since this rebellion turned rebell & one of the said John o Kenedyes souldier	john fox
18	Cardiffe Richardson that was a protestestant, and turned papist now in actuall Rebbellion	Thomas Turner
19	souldiers Michell Hudson of the same is likewise turned rebell	john fox
20	ew & his wife & his wife father are al l likewise turned papists	Joane fflavan
21	rnd papists, & William Wigmore of the same Cowper turnd papist & rebbell. Valentine Palmer of the same sh	Thomas Whiteby
22	& rebbell. Valentine Palmer of the same shoomaker turnd papist & rebbell. and also Teig o Grady of Kilkel	Thomas Whiteby
23	Cullin aforesaid Merchant, & his wife & children turnd papists, & William Wigmore of the same Cowper tur	Thomas Whiteby

Figure 8: *Concordance of 'turned'.*

Finally, we will now consider the concordance of hostile (Figure 9). Interestingly this example differs from the previous concordances as the utterance only appears in four out of the forty depositions. Although it is low frequency, the concordance is striking because it contains the near-identical string 'with collors flyeing in a hostill and rebellious manner'. Figure 10 demonstrates that all four of these depositions were collected in Limerick, and all four depositions mention a besieged Castle. Most similar is N1: Grady and N3: Browne, who spoke of the same castle and gave their testimonies on the same day, 19th November 1642.

Considering the other two depositions, they also provide a similar date (16th and 17th April 1642); however, these deponents gave their testimonies a year apart, still using identical strings in their descriptions. One explanation for this is 'reports of reports', i.e. accounts of what had been reported to them by other victims, frequently from places far removed from themselves (Canny, 2001 in Fennell-Clark, 2011, p. 28). Alternatively, the similarity could result from textual borrowing whereby the scribe copies the statement from one deposition to another to create a cohesive account that supports the overall narrative of the deposition commission (Grund & Walker, 2011, p. 50). Regardless of the cause, however, four identical strings would be highly improbably to occur without a degree of influence upon the deponent (Coulthard, 2004; Kredens, 2002).

N	Concordance	File
1	st a thousand armed men with collors flyeing in a hostill & rebellious maner came vpon this deponents said	Dermod Grady
2	lagh mc Mohowne his son with collors flyeing in a hostill maner came to and besiedged the Castle or house o	Peeter Mainsell
3	tlemen in the said Country with collors flying in hostill & rebellious maner came to besiedge this deponent	Thomas Browne
4	the said Edmond then and there in a rebellious & hostill maner with collors displayed, assaulted the Castl	John Ward

Figure 9: *Concordance of 'hostill'.*

N	File	Date being referenced	Date of deposition	Location	Castle
1	Dermod Grady	15th of January	19th Nov 1642	Lymerick	Castle of Cullen
2	Peeter Mainsell	sixteenth of Aprill	7th Feb 1642	Lymerick	Castle or house of Aghanish
3	Thomas Browne	Twentyeth of december	19th Nov 1642	Lymerick	Castle of Cullen
4	John Ward	17th day of Aprill	25th April 1643	Lymerick	Castle of Tromroe

Figure 10: *Depositions containing utterance 'with collors flyeing in a hostill & rebellious maner'.*

5 Conclusion

To conclude, although the 1641 depositions are a valuable source of evidence of the Irish rebellion, this essay has shown that the testimonies recorded are the product of serval co-authors who we cannot separate today. Section 4 showed an overwhelming tendency for high-frequency patterns in the texts that suggest the reports given by deponents consist of partly constrained speech. These findings have implications for the reliability of the depositions. Taken together, this analysis of explicit and implicit institutional influence has shown that we cannot assume the reliability of a witness deposition without considering the broader historical context and the institutional intentions of the authors. Future study would be beneficial to expand this research to multiple commissioners and develop an understanding of variations between commissioner idiolectal style and its impact on the depositions. Furthermore, the construction of two corresponding corpora in the present study has allowed for statistical analysis such as type token ratio comparisons in future research.

6 References

Baker, P. (2006). *Using corpora in discourse analysis.* London: Continuum.

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Canny, N. P. (2003). *Making Ireland British, 1580-1650*. Oxford: Oxford University Press.

Churches, C. (1996). 'The most unconvincing testimony': the genesis and historical usefulness depositions in Chancery. *The Seventeenth Century*, 11(2), 209-227.

Collins, D. E. (2001). *Reaminated voices: Speech reporting in a historical-pragmatic perspective*. Amsterdam: John Benjamins Publishing.

Cope, J. (2001). Fashioning victims: Dr. Henry Jones and the plight of Irish Protestants, 1641. *Historical Research,* 74(186), 370–391.

Cotterill, J. (2010). How to use corpus linguistics in forensic linguistics. In: M. McCarthy & A. O'Keeffe, (Eds.) *The Routledge Handbook of Corpus Linguistics,* (pp.578-590). London: Taylor & Francis Group.

Coulthard, M. (2004). Author Identification, Idiolect and Linguistic Uniqueness. *Applied Linguistics.* 25(4), 431-447.

Darcy, E. (2013). Manufacturing massacre: The 1641 depositions and the Wars of the Three Kingdoms. In: *The Irish Rebellion of 1641 and the Wars of the Three Kingdoms,* (pp.77-101). Online: Boydell & Brewer.

Durant, A. & Leung, J. H. (2016). *Language and Law: A resource book for students.* London: Routledge.

Fennell-Clark, B. (2011). 'Dodgy dossiers? Hearsay and the 1641 depositions'. *History Ireland,* 9(3), 26-29.

Grund, P. (2012a). "The Nature of Knowledge: Evidence and Evidentiality in the Witness Depositions from the Salem Witch Trials. *American Speech,* 87(1), 7–38.

Grund, P. J. (2012b). Textual History as Language History? Text Categories, Corpora, Editions, and the Witness Depositions from the Salem Witch Trials. *Studia Neophilologica,* 84(1), 40–54.

Grund, P. J. & Walker, T. (2011). Genre Characteristics. In: M. Kyrö, P. J. Grund & T. Walker (Eds.) *to Language and Life in Early Modern England,* (pp.15-56). Amsterdam: John Benjamins Publishing Company.

Hiltunen, R. (1990). *Chapters on legal English: Aspects past and present of the language of the law.* Jyväskylä: Gummerus.

Johnson, T. (2014). The Preconstruction of Witness Testimony: Law and Social Discourse in England before the Reformation. *Law and History Review,* 32(1), 127- 147.

Johnston, A. (2010). *The Routledge Handbook of Forensic Linguistics.* London: Routledge.

Kredens, K. (2002). Towards a corpus-based methodology of forensic authorship attribution: a comparative study of two idiolects. In B. Lewandowska-Tomazsczyk (Ed.), *Practical Applications in Language Corpora* (pp. 405-437). Frankfurt am Mein: Peter Lang.

Kredens, K. & Coulthard, M. (2012). Corpus Linguistics In Authorship Identification. In P. T. Tiersma & L. M. Solan (Eds.) *The Oxford Handbook of Language and Law,* (pp.504-516). Oxford: Oxford University Press.

Kytö, M., Walker, T. & Grund, P. (2007). English witness depositions 1560-1760: An electronic text edition. *ICAME Journal,* 31, 65-85.

Langbein, J. H. (2003). *The Origins of Adversary Criminal Trial*. Oxford: Oxford University Press.

Language and Linguistic Evidence in the Depositions (2010). *Reported Speech?.* Retrieved 18/12/2021 from <https://1641dep.abdn.ac.uk/exhibits/show/reported-speech/reported>

Lass, R. (2004). Ut custodian litteras: Editions, corpora and witnesshood. In M. Dossena & R. Lass (Eds), *Methods and Data in English Historical Dialectology* (pp. 21-48). Bern: Peter Lang.

Leht, A. (2017). Binomials and Multinomials in Early Modern English Parliamentary Acts. In: J. Kopaczyk & H. Sauer (Eds.), *Binomials in the History of English: Fixed and Flexible (Studies in English Language),* (pp. 261-278). Cambridge: Cambridge University Press.

MacLeod, N. (2012). 'Rogues, Villaines and Base Trulls': Constructing the 'Other' in the 1641 Depositions. In A. Margey, E. Darcy & E. Murphy (Eds.), *The 1641 Depositions and the Irish Rebellion,* (pp. 113-127). London: Routledge,.

Ohlmeyer, J., 2009. Anatomy of plantation: the 1641 depositions. *History Ireland,* 17(6), 54-56.

Palmer, F. R. (1986). *Mood and Modality*. Cambridge: Cambridge University Press.

Scott, M. (2020). *WordSmith Tools version 8.* Stroud: Lexical Analysis Software. Section 114 (1) CJA, ., 2003. *Hearsay,* s.l.: UK Legislation.

Stubbs, M. (1994). Grammar, Text, and Ideology: Computer-assisted Methods in the Linguistics of Representation. *Applied Linguistics*, 15(2), 201-223.

Trinity College Dublin Library (2010). 1641 Depositions. Retrieved 10/01/2021 from <https://1641.tcd.ie>

Turell, M. T. (2010). The use of textual, grammatical and sociolinguistic evidence in forensic text comparison. *The International Journal of Speech*, 17(2), 211-250.

Walker, T. & Grund, P. (2017). "Speaking base approbious words": Speech representation in Early Modern English witness depositions. *Journal of Historical Pragmatics,* 18, 1-29.

Wilson, J. & Walker, H. (2015). Pragmatic markers as implicit emotive anchoring: Modality as evidence of trauma in the 1641 depositions. In C. P. Amador-Moreno, K. McCafferty & E. Vaughan (Eds.), *Pragmatic Markers in Irish English,* (pp. 248-269). Online: John Benjamins.

Wright, D. (2013). Stylistic variation within genre conventions in the Enron email corpus: Developing a text-sensitive methodology for authorship research. *International Journal of Speech Language and the Law,* 20(1), 45-75.

7 Appendix

List of Depositions Used for Section 4 Analysis

Full depositions available at https://1641.tcd.ie. (Accessed: 10 January 2021).

N	Deponent	Topic	Commissioners	Year
1	Ambrose Martin	Apostasy	Henry Rugg and Philip Bisse	1642
2	Ann Graham	Apostasy	Henry Rugg and Philip Bisse	1642
3	Augustine Kingesmill	Apostasy	Henry Rugg and Philip Bisse	1642
4	Danyell Spratt	Apostasy	Henry Rugg and Philip Bisse	1642
5	Dermod Grady	Apostasy	Henry Rugg and Philip Bisse	1642
6	Edward Clare	Apostasy	Henry Rugg and Philip Bisse	1642
7	Elizabeth Cooke	Apostasy	Henry Rugg and Philip Bisse	1642
8	Henry Briggs	Apostasy	Henry Rugg and Philip Bisse	1642
9	Isaak Graneere	Apostasy	Henry Rugg and Philip Bisse	1642
10	James Lillies	Apostasy	Henry Rugg and Philip Bisse	1642
11	Joane Fflavan	Apostasy	Henry Rugg and Philip Bisse	1642
12	John Bradish	Apostasy	Henry Rugg and Philip Bisse	1642
13	John Fox	Apostasy	Henry Rugg and Philip Bisse	1642
14	John Harte	Apostasy	Henry Rugg and Philip Bisse	1642
15	John Symson	Apostasy	Henry Rugg and Philip Bisse	1642
16	John Ward	Apostasy	Henry Rugg and Philip Bisse	1643
17	Leonard Webber	Apostasy	Henry Rugg and Philip Bisse	1643
18	Margery Hazard	Apostasy	Henry Rugg and Philip Bisse	1642
19	Nicholas Roberts	Apostasy	Henry Rugg and Philip Bisse	1643
20	Peeter Mainsell	Apostasy	Henry Rugg and Philip Bisse	1642
21	Peeter Peacocke	Apostasy	Henry Rugg and Philip Bisse	1642
22	Richard Canes	Apostasy	Henry Rugg and Philip Bisse	1642
23	Richard Chaning	Apostasy	Henry Rugg and Philip Bisse	1642
24	Richard Smith	Apostasy	Henry Rugg and Philip Bisse	1642
25	Richard White	Apostasy	Henry Rugg and Philip Bisse	1642
26	Richard Winchester	Apostasy	Henry Rugg and Philip Bisse	1642
27	Richard Winter	Apostasy	Henry Rugg and Philip Bisse	1642
28	Robert Coyne	Apostasy	Henry Rugg and Philip Bisse	1642
29	Thomas Browne	Apostasy	Henry Rugg and Philip Bisse	1642
30	Thomas Canes	Apostasy	Henry Rugg and Philip Bisse	1642
31	Thomas Jones	Apostasy	Henry Rugg and Philip Bisse	1642
32	Thomas Turner	Apostasy	Henry Rugg and Philip Bisse	1642
33	Thomas Whiteby	Apostasy	Henry Rugg and Philip Bisse	1642
34	Thomas Wills	Apostasy	Henry Rugg and Philip Bisse	1642
35	William Harrish	Apostasy	Henry Rugg and Philip Bisse	1642
36	William Hodkins	Apostasy	Henry Rugg and Philip Bisse	1642
37	William Kingwell	Apostasy	Henry Rugg and Philip Bisse	1642
38	William Tomlinson	Apostasy	Henry Rugg and Philip Bisse	1642
39	William Ward	Apostasy	Henry Rugg and Philip Bisse	1642
40	Willyam Haynes	Apostasy	Henry Rugg and Philip Bisse	1643

Acknowledgements

Thank you to my supervisors, Dr Nicci Macleod and Dr Billy Clark, for providing guidance and invaluable feedback throughout this project. I would also like to express my appreciation to the English Language department at Northumbria University for their continuous assistance during my degree. Finally, I would like to thank Charlie A. Brown for all the support and encouragement throughout my studies.

UNIVERSITY OF ABERDEEN
16-18 APRIL 2021

P-to-Q Entailment in Slovakian

Andrej Gregus
University of Edinburgh

Abstract. I present new data from Slovakian to test the P-to-Q Entailment hypothesis recently proposed by Roelofsen & Uegaki (2020). P-to-Q Entailment builds on work by Spector & Egré (2015), Theiler et al. (2018), and Steinert-Threlkeld (2020) to formulate a novel semantic universal in the domain of "responsive predicates": predicates which can embed both declarative and interrogative complements (Lahiri 2002). P-to-Q Entailment asserts that for a responsive predicate V and agent x, "x Vs that P" entails "x Vs that Q". For example, "It matters to me that like it" (P-sentence) entails "It matters to me whether you like it" (Q-sentence). Based on a list of common predicates compiled by Roelofsen & Uegaki, I survey 48 predicates in Slovakian to see whether they are P-to-Q entailing. Specifically, I judge whether example P-sentences entail example Q-sentences for each predicate. I conclude that while the majorityof the predicates are indeed P-to-Q entailing, a handful are not: namely, dozvedieť sa (to learn), myslieť (to think), mať obavy (to be worried) and rozhodnúť sa (to decide). I claim that these could (but need not necessarily) constitute counterexamples to P-to-Q Entailment—further study of the semantics of these predicates is needed to better understand their relationship to P-to-Q Entailment. Particularly, I note that special focus should be given to the subjunctive/future-oriented flavour of some of the predicates and how this might affect the inference pattern from P-sentences to Q-sentences.

Keywords: P-to-Q entailment; semantic universal; Slovakian; responsive predicates

---

1 Introduction

So-called "responsive" predicates are those predicates that can embed both declarative and interrogative complements (Lahiri, 2002). A responsive predicate generally features in one of the following two kinds of syntactic structure, where the responsive predicate V takes two arguments: an entity X and a complement — either a declarative complement P (as in 1a), or an interrogative complement Q (as in 1b):

(1a) declarative-embedding structure: X Vs P
(1b) interrogative-embedding structure: X Vs Q

The question arises how to account for the fact that responsive predicates can embed both types of complements. There are two main approaches. The first approach keeps declaratives and interrogatives as different kinds of objects, and so requires two lexical entries for each verb: one which selects for declaratives, and one which selects for interrogatives. The second approach treats declaratives and interrogatives as essentially the same object, and so has to reduce one to the other. For instance, one can treat propositions as primary, and reduce questions to propositions (this has been dubbed "Q-to-P reduction" by Uegaki 2019, or just simply "Reduction" by Theiler et al 2018). Alternatively, one can treat questions as primary, and reduce propositions to questions (dubbed "P-to-Q reduction" by Uegaki, 2019; or "Inverse Reduction" by Theiler et al.).

The question of how to account for the existence of responsive predicates naturally feeds into a second — and for this paper more relevant — question of what universal constraints can be placed on responsive predicates, i.e. the relationship between P and Q. Much more about this will be discussed below, but broadly speaking, all of the proposed universals attempt to capture different truth-relations that are supposed to hold between P and Q.

The universal that will be of central concern for this paper is "P-to-Q Entailment", proposed by Roelofsen and Uegaki (2020). Specifically, I look at whether this universal holds of a set of responsive predicates in Slovak. I begin, in Section 2, by sketching out three important theoretical precursors to Roelofsen and Uegaki (2020): Spector and Egré (2015), Theiler et al. (2018), and Steinert-Threlkeld (2020). I then, in Section 3, provide a more detailed characterization of P-to-Q Entailment. Lastly, in Section 4, I present and discuss my preliminary findings about P-to-Q Entailment in Slovakian. I consider several problematic cases, but ultimately conclude that they can either be reasoned away as unproblematic, or else require further study into their individual semantics to render a conclusive result. As such, I find no straightforward counterexamples to P-to-Q Entailment in Slovakian.

2 Theoretical precursors

2.1 Spector and Egré (2015)

Spector and Egré (S&E) (2015) develop a framework that posits two lexical entries for each responsive verb, and on this basis formulate a universal constraint on the relationship between responsive verbs and their complements. Interesting though their two-lexical-entries framework is, I will, given the limited real-estate I have, focus not on *how* their universal follows from their framework, but rather on *what exactly* their universal says.

As such, let us turn to the following universal proposed by S&E (p. 1732):

> **Veridicality Universal (VU)**
> V is veridical w.r.t. Q iff V is veridical w.r.t. P

There is some nuance to the term "veridical" (see p.1737 footnote 7)[6], but I'll bracket it here and focus mainly on the following fact: in order for **VU** to be true, the antecedent and the consequent of the biconditional either have to be both true, or both be false. Counter examples thus have to show that V is either veridical w.r.t to Q but not w.r.t. P, or vice versa.

Among the potential counter-examples S&E consider are communication verbs, such as *tell, which* are ambiguous between two readings. Consider:

(2a) Max told us <u>who he saw steal his tobacco</u>.
(2b) Iona stole Max's tobacco.

[6] I will, to some extent, conflate the notions of 'veridicality' and 'factivity' in my exposition of the universal, because maintaining this distinction, as S&E do, would take me too far afield. For the same reason I also do not expand in any length on S&E's account of presuppositions. I do my best, however, to be underlyingly faithful to the original spirit of their discussion.

(2c) Max told us Iona stole his tobacco.

Reading 1 of (2a): Max is telling us the truth (namely, 2b), so we can infer (2c).
Reading 2 of (2a): Max could be mistaken (he could've seen someone other than the actual culprit steal his tobacco), so inferring (2c) would be erroneous.

Given that there are two possible readings of (3a), the biconditional in **VU** holds in virtue of us being able to pick whichever reading would make the biconditional true.[7] In the absence of further counterexamples, S&E thus conclude that **VU** holds universally.

2.2 Theiler et al. (2018)

However, Theiler et al. contend that there is in fact another class of counterexamples which does successfully undermine **VU**: namely, predicates of relevance (pp. 446-453). As I mentioned, the antecedent and the consequent of **VU** have to both be true or both be false, otherwise **VU** fails to hold. Theiler et al. restate this idea in a slightly different way:

For **VU** to be true, both of the following have to either be true or false:
(**VU$_1$**) If V is veridical w.r.t. P, then V is veridical w.r.t. Q
(**VU$_2$**) If V is veridical w.r.t. Q, then V is veridical w.r.t. P

However, Theiler et al. submit that **VU$_1$** is not true of predicates of relevance like *matter*.

Consider:

Veridicality w.r.t. P
(3a) It matters to Max that Iona stole his tobacco.
(3b) Iona stole Max's tobacco.

Veridicality w.r.t. Q
(4a) It matters to Max who stole his tobacco.
(4b) Iona stole Max's tobacco.
(4c) It matters to Max that Iona stole his tobacco.

While the inference from (3a) to (3b) is valid, the inference from (4a) to (4c) is not, because Max might fail to know that it was indeed Iona who stole his tobacco. Thus, *matters* is veridical w.r.t. P (the antecedent of **VU$_1$** is true), but it is not veridical w.r.t. Q (the consequent of **VU$_1$** is false). Thus, **VU$_1$** fails to hold, and **VU** as a whole is undermined.

In addition to discounting **VU**, Theiler et al. formulate another universal:

Clausal Distributivity (CD)[8]

[7] This framing of their conclusion is my own, but I believe it puts their point more transparently.
[8] Their formulation of the universal (found on p. 448) is a bit involved, so I have opted instead to borrow Roelofsen and Uegaki (2020)'s formulation.

x Vs Q iff there is an answer p to Q such that x Vs p

However, as pointed out by Roelofsen and Uegaki (R&U) (2020), predicates of relevance are equally a counterexample to **CD**. To show this, we can re-run the same example which I used above to discount $\mathbf{VU}_1$, but the essential common point is this: biconditional formulations of a universal on responsive predicates (which both **VU** and **CD** are) fail because the Q-to-P entailing direction fails[9]. Thus, to foreshadow Section 3 to come, in order to arrive at a universal which can overcome these counterexamples, we have to discount the problematic Q-to-P direction and restrict ourselves instead to the P-to-Q direction.

2.3 Steinert-Threlkeld (2020)

Before we turn to P-to-Q entailment, though, I would like to very briefly consider an interesting contribution by Steinert-Threlkeld in support of the stronger, bidirectional **VU**. He devised a computational experiment which aimed to show that verbs which satisfy **VU** **'**are easier to learn than those that do not' (Steiner-Threlkeld, 2020, p. 139). Unlike the previous two proposals, which have focused on formulating a rule (entailment pattern) that is meant to hold universally and examining *whether* it holds, Steinert-Threlkeld offers an explanation of *why* such a pattern should hold. Thus, **VU** and **CD**, though liable to counterexamples, have the empirical backing of a plausible justification for their existence.

3 P-to-Q Entailment

Picking up on the concluding idea of **Section 2.2**, in order to arrive at a more feasible universal constraint on clause embedding we have to get rid of the problematic Q-to-P direction of entailment and limit ourselves to the P-to-Q direction. As a result, Roelofsen and Uegaki (2020) propose P-to-Q Entailment as the best candidate universal on responsive predicates, given that it is able to overcome all of the counterexamples that undermined its predecessors. I will consider here the two counterexamples mentioned in the previous Section of this paper (for a more detailed list see R&U pp. 9-15) by assessing whether the following schema holds:

Schema for P-to-Q Entailment (PQE)
x Vs that P ⇒ x Vs that Q (provided Q is exhaustivity-neutral[10])

Starting with predicates of relevance, the inference from (5a) to (5b) seems valid:

(5a) It matters to Max that Iona stole his tobacco.
(5b) It matters to Max whether Iona stole his tobacco.

Likewise with the inference from (6a) to (6b) in regard to communication verbs:

(6a) Max told us that Iona stole his tobacco.

[9] This is slightly simplified. The refutation of **VU** in Roelofsen & Uegaki (2020), following Theiler et al. (2018), construes **VU** as pertaining to *exhaustivity-neutral* interrogative complements (i.e. complements like polar questions and *whether*-interrogatives, whose answers are taken from two-membered sets), whereas my example does not.
[10] Seeing as this is an explicit condition on **PQE**, I will adjust my examples to be exhaustivity-neutral.

(6b) Max told us whether Iona stole his tobacco.

Although R&U hold **PQE** to be robust, they concede that there are at least a handful of *prima facie* counterexamples that seem to violate **PQE**: namely, Buryat 'hanaxa', Turkish 'bil', and Tagalog 'magtaka'. However, they conclude that further study of these is required to provide a more wholistic picture. One result of such study might be that the semantics of these predicates can be re-conceptualized in a way that does conform to P-to-Q Entailment. Alternatively, P-to-Q Entailment could be modified such that it is able to successfully account for these counterexamples. Lastly, to reinforce the empirical robustness of P-to-Q Entailment, it would be useful to research into *why* **PQE** holds, in the same way that Steinert-Threlkeld (2020) argued **VU** holds due to ease of learnability.

4 P-to-Q Entailment in Slovakian

Having established the theoretical background to **PQE**, the following Section will present my preliminary findings from Slovakian and argue that there are no straightforward counterexamples to **PQE** in Slovakian. I will begin by outlining the methodology (Section 4.1) and then move on to discuss the results of my research (Section 4.2). I present several problematic predicates that resist a straightforward **PQE** judgement, either because of translation issues, the examples they appear in, or the complex presuppositional semantics of the declarative that seem not to transfer over into the interrogative. However, I conclude that none of these cases are straightforward counterexamples to **PQE**, and that further study is needed to resolve the cases which encode presuppositions.

4.1 Methodology

I have taken the list of paradigmatic responsive predicates graciously provided to me by Uegaki (2020, personal correspondence) as my departure point and translated them into Slovakian[11]. This generates a table with the following structure:

Table 1: *Layout of responsive predicate table.*

Category of predicate	Predicate V in English	Predicate V in Slovak	x Vs P	x Vs Q	Does "x Vs P" entail "x Vs Q"?

I have chosen 'že' as the closest Slovak equivalent of 'that', and 'i' as the closest Slovak equivalent of *whether.* Admittedly, some of the translations are a bit strained, but I can conceive of a scenario, however niche, when they could be plausibly uttered[12]. The example embedded clause I have used for P is 'že prišiel domov' [that he came home], and the one I have used for Q is 'či prišiel domov' [whether he came home]. In some cases, for the sake of readability (and to eliminate ambiguity when making judgements regarding entailment patters) I have made the matrix subject and the embedded subject differ in gender, and all the

[11] Translations are my own, though some have been corroborated by the go-to dictionary for Slovak-English translation found here: https://slovniky.lingea.sk/anglicko-slovensky.
[12] All grammaticality/well-formedness judgements are mine.

matrix verbs (the responsive predicates V) be in present tense and all the embedded verbs be in the past tense.

Here is a sample entry in the table:

Table 2: *Example entry in the responsive predicate table.*

assessment	*accept*	prijať	Prijímajú, že prišiel domov.	Prijímajú, či prišiel domov.	YES

And here is the translation of each of the two sentences from columns 4 and 5:

Table 3: *Sample translations of P- and Q-sentences.*

Declarative

Prijímajú	že	prišiel	domov
accept-3rd-PL	that	come-PAST-3rd-SG-MASC	home
They accept that he came home.			

Interrogative

Prijímajú	či	prišiel	domov
accept-3rd-PL	whether	come-PAST-3rd-SG-MASC	home
They accept whether he came home.			

4.2 Results & Discussion

Below is the complete, alphabetically ordered table of my findings:

Table 4: *Complete table of all responsive predicates tested in Slovakian.*

Category of predicate	**Predicate V in English**	**Predicate V in Slovakian**	**x Vs P**	**x Vs Q**	**Does "x Vs P" entail "x Vs Q"?**	**Notes/reason for judgement in previous column**
assessment	accept	*prijať*	Prijímajú, že prišiel domov.	Prijímajú, či prišiel domov.	1	
doxastic	agree with X [stative:	*súhlasiť*	Súhlasia, že prišiel domov.	Súhlasia, či prišiel domov.	1	

	to be in agreement]					
communication	announce	*vyhlásiť*	Vyhlasuje, že prišiel domov.	Vyhlasuje, či prišiel domov.	1	
communication	argue	*tvrdiť (?)*	Tvrdí, že prišiel domov.	*Tvrdí, či prišiel domov.	NA	V is not responsive
inquisitive	ask	*pýtať sa*	*Pýta sa, že prišiel domov.	Pýta sa, či prišiel domov.	NA	V is not responsive
communication	assert	*vyhlásiť*	Vyhlasuje, že prišiel domov.	Vyhlasuje, či prišiel domov.	1	
doxastic	assume	*predpokladať*	Predpokladá, že prišiel domov.	*Predpokladá, či prišiel domov.	NA	V is not responsive
doxastic	be certain	*byť si istý*	Je si istá, že prišiel domov.	Je si istá, či prišiel domov.	1	
doxastic	be convinced	*byť presvedčený*	Je presvedčená, že prišiel domov.	*Je presvedčená, či prišiel domov.	NA	V is not responsive
inquisitive	be curious	*byť zvedavý*	*Je zvedavá, že prišiel domov.	Je zvedavá, či prišiel domov.	NA	V is not responsive
emotive factive	be happy	*byť šťastný*	Je šťastná, že prišiel domov.	*Je šťastná, či prišiel domov.	NA	V is not responsive
doxastic evaluative	be right	*mať pravdu*	Má pravdu, že prišiel domov.	*Má pravdu, či prišiel domov.	NA	V is not responsive
emotive factive	be surprised	*byť prekvapený*	Je prekvapená, že prišiel domov.	*Je prekvapená, či prišiel domov.	NA	V is not responsive

doxastic	be unaware	*nebyť si vedomý*	Nie je si vedomá, že prišiel domov.	Nie je si vedomá, či prišiel domov.	1	
bouletic: negative be worried		*mať obavy*	Má obavy, že prišiel domov.	Má obavy, či prišiel domov.	???	
doxastic evaluative	be wrong	*mýliť sa*	Mýli sa, že prišiel domov.	*Mýli sa, či prišiel domov.	NA	V is not responsive
doxastic	believe	*veriť*	Verí, že prišiel domov.	*Verí, či prišiel domov.	NA	V is not responsive
predicates of relevance	care	*záležať (niekomu) (na niečom)*	Záleží mi (na tom), že prišiel domov.	Záleží mi (na tom), či prišiel domov.	1	
communication	claim	*vravieť*	Vraví, že prišiel domov.	Vraví, či prišiel domov.	1	
communication	complain	*sťažovať sa*	Sťažuje sa, že prišiel domov.	*Sťažuje sa, či prišiel domov.	NA	V is not responsive
decision	decide	*rozhodnúť sa*	Rozhodla sa, že prišiel domov.	Rozhodla sa, či prišiel domov.	???	
directive	demand	*domáhať sa*	*Domáha sa, že prišiel domov.	Domáha sa, či prišiel domov.	NA	V is not responsive
assessment	deny	*poprieť*	Popiera, že prišiel domov.	Popiera, či prišiel domov.	1	
dubitative	doubt	*pochybovať*	Pochybuje, že prišiel domov.	Pochybuje, či prišiel domov.	1	
doxastic	expect	*očakávať*	Očakáva, že prišiel domov.	*Očakáva, či prišiel domov.	NA	V is not responsive

communicat ion	explain	*vysvetliť*	Vysvetluje, že prišiel domov.	Vysvetluje, či prišiel domov.	1	
bouletic: negative fear		same as "be worried"	doxastic: change of state	forget	*zabudnúť*	Zabudla, že prišiel domov.
Zabudla, či prišiel domov.	1					
bouletic	hope	*dúfať*	Dúfa, že prišiel domov.	Dúfa, či prišiel domov.	1	
communicat ion	inform	*oznámiť*	Oznámila, že prišiel domov.	Oznámila, či prišiel domov.	1	
inquisitive	inquire	same as "ask"	inquisitive	investigate	*skúmať*	*Skúma, že prišiel domov.
Skúma, či prišiel domov.	NA	V is not responsive				
doxastic	know	*vedieť*	Vie, že prišiel domov.	Vie, či prišiel domov.	1	
doxastic: change of state	learn	*dozvedieť sa*	Dozvedela sa, že prišiel domov.	Dozvedela sa, či prišiel domov.	???	
directive	order	*prikázať*	*Prikazuje, že prišiel domov.	*Prikazuje, či prišiel domov.	NA	V is not responsive.
bouletic	pray	*modliť sa*	*Modlí sa, že prišiel domov.	Modlí sa, či prišiel domov.	???	
bouletic	prefer	*preferovať*	Preferuje, že prišiel domov.	*Preferuje, či prišiel domov.	NA	V is not responsive.
directive	propose	*navrhnúť*	*Navrhuje, že	*Navrhuje, či	???	

			prišiel domov.	prišiel domov.		
ratification	prove	*dokázať*	Dokázala, že prišiel domov.	Dokázala, či prišiel domov.	1	
emotive factive	regret	*mrzieť*	Mrzí ju, že prišiel domov.	*Mrzí ju, či prišiel domov.	NA	V is not responsive
perception	see	*vidieť*	Vidí, že prišiel domov.	Vidí, či prišiel domov.	1	
doxastic	suspect	*tušiť*	Tuší, že prišiel domov.	Tuší, či prišiel domov.	1	
communication	tell (someone)	*povedať (niekomu)*	Povedala (mi), že prišiel domov.	Povedala (mi), či prišiel domov.	1	
doxastic	think	*myslieť*	Myslí, že prišiel domov.	*Myslí, či prišiel domov.	???	
bouletic	want	*chcieť*	*Chce, že prišiel domov.	*Chce, či prišiel domov.	NA	V is not responsive.
communication: manner of saying	whisper	*pošepkať*	Pošepkala, že prišiel domov.	Pošepkala, či prišiel domov.	1	
inquisitive	wonder	*zaujímať (niekoho)*	Zaujíma ma, že prišiel domov.	Zaujíma ma, či prišiel domov.	1	
communication: manner of saying	write	*písať*	Píše, že prišiel domov.	Píše, či prišiel domov.	1	

Predicates that are unambiguously responsive abide by **PQE** and receive a 1 in the judgement column. However, predicates which are not straightforwardly responsive (i.e., they are either rogative or antirogative) receive a judgement of NA (not applicable), because **PQE** does not apply to them.

Let us now turn to a handful of problematic verbs, highlighted in grey in the table. These are, in the order that they appear in the table:

Table 5: *Potentially problematic predicates.*

bouletic: negative	be worried	*mať obavy*	Má obavy, že prišiel domov.	Má obavy, či prišiel domov.	???
decision	decide	*rozhodnúť sa*	Rozhodla sa, že prišiel domov.	Rozhodla sa, či prišiel domov.	???
doxastic: change of state	learn	*dozvedieť sa*	Dozvedela sa, že prišiel domov.	Dozvedela sa, či prišiel domov.	???
bouletic	pray	*modliť sa*	*Modlí sa, že prišiel domov.	Modlí sa, či prišiel domov.	???
directive	propose	*navrhnúť*	*Navrhuje, že prišiel domov.	*Navrhuje, či prišiel domov.	???
doxastic	think	*myslieť*	Myslí, že prišiel domov.	*Myslí, či prišiel domov.	???

4.2.1 Modliť sa (to pray) and navrhnúť (to propose)

The Slovakian counterparts to pray and propose — 'modliť sa' and 'navrhnúť' — are ill-formed with the complementizers 'že' or 'či'. Instead, these verbs require the subjunctive complementizer 'aby', which (as far as I am aware) does not have a direct counterpart in English. The closest approximation would be "so that" or "in order that". The predicate 'navrhnúť' (propose) only works with 'aby', where the resulting complement clause is declarative. Since this is the only construction it can figure in, the predicate is not responsive in Slovakian, and thus is not subject to **PQE** judgements (since these apply exclusively to responsive verbs).

The predicate 'modliť sa'' [pray], however, is not as easy to reason away. The declarative version requires 'aby' (rather than 'že') to be ill-formed, but, unlike 'navrhnúť' [propose], it is also well-formed in the interrogative version with 'či'). So the question becomes: Does 'Modlí sa, aby prišiel domov' entail 'Modlí sa, či prišiel domov'?

The issue is that the declarative 'aby' version has a future-oriented, wishful reading ("She prays that he would, in the future, come home"), whereas the interrogative 'či' version has a past tense, evaluative reading ("She prays about whether he had come home by now"). Thus, there is a mismatch in the tense that is projected by the predicate in the two sentences, which makes the judgement of whether P entails Q unstraightforward at best, and impossible at worst (how can a future event imply a past tense event?)[13].

4.2.2 Dozvedieť sa (to discover)

[13] I am assuming that, in order for us to assess whether **PQE** holds, the sentences have to be consistent in their tense-readings. If I am mistaken in making this assumption, then I believe *modliť sa* is by far the best candidate out of all the ones I discuss to mount a compelling challenge to **PQE**.

The next predicate, ‘dozvedieť sa’, I have flagged simply because it is not an exact translation of the English "learn". ‘Dozvedieť sa’ has the much wider reading of "learn new things", "discover", rather than the more restricted "absorb new information" reading that is more salient to the English version. If we grant my translation, however, **PQE** holds.

4.2.3 Myslieť (to think)

My example interrogative-embedding sentence with ‘myslieť’ [think] is ill-formed, but both the declarative and the interrogative example sentences can be modified by adding ‘už’ (which, roughly, means “already”) to make the latter grammatical:

Table 6: *Comparison of P- vs. Q-sentence embedding myslieť (‘think’) predicate.*

	Declarative	**Interrogative**
Old version	*Myslí, že prišiel domov.*	**Myslí, či prišiel domov.*
New version with "už"	*Myslí, že už prišiel domov.* (She thinks that he already came home.)	*Myslí, či už prišiel domov*. (She thinks whether he already came home.)

Adding ‘už’ makes the interrogative read like indirect speech, thus making it sound grammatical. To sum up: on the old version, ‘myslieť’ was not responsive so **PQE** did not factor in, and on the new version, **PQE** holds. Neither version thus poses a problem to **PQE**.

4.2.4 Mať obavy (be worried)

This predicate is problematic because the declarative version has the presupposition that the matrix speaker does not *wish* for the embedded subject to have come home, whereas the interrogative version has no such presupposition (it connotes straightforward worry about whether he did or did not come home). As such, since the wishing-for-a-negative-outcome presupposition is, at least in my judgement, absent from the interrogative, more research needs to be done into this predicate to unpack the semantics of its presupposition structure. As it stands, however, **PQE** is inconclusive on this case.

4.2.5 Rozhodnúť sa (to decide)

On its declarative version, this verb has performative force — as though the matrix speaker is deciding to undertake the action to come home. The interrogative version, on the other hand, creates the sense that the speaker is looking back and reflecting whether or not the embedded speaker has or has not come home. As such, this verb, just as ‘mať obavy’ in the previous Section, requires further study to be able to conclusively affirm or deny (or deem as inapplicable) **PQE**[14].

[14] Again, as in the case *modliť sa*, I am assuming that these issues can be remedied once the semantics of ‘mať obavy’ and ‘rozhodnúť sa’ are better understood. I am choosing to say, ‘this needs further investigation’ rather than saying, ‘this is a

5 Conclusion

This paper has hoped to serve a two-fold purpose: firstly, to present the P-to-Q Entailment universal and overview the literature that led up to it (Sections 2 and 3); and, secondly, examine whether P-to-Q Entailment holds in Slovakian (Section 4). The preliminary findings show that, within the verbs that I have tested for, there are no straightforward counter examples to P-to-Q Entailment.

I did, however, flag a handful of verbs which I believed deserve more detailed commentary. The first category includes verbs that require a subjunctive complementizer, which either disqualifies them from being responsive verbs (‘navrhnúť’) or else gives them differently tensed readings which are hard to evaluate whether or not they are P-to-Q Entailing (‘modliť sa’). The second category consists of verbs whose translations are either tentative (‘dozvedieť sa’), and so an affirmative P-to-Q Entailment judgement is made on shaky ground; or which require a slight modification to the example they feature in to make them responsive, though they ultimately nevertheless come out as P-to-Q entailing (‘mysliet’’). The last category are verbs whose declarative clauses have presuppositions (‘mať obavy’*)* or performative force (‘rozhodnúť sa’) which does not carry over into their interrogative counterparts, and as such require further study to unpack their semantics. The upshot of this paper, then, is that I have found no straightforward counterexamples to P-to-Q Entailment; I have only found cases which require further study and thus yield an inconclusive P-to-Q Entailment judgement.

6 References

Lahiri, U. (2002). *Questions and Answers in Embedded Contexts*. Oxford University Press.

Spector, B., & Egré, P. (2015). A uniform semantics for embedded interrogatives: An answer, not necessarily the answer. *Synthese,* 192(6), 1729–1784.

Steinert-Threlkeld, S. (2020). An explanation of the veridical uniformity universal. *Journal of Semantics* 37, 129–144.

Theiler, N., Roelofsen, F., & Aloni, M. (2018). A uniform semantics for declarative and interrogative complements. *Journal of Semantics,* 35(3), 409–466.

Roelofsen, F. & Uegaki, W. To appear. “Searching for a universal constraint on the possible denotations of clause-embedding predicates.” *Proceedings of SALT 30*.

Uegaki, W. (2019). “The semantics of question-embedding predicates.” *Language and Linguistics Compass*, 13(1).

counterexample to **PQE**’, because it is quite possible that there is something more complex going on in these verbs that, once better understood, might be elucidated in a way that is compatible with **PQE**.

Where is the Surfer? Where is the Hipster?: Spatial Relations in Southern Californian English vs. Pacific Northwestern English

Jaidan McLean
University of Oregon

Abstract. Looking at linguistic research on American English dialects there is no ignoring that the West is an under-researched region. In an effort to compensate for the lack of research, this pilot-sized empirical study will analyse semantic differences using figure-ground relations from Southern Californian English and Pacific Northwestern English. A figure-ground (FG) relation is where one entity, the figure, is being located in relation to another entity, the ground (e.g., 'the apple is on the table', 'apple' is the figure being located in relation to table, the ground) (Bowerman and Pederson, 1992). FG relations have been discussed extensively across languages, but not across dialects. Following the cross-linguistic approaches of Levinson and Meira (2003), this study uses FG relations to further the distinction between these two English dialects. The project used the BowPed (Bowerman and Pederson, 1992) pictures series that was created to elicit FG relation responses to gather data. Responses were collected from 10 speakers, five born and raised in the Los Angeles area, and five in the Seattle area. The results found 10 instances of split variation between the two regional dialects, falling on the locative used in the FG relations (e.g., 'the boat is in the water' vs. 'the boat is on the water'). The 10 differing locatives make an important addition to the discussion of West Coast English and pose the need for more research.

Keywords: topological relations; spatial semantics; regional dialects; American English

---

1 Introduction

The majority of approaches to researching spatial relations have been cross-linguistic, in order to compare and analyse the spatial cognition of separate languages and cultures, such as Levison and Meira's (2003) work. This paper will be following the approach to eliciting spatial relations as the Levinson and Meira (2003) publication, but with a regional dialectal twist. By narrowing the scope from cross-linguistic to within language analysis, I hope to expand the view of this type of semantic study to include more within-language variational approaches.

The spatial relations analysed here will be simple figure-ground (FG) relations, which is where one entity, the figure, is put in relation to another entity, the ground (Bowerman & Pederson, 1992). For example, in the phrase 'the apple is on the table', 'apple' is the figure entity that is being located in relation to the ground entity 'table'. In order to examine FG relations, a stimulus set called the 'BowPed' was designed with images that will elicit specific spatial relation responses (Bowerman and Pederson, 1992).

In taking the BowPed and same pilot-sized study approaches of Levinson and Meira (2003), this paper specifically elicited data from two Western English regional dialects, Southern Californian and Pacific Northwestern. Looking at responses from five participants born and raised in Los Angeles and five from Seattle, I hope to create more literature in the linguistic diversity of Western English regional dialects. There is no secret that American English dialects, for example, the South versus the North are going to have

great differences in linguistic variables, but what about what has been deemed 'the West?'. In an effort to further research and diversify the West region's English dialects, this paper will be focusing on answering whether Western English dialects differ on the basis of spatial relations.

2 Literature Review

Just as mentioned in the introduction, the work of Levinson and Meira (2003) is a prominent influence in this paper and in other research on spatial relations because of its significant findings in the nine languages analysed (Bohnemeyer et al, 2015). This work showed that FG relations differ quite a lot in languages. Although they differ significantly across language, there were four or so conceptual categories that were a theme in each language: containment, support, non-contiguity, and higher complexity (Levinson & Meira, 2003). These categories then branch into more specific concepts that were dependent on the language of topological relations received from participants.

The topological relations received from participants were elicited through the use of the BowPed, a stimulus set designed to test a language's spatial relations and potentially give insight to its speaker's spatial cognition (Bowerman & Pederson, 1992). This stimulus set will be expanded on later, as it is the same set I used in eliciting data for this study.

My participants for this study are from the Los Angeles and Seattle areas, two metropolitan areas that serve as significant locations for the regional dialect they represent. I chose these two locations because they are geographically at opposite ends of the West Coast, so I would like to assume they have the most potential in having varying topological relation responses. Along with this, other works have shown variation in Western English, specifically in vowel shifts. Literature such as Crosby and Dalola (2020) studying BAG-raising in Washington, show that there is vowel variation in these Western states. Along with this, there has been work done that shows linguistic variation beyond vowels (see Fridland et al., 2016), which beg the question for more research outside of vowel variation in Western English.

So if these regional dialects have more and more research showing that the Western English regional dialects are indeed linguistically different, how come a highly used American English map does not display any differences in the north and southern Western English? The map in question is the 'The Atlas of North American English' figure, shown below, where the entirety of 'The West' spans (geographically) from Montana down to New Mexico, and over to the Pacific Ocean (Labov et al., 2005).

Figure 1: *Atlas of North American English (Labov et al, 2005) American English Regional Dialects Map.*

3 Research Questions

In an effort to create more research on distinguishing Western English dialects, as well as more diversity in the research on spatial relations, there are two specific inquiries made going into this study. The first is will the FG relations between the Southern Californian English speakers and the Pacific Northwestern English speakers vary, specifically in their locative use? For example, 'the cup is on the table' versus 'the cup is above the table' where the locatives here, 'on' and 'above', are signifying a difference in spatial relation even though the figure 'cup' is still being located relative to the ground 'table'.

Secondly, if it is the case that participants in the two locations respond with a differentiating locative in their FG relation responses, is there a significant enough variation that could imply a need for further research in spatial relation between regional dialects, as opposed to between languages? Going into this project the hypothesis was that locatives between the two Western English dialects would vary, however to what extent was to still be determined.

4 Methodology

Throughout introducing this paper I have mentioned the BowPed and its design to elicit FG relations. More specifically though, the stimulus set was created with 71 images that include one specific highlighted item that lends itself to elicitation of certain locatives and FG relations. Below is an example image from the BowPed that I used to elicit relations of 'around' and 'on' distinctions. Along with the BowPed image, a simple spatial language question of 'Where is the __?' is used, where the blank refers to the highlighted item. The highlighted item is the posited figure that participants are then asked to put in relation to the ground, through the 'where' question. Also in this example image is the University of Oregon logo and a

text box for the participants' response, because the image shows the format in which I surveyed my participants. I created a survey through a program offered by the University of Oregon called Qualtrics. Qualtrics is an anonymous surveying platform that I was able to utilize in collecting responses easily from my participants living in the varying West Coast locations. The varying locations include the Los Angeles and Seattle areas, where I surveyed five participants from each. These participants were born and raised in the areas and have therefore grown up primarily around Southern Californian English or Pacific Northwestern (PNW) English. All of my participants are monolingual English speakers, and are friends and family that I have known from my time living in both of these locations. I was therefore able to easily text them a link to the Qualtrics survey with the BowPed stimuli along with instructions to "just say whatever comes to mind first as if I were someone who has not seen the picture and was asking you to answer 'Where is the (insert highlighted item here)?'.

Figure 2: *Example image of the BowPed used in Qualtrics.*

Once all 10 participants completed the survey, I was able to collect their responses into a spreadsheet for analysis. After organizing the responses with their corresponding BowPed image and question in the spreadsheet, I began to colour code. In Appendix One, there is a filtered example of the spreadsheet and the colour coding I used. I first coded for the six conceptual categories that Levinson and Meira proposed for one of the languages they discussed, Tiriyó, which included an expansion from the four I previously mentioned. The six categories I coded for are (with an example BowPed response):

IN, *the apple is in the bowl*
ON, *the apple is on the table*
ATTACHMENT, *the balloon is tied to the stick*
UNDER/OVER, *the ball is under the chair*
PROXIMITY, *the tree is next to the church*
MIDST, *the arrow is through the apple*

I underlined each locative used here, because this was the key component I was looking for in my coding. One issue that did occur was responses including two FG relations, such as 'the owl is in the hole in the tree', in which I chose to only code the first locative and its relative entities since this is ultimately the participant's initial thought. After coding the spreadsheet, I created two etic grids which allowed me to visualize the data a little better for each region. Appendix Two includes the Southern Californian English grid and Appendix B includes the PNW one. The etic grids were derived from the majority-ruling FG relation and its locative used, so if three of the five said 'the cloud is above the mountain versus two saying the cloud is over the mountain', then the majority response was used for further analysis.

5 Findings

After creating the etic grids and comparing them to the spreadsheet data, there were actually10 BowPed images that showed a split in the responses for each dialect. These 10 images can be found in Appendix Four, but the specific responses were:

Table 1: *Cases of disagreement for FG responses in each dialect.*

BowPed Question	Southern Californian	Pacific Northwestern
6. Where is the dog?	outside the doghouse	next to the doghouse
11. Where is the boat?	on the water	in the water/on the water
20. Where is the balloon?	tied to the stick	on the stick
26. Where is the crack?	in the cup	on the cup
36. Where is the cloud?	over the mountain	above the mountain
46. Where is the bandana?	on the person's head	around the person's head
48. Where are the rain drops?	outside the window	on the window
66. Where is the strap?	on the bag	attached to the bag/on the bag
67. Where is the owl?	in the tree	inside the tree
71. Where is the dog?	in the doghouse	inside the doghouse

For numbers 11 and 66 (which correspond to the BowPed image's number), the reason why Pacific Northwestern English has two responses is because there was a tie. I had some responses that were thrown out because they either did not include a FG relation or did not include a locative, which meant I had a tie with the remaining four eligible responses. Rather than tossing these variations out, I kept that as part of

the 10 because I feel that there even being a tie is somewhat significant and is something that I would like to further expand on going forward with this study.

Even though there were 10 significant variations in the locatives used, I will be focusing on three of them to discuss further. I am choosing these three because in coding and finding these data patterns, I realized I needed two additional conceptual categories to the six I adapted from Levinson and Meira (2003). I added the category of AROUND and INSIDE/OUTSIDE because of the responses for numbers 46, 67, and 71.

Figure 3: *BowPed image number 46 used to elicit "around" versus "on" distinctions.*

Number 46, as seen in the image below, asks ‘Where is the bandana?’ in which the 3/5 Southern Californian speakers said, ‘on the person’s head’, while 5/5 PNW speakers said, ‘around the person’s head’. Due to every participant from Seattle specifically using the locative ‘around’ rather than ‘on’ I believe there are more specific justifications for each locative’s use. Here, ‘around’ means that the figure is in contact with the ground, but more importantly it is encircling the ground. The encircling distinction leads me to believe that this category is separate from “on” in the cognitive stance of PNW English speakers.

Figures 4 and 5: *BowPed images number 67 and 71 used to elicit 'in' vs. 'inside' distinctions.*

The two other images, numbers 67 and 71, brought up a different distinction that involves the concept of containment. 67 asked, 'Where is the owl?' to which 4/5 Southern Californian participants responded, 'in the tree' while 3/5 PNW participants responded, 'inside the tree'. This same 'in' versus 'inside' pattern is present in number 71's 'Where is the dog?' to which Southern Californians preferred 'in the doghouse' and PNW 'inside the doghouse'. Both of the locative 'in' and 'inside' are in relation to the concept of containment, however it seems that the 'inside' locative may have a sense of encompassing or enclosing the figure, that PNW English speakers are taking into account more than the Southern Californian English speakers.

After these three images stood out, the project went back to the PNW English speakers to collect negative evidence on the less specific FG relations given by the Southern Californian participants. When I asked the PNW participants who had all said for number 46 that 'the bandana is around the person's head', if saying 'on the person's head' was acceptable, the majority said yes, with one exception. The exception being that 'on' does not specify the bandana is tied around and misses the concept of being a part of (or attached) to the head, but rather it is possible with saying 'on' that the bandana was simply placed upon the person's head. This specification seems to imply a semantic distribution of using a locative construction expressed with a verb and preposition in place of a simple locative construction consisting of just a preposition, that PNW English speakers seem to prefer.

Similarly, when asked if 'the owl is in the tree or the dog is in the doghouse' is an acceptable response to number 67 and 71, the PNW participants said yes. Once again though, the three who responded with 'inside' specified that they would not prefer 'in' only because of the encompassing factor once again, where the ground entity given is the location that the figure is enclosed by. Even more specifically one PNW participant described that simply saying 'in' for something like image 67 could refer to the owl being in the branches of the tree or amongst the leaves, which would not match the FG relation of the owl and the entirety of the tree trunk.

As for the SoCal participants, when I followed up with these speakers about whether ‘around’ and ‘inside’ are acceptable in these same FG relations, all five agreed those are perfectly acceptable substitutes. Two participants did note specifically that they dispreferred ‘inside’ for image 71 because the dog is still visible and has its paws not completely enclosed. One particularly said that had maybe just the dog’s tail been visible, they would have preferred ‘inside’ to ‘in’ since the majority of the dog would not be visible. Regarding the image 20 though, no participant had a specific reason as to why ‘tied to’ was not their first response. Although, one SoCal participant did joke that they ‘just use ‘on’ for everything’ referring to their locative construction choice of ‘on’ to ‘tied to’. Even though it was said in a joking manner, this could actually hint at a pragmatic difference in the Western English dialects, particularly Grice’s maxim of quantity (Grice, 1975).

6 Discussion

Although I hypothesized that there would be variation between the two regional dialects, I was not expecting this much to occur. I expected maybe one or two images would elicit significantly different responses between Southern Californian English and Pacific Northwestern English, but not that 10 images would. However, due this pilot-sized pool of 10 participants, I do still have to inquire whether these findings truly show a larger, more cognitive difference in semantic FG relations, or if these findings are just showing how flexible English locatives are. In furthering this study, I would prefer to sample from a large group and a more varied group of PNW and Southern Californian participants.

Returning to my research questions, my first one asking whether the FG relations expressed would show any difference in locatives seems to be quite true. Having so much locative variation between the two regional dialects, with 10 instances, I would argue emphasizes that there is greater linguistic variation occurring in Western English. Maps cannot simply be cut out to one large area, but rather need to be updated to show that regional dialects are present.

Along with solidifying more non-vowel research on regional Western English dialects, I feel that this variation opens up a new avenue for approaching spatial relation work. The amount of variation occurring in the locatives used for expressing FG relations of these within a language leads me to question how stable the data from cross-linguistic studies can be. If there seems to be a fair bit of semantic — or even pragmatic — differences occurring within a language, then should that be taken into account more when making claims about cross-linguistic spatial language variation?

I note that perhaps even pragmatic differences are occurring, based on the negative evidence collected from the SoCal participants. As mentioned, Grice’s maxim of quantity, where a speaker tries to be as informative as possible while also not giving more information than actually needed, could be the linguistic explanation for the SoCal participant’s comment (Grice, 1975). The generalization of ‘on’ in the Southern Californian English spatial relations could be motivated by this need for answering as detailed as possible in as little speech as possible. Similarly, this led me to think about the specificity of the PNW speakers and that perhaps a similar explanation relating to Grice’s maxim of quality, where a speaker tries to be truthful and not give responses not supported by some evidence (Grice, 1975). In other words, in giving a specific ‘tied to’ response in place of the SoCal’s ‘on’, the PNW participants are drawing on the fact that the balloon is attached and that it is a truthful expression.

Ultimately, these sorts of conclusions are more of a discussion than concrete findings due to the size and depth of the study. Perhaps this pragmatic-semantic-regional trifecta is something that can be further researched by linguists in regards to dialectal spatial language, because it is something that 10 people alone

cannot necessarily dispute. Regardless of the size of the study though, it is apparent that the need for more Western English research is real since more and more linguistic variation is being uncovered.

7 References

Bohnemeyer, J., et al. (2015). The Contact Diffusion of Linguistic Practices, *Language Dynamics and Change*, 5(2), 169-201. doi: 10.1163/22105832-00502002.

Bowerman, M. and Pederson, E. (1992). Topological relations picture series. In Stephen C. Levinson (ed.), *Space stimuli kit* 1.2, 51. Nijmegen: Max Planck Institute for
Psycholinguistics.

Crosby, D. and Dalola, A. (2020). Begging for bags: BAG-raising and prescriptive ideologies in Spokane Washington. Proceedings of the Linguistic Society of America, 5(1): 354-368.

Fridland, V., Kendall, T., Evans, B., and Wassink, A. B. (2016). *Speech of the Western States. Volume 1: The Coastal States*. Publication of the American Dialect Society, 101. Durhan, North Carolina: Duke University Press.

Grice, P. (1975). Logic and Conversation. In Cole, P. and Morgan, J. (eds.) *Syntax and semantics. 3: Speech Acts*. New York: Academic Press, 41-58.

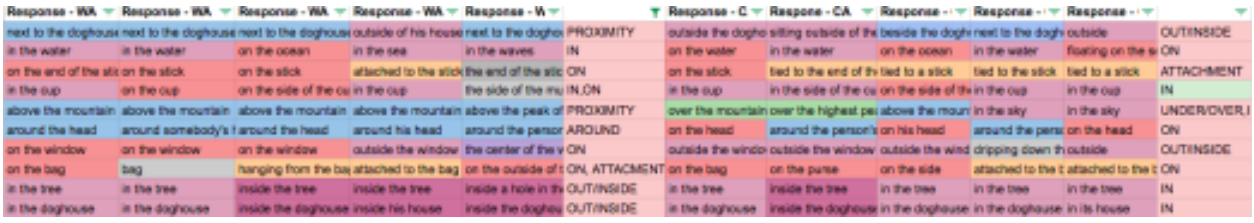
Labov, W., Ash, S., and Boberg, C. (2005). *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin and New York: Mouton de Gruyter.

Levinson, SMeira, S., and Max Planck Institut fur Psycholinguistik Nijmegen, N.L.C.G. (2003). 'Natural Concepts' in the Spatial Topological Domain--Adpositional Meanings in
Crosslinguistic Perspective: An Exercise in Semantic Typology. *Language* 79(3), 485-516.

8 Appendices

8.1 Appendix One

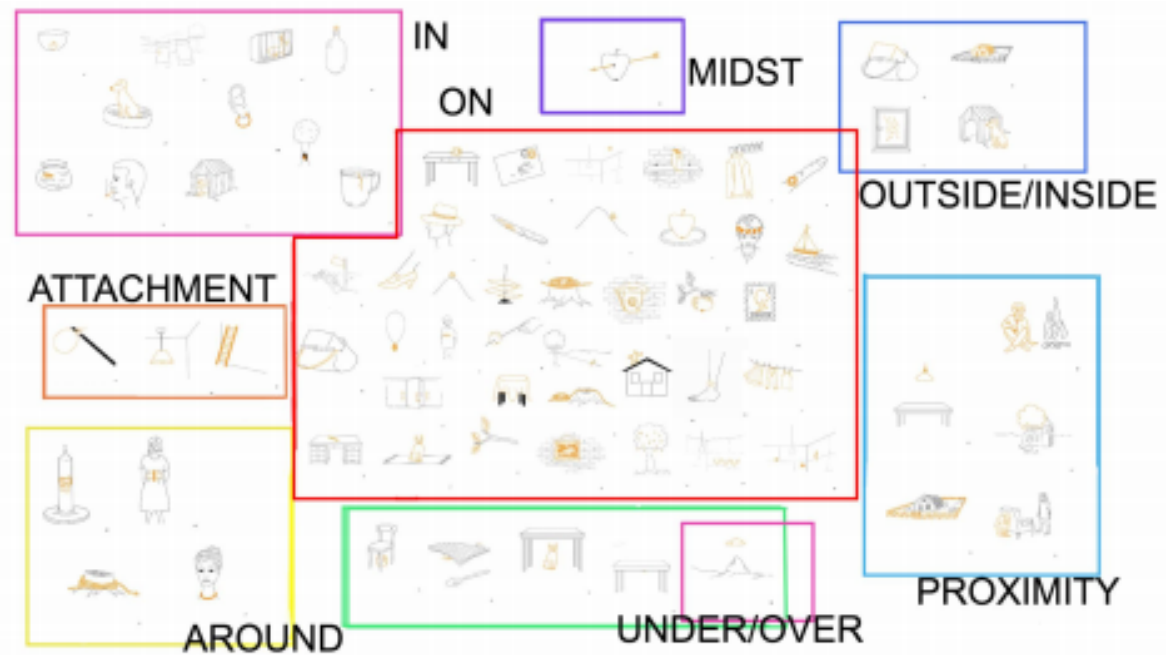
Color-Coded Spreadsheet Example



Response - WA	Response - WA	Response - WA	Response - WA	Response - W		Response - C	Respons - CA	Response -	Response -	Response -	
next to the doghouse	next to the doghouse	next to the doghouse	outside of his house	next to the doghou	PROXIMITY	outside the dogho	sitting outside of the	beside the dogh	next to the dogh	outside	OUT/INSIDE
in the water	in the water	on the ocean	in the sea	in the waves	IN	on the water	in the water	on the ocean	in the water	floating on the s	ON
on the end of the stic	on the stick	on the stick	attached to the stick	the end of the stic	ON	on the stick	tied to the end of th	tied to a stick	tied to the stick	tied to a stick	ATTACHMENT
in the cup	on the cup	on the side of the cu	in the cup	the side of the mu	IN,ON	in the cup	in the side of the cu	on the side of the	in the cup	in the cup	IN
above the mountain	above the mountain	above the mountain	above the mountain	above the peak of	PROXIMITY	over the mountain	over the highest pe	above the moun	in the sky	in the sky	UNDER/OVER,I
around the head	around somebody's h	around the head	around his head	around the person	AROUND	on the head	around the person's	on his head	around the pers	on the head	ON
on the window	on the window	on the window	outside the window	the center of the v	ON	outside the windo	outside the window	outside the wind	dripping down th	outside	OUT/INSIDE
on the bag	bag	hanging from the ba	attached to the bag	on the outside of t	ON, ATTACMENT	on the bag	on the purse	on the side	attached to the t	attached to the t	ON
in the tree	in the tree	inside the tree	inside the tree	inside a hole in th	OUT/INSIDE	in the tree	inside the tree	in the tree	in the tree	in the tree	IN
in the doghouse	in the doghouse	inside the doghouse	inside his house	inside the doghou	OUT/INSIDE	in the doghouse	inside the doghouse	in the doghouse	in the doghouse	in its house	IN

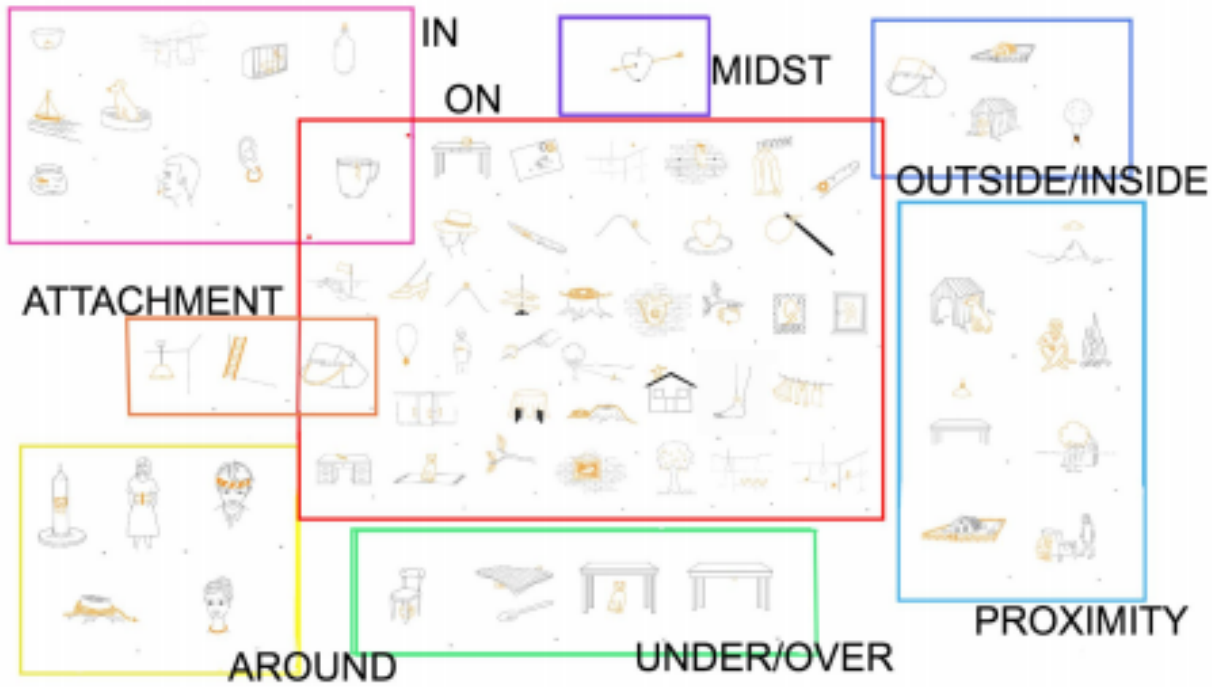
8.2 Appendix Two

Southern Californian English Etic Grid



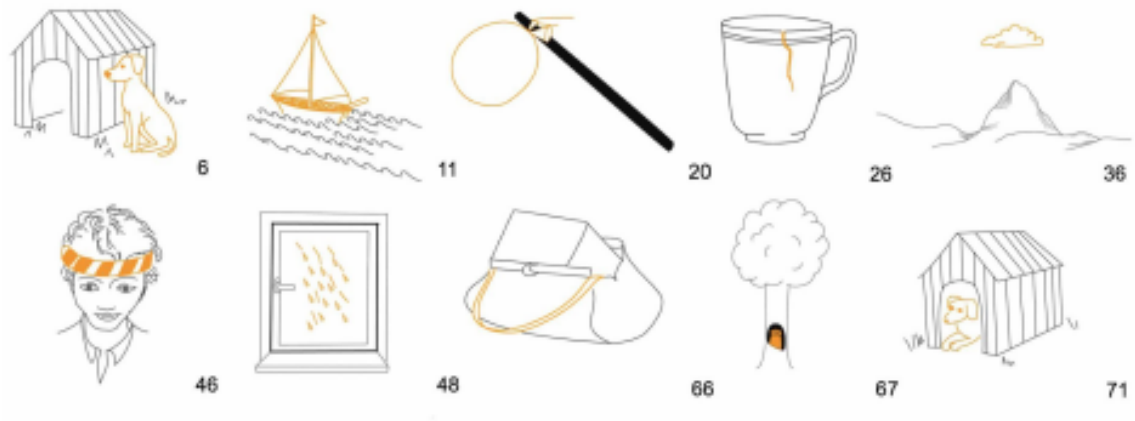
8.3 Appendix Three

Pacific Northwestern English Etic Grid



8.4 Appendix Four

Variation in BowPed Responses



The Discursive Construction and Performance of Non-Binary Identity

Angel Garmpi
University of Edinburgh

Abstract. This study examines the discursive construction and performance of non-binary identities in the context of lived experience narratives. It uses data from semi-structured interviews with seven 18–25-year-old AFAB participants who identified as non-binary. The study uses a broad thematic analysis (Bradford et al. 2019) to contextualise a discursive analysis (Corwin 2009) into how participants manipulate the semiotics (Silverstein 2003; Jaffe 2016; Gal 2016) of implicitly and explicitly gendered lexicon (following Zimman 2014; Zimman 2017a). Results show that participants sought to legitimise their self-identification by separating identity and embodiment (Zimman 2017a), and to create a non-binary inclusive ideology legitimising their identities (Corwin 2009; Darwin 2017). As the discursive construction of non-binary identities was also shown to be informed by the complex interaction of gender, embodiment, and sexuality (Connel 1995; Cameron 1998; Kiesling 2002; Eckert 2011; Zimman 2013) through multiple instances of iconisation and indexicality (Gal 2016; Jaffe 2016), this analysis also provides insight into the construction of normative gender ideology (West & Zimmerman 1987, 2009; Butler 1993). It is argued that participants engaged with this ideology to discursively construct and perform their identities, both resisting and affirming it while ultimately seeking to emphasize individuality and personhood against the restrictions of binary gender. Thus, this study concludes that its findings indicate a need for more research within a sociocultural linguistics framework (Bucholtz & Hall 2016), where embodiment and the physical body are seen as central to the production, perception and social interpretation of language.

Keywords: discourse (analysis); gender; identity; indexicality; sociocultural linguistics

---

1 Introduction

Non-binary individuals may identify as both male or female, as neither, or as being altogether outside the spectrum of binary gender (Darwin, 2017). Their experiences are different to those of individuals with binary trans identities (Goldberg & Kuvalanka, 2018; Bradford et al., 2019), while their identities have been shown to be the sites of intense semiotic and stylistic activity (Corwin, 2009; 2017; Darwin, 2017), and to involve collective co-construction especially in online communities (Darwin, 2017; Opperman, 2019). Sociolinguistic research has shown that language is central in the construction of non-binary identities. Non-binary individuals strategically employ or resist stereotypically gendered prosodic cues in accordance with their own identity (Corwin, 2009; Shar & Nycz, 2018; Steele, 2019). Furthermore, research has shown that language is crucial for the self-determination of non-binary individuals. The treatment of pronouns as not given until confirmation, the use of gender-neutral pronouns, and the proliferation of different gender identity descriptors (such as agender, genderqueer, genderfluid) (Corwin, 2009; Darwin, 2017; Zimman, 2017a) forefronts individuality and creativity, and broadens the linguistic realm for self-identification (Zimman, 2017a). Furthermore, analysis of non-binary individuals' narratives (Corwin, 2009)

showed that they manipulate and negotiate culturally recognisable gender categories in order to construct a non-binary identity against the wider context of normative, binary gender.

However, to date, there has been no systematic research into how non-binary identities are performed (Butler, 1990; 1993) in narrative discourse. Narrative is crucial for examining the construction of identity in general and of gender in particular because it allows individuals to construct edited versions of themselves and forward the most salient parts of their identities (Thornborrow & Coates, 2005; Benwell & Stokoe, 2006). Therefore, this study aims to address this gap in the literature. It uses a broad thematic analysis (Bradford et al., 2019) in order to contextualise a discursive analysis (Corwin, 2009) of the narratives of seven non-binary individuals. As the way in which speakers manipulate indexical meanings can be considered a performative use of language (Barrett, 2009; Calder, 2019) and indexicality depends on the cultural and so ideological context of a society (Ochs, 1996), I argue that examining the indexicality (Silverstein, 2003; Jaffe, 2016; Gal, 2016) of implicitly and explicitly gendered lexicon (following Zimman, 2014; Zimman, 2017a) in narrative discourse can reveal both how non-binary gender is performed, but also point to the wider ideological framework of gender and embodiment.

The results show that the semiotic manipulation of implicitly and explicitly gendered lexicon in narrative discourse is crucial for the construction and performance of non-binary identity. Through it, participants legitimised their self-determination and gender self-identification (Zimman, 2017a), but also created a non-binary inclusive ideology able to legitimise the experiencing and expression of their identities (Corwin, 2009; Darwin, 2017). In narrating their experiences as non-binary individuals, participants also highlighted how their gender identity interacted with, or was informed by, embodiment and sexuality. These identities were related through multiple instances of iconisation and indexicality (Gal, 2016; Jaffe, 2016) in the narrative discourse, which reveals crucial links between gender, embodiment, embodied practices, and sexuality in normative gender ideology (West & Zimmerman, 1987, 2009; Butler, 1993). I argue that this analysis ultimately shows that in discursively constructing and performing their non-binary identities, participants engage with normative gender ideology. Therefore, in contributing to the literature examining how non-binary identities are discursively constructed and performed, this study also offers crucial insights into how normative gender ideology is constituted. I argue that this shows the need for more research within a sociocultural linguistics framework (Bucholtz & Hall, 2016), where embodiment and the physical body are seen as central to the production, perception, and social interpretation of language.

2 Literature Review

Early work on language and gender operationalised gender as a pre-defined, binary category socially equivalent to a speaker's sex. In feminist work, differences in the speech of so defined 'women' and 'men' were analysed as reflecting social differences and power imbalances between these categories of speakers (Lakoff, 1973; Cameron, 1992). Subsequent sociophonetic (Podesva & Kajino, 2014) research on language, gender and sexuality adopted a broadly Labovian (after Labov, 1966) or 'first-wave' (Eckert, 2012) analytical perspective, in that pre-defined social categories, such as gender, were seen as possible predictors of linguistic variation.

In contrast, the present paper adopts a 'third-wave' perspective (Eckert, 2012), which theorises language as what creates, and not only describes, the social context (Fairclough, 1989; Cameron, 1998; Gal, 2016; Jaffe, 2016). Consequently, in this perspective the language of speakers is analysed as what dynamically creates, not merely reflects, their social identities. In the research of language and gender, in particular, this is encoded in the concept of gender performativity. Originating in the work of postmodernist

queer theorist Judith Butler (Butler, 1990; 1993), gender performativity challenges the perception of gender as a constantive, descriptive social category that follows from a person's sex. Instead, it proposes that gender is a kind of 'doing', achieved when an individual repeats acts associated with appropriate gendered behaviour in appropriate contexts. Crucially, gender performativity theory positions discourse as central in the constitution and performance of gender (Butler, 1993). That is, statements of gender, such as 'It's a girl' do not merely describe pre-existing gender categories but create them, as well as the gendered expectations that accompany them (Livia & Hall, 1997). Therefore, 'third-wave' sociolinguistic research into how speakers use language to construct a gender identity can also be considered as research into the performativity of gender in language.

Research on the performative uses of language can add to the knowledge of how linguistic means constitute social meaning, as well as reveal the socio-cultural norms regulating gender. Butler (1993) argues that subversive gender performances in particular, such as drag, can maximally expose the performative nature of gender. That is because gender performativity reproduces the socio-cultural norms which regulate it. Thus, these norms become congealed over time, obscuring the performative nature of gender and thus producing the illusion that gender is 'natural' (Butler, 1990). However, while 'subversive' gender identities which have been extensively studied include drag queens (Barrett, 2009; Mann, 2011; McKinnon, 2017; Barrett, 2017; Calder, 2019), and transgender individuals (Zimman, 2013; 2017b; Cashman, 2018; Borba, 2019), non-binary identities remain understudied in sociolinguistics. Therefore, this paper uses a third-wave sociolinguistic framework in order to research how non-binary individuals use language to perform non-binary gender in discourse.

Existing research shows that non-binary identities merit research in their own right. Firstly, their experiences have been shown to differ to those of individuals with binary trans identities (Goldberg & Kuvalanka, 2018; Bradford et al., 2019). Specifically, non-binary identities encompass a broad range of identities, provide room for gender fluidity, and resist narratives of both binary gender and transnormativity, while non-binary individuals face unique challenges in terms of communicating and legitimising their identity in different settings, both within and outwith the LGBTQ+ community. In addition, non-binary identities have been shown to be the sites of intense semiotic and stylistic activity (Corwin, 2009; 2017; Darwin, 2017), and to involve collective co-construction especially in online communities (Darwin, 2017; Opperman, 2019), where non-binary individuals engage in the promotion of literacy around and visibility of, their identities, often challenging binary gender assumptions.

In particular, sociolinguistic research has shown that language is central in the construction of non-binary identities. Sociophonetic work suggests that non-binary individuals strategically employ or resist stereotypically gendered prosodic cues in accordance with their own identity. Corwin (2009) found that genderqueer individuals in her study tended to combine stereotypical prosodic indicators of masculinity and femininity (small pitch ranges and frequent high-rise terminals respectively) to index a non-binary identity, whereas Shar and Nycz (2018) showed that intonation in masculine-aligning, as opposed to feminine and neutral aligning genderqueer individuals follows patterns shown to generally index masculinity. Similarly, Steele (2019) showed that non-binary speakers selectively used some prosodic resources but not others to perform gendered speech styles (such as a higher F0 to index femininity). This research also suggested that Black non-binary speakers used these linguistic variables differently than their white counterparts, which foregrounds the need for an intersectional approach to sociolinguistic research on non-binary identities (Crenshaw, 1989).

Language has also been identified as an important tool for the self-determination of non-binary identities. Pronouns are a salient example (Corwin, 2009; Darwin, 2017; Zimman, 2017a). Non-binary

individuals may alternate between binary pronouns, or use gender-neutral pronouns, such as singular they/them, or coined alternative pronouns such as ‘ze/hir’, or ‘xe/xem’. In addition, they treat pronouns as non-given, until the person in question is asked about the ‘pronouns they use’. Pronoun etiquette is particularly interesting, linguistically, because even knowing an individual’s gender identity is not seen as sufficient for determining their pronouns (Zimman, 2017a). Zimman argues that this division of gender identity terms from gendered pronouns creates linguistic fragmentation, in that each level of gendered language offers a separate realm for self-identification. Gender identity labels are another example of linguistic innovation and creativity. There is a proliferation of different labels within the wider non-binary community, including for example agender, genderqueer, genderfluid, and others (Darwin, 2017). Crucially, these labels are often used as fuzzy concepts, in that how a person self-identifies is perceived as a matter of ‘experiential legitimisations, primarily grounded in affect’ (Zimman, 2017a, p. 233). That is, a matter of how expressing an identity and having it recognised by others ‘feels’. Therefore, this research suggests that linguistic innovation is crucial in the work of identity expression for non-binary individuals. At the same time, however, self-determination and its linguistic expression are in turn in a tenuous relationship with embodiment.

In the framework of sociocultural linguistics (Bucholtz & Hall, 2016) embodiment and the physical body are seen as central to the production, perception and social interpretation of language. As such, the language of individuals with particular kinds of bodies can license or challenge different social meanings (Calder, 2019). The unique challenges regarding the embodiment of non-binary identities can best be understood through the ‘doing gender’ model (West & Zimmerman, 1987). It argues that people are accountable to other’s gender reference points, and therefore that the successful achievement of gender depends on others' recognition of that gender. However, as the physical body and sex categories are fundamentally a part of ‘doing’ gender (West & Zimmerman, 2009), the recognition of gender is a process relying heavily on visual cues (Darwin, 2017), which includes embodied cues beyond those which are gender expression choices. Messerschmidt (2009) then argues that a primary struggle of non-binary embodiment is that strangers often miscategorise non-binary individuals as masculine girls or feminine boys, as a result of not recognising non-binary gender when they see it. Research has highlighted some of the ways in which non-binary individuals use gendered embodied cues alongside linguistic ones such as those outlined above against this wider context, in order to express their gender identity and also make it recognisable to others. Darwin (2017) conducted an online ethnography of a genderqueer community on Reddit from an interactionist and discourse analytic perspective. In analysing the posts and selfies of different genderqueer users, she found that there was significant variation in how they ‘did gender’: some individuals strategically did binary gender in terms of embodied cues such as clothing and posture to ‘pass’ as their desired binary gender each time, whereas those who resisted binary gender attribution employed androgyny in their expression. Darwin argues that non-binary individuals therefore challenged binary gender accountability, as well as tried to shift binary gender ideology to recognise them. Similarly, Corwin’s (2009) ethnographic research showed that genderqueer individuals agentively chose the gendered signs (such as clothing, use of makeup, posture, gesture) they were displaying day-to-day in order to express their gender fluidity or androgyny. Thus, this research supports the relevance of an integrated, sociocultural linguistic framework to the research on the language of non-binary individuals, as it shows that language use, and linguistic self-determination of non-binary individuals is both shaped and constrained by embodiment and gender expression (Bucholtz & Hall, 2016; Zimman, 2017a).

Therefore, existing research shows that the construction of non-binary identities involves both linguistic means, from prosodic cues to pronouns and gender identity terms, and embodied resources.

However, there has been little research into how non-binary identities are performed through narrative. Narrative is crucial for examining identity construction because it allows individuals to construct edited versions of themselves, and forward the most salient parts of their identities (Benwell & Stokoe, 2006). Furthermore, Thornborrow and Coates (2005) argue that narrative is key to the construction of gender and link narrative discourse to gender performativity. For example, Coates (2005) showed that how male participants in her study were willing to co-construct narratives with their female partners as a performance of hegemonic masculinity through heterosexuality. Even though these narratives were told by more than one narrator, Ward (2019) argues that gender construction in narratives can also be seen in interview discourses. Specifically, in her study with trans participants, she shows that overt gender performativity can be accessed via specific questions on the participants' gender identities, and covert performativity through an analysis of how participants tell their stories and relate to the wider narratives, such as those of binary gender and transnormativity. With regards to non-binary identities, Corwin (2009) examined narrative extracts with an interview with a genderqueer individual, 'Atlas', from a discourse analytic perspective. She showed that Atlas manipulated culturally recognisable gender categories and voiced dialogues where their voice and an Other's voice negotiated identity, in order to construct a non-binary identity against the wider context of normative, binary gender. However, to date, there has been no systematic narrative discourse analysis for non-binary identities. Therefore, this paper aims to contribute to filling this gap in the literature by examining how non-binary identities are performed in the discourse of lived experience narratives.

Specifically, the scope of the present discourse analysis focuses on implicitly or explicitly gendered lexicon, and how it serves gender performativity in narrative. Explicitly gendered lexicon includes words which semantically entail gender (e.g., as 'woman', 'girl', 'female', 'feminine'), whereas implicitly gendered lexicon involves terms which index certain kinds of gendered personalities (e.g., 'dude', 'bitch', 'beautiful' etc.) (Zimman, 2017a). This lexicon can be seen as both describing and creating different kinds of social categories. In the case of gender, however, this lexicon also evokes embodiment and different kinds of bodies. That is, gendered terminology is not only constitutive of the social category of gender, but also relies on the discursive construction of sex (see Butler, 1993). For example, in analysing the use of genitalia terminology in an online community of transgender men, Zimman (2014) showed that sex is not a 'natural' category vis-à-vis the social nature of gender, but is also shaped in the discourse.

Lexica pertaining to social categories have been used in discourse analysis, in particular under the scope of membership categorisation analysis (MCA; Sacks, 1992). MCA seeks to examine how speakers in a society categorise themselves, and therefore how the social world is organised through the particular actions or characteristics linked to social categories (Stokoe, 2003). Crucially, MCA emphasises that social category terminology is inferentially rich, meaning that they are stores of social knowledge for the culture that uses them (Sacks, 1992). For example, categories such as male and female hold assumptions about the physical body, as well as cultural and social assumptions based on how people perceived as having these bodies are supposed to act and what role they are supposed to play in society (Ward, 2019). From a sociolinguistic perspective, this suggests that implicitly or explicitly gendered lexicon has a rich indexical potential. As the way in which speakers manipulate indexical meanings can be considered a performative use of language (Barrett, 2009; Calder, 2019), I propose that examining the indexicality of gendered lexicon in narrative discourse can offer valuable insights into the performativity of gender and the construction of gender identity.

Indexicality ('pointing, contiguity, co-occurrence' (Gal, 2016, p. 120)) is what mediates between language and gender as linguistic signs (e.g., words, pronunciations, syntactic constructions) become

gendered through indirect association with social meanings (Barrett, 2009), and especially because only few linguistic features directly indexing gender (e.g., third-person pronouns and address titles) (Ochs, 1992). Furthermore, linguistic signs are indexically mutable (Eckert, 2012), that is, able to change meaning over time, and to have multiple meanings, organised in an 'indexical field' (Eckert, 2008). How a linguistic sign comes to index social meaning has been described through the 'indexical order' (Silverstein, 2003): where at first a linguistic form is linked to a population or a stance, it can later acquire additional social meaning (higher-order indexicals) outside the initial context, and so for example index membership of a population, or stances attributed to it. Once in place, these higher order indexicals can dominate the indexical interpretation of a linguistic sign (Jaffe, 2016). Thus, the indexicality of gendered terminology can offer a dynamic access into the social space of gender.

Crucially, an analysis of indexicality also potentially allows for an examination of the ideology of the social context within which it operates. Firstly, the indexical valence of a form derives from a history of usage and cultural expectations (Ochs, 1996). Furthermore, indexicality is one of the ways in which signs can be related to the objects they represent. If indexical association becomes conventionalised, then the relationship of the sign to its object becomes naturalised or iconic, in a process called iconisation or rhematisation (Jaffe, 2016; Gal, 2016). This process is ideological, in that whether a sign is interpreted as an index, icon, or both depends on the ideological framework of the interpretant (Jaffe, 2016). Thus, as indexicality depends on the cultural and so ideological context of a society, examining the indexicality of gendered lexicon can reveal both how non-binary gender is performed, but also aspects of the wider ideological framework surrounding gender and embodiment.

Finally, this study aims to further enrich the scope of its analysis of narrative discourse in the following two ways. Firstly, by adding sexuality to its joint consideration of gender and embodiment. Butler's heterosexual imperative asserts that normative gender expressions and the discursive construction of binary sex are further founded on heteronormative sexuality (Butler, 1993). This has also been shown linguistically (Connel, 1995; Cameron, 1998; Kiesling, 2002; Eckert, 2011; Zimman, 2013). For example, Eckert (2011) showed how sound change in girls of preadolescent communities coincided with the emergence of the heterosexual market, arguing for an integrated approach to gender and sexuality. Thus, it can be argued that non-normative gender identities, such as non-binary identities, will also be the locus for the contestation and re-imagination of sexuality and desire, and that this will be reflected in their performative use of gendered terminology. Secondly, this study aims to further enrich the scope of its analysis of narrative discourse by contextualising it against a broad thematic analysis. As gender is only a part of the identities of the participants, narrative analysis can also allow an examination of their lives 'holistically through the stories they tell' (Benwell & Stokoe, 2006, p. 143). Thematic analysis can therefore supplement a discursive analysis of narratives, by shedding light on the wider context in which these identities are constructed. Existing research on genderqueer individuals' lived experience narratives led to the identification of three main themes. Namely: how participants subjectively experienced their identities, how they experienced their identities within the broader social context and narratives of transnormativity and binary gender, and finally, how they navigated these master narratives and created their own about being genderqueer (Bradford et al., 2019). The present research will therefore also consider the broader themes emerging in the narratives of non-binary individuals in order to provide a context against which to carry out discourse analysis.

3 Methods

3.1 Participants

The present research and methodology have been approved by the Research Ethics Committee of the School of Philosophy, Psychology, and Language Sciences at the University of Edinburgh.

Participants were recruited through a mixture of online advertising and word of mouth, or through the author's own personal connections. Online advertising comprised posts on Twitter and Facebook, primarily in the relevant groups or pages of the University of Edinburgh's LGBTQ+ society (PrideSoc). The aim was to recruit participants who explicitly identified with a non-binary identity (non-binary, genderfluid, genderqueer etc.), and who were at the time of the study, students between 18 – 25 years old.

The first interview conducted was a pilot which has not been included in the analysis. Nine subsequent interviews were conducted in Edinburgh over the span of 5 months (June 2019 – October 2019). Of these, two (n=2) have been excluded from the present analysis because neither participant primarily self-identified with a non-binary identity descriptor.

Of the seven participants included in the present study, all were assigned female at birth. They ranged in age from 20 to 24 (M= 21.3, SD=1.6). Six (n=6) out of seven participants were white, one (n=1) had a Latinx heritage and background (Jem). With regards to educational background, all were studying or had studied at University level. Participants were asked where they grew up, instead of their nationality. As the study focuses on narratives of lived experience, this was done to avoid conflating nationality with each participants' individual experiences. Three out of seven participants (n=3) reported growing up in the USA, with the rest (n=4) having grown up in European countries. When asked to identify their sexual orientation and attraction (distinction following Bradford et. al., 2019), participants gave 5 unique orientation terms. Five out of seven (n=5) expressed an identification with a descriptor on the asexual spectrum. Complete demographic information can be seen in Table 1.

Table 1: *Participants' demographic information.*

Name	Pronouns	Age (years)	Gender Identity	Orientation	Attraction	Grew up in
Sam	They/them	22	Non-binary/ questioning	Questioning/ femme people	Demisexual	Netherlands
Alex	They/them or she/her	21	Non-binary	Gay	On asexual spectrum, demiromantic	Greece
Jem	They/them	20	Non-binary genderfluid	Queer	Aromantic asexual	USA
Adam	He/him and they/them	21	Non-binary trans guy/ non-binary transmasculine	-	Aromantic asexual	USA
Luuk	They/them	20	Non-binary/ agender	Bisexual (questioning)	-	Netherlands

Annie	She/her	24	Non-binary genderfluid	Straight	On asexual spectrum	Wales
Leslie	They/them	21	Non-binary/ gender non-conforming	Gay	-	USA

It is noted here that the descriptors used in the above table are those the participants themselves used during their interviews. All names are pseudonyms which respect the gender identity of the participants.

3.2 Procedure

For the purposes of this study narratives of personal experience were collected via interviews (following Corwin, 2009; Ward, 2019). Qualitative analysis of interview data can offer unique 'insight into the intentions, feelings, purposes and comprehensions of the interviewee', as well as how they interpret themselves and the topic in question (Cruickshank, 2012, p. 42). Cameron (2001) argues that the research interview is a speech event in itself, and thus a unique situation in which participants may present themselves as a certain kind of person. Thus, qualitative analysis of interview data is a valuable tool for examining gender identity construction, as long as the context in which the data was elicited is taken into account (Ward, 2019).

Semi-constructed interviews were carried out with each participant individually. Interviews were recorded using a Zoom H2n portable recorder. They lasted between 28 and 80 minutes (M= 52 SD= 17.8). The author had a script of questions available at all times (see Appendix Four), but elaboration and free discussion was encouraged, and practiced, in all interviews. This aimed to create a safe, friendly environment, where participants felt comfortable to express and discuss their gender identity. It also allowed participants to avoid any topics they might feel uncomfortable with. Therefore, this methodological choice also helped mitigate the observer effect. Furthermore, in order to allow participants time to settle into the recording process smoothly, demographic questions were asked first, followed by general, conversation-starting questions (for example about participants' experiences of the city of Edinburgh). Each participant was fully briefed and debriefed before and after each interview.

Participant observation in friend or university groups of non-binary individuals was also attempted, in order to supplement the interview data with ethnographic work. However, this was not possible as, for reasons which the author respects, the group contacted preferred not to attempt this.

Lastly, a note about the author's own positionality vis-à-vis the participants. Born in a conservative, traditional background, my personal journey to accept my LGBTQ+ identity has also made me sensitive to some of the struggles and issues facing the wider non-binary community. However, I do not claim an 'insider' status and therefore recognise any limitations of my perspective. To counteract this, I did my best to be authentic and respectful, as well as reflective of my own position, relative privilege, and unconscious bias.

3.3 Data Analysis

The author orthographically transcribed the interview recordings in full, using ELAN (version 5.8). The transcription conventions used are adapted from Corwin (2009) (see Appendix One). All excerpts with non-

binary related content were compiled into a 28,000-word corpus. This was parsed thematically, in order to help establish the wider context in which non-binary participants constructed their identities. The main 3 themes which emerged broadly correspond those identified by Bradford et al. (2019). They are:

(1) Subjective experiences of the non-binary self.
(2) Non-binary identity as experienced alongside embodiment, embodied practice, and sexuality.
(3) Recasting the normative: emergent non-binary gender ideology.

It is noted that each participant gave a different weight to each theme overall.

The present analysis aims to examine how participants discursively constructed their non-binary identities. Specifically, the focus is on implicitly or explicitly gendered lexicon, and how it serves gender performativity in narrative (following Zimman, 2014). Explicitly gendered lexicon includes words which semantically entail gender (e.g., as 'woman', 'girl', 'female', 'feminine'), whereas implicitly gendered lexicon involves terms which index certain kinds of gendered personalities (Zimman, 2017a). For the purposes of this paper these include complimentary terms (e.g., 'beautiful', 'handsome'), insult terms (e.g., 'bitch', 'asshole'), and address forms (e.g., 'dude'), as well as words for stereotypically gendered practices or items (e.g., 'makeup', 'bra'), and terms describing the physical body (e.g., 'boobs', 'chest').

Excerpts demonstrating the three main themes were selected for discursive analysis. Those rich in implicitly or explicitly gendered lexicon were further analysed semiotically. As the way in which speakers manipulate indexical meanings can be considered a performative use of language (Barrett, 2009; Calder, 2019) and indexicality depends on the cultural and so ideological context of a society (Ochs, 1996), I argued that examining the indexicality (Silverstein, 2003; Jaffe, 2016; Gal, 2016) of gendered lexicon can reveal both how non-binary gender is performed, but also point to the wider ideological framework of gender and embodiment.

It is emphasised that this research seeks to uncover some of the linguistic means which non-binary individuals may use to perform their identity. It does not aim to generalise on how identity is constructed among non-binary individuals more broadly.

4 Results

4.1 Subjective experiences of the non-binary self

Participants emphasised that their subjective experiencing of their gender was the locus of their gender identity, and distinguished this identity from gender performance, corroborating Zimman (2017a). Firstly, participants often expressed their subjective experience of their non-binary identities in relation to binary categories. For example, in Extracts 1 and 2, Annie and Alex describe how they self-identify. Both reject belonging to binary categories of gender, which Annie describes as 'boxes' (Extract 1). They juxtapose binary categories to their own feelings of fluidity: they 'move' between such categories (Extract 1), are 'in the middle' (Extract 2), blend them ('androgynous', Extract 1), or 'do whatever' altogether outside of a 'spectrum' (Extract 2). This fluidity and indeterminacy are further underscored by gender identity descriptors such as 'genderfluid', used by Annie (Extract 1). Thus, Annie and Alex discursively create a malleable, navigable spectrum out of a binary distinction.

Furthermore, participants described their identity in terms of their feelings. This is already seen in Extracts 1 and 2. It also expressed by Luuk (Extract 3), when asked what the descriptor 'non-binary' means to them. Luuk explains that identifying as non-binary allowed them to express the 'vagueness' in their subjective experience of their gender identity, removing the pressure to 'have a super concrete perception' of themselves. Luuk relates this perception of themselves to knowing 'how [they] feel', which they 'sometimes' find 'difficult'. Thus, it emerges that the linguistic performance of self-determination heavily relies on placing gender experience in the affective space (Zimman, 2017a). For all participants in this study, what defines their gender identity is their subjective experiencing of gender. Accordingly, I will use the term affective gender to invoke this understanding of gender identity as rooted in individual affective experience.

Participants also emphasised a distinction between gender identity and gender performance. An example of this is Jem's discussion of their genderfluidity (Extracts 4 and 5). In Extract 4, Jem describes how they understand and accommodate the manifestations of their genderfluidity based on 'subtleties'. For example, a day when Jem does not want to 'look down' when they change, signifies the need to wear a binder. Jem also attributes their reactions to being 'casually misgendered' to how 'strongly' they are 'feeling gender' at the time. Thus, Jem emphasises how their identity is rooted in their subjective experience. In Extract 5, Jem voices a hypothetical dialogue with those who misunderstand their genderfluidity ('people are like, 'how do you decide?'). They parody, and thus delegitimise, the notion by voicing the hypothetical thought-process of choosing their gender as flippant and superficial (as 'waking up and going 'hm, I think I'll be a boy today!''). By establishing affective gender as outside of their conscious control, Jem legitimises the separation of their identity from their embodied performance ('I have no decision-making abilities in what my gender is, I get decision-making abilities in how I'm going to present myself'). In doing so, Jem reframes gendered embodied practices such as binding and wearing lipstick (Extract 5) as agentive decisions. The distinction between gender identity and gender performance is therefore shown to be reflected in the discursive construction of non-binary self (Zimman, 2017a). Embodiment and identity are distinct levels of meaning for the non-binary self, echoing Zimman's (2017a) linguistic fragmentation. This allows embodied cues to be reimagined as indexical of one's affective gender: wearing a binder indexes that the individual 'feels' like a boy, but not that they *are* one (Jaffe, 2016; Gal, 2016). However, embodiment remains fundamentally implicated in both the subjective experience of gender identity (providing clues to fluidity, Extract 4), and its expression (using different gendered embodied cues, a lipstick or a binder, to index affective gender manifestations, Extract 5). This suggests that the physical body potentially informs 'doing gender', even when that is 'doing 'affective' gender' (West & Zimmerman, 2009).

Lastly, participants discussed their subjective experiencing of their gender vis-à-vis other people's perceptions of them. Leslie (Extract 6) narrates realising that their own, non-binary interpretation of their gender performance ('I think I look petty androgynous') is not always shared by others ('then people being like 'miss' or something'). Leslie expresses their incredulity at this by addressing the hypothetical person who misgendered them ('how do you know that about me?'). By exposing the unreliability of every-day gendered assumptions relying on stereotypical visual cues, such as address terms, Leslie reinforces the distinction of gender identity and embodiment. Sam (Extract 7) describes a similar experience, where they are misgendered by those around them ('how can you still see me as a woman, despite how I feel when I see myself?') even when they 'think of [themselves] as more masculine'. In the binary ideological framework surrounding non-binary individuals such as Leslie and Sam, the embodied cues which for them index 'androgynous' or 'more masculine' do not have the potential to index 'non-binary', but only 'miss',

and ‘woman’. Therefore, embodiment can constrain the recognition of the non-binary self (Messerschmidt, 2009; Darwin, 2017). This invokes the ‘doing gender’ model, as it shows the successful achievement of gender outside the self and subjective experience depends on others' recognition of that gender (West & Zimmerman, 1987).

Therefore, the experiencing of the non-binary self is crucial in the discursive construction a non-binary identity. This experience is shown to comprise the distinct but interlinked levels of affective gender and gender performance. Specifically, the possible misinterpretations of an individual’s identity based on their embodiment point to how non-binary identities are not only subjectively experienced, but also called to exist within a broader societal network of identities. The analysis of the examples in the second main theme below shows how non-binary individuals semiotically (re)interpret terminology pertaining to gender, embodiment, and gendered practices in creating a non-binary self, and how this process is shaped by normative ideology.

4.2 Non-binary identity as experienced alongside embodiment, embodied practice, and sexuality

In talking about their experiences as non-binary individuals, participants discussed how their gender identity interacted with, or was informed by, embodiment and sexuality. Their discourse on gender and embodiment showed how these identities were related through multiple instances of iconisation and indexicality (Gal, 2016; Jaffe, 2016). These semiotic processes further reveal how the links between gender identity, embodiment, and sexuality are organised in normative ideology. Thus, it is shown that by engaging with the semiotic processes while constructing their non-binary identities, non-binary individuals ultimately engage with normative gender ideology.

Firstly, participants described experiencing their gender identity in terms of experiencing their body. This has already been suggested in Extract 4, where Jem described understanding the manifestations of their genderfluidity by also paying attention to their feelings about their body (for example by wearing a binder when feeling uncomfortable with that area of their body). In Extract 8, Annie explicitly discusses how she feels about her body. She did ‘not like people mentioning’ any ‘feminine’ parts of her body. In characterising her ‘hips’ and ‘boobs’ as ‘feminine’ parts on a ‘female’ body, Annie expresses her understanding of her physical body in terms of gender, a cultural construct. In doing so, she emphasises how the physical body is not ‘natural’, but reimagined as iconic of normative gender. As such, a ‘female’ body is ‘feminine’, and ‘feminine’ is iconic of a female body. This is an instance of Rhematisation, as ‘feminine’ and the physical reality of a female body are seen as naturally occurring together and at some level resembling each other (Gal, 2016). In saying that she wanted a ‘very androgynous body’ with no ‘female or male aspect to it’ therefore, Annie is not simply rejecting the physical reality of ‘male’ and ‘female’ bodies, but resisting the gendered iconisation of her own.

Similarly, in Extract 9 Luuk explains they ‘don’t like’ being perceived as a woman ‘because of the way [their] body looks’. However, they ‘don’t know’ if they ‘hate [their] body per se’. They describe an imaginary scenario of being ‘on [their] own on an island’, and how they do not know if they would then ‘hate’ their body because they ‘have breasts or whatever’. The juxtaposition of the two situations reveals their core difference: if Luuk was solitary on an island, no one would perceive them wrongly (‘as a woman’) because of their body. Furthermore, it reveals that this perception involves an iconisation between Luuk’s body (‘how it looks’, their ‘breasts or whatever’) and being a ‘woman’ (Gal, 2016; Jaffe, 2016). This iconisation reflects a binary gender ideology, where Luuk’s body can only be perceived as that of a

woman's, and not a non-binary individual's. Luuk's use of a passive without an expressed agent ('Because of the way my body looks**,** I'm perceived as a woman'), presents that ideology as prevalent and exceptionless, enough for Luuk to 'hate' their body even though they 'don't like' its iconisation as that of a 'woman's'. Luuk thus is both negatively affected by this ideology, and resists it. Therefore, this narrative overall suggests that Luuk feels they cannot reclaim the semiotics of their embodiment in real life. These examples illustrate that embodiment (as referring to sex, the physical body) is part of the experience of gender, and that this gendering of embodiment is mediated through normative (binary) gender ideology (West & Zimmerman, 1987; Zimman, 2014). As such, embodiment is not simply an objective physical or biological reality but a cultural construct itself (Butler, 1993).

Another facet of embodiment includes gendered embodied practices, such as makeup, hairstyle, and clothing (Darwin, 2017). Participants described how engaging with such practices was related to performing their non-binary gender. Therefore, they also revealed how these practices were implicated in normative gender ideology. In Extract 10, Sam discusses the association between makeup and 'women and girls'. Their narrative illustrates the indexical order of this embodied practice: the action of wearing makeup (a first order indexical) indexes the stance of 'wanting to look beautiful' (a second order indexical) (Silverstein, 2003). In turn, Sam says that 'the whole makeup thing' would be perceived as 'feminine'. This grouping of action and stance ('whole makeup thing') can be understood as a manifestation of the coherence prevalent in indexical fields (Jaffe, 2016). This entails that since makeup can index a stance ('wanting to look beautiful'), this stance will dominate the semiotic interpretation of makeup, even if indexing that stance was not the aim of the individual themselves. Indeed, Sam describes their wish to wear makeup not in terms of beauty, but agentive experimentation (to 'change how [they] look a little bit'). However, 'that whole thing' was 'so associated with women and girls' that '[Sam] was just like, 'nope!'. Not doing that'. That Sam chose to opt out of wearing makeup altogether suggests they felt they could not challenge the indexical valence of this embodied practice (Ochs, 1996). This indicates the prevalence of the normative ideology within which makeup indexes femininity and 'wanting to look beautiful'.

Another embodied practice many participants identified as gendered was hair length. In Extract 11 Alex describes their experience of cutting their hair short. Whereas Alex 'felt really good' about having short hair and experimenting with different styles, there was a 'pressure' from their family against that ('you're still a woman and you need to have something on your head'). That for Alex's family a woman having short hair would challenge her categorisation as a woman indicates an iconisation between hair length and binary gender. Alex describes that even when not thus incompatible with 'being a woman', short hair was still indexical of masculinity: 'other people labelled it for me, I guess, that it was more masculine'. This indicates that in binary gender ideology, hair is important for a gendered interpretation of embodiment. For Alex however, hair is independent of their gender identity: 'I didn't feel [short hair] (…) was more er, masculine or something'; 'I feel like maybe I want to grow it a bit again, but not because (…) it will be again more feminine or something'. By resisting hair as iconic or indexical of gender, Alex forefronts their agency, and challenges the existing indexical field of that embodied practice (Eckert, 2008).

Gendered embodied practices are important for the performance of non-binary identity in that they can influence the recognition of that identity by others (Darwin, 2017; West & Zimmerman, 1987). The following example shows that this recognition can rely not on legitimising the perception of the self as non-binary per se, but on de-legitimising binary interpretations of an individual's gender and embodiment. Jem (Extract 12) describes having days where they 'make it a point to try and confuse as many people as possible!'. To do so, they may put on a 'leather jacket', 'binder', and 'bright red lipstick'. That is, to avoid being gendered normatively, Jem employs practices which do not, in normative gender ideology, index a

single binary gender category. Jem therefore exploits the indexical potential of these embodied practices from a binary ideology perspective in order to perform their non-binary gender. Jem adds 'sometimes it's like, I cannot stand the idea of being perceived as female, so I'm just gonna give you 'what the fuck' instead'. Therefore, by manipulating the use of gendered embodied practices, Jem is not only challenging these practices' ability to index their gender, but also their kind of body. This is possible because the physical body is already potentially iconic of gender, a 'naturalised' cultural construct, in normative gender ideology and discourse: as the body itself becomes gendered and how embodied practices have the potential to index not only gender but particular kinds of bodies (Butler, 1993; Zimman, 2014). Thus, it is shown that the discursive construction of non-binary gender identity involves embodiment and embodied practice, and how these are constituted in normative ideology.

Furthermore, participants' narratives showed the implication of embodiment in the constitution of heteronormative sexuality. In Extract 13 Leslie discusses shopping for a bra: it's 'so intense of an experience' where there is 'so much going on', including an 'explosion of pink', and the slogans 'sexy', 'please your man'. Leslie's narrative shows that in the context of the store, shopping for a bra not only indexes normative femininity, but establishes this femininity as indexical of heteronormative sexuality. Leslie's concluding remark 'I just want a sports bra!' communicates their frustration at the prevalence of these normative ideological processes, and expresses a refusal to participate in them. This shows how the performance of normative gender relies on the performance of normative sexuality, as per the 'heterosexual imperative' (Butler, 1993; Connel, 1995, Cameron, 1998; Kiesling, 2002; Eckert, 2011; Zimman, 2013).

The following examples illustrate how the performance of non-binary identity can therefore also be the locus for the contestation of normative sexuality and desire. In Extract 14, Luuk discusses their sexual orientation. At the time of the interview, Luuk identified as bisexual. However, they used to think they were a lesbian, because they 'wouldn't feel comfortable in a relationship with a man if [they] were perceived as a woman'. When prompted to elaborate, Luuk explained that for them a relationship is about how 'the person relates to you, how do you relate to a person'. For Luuk, being perceived as a woman by a 'straight man' would be more uncomfortable than being perceived as a woman by their girlfriend at the time. They suggest that the reason is partly 'these notions of how a woman should be in a relationship', but state that it is mostly about 'how the other person perceives [them]'. Being in a heterosexual relationship, and especially with a heterosexual man, would for Luuk result to a particular perception of their gender, but also of their behaviour and expectations placed on them. This illustrates the link between normative, binary gender and heterosexuality. Furthermore, 5 out of 7 participants in this study identified or expressed an affiliation with some identity on the asexual spectrum. In Extract 15, Adam, who identifies as aromantic asexual, describes how refusing to embody femininity in high school was a strategy for them to avoid participation in heteronormative desire. Adam describes that being aromantic asexual meant that the 'nightmare scenario' in high school would be someone finding them attractive. They explain that for that reason they 'swerved hard' into a 'masculine', 'tomboyish' presentation, because they felt at the time that by doing so, they would be 'unappealing' to 'cis boys'. Therefore, this analysis suggests that the discursive construction of non-normative gender identities also involves the construction of a non-normative and/or non-heterosexual sexuality and desire. In turn, this shows the interdependency of binary sex and gender and normative heterosexuality (Connel, 1995; Cameron, 1998; Kiesling, 2002; Eckert, 2011; Zimman, 2013).

Thus, this analysis shows how the discursive construction of non-binary identities is informed by the complex interaction of gender, embodiment, and sexuality. In their narratives, participants (re)interpret the indexicality and iconisation of implicitly and explicitly gendered lexicon to perform their gender, and so also reveal the links between gender, embodiment, and sexuality in normative gender ideology. The

examples in the third main narrative theme show how non-binary individuals reimagine a gender ideology within which their gender identity can be recognised (Corwin, 2009; Darwin, 2017).

4.3 Recasting the normative: emergent non-binary gender ideology

The discursive and semiotic analysis in the previous two main themes showed how participants draw on their subjective experience, and embodiment, embodied practice, and sexuality as these are constituted in normative ideology, to construct their non-binary identities. In the third main narrative theme, participants discussed how they wished gender in general, as well as their non-binary identities in particular, were understood and accommodated. They did so by addressing normative ideology through the (re)interpretation of the semiotic processes in the discourse pertaining to gender, embodiment, and embodied practices. As a result, participants constructed a gender ideology inclusive of them (Corwin, 2009; Darwin, 2017).

In describing a non-binary inclusive ideology, some participants revisited the distinction between gender identity and gender performance (Zimman, 2017a). For example, Adam (Extract 16) expresses that for trans and non-binary people ‘everything is very individualised’, meaning that ‘everyone’s gonna have their own experience of gender’. They explain that ‘if a trans guy is effeminate, or if a trans girl is masculine, or if a non-binary person, you know, tends to present very strongly one way or the other, or you know, whatever, that doesn’t negate their identity’. Therefore, Adam extends the subjectively experienced distinction between gender identity and gender performance to a gender ideology able to accommodate this diversity in experience. This is achieved through Adam’s discursive choices. Adam uses examples (such as ‘effeminate trans guy’) which have identities and presentations at opposing ends of a binary gender spectrum. In doing so, Adam seemingly accepts that there are embodied practices or traits gendered in a binary way. However, Adam’s examples expand the indexical field of these practices (Eckert, 2008). By asserting that gender performance does not negate gender identity, Adam introduces the possibility that normatively gendered traits are not automatically indexical of their corresponding binary gender identity. Rather, these traits are presented as part of a broader indexical field, by becoming able to index more than the binary gender identities. Thus, Adam’s discourse shapes an ideology where ‘effeminate’ can index trans guy, ‘masculine’ can index trans girl, and a normatively binary presentation can index a non-binary person without semiotic contradictions. Therefore, the creation of a non-binary inclusive gender ideology can exploit binary discourse in order to create more semiotic space for non-normative identities.

Some participants took non-binary inclusive ideology a step further, to a complete semiotic separation of embodiment and identity. In Extract 17 Jem describes the meaning of ‘non-binary’. They assume the position of an expert, by using the second person (‘you’) and expanding on what they characterise as the ‘textbook definition’ of the term ‘non-binary’. Specifically, Jem explains that ‘not aligning with, you know, the traditional male or female’ does not mean ‘you can’t use he/him or she/her, or you can’t present more femininely or more masculinely, there is no one way to be non-binary’. Therefore, Jem firstly challenges the potential of pronouns, one of the few direct indexes of gender to index gender identity (Ochs, 1992). This creates linguistic fragmentation, in that each level of gendered language offers a separate realm for self-identification (Zimman, 2017a). Furthermore, Jem does not challenge the binary gendering of embodied performances per se, but instead challenges their ability to index gender identity altogether. As there is ‘no one way’ to perform non-binary gender, any indexical links between embodied practices and gender identity are broken: the individual has the agentive power to use gendered traits and embodied practices to index a particular kind of gender expression, and ‘only’ expression. By contrast,

identity is determined by an individual's affective space. This is clearly shown by Jem dramatizing a hypothetical dialogue between themselves and a person questioning their gender. Jem voices a persona that is knowledgeable about non-binary gender, by asking the other hypothetical person how they feel ('Do you feel like a boy?', 'Do you feel like a girl?'). Then, based on that person's answers to these questions ('No, not really', 'No'), they prescribe that that person can then identify as non-binary ('Okay cool, enby!'). Thus, through linguistic fragmentation and the reinterpretation of the semiotics of identity and embodiment, Jem discursively constructs a non-binary inclusive gender ideology which legitimises self-identification.

An ideology of self-identification for some participants included dismantling 'biological essentialism' (Extract 18). In Extract 18, Luuk talks about how they wished society saw gender. Luuk wishes people 'would not think that clothes say anything about your gender', and that 'genitalia or chromosomes' would not 'matter that much'. Luuk does not only separate gendered identity from gender performance ('clothes'), but from the physical body as well. This invokes the naturalisation of embodiment as part of doing gender (West & Zimmerman, 1987; Butler, 1993). Therefore, it is seen how non-binary individuals address normative ideology in their narrative discourse and through the semiotic(re)interpretation of gendered lexicon, in order to construct a gender ideology inclusive of them (Corwin, 2009; Darwin, 2017).

However, it can be argued that the construction of this non-binary inclusive ideology ultimately seeks to more generally emphasize the agency and personhood of individuals in the face of assumptive binary gender ideology. This can be seen in Extract 19, where Annie discusses the acceptance and accommodation of non-binary individuals. She identifies the presence of non-binary toilets in University buildings as a positive development, and describes it as 'that's just an example of, like, it's less stress, it's just a toilet, like, anyone can use it, why does it have to be, like, segregated like that?'. Annie therefore describes performing gender, by for example choosing which gender segregated toilet to use, as an unnecessary pressure. She adds that if gender is removed from such contexts, 'everything is kind of like that level, of just like, 'oh, it's for a person'', and then 'people stop feeling like people trying to put them in this category they don't quite feel happy with, all of the time'. Therefore, for Annie the binary gendering of otherwise equitable spaces such as toilets and the resulting demand to perform gender in these spaces, is stressful and unnecessary because it undermines the personhood of the individuals that use them. Instead, it reduces people to categories, and furthermore forces them to constantly repeat their belonging to these categories. Annie's inclusive language ('everything', 'people') therefore implies that this binary gendering is not only an unnecessary pressure for non-binary people, but people in general, regardless of their gender identity. Annie's narrative therefore criticises and questions one of the instances of binary gendering in every-day life, and offers an emphasis on agency and personhood as an alternative.

Thus, this analysis firstly shows the main themes emerging in non-binary individuals' narratives on identity in lived experience. Through these, it reveals some of the linguistic means involved in the discursive construction of non-binary identities. Through linguistic fragmentation, discursive choices, and the (re)interpretation of the semiotics of implicitly and explicitly gendered lexicon, participants in this study affirm the primacy of self-identification, and create a non-binary inclusive ideology able to legitimise both the experiencing and expression of non-binary identities. However, in doing so they engage with normative gender ideology, revealing crucial links between gender, embodiment, embodied practices, and sexuality. Therefore, this analysis reveals not only how non-binary identities are constructed, but also of how normative gender ideology is constituted.

5 Discussion

The findings of the present study overall corroborate existing research on the lived experience narratives and discursive construction of non-binary identities (Corwin, 2009; Darwin, 2017; Zimman, 2017a; Bradford et al., 2019). In doing so, they also emphasise the relevance a sociocultural linguistics framework in examining the linguistic construction of non-binary identity (Bucholtz & Hall, 2016).

The discursive analysis of the narratives in this study focused on implicitly and explicitly gendered lexicon. Explicitly gendered lexicon is defined as words that semantically entail gender (e.g., as 'woman', 'girl', 'female', 'feminine'), while implicitly gendered lexicon refers to terms which index certain kinds of gendered personalities (Zimman, 2017a). The latter includes complimentary terms (e.g., 'beautiful', 'handsome'), insult terms (e.g., 'bitch', 'asshole'), and address forms (e.g., 'dude'), as well as words for stereotypically gendered practices or items (e.g., 'makeup', 'bra'), and terms describing the physical body (e.g., 'boobs', 'chest'). Taken together, implicitly and explicitly gendered lexicon can be seen as both describing and creating different kinds of social categories. Such lexica been used in discourse analysis, particular under membership categorisation analysis (MCA; Sacks, 1992), in order to examine how speakers in a society categorise themselves, and therefore how the social world is organised through the particular actions or characteristics linked to social categories (Stokoe, 2003; Ward, 2019). From a sociolinguistic perspective, this suggests that implicitly or explicitly gendered lexicon has a rich indexical potential. As the way in which speakers manipulate indexical meanings can be considered a performative use of language and indexicality depends on the cultural and so ideological context of a society (Ochs, 1996; Barrett, 2009; Calder, 2019). I examined how this gendered lexicon was used to perform non-binary identity in the narrative discourse of non-binary individuals. The findings show that participants interpreted or manipulated the indexicality of gendered lexicon in order to perform their non-binary identities, as well as construct a non-binary inclusive gender ideology (Silverstein, 2003; Jaffe, 2016; Gal, 2016). For example, participants' discursive choices revealed an iconisation between gender identity and embodiment (Extracts 8, 9), which was further shown by how participants both affirmed (Extract 10) and challenged the ability of embodiment to index gender identity (Extracts 16, 17) (Darwin, 2017). Furthermore, participants manipulated the indexicality of discourse pertaining to embodiment in order to legitimise the recognition of their non-binary gender presentation (Extract 12) (Corwin, 2009; Darwin, 2017). The lexicon around gender and embodiment, however, was also shown to be implicated in indexing desire and sexuality (Extracts 12, 13, 14) (Connel, 1995; Cameron, 1998; Kiesling, 2002; Eckert, 2011; Zimman, 2013). At the same time, participants reimagined explicitly gender lexicon (words such as *boy, girl* etc.) and gender identity labels as fuzzy concepts which could be defined only by the individual through 'experiential legitimisations, primarily grounded in affect' (Zimman, 2017a, p. 233) (e.g., Extracts 5, 16, 17) and challenged the ability of pronouns, one of the few direct indexes of gender to index gender identity (Ochs, 1992). By treating each level of gendered language as a separate realm of self-identification in what Zimman (2017a) describes as 'linguistic fragmentation', participants legitimised the experiencing and expression of non-binary identities, and emphasised their self-determination. Thus, the findings show how participants' use and semiotic interpretation and manipulation of gendered lexicon in narrative discourse serves both the construction and performance of non-binary identity. At the same time, this examination of gendered lexicon in the discourse reveals how gender identity, embodiment, and sexuality are semiotically linked in normative ideology more broadly. This is clearly illustrated in the thematic analysis which served to contextualise the discursive analysis of the narratives in this corpus.

Thematic analysis revealed three major themes, which corresponded to those identified by Bradford et al. (2019). The first narrative theme showed how non-binary identity is subjectively experienced.

Participants discursively created a malleable, navigable gender identity spectrum out of a binary distinction, and located the experiencing of their identity in the affective space (e.g., Extracts 1, 2, 3), which they separate from embodiment and embodied performance (e.g., Extract 4, 5). They emphasised this separation as what legitimises their identity, but also as a frequent cause of misgendering, as this separation is not always obvious or understood by others (Extract 6, 7). Therefore, these findings support that self-identification can offer a powerful form of self-expression against a normative gender ideology where 'an individual's gender role, identity and expression [are] based on particular readings of their body' Zimman (2017a). However, it was shown that embodiment was nonetheless crucial for non-binary identities. It was implicated in the subjective experiencing of genderfluidity (Extract 4), but also, crucially, in the performance of non-binary gender and its ability to be recognised by others, as shown in the second narrative theme (Darwin, 2017).

In the second narrative theme, the performance and recognition of non-binary identity were shown to interact with embodiment, gendered embodied practices, and sexuality as these are constituted in normative gender ideology. Participants described how their physical bodies (Extracts 8, 9) or certain gendered embodied practices such as makeup and hair length (Extracts 10, 11) could constrain the interpretation of their gender by others in a binary way. This for some participants created negative or uncomfortable feelings about their body (Extracts 8, 9), or led them to avoid certain embodied practices (Extract 10). Other participants responded to binary constraints by removing them from their own interpretations of embodied practices (Extract 11), or by exploiting the limitations of the binary gendering of embodiment in order to delegitimise binary interpretations of their own gender (Extract 12). Therefore, this analysis shows that although the subjective experiencing of non-binary gender is affective and distinguished from embodiment, the successful performance of non-binary gender for the self, remained in many ways remains dependent on, and strove to achieve, others' recognition of that gender (Darwin, 2017). This further supports that the physical body and embodied performances are a fundamental part of 'doing' gender (West & Zimmerman, 1987).

The present study suggests that recognising non-binary gender in embodiment and embodied performance remains rare in society more widely (Messerschmidt, 2009). This, I argue, points to the prevalence of normative gender ideology, but also to its dependence on the 'naturalisation' of embodiment as a given fact and not a cultural construct (Butler, 1990; 1993). Participants' narratives also highlighted how the performance of binary gender implied heteronormative sexuality (Extracts 13, 15), or vice versa (Extract 14), echoing the heterosexual imperative (Butler, 1993). Participants (Extracts 14, 15) discussed how their gender influenced their or others' perception of their sexuality, with 5 out of 7 identifying or affiliating with the asexual spectrum. While there has been linguistic research on how the performance of normative gender relies on the performance of normative sexuality, the present findings show specifically how on-binary identities can also involve reimagination of or resistance to normative desire and heterosexuality (Connel, 1995; Cameron, 1998; Kiesling, 2002; Eckert 2011; Zimman, 2013).

The third narrative theme demonstrated that subjectively experiencing and performing a non-binary identity involved constructing a non-binary inclusive gender ideology in which that identity can be made meaningful. Participants emphasised the affective nature of gender identity and the separation of gender identity and embodiment (Extracts 16, 17, 18), and imagined a non-binary society (Extracts 18, 19). This follows research showing that non-binary individuals used narrative to create new social forms and gender ideologies (Corwin, 2009; Darwin, 2017). Furthermore, I argue that, within the sociocultural context of the participants represented in this corpus, this non-binary inclusive gender ideology ultimately sought to more generally emphasize individuality and the personhood of individuals. In this, however, the complete

removal of gender identity from the physical and social reality of gender invokes the caveat that self-determination can therefore also serve to obscure how the agency to self-identity is 'systematically limited along predictable lines of power and oppression' (Zimman, 2017a).

Therefore, this thematic analysis revealed some of the most salient aspects of experiencing a non-binary identity, and in so doing helped contextualise how discourse was crucial for non-binary gender performativity (Thornborrow & Coates, 2005). This more generally highlights the usefulness of using narrative to examine identity construction (Benwell & Stokoe, 2006). It also underscores how examining the use of gendered lexicon in discourse can reveal how gender identity is, in both performance and ideology, linked to embodiment, embodied practices, and sexuality (Zimman, 2004; 2017a). This supports that embodiment and the physical body are central to the production, perception and social interpretation of language, and therefore the linguistic performativity of gender. Therefore, the findings of the present study ultimately suggest the relevance of a sociocultural linguistic framework in examining the linguistic construction of non-binary identity in particular, as well as gender identity more broadly (Bucholtz & Hall, 2016).

However, this study was limited in that it examines some of the linguistic means through which non-binary identity can be performed in narrative discourse, within only a homogeneous group of primarily white, university-educated, AFAB participants in their early twenties. This entails that the results are limited to the particular characteristics of and sociocultural context relevant to this sample, and as such are not intended to, and should not, be generalised to all non-binary individuals. More research on non-binary identities is needed, especially research including intersectionality in its analysis (Crenshaw, 1989). For example, there is need for research on non-binary individuals who were assigned male at birth, as well as on non-binary individuals from other, non-white racial backgrounds. Such research can benefit from a systematic comparison of non-binary identity and ideology construction in different groups, and how this relates to existing linguistic variation and the variations in the normative gender ideology of these groups. Similarly, future research could examine the linguistic identity construction of older non-binary individuals, and track the changes in linguistic performativity in different generations. Finally, there is need for ethnographic work in order to research how non-binary individuals construct their identities in interaction, and how meaning-making, gender performativity, and ideology reimagination is different in the setting of non-binary communities.

6 Conclusion

The present study examined the discursive construction and performance of non-binary identities in the context of lived experience narratives. It aimed to address the gap for a systematic research on non-binary gender performativity in narrative discourse by using a broad thematic analysis to contextualise and enrich the findings of discursive analysis (Butler, 1990; 1993; Corwin, 2009; Bradford et al., 2019; Corwin, 2009). As the way in which speakers manipulate indexical meanings can be considered a performative use of language (Barrett, 2009; Calder, 2019) and indexicality depends on the cultural and so ideological context of a society (Ochs, 1996), I argued that examining the indexicality (Silverstein, 2003; Jaffe, 2016; Gal, 2016) of implicitly and explicitly gendered lexicon (following Zimman, 2014; 2017a) in narrative discourse would reveal both how non-binary gender is performed, but also show how normative gender ideology is constituted.

The results showed that implicitly and explicitly gendered lexicon in narrative discourse was crucial for the construction and performance of non-binary identity. Participants discursively separated gender

identity from embodiment in their subjective experiencing of gender (Zimman, 2017a) through the semiotic manipulation of lexicon pertaining to gender and embodiment. By doing so they could both legitimise their subjective experiences, but also construct an ideology in which non-binary gender is included and recognised (Corwin, 2009; Darwin, 2017). This analysis also revealed how the discursive construction of non-binary identities was informed by the complex interaction of gender, embodiment, and sexuality (Connel, 1995; Cameron, 1998; Kiesling, 2002; Eckert, 2011; Zimman, 2013). In showing how these identities were related through multiple instances of iconisation and indexicality (Gal, 2016; Jaffe, 2016), the analysis also showed how they are interrelated in normative gender ideology (West & Zimmerman, 1987; 2009; Butler, 1993). Thus, in discursively constructing and performing their non-binary identities, participants also engaged with this normative gender ideology. While they both affirmed and challenged it, I argue that they ultimately sought to emphasize individuality and personhood against the restrictions of binary gender. However, their discursively achieved removal of gender identity from the physical and social reality of gender invokes the caveat that self-determination can also serve to obscure how the agency to self-identity is 'systematically limited along predictable lines of power and oppression' (Zimman, 2017a), where particular kinds of bodies can license or challenge different social meanings (Calder, 2019). Thus, this study contributes to the literature examining how non-binary identities are discursively constructed and performed, but also offers crucial insights into the constitution of normative gender ideology and its relation to embodiment and sexuality. I therefore conclude by arguing that this shows the need for more research within a sociocultural linguistics framework, where embodiment and the physical body are seen as central to the production, perception and social interpretation of language (Bucholtz & Hall, 2016).

7 References

Barrett, R. (2009). Language and Identity in Drag Queen Performances. In N. Coupland and A. Jaworski (Eds.), *The New Sociolinguistics Reader,* 250–257. New York: Palgrave Macmillan.

Barrett, R. (2017). Fierce Fish Who Pee: Indexicality and Identity Among African American Drag Queens. In *From Drag Queens to Leathermen.* Chapter 2. Oxford University Press: Oxford Press Scholarship Online.

Benwell, B. & Stokoe, E. (2006). *Discourse and Identity*. Edinburgh: Edinburgh University Press.

Borba, R. (2019). The interactional making of a "true transsexual": Language and (dis)identification in trans-specific healthcare. *International Journal of the Sociology of Language* 2019(256). 21–55. DOI:

Bradford, N. J. et al. (2019). Creating gender: A thematic analysis of genderqueer narratives. *International Journal of Transgenderism,* 20(2-3), 155–168.

Bucholtz, M. & Hall, K. (2016). Embodied sociolinguistics. In N. Coupland (Ed.), *Sociolinguistics: Theoretical Debates.* 173–197. Cambridge: Cambridge University Press.

Butler, J. (1990). *Gender Trouble: Feminism and Subversion of Identity.* New York: Routledge.

Butler, J. (1993). *Bodies that matter: on the discursive limits of sex*. New York: Routledge.

Calder, J. (2019). From Sissy to Sickening: The Indexical Landscape of /s/ in SoMa, San Francisco. *Journal of Linguistic Anthropology* 29(3). 332–358

Cameron, D. (1992). *Feminism and linguistic theory*. 2nd edn. Basingstoke: Palgrave.

Cameron, D. (1998). Performing Gender Identity: Young Men's Talk and the Construction of Heterosexual Masculinity. In J. Coates (Ed.), *Language and gender: A Reader*, 270–284. Oxford: Blackwell.

Cameron, D. (2001). *Working with Spoken Discourse*. London: Sage.

Cashman, H. R. (2018). Narrating the interSection: body, time, space and transition in one queer life. *Gender and Language*, 12(4), 416–436. DOI: 10.1558/genl.34633

Coates, J. (2005). Masculinity, Collaborative Narration and the Heterosexual Couple. In J. Thornborrow and J. Coates (Eds.) *The Sociolinguistics of Narrative*. 89 –106. Philadelphia: John Benjamins.

Connel, R. (1995). *Masculinities.* Berkeley, CA: University Press.

Corwin, A. I. (2009). Language and gender variance: constructing gender beyond the male/female binary. *Electronic Journal of Human Sexuality*, 12.

Corwin, A. I. (2017). Emerging genders: semiotic agency and the performance of gender among genderqueer individuals. *Gender and Language*, 11(2), 255–277.

Crenshaw, K. (1989). Demarginalizing the InterSection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, 1(8).

Cruickshank, J. (2012). The Role of Qualitative Interviews in Discourse Theory. *Critical Approaches to Discourse Analysis across Disciplines*, 6 (1), 38 –52.

Darwin, H. (2017). Doing Gender Beyond the Binary: A Virtual Ethnography. *Symbolic Interaction*, 40(3). 317–334.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), 453–476

Eckert, P. (2011). Language and Power in the Preadolescent Heterosexual Market. *American Speech*, 86(1). 85–97.

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, (41), 87–100.

Fairclough, N. (1989). *Language and Power*. Essex: Longman.

Gal, S. (2016). Sociolinguistic Differentiation. In Nikolas Coupland (ed.), *Sociolinguistics: Theoretical Debates.* 113–135. Cambridge: Cambridge University Press.

Goldberg, A. E. & Kuvalanka, K. A. (2018). Navigating identity development and community belonging when "there are only two boxes to check": An exploratory study of nonbinary trans college students. *Journal of LGBT Youth*, 15(2), 106–131.

Jaffe, A. (2016). Indexicality, stance, and fields in sociolinguistics. In Nikolas Coupland (Ed.), *Sociolinguistics: Theoretical Debates.* 86–112. Cambridge: Cambridge University Press.

Kiesling, S. F. (2002). Playing the Straight Man: Displaying and Maintaining Male Heterosexuality in Discourse. In K. Campbell-Kibler, R. J. Podesva, S. J. Roberts, and A. Wong (Eds.), *Language and Sexuality, Contesting Meaning in Theory and Practice*, 249–266. Stanford, California: CSLI Publications.

Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Lakoff, R. (1973). Language and woman's place. *Language in Society*, 2(1), 45–79.

Livia, A. & Hall, K. (1997). *Queerly phrased: Language, gender, and sexuality*. New York: Oxford University Press.

Mann, S. L. (2011). Drag Queens' Use of Language and the Performance of Blurred Gendered and Racial Identities. *Journal of Homosexuality: Language Matters*, 58(6–7), 793–811.

McKinnon, S. (2017). "Building a thick skin for each other": The use of 'reading' as an interactional practice of mock impoliteness in drag queen backstage talk. *Journal of Language and Sexuality*, 6(1), 90–127.

Messerschmidt, J. W. (2009). 'Doing Gender': The Impact and Future of a Salient Sociological Concept. *Gender & Society* 23*(1)*, 85–88.

Ochs, E. (1992). Indexing gender. In Alessandro Duranti and Charles Goodwin (eds.), *Rethinking Context: Language as an Interactive Phenomenon.* 335–358. Cambridge: Cambridge University Press.

Ochs, E. (1996). Linguistic resources for socialising humanity. In John Gumperz and Stephen Levinson (eds.), *Rethinking Linguistic Relativity*. 407–437. New York: Cambridge University Press.

Opperman, M. M. et al., (2019). *#Nonbinary, #Genderqueer: Non-binary Gender World-making through the Networked Counterpublic of Tumblr*. Texas A&M University: Dissertation. ProQuest Dissertations Publishing.

Podesva, R. J. & Kajino, S. (2014). Sociophonetics, gender, and sexuality. In S. Erlich, M. Meyerhoff, and J. Holmes (Eds.), *The Handbook of Language, Gender, and Sexuality*, 2nd edn. 103–122. Chichester, West Sussex: Wiley Blackwell.

Sacks, H. (1992). *Lectures on Conversation*. Oxford: Blackwell.

Shar, E. & Nycz, J. (2018). *Expressing Gender and Queer Identity with Intonation: Variation Amongst Masculine, Neutral, and Feminine Aligning AFAB Queer Speakers in DC*. Georgetown University: Thesis. ProQuest Dissertations Publishing.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language and Communication*, 23(3–4). 193–229.

Steele, A. J. (2019). *Non-binary speech, race, and non-normative gender: Sociolinguistic style beyond the binary.* Ohio State University: MA Thesis.

Stokoe, E. H. (2003). Doing Gender, Doing Categorisation: Recent Developments in Language and Gender Research. *International Sociolinguistics*, 2 (1), 1–12.

Thornborrow, J. & Coates, J. (2005). *The Sociolinguistics of Narrative*. Philadelphia: John Benjamins.

Ward, K. (2019). *Building an identity despite discrimination: a linguistic analysis of the lived experiences of gender variant people in North East England*, University of Sunderland: PhD Thesis. ProQuest Dissertations and Theses.

West, C. & Zimmerman, D. H. (1987). Doing Gender. *Gender & Society* 1(2), 125–151.

West, C. & Zimmerman, D. H. (2009). Accounting for Doing Gender. *Gender & Society*, 23(1), 112–122.

Zimman, L. (2013). Hegemonic masculinity and the variability of gay-sounding speech: The perceived sexuality of transgender men. *Journal of Language and Sexuality*, 2(1), 1–39.

Zimman, L. (2014). The Discursive Construction of Sex: Remaking and Reclaiming the Gendered Body in Talk about Genitals among Trans Men. In L. Zimman, J. Davis, and J. Raclaw (Eds.), *Queer Excursions*. Chapter 2. Oxford University Press: Oxford Press Scholarship Online.

Zimman, L. (2017a). Trans people's linguistic self-determination and the dialogic nature of identity. In E. Hazenberg & M. Meyerhoff (Eds.), *(Re)Presenting Trans: Linguistic, Legal, and Everyday Perspectives*. 226–248. Wellington, New Zealand: Victoria University Press.

Zimman, L. (2017b). Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants. *Linguistics*, 55(5), 993–1019.

8 Appendices

8.1 Appendix One: Transcription Conventions

Transcription conventions, adapted from Corwin (2009).

::: colons indicate lengthening of the sound preceding the colons
( ) parentheses indicate the author's description of affect/gesture
[ ] square brackets indicate overlapping speech
word underlining indicates emphatic speech

8.2 Appendix Two: Interview Extracts

This Section contains those extracts not presented in the main text, or those not presented in the main text in full.

Extract 2: Alex (they/them or she/her, non-binary)

0421 Like some people are like, I don't feel like

0422 A man or I don't feel like a woman or something

(…)

0790 So yeah that…. U:::m and then I felt like, okay then

0791 I'm non-binary, that, yes, because I'm not really

0792 I really feel like, either in the middle o:::r

0793 In the-

0794 Not in a spectrum at all, just like... doing whatever

Extract 4: Jem (they/them, non-binary genderfluid)

0726 Those kind of

0727 Subtleties, like if I:::

0728 Wake up and I find myself not wanting to look down when I change,

0729 I'm like, 'mm, maybe I should put on a binder today!'

0730 U:::h

0732 You know, I get

0733 Casually misgendered and I'm like, meh, whatever, or I'm, it, you

0734 know

0735 **Tanks** my mood

0736 Um

0737 Kind of

0738 Is more of an indicator of how strongly I'm kind of feeling gender

0739 at the time

Extract 12: Jem (they/them, non-binary genderfluid)
1367 You know sometimes I specifically have days where I'm like, I'm

1368 gonna make it a point to try and confuse as many people as

1369 possible!

1370 So, you know, those are the days where I

1371 Put on my, you know

1372 My

1373 Leather jacket, and

1374 My binder, and

1375 Bright red

1376 You know, bright red lipstick, and (laughs)

(…)

1382 Sometimes it's like, I

1383 Cannot stand the idea of being perceived as

1384 F:::emale, so I'm just gonna give you

1385 'What the **fuck'**, instead (…)

Extract 14: Luuk (they/them, non-binary/agender)
614 Luuk: Before I used to think that I was like

615 Maybe a lesbian, but I think that is because

616 I wouldn't feel comfortable being in a relationship

617 With a man if I were perceived as a woman

618 But if

619 My

620 F:::

621 Possible like, boyfriend would

622 See me as like

623 Not a woman, as like nonbinary, or whatever

624 I think that I would like

625 Like

626 Date a man like if the

627 Occasion arises, I suppose

628 So I think that like, maybe I'm bi

(…)

633 Int: Can I ask why

634 Why you think you felt

635 Or feel that way, that if you were read as a woman in a relationship

636 you wouldn't want that?

(…)

650 Luuk: If you're like transgender that doesn't necessarily mean that you're

651 gay or straight, or whatever

652 So it's to nothing to do with

653 Each other, it's separate, but for me, like

654 Relationship

655 Is

656 About how

657 Also like, about how the person relates to you, how do you relate to

658 the person

659 And I think that

660 Gender matters in that, I mean

661 I don't know, like

662 Yeah

663 I just

664 Yeah, maybe, I mean, also just like

665 I suppose these notions of how a woman should be in a relationship

666 But not just that like, I don't know, I-I

667 For me, it

668 Is a factor in a relationship how

669 The other person perceives me of course, like

I've got a girlfriend now, but like

Also like

Now I identify

As nonbinary, if she would still see me as a woman I would also be

Uncomfortable

With that, but I would be

More uncomfortable like being in a relationship with a

Straight man who sees me as a woman

Extract 18: Luuk (they/them, non-binary/agender)

Like, I feel like people

Would be like

Less

Biologically essentialist or whatever, you know?

They would

Not

Yeah, I mean

They would not

Think that like

Clothes say anything about your gender, or that

Like, your gender assigned at birth or

Like maybe there would not

Wouldn't be gender assigned at birth, like maybe

People would just like

Let their children like choose (chuckles softly) or whatever

But like

They wouldn't

Think that like genitalia or chromosomes

Would matter that much

8.3 Appendix Three: Information Sheet for Participants and Consent Form

Modified so that personal contact details and information identifying the researchers are removed:

Information sheet for participants

Study title:	Gender Identity and Language

We're asking for your help in a study of the experiences and language of people who identify as gender non-conforming. If you agree to help us, we will be spending time with you and your peers, participating in the things you're doing but also taking some observational notes. We will also ask you to participate in a conversation, where you'll be asked some questions about your experiences, spanning from when you were younger until the present, and how these intersect with your gender identity and, also, your sexual orientation. Your answers will be audio-recorded. This conversation will last about one to one-and-a-half hours. We'll use the audio recordings to later write down our conversation, so that we can study it in more detail. No one will ever hear any part of the audio recording, itself, other than the two researchers. Your name, as well as any other names (e.g., of family, friends etc.) which you may mention during our conversation will never be used in the study. No one will be able to identify you from anything you say. I'm going to tell you more about the study if you are happy to go on. You can ask questions at any time.
I now need to explain what your rights are, what will be done with what we record, and what the benefits and risks might be. Then, I will ask you if it's OK with you for us to use our observations in various ways. If you are happy with any of these ways, then you can sign/mark the paper where indicated to show that you agree to help us and to let us use what you say in these ways. You will always be able to ask any questions you may have, both now, as well as during, and after, our conversation, should you decide to participate.

Compensation. This project is an honours dissertation and we are unfortunately unable to offer any monetary compensation. However, we hope that you will benefit from knowing that you have helped to aid research on the experiences of gender non-conforming people.

Risks. We don't know of any way that helping us with this study could harm you.

Privacy and what we will do with the information we gather. All the information we collect during the course of the research will be handled in a way that follows the Data Protection Law. For example, we normally keep recordings and descriptions separately from any details that would make it possible for other people to identify you personally. So, we might say that this experience comes from an undergraduate student of nationality/ethnicity A, for example, but we will not give any details which could be used to identify you, such as address/neighbourhood/school you attended etc. If some things you tell us are personal stories that you want to label as belonging to you or to your family, you can ask us to use your name. The choice is up to you. You can change your mind at any time.

We will use the information we gather for our study for research and teaching purposes: we might write about what we learn or talk about it with other researchers, or with the general public in TV or radio programmes. We might share the information we gather with other researchers, but we will never share any

information that would allow them to identify you personally. Audio recordings will never be shared. We will make sure that the information we gather is safe by using a storage service our University provides or by keeping it safely locked away at the University.

What are my data protection rights? The University of Edinburgh is known was 'a Data Controller' for the information you provide. You have the right to access any information that is held about you. You can exercise this right in accordance Data Protection Law. You also have other rights including rights of correction (if you think there is an error), erasure (if you want your information thrown away) and objection (if you disagree with something). For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

Your choice at all times. Even if you are happy about participating now, you can pull out of the study at any time and for any reason, even during our conversation, at any point up until the 31st of December 2019 (after which your data will be used for analysis). You are especially encouraged to feel free to not answer any questions you do not wish to answer, for any reason, no questions asked.

Contacting us. We are researchers at the University of Edinburgh in Scotland. If you have questions or comments about your rights, you can phone the people we work with (+44) 131 651 5510 or email them at lel.ethics@ed.ac.uk.

Would you like to ask me any questions now? Is it OK to go on?

Thank you for your help! Now I'm going to ask you if you are happy for us to use what we learn from you in various ways.

Participant consent and agreement to data usage

Study title:	Gender Identity and Language

PLEASE MARK EITHER 'YES' OR 'NO' FOR EVERY STATEMENT BELOW:

Consent for participation:	**Yes**	**No**
I consent to take part in the above study, including observation and audio recording.		
Agreement to data usage requests:	**Yes**	**No**
I agree to allow this session to be **audio recorded**, and I understand that my voice only be heard by the researchers on this project.		

Participant name		Participant signature		Today's date

8.4 Appendix Four: Interview Script

The author had a script of questions available at all times during the interview process, but elaboration and free discussion was encouraged, and practiced, in all interviews. The script below was meant as a starting point: not all questions presented below were asked in all interviews, and there are questions which were particular to the flow of the discussion in each interview (and are therefore not included here).

'Warm up' questions

- How are you feeling today?
- How old are you?
- How long have you been living in Edinburgh?
- What do you like most about the city?
- Are you an Undergraduate/Masters/PhD student? What are you studying?

Background questions:

- Where did you live before you came to Edinburgh?
- Did you grow up there? / Where did you grow up?
- How is/was [place X]? Anything you are particularly fond of – or not – about [place X]?
- What was your childhood like?
- What was your family like?
- How was school like for you?
- Did you feel there were any expectations placed on you growing up?

Gender Identity:

- What pronouns do you use?
- How do you identify yourself?
- What does the identity descriptor you use mean for you/how would you explain it?
- How did you come by this term, what was your journey in understanding your identity?
- How did you perceive your gender when you were younger? How did you feel? How did you like to express yourself/your identity? (e.g., what did you like to do/wear)
- How do you think your environment/family perceived your gender?

- Were you ever told you couldn't or shouldn't do, like, wear, or look like [X], because of your gender? And if yes, why? Was there anything you wanted to do/wear/like but felt you couldn't?
- What was your first experience involving a NB – or trans – [i.e., gender nonconforming] person? (e.g., someone in your environment, a friend, online, in a film/series, or even yourself etc.) How did you feel? How was this experience felt/treated by your environment, if it was?
- How would you describe your gender presentation?
- Did you ever feel you had to be or look a certain way at any point in this journey, or that something was holding you back? Do you feel so now?
- Are you 'out'/ Did you want to 'come out'? Why/why not, to whom. If you did: what felt important to you about this experience? Did you feel any pressure of any sort (e.g., to explain your identity to your family)? How did it go down, in the end, what happened? What is the situation now (e.g., with family)?
- Do you have any communities that you belong to?
- Have you ever felt pressured by any community (e.g., by expectations), and in what way(s)?
- How has it been for you being non-binary in Edinburgh?
- When do you wish people, or society in general saw gender?

Sexual orientation and embodiment

- How do you identify with regards to sexual orientation?
- What was your journey with regards to exploring/understanding this aspect of yourself?
- Are you 'out'/do you want to come out?
- How has discovering this part of yourself differed to your journey to discover your gender identity?
- Do you feel your sexual orientation interacts with your gender identity? In what way(s)?
- Have you ever been told you cannot identify with a particular orientation because of your gender, and vice versa?
- How do you feel about your body? How is it most comfortable to talk about your body?
- Any terms that you prefer to use when referring to your body, any you don't want used at all? Why?
- (if applicable) How do you feel validated or invalidated with regards to your gender identity in an intimate relationship?

Acknowledgements

I would first and foremost want to express my gratitude to all the individuals who agreed to participate in this study. Thank you for trusting me, and for sharing your stories and experiences with me. I am honoured that I got to know you, and that I can represent your voices through my research. I am wholeheartedly grateful to my supervisor, Lauren Hall-Lew. If I am a better researcher, it is because she always showed me a new perspective, especially in things I took for granted. And because a little more than a year ago I could never have believed that I would be where I am today – but she did. Lastly, thank you to the people who truly were beside me, throughout this long journey: you are 'the bright side of life' for me.

Minimax Feature Merge: The Featural Linguistic Turing Machine

Louis Van Steene
University of Cambridge

Abstract. In Minimalist syntax, linguistic expressions are typically modelled as being projected from a set of lexical items, themselves composed of three independent kinds of features (phonological, syntactic and semantic/pragmatic). The nature of syntactic features has perpetually been confused, and yet they remain the foundation of much of syntactic theory. I contest that an alternative architecture may be preferable in terms of explanatory power within the purview of mathematical biolinguistics, as described by Watumull (2012; 2013; 2015). Namely, I contest that, rather than being the driving force behind syntax, the lexicon is instead distributed amongst the interfaces in the form of non-generative lookup tables, taking Scheer's (2020) view to the logical conclusion, in parallel to DM. Syntax combines syntactic primitives freely except as constrained by the interfaces; these features are atomic, arbitrary (substance-free) computational symbols comprising the set F with cardinality at least one. Following Watumull (2015), language is considered as a mathematical structure, abstracted from its neurological substrate. This structure is isomorphic to a Turing machine, in turn isomorphic to the simplest group-theoretical object, the free magma. The central motivation is the concept of optimality captured in the minimax principle, in turn minimising the burden of the innate first factor and maximising the role of the mathematical laws and heuristics that comprise Chomsky's (2005) third factor. The ultimate aim is to begin to meet the prerequisites of explanation as defined in biolinguistics—learnability and evolvability—by formalising a theory of syntax and its place in the linguistic architecture from the ground up.

Keywords: syntax; minimalism; biolinguistics; features; interfaces; group theory

---

1 Introduction

Formalising the primitive symbols and procedures of linguistic computation is a central goal of linguistic theory, and a prerequisite for understanding the nature of language and its place within the human mind and brain. It is important theory-internally, in order to make it possible to evaluate a theory objectively. It is also important theory-externally, as one evaluates how language fits into a wider theory of cognitive science. In the spirit of the biolinguistic programme, interweaving a number of disciplines is essential to revealing fundamental facts about the nature of the human capacity for language.

In pursuit of this goal, this dissertation constitutes an investigation into the fundamental building blocks of syntactic structures. In Chomsky's (2000; 2001) typical formalisation, an I-language (internal, intensional, individual) L is taken as a device that generates a set of expressions *Exp = hP F, LFi*, i.e., mappings between phonological and logical form. Chomsky denotes the fundamental building blocks of this system as features, which are the properties of language which enter into PF, LF, and the computational system L that generates them. L is a state of the faculty of language FL, 'a component of the human mind/brain dedicated to language' (Chomsky, 2000, p. 89).

(1) An expression *Exp* is an ordered pair *hP F, LFi* where PF and LF are legible inputs to the interface between the syntactic and sensorimotor SM or conceptual-intentional C-I modules, respectively. The syntactic module L generates an infinite set of expressions *{Exp}*.
(2) The Strong Minimalist Thesis (SMT): Syntax is an optimal solution to legibility conditions imposed by the phonological and semantic modules.
(3) UG ('universal grammar') is the initial state of FL.

Significantly, constraining the output to objects of the form *Exp*, in line with the Strong Minimalist Thesis (2), entails a 'Spellout' procedure to be within the syntactic component; namely, a procedure which converts syntactic structures into representations legible by sensorimotor (SM) and/or conceptual-intentional (C-I) systems, respectively. To be explicit about the exact nature of the architecture of grammar requires a clear elaboration of the nature of the interfaces to 'narrow syntax' (NS). This represents a terminological divergence from Chomsky (2001), for whom '[n]arrow syntax maps a selection of choices from Lex to LF' (Chomsky, 2001, p. 11), and one more in line with Boeckx's (2014) conception, which devolves more power to phenomena arising from the interfaces. As such, I adopt more traditional labels for the components of grammar: we have the syntactic component, which is the primary focus of this dissertation; the phonological component, connected to the syntactic component via the phonological interface; and, the semantic-pragmatic component, connected to the syntactic component via the semantic pragmatic interface. The complexities of the latter two components lie well beyond the scope of this work; where reference is made to their inner workings, it is cursory in nature—there is a certain independence afforded to respective theories of each. The syntactic component and the nature of the interfaces in the technical, modular sense are the focus here. Thus, the 'standard working assumption [] that performance systems are external to FL' (Chomsky, 2000, p. 90) is maintained. Section 3 focuses on the aforementioned architectural considerations and proposes a modular architecture, adopting insights from various theoretical approaches.

Another, related issue with the theory presented by Chomsky (2000, i.a.) is in the definition of the lexicon Lex. Chomsky (2001, p. 10) adopts only a vague notion, in which 'Lex [is] in principle 'Bloomfieldian', a 'list of exceptions' that provides just the information required to yield the interface outputs and does so in the best way, with least redundancy and complication'. In a system which maps features to expressions, as above, the organisation of features needs to be established in a clearer way. This is the focus of Section 4, which looks more deeply at the operation at the heart of L, Merge, and how it must operate on individual features. The second operation Chomsky (2001, p. 10) assumes—which 'assembles [features] to lexical items LI of [Lex], the LIs then entering into computations as units'—is thus incorporated into Merge, making the traditional structure of Lex obsolete.

To open, however, a full framing of the research programme to which this dissertation belongs, 'mathematical biolinguistics', is provided in Section 2, which also sets the philosophical stage for the discussion. This is followed by a description of an alternative to the lexicocentric 'Y-model' architecture of grammar in Section 3. Section 4 then focuses on the internal machinery of the syntactic component of this architecture. The ultimate goal is to take a step towards overcoming the granularity mismatch problem, as described by Poeppel and Embick (2005), by proposing a sufficiently fine-grained approach to syntactic structure building and placing this component within a coherent architecture of the language faculty and its place in the mind/brain.

2 Mathematical biolinguistics

2.1 An overview

'Biolinguistics' refers to the study of 'language as a natural phenomenon—an aspect of [man's] biological nature, to be studied in the same manner as, for instance, his anatomy' (Lenneberg, 1967, p. vii). The Minimalist approach, detailed in Section 1, naturally meshes with this paradigm. Furthermore, 'issues of computation [are] central to the biolinguistic enterprise' (Di Sciullo & Boeckx, 2011, p. 6), since it is this computation which is biologically implemented. Issues of computation, and the mathematical theory surrounding it, have been a central part of the history of generative grammar[15]. This dissertation belongs to a subset of the biolinguistic enterprise, explicitly demarcated by Jeffrey Watumull (2012; 2013; 2015): 'mathematical' biolinguistics.

In mathematical biolinguistics, the focus does not lie on the precise neurobiological system – rather, the object of study is the mathematical object which is somehow physically encoded in biological wetware. The formalisation of language into its mathematical fundamentals involves, as Watumull (2015, p. 6) states, the formalising of the objects, functions, and relations of language. The mathematical formalisation is isomorphic to the mental representation, which in turn is isomorphic to its neural implementation, as detailed by Gallistel (2001) in his elaboration of the psychology of mental representations. In other words, '[w]hat matters in representations is form, not substance' (Gallistel, 2001, p. 9692). This dissertation proposes a formalisation of the objects that enter into syntactic computation (features) and the function that combines them (Merge). Note that mathematical biolinguistics is not entirely commensurable with the field of biolinguistics generally (the latter as defined by Boeckx (2013) and Martins & Boeckx (2016)), as indicated directly by Martins and Boeckx (2016, f.n. 4) where 'another author' in this footnote is referring to Jeffrey Watumull[16]. The distinction as relevant within the context of this dissertation, however, is mainly one of focus. Mathematical biolinguistics focuses almost entirely on 'software' concerns, representing a departure from biolinguistics proper which focuses on precise questions regarding the form and evolution of the biological substrate. This effectively follows from Deutsch's (2011) logic: 'if you can't program it, you haven't understood it' (Deutsch, 2011, p. 146, emphasis original). Importantly, in this context, the 'program' must have the quality of 'strong generativity', in other words in must generate the hierarchical structure of language in the same way as the mind/brain, rather than merely equivalent strings—this is what gives the 'program' explanatory adequacy. Indeed, '[t]he understanding of mental software [] precedes and conditions the understanding of neurophysiological hardware' (Watumull, 2012, p. 232), a point which lies at the heart of this Section.

Nevertheless, mathematical biolinguistics is certainly still worthy of the 'biolinguistics' moniker, not least because of the shared central concern of the two enterprises: the 'austere' requirements of learnability and evolvability (Chomsky, 2019b). Both of these constraints are given focused attention as part of Section 4; the latter is expanded upon in the following SubSection.

2.2 Evolvability, economy and the Minimax Theorem

[15] Compare Tomalin (2006) for a historiographical study of the origins and mathematical and philosophical bases of generative grammar.

[16] 'It is worth noticing that Behme (2015) is a continuation of an ongoing discussion with another author, which revolves around what is called 'Chomsky's biolinguistic ontology'—despite the far more ambitious and generic title—and as such qualifies for what we have identified as 'biolinguistics as a Chomskyan enterprise'' (Martins & Boeckx, 2016, f.n. 4). Watumull (2013, p. 302) further confuses the matter as 'my [Watumull's] work and the ontology it assumes are not representative of all biolinguistic research'. Compare Section 2.5 for further analysis of these papers.

The requirement of evolvability deserves some further attention here. It at the very least intersects with computational (and mathematical) optimality, as elaborated by Watumull and Chomsky (2020) and Roberts et al. (in press).

The assumed theoretical basis of the discussion of evolution is grounded in Balari and Lorenzo's (2012) approach, what they refer to as the 'dual nature of language' (Balari & Lorenzo, 2012, p. 5). A 'grammar' consists of an integral part of the human organism combined with a cultural invention. Phenomena like language change belong to the latter source. The former is equivalent to what Balari and Lorenzo call the Central Computational Complex (CCC): a computational system coupled with a collection of interfaces to the peripheral or external systems that it subserves. This dissertation constitutes a preliminary theory of the CCC— although I instead predominantly use the more traditional term 'faculty of language' (FL) to refer to the object of study, in a somewhat looser sense than Chomsky's (2000; 2001) since I include architectural considerations as in Section 3.

In sum, the general idea of 'optimality' is interpreted as equivalent to the Leibnizian sense of economy—a minimax optimisation (see Roberts & Watumull, 2015). In short, the Minimax Theorem, originally formulated by von Neumann (1928), is as such: 'minimise the maximum loss and maximise the minimum gain' (Watumull, 2015, p. 48). Precisely how this applies to language is detailed in Section 4. From the perspective of evolvability, this property is presumed, going 'beyond explanatory adequacy' (Chomsky, 2004), to have resulted from the evolutionary phenomenon of convergence and its effects on the CCC. An intersecting sense is provided by the theory of computational complexity which naturally arises in the context of computational cognitive science and which is the topic of the following SubSection.

2.3 Computability theory and cognitive science

The research programme pursued here is subsumed within a Gallistelian approach to cognitive science, predominantly as summarised by Gallistel and King (2010). The key ideas as relevant in this work are the concept of 'computability' as subsumed by the Turing machine, and the related notion of an 'effective procedure'.

The Turing machine is a definition of computation — one could see it (non-technically) as the 'Platonic form' of computation. Formulated by Turing (1936) as a means of solving the 'Entscheidungsproblem' (Decision Problem), it was proven that the Turing machine can compute everything that is computable, and only that which is computable — the Turing machine is equivalent to the computable functions. As such, it is a foundational component of computability theory (also known as recursion theory). To be clear, '[a] Turing machine is not a physical device—nor an intended model/simulation thereof — but rather a mathematical abstraction representing the functional conditions necessary and sufficient for any system — including the brain — to be computational' (Watumull, 2012, p. 233). 'The Turing machine is a mathematical abstraction rooted in a physical conception' (Gallistel & King, 2010, p. 125); it has three essential components (see Turing, 1936): a symbolic memory (the 'tape'), an ability to read from and write to this memory (the 'read/write head'), and a finite-state processor whose actions are determined by the current state of the processor coupled with the symbol currently being read. The Turing machine renders precise the notion of effective procedure — an algorithm which acts on symbols and produces other symbols. A simplistic way of implementing a procedure is with a lookup table, a one-to-one mapping between inputs and outputs, which quickly becomes impossibly large (and thus time-consuming to traverse) with complex functions. Much investigation into computation is the search of

procedures which are compact, or generative, meaning 'the number of bits required to communicate them (the bits required to encode the algorithm) is many orders of magnitude smaller than the number of bits required to communicate the look-up table for the function realized by the procedure' (Gallistel & King, 2010, p. 301). Generativity 'makes it possible for finite symbolic resources to pick out any referent from an infinitude of possible referents' (Gallistel & King, 2010, p. 81), allowing for d-infinity: '[a] finite system that in principle can [strongly] generate an infinite set of hierarchically structured expressions by recursively combining discrete elements' (Watumull, 2012, p. 224). The study of the computational mechanisms of language is thus the search for an effective procedure which strongly generates language[17]. This requires a consideration of the 'space' and 'time complexity' of the algorithms involved (see Mobbs, 2015).

2.4 The mathematical hypotheses and the linguistic Turing machine

There are a number of hypotheses central to Watumull (2015), the work which this dissertation extends, and consequently it is worth recapitulating them:

(1) The mathematical hypotheses
(i) The External Reality Hypothesis (ERH)
'There exists an external physical reality completely independent of us humans.' (Tegmark, 2014, p. 254, emphasis original)

(ii) The Mathematical Universe Hypothesis (MUH)
'Our external physical reality is a mathematical structure.' (Tegmark, 2014, p. 254, emphasis original)

(iii) The Mathematical Language Hypothesis (MLH)
'FL is a mathematical structure' (Watumull, 2015, p. 2).

This strong form of the MLH is in some sense not essential for much of the discussion to hold — Tegmark's structural realism (see Cohen (2008) for a philosophical approach) constitutes a somewhat extreme position; however, it can be used to justify a plausible biolinguistic ontology (Watumull, 2013; Watumull, 2015) contra Postal (2009). Furthermore, as Tegmark (2014) details, his MUH is built upon an empirical as well as a philosophical basis and is as a result entirely falsifiable. Even if this were to occur, it would not invalidate the MLH or findings contingent on the MLH, thanks to the implicational logic of the situation. Just like the MUH, the MLH has an empirical foundation — namely that of the findings of generative grammar (see Chomsky et al., 2019) — and if nothing else, it is a useful model. In other words, the mathematical baby does not need to be thrown out with the bathwater — rather, the bathwater could even be kept and sold on for profit. Nevertheless, the hypotheses in Example 4 will be assumed in their full force in the present context, in line with Watumull (2015).

[17] Contrasting with 'weak generativity', which only generates strings, rather than 'hierarchically structured expressions' (see Watumull, 2012).

The MLH needs to be elaborated further — namely by defining the mathematical structure in question. Watumull (2012; 2015) defines this as the 'linguistic Turing machine' (LTM). The LTM is defined as such[18]:

(1) The linguistic Turing machine LTM is the quintuple (Q, Γ, δ, $\#S$, $\#H$) Q: states
Γ: symbols (not including the blank symbol, see Sipser, 2012, p. 168)
δ: transition function, from one state/symbol pair to another
$\#_S$: start symbol
$\#_H$: halt symbol[19]

The LTM encodes the computable function f_{MERGE}, defined as in Example 6[20]: (6) $f_{MERGE}(X, Y) = X, Y$ for $X, Y \in \Gamma$.

The set of symbols Γ is the set of syntactic objects (SOs), the union of the urelements (LIs) with the result of the recursive application of f_{MERGE} on the LIs. The idea of 'recursive application' is crucial and is facilitated in the LTM by the 'tape', the persistent memory of the Turing machine. The output of each step in the computation is stored on the tape, which serves as the input to the next step. Further consequences of this are discussed in Section 4.

The Turing Program for Linguistic Theory (TPLT) (Watumull, 2012) is a label for the investigation of LTM and thus constitutes a subset of mathematical biolinguistics. In particular, the goal is to demonstrate that '[t]he mathematical universality of the Turing machine — the abstract system of computation — implies that it is not only relevant to biolinguistic research, but intrinsic to linguistic — indeed any cognitive-neurobiological — computation' (Watumull, 2012, p. 222, emphasis original). This is self-evident under a Gallistelian approach, thus understanding exactly what kind of computation is needed for language is critical to the programme. Section 4 is thoroughly grounded in TPLT, as a reconceptualization of Watumull's (2015) LTM using features, rather than LIs, as primitives. The majority of Watumull's proofs hold despite this adjustment, as they are not contingent on the *urelements*, the indivisible constituents of sets. There is not room for a thorough mathematical breakdown of this; however, some theoretical consequences of this proposal are discussed in Section 4. For now, the point is that this objective clearly aligns with the central challenge proposed by the TPLT: '[t]o precisify (formalise) the definitions of linguistic primitives in order that 'linking hypotheses' (not mere correlations) to as yet undiscovered neurobiological primitives can be formed' (Watumull, 2012, p. 227). These 'as yet undiscovered neurobiological primitives' align with the Gallistelian search for the neural mechanisms of computation, which, as Gallistel and King (2010) speculate, may be sub-neuronal. Crucially, under the MLH and within the TPLT, the precise formulation of the LTM is not 'play[ing] mathematical games', but rather 'describ[ing] reality' (Chomsky, 1975, p. 81) 'the theory needs to be mathematical because the phenomenon is mathematical' (Watumull, 2012, p. 229).

Finally, note that the treatment of FL as a mathematical structure is implicit in, or at least compatible with, much of the work that falls under the umbrella of biolinguistics. The search is for a biologically implement(ed/able) generative grammar, an invertible function which maps between PHON and SEM. The

[18] Equivalent formulations are possible. This particular formulation has been adapted from Watumull (2012) for expository purposes.
[19] See Watumull (2015) for proofs and derivations using his formulation of the LTM.
[20] A terminological note is in order: I use f_{MERGE} specifically to refer to the mathematical function as defined in Example 6, or as defined in Section 4 below. The term 'Merge' will be used to describe the binary set-forming operation more generally, for instance as considered in the wider Minimalist literature.

function must be biologically implementable because it is evidently biologically implemented, as decades of empirical evidence provided in support of generative grammar have more or less proven. This provides us with necessary constraints on what this system can be like—which can be derived from the essential properties of a biological system: learnability and evolvability. These two parallel constraints find a special case in language but are considered essential throughout the evo-devo programme. Central to mathematical biolinguistics and the TPLT is the Gallistelian notion that the brain is a 'biological computer', subject to the same universals of computation uncovered by Turing. While the specifics of language must be constrained biologically, notably at the interfaces, its core inherits the power and constraints of the computational system underlying it — the LTM. The architecture that enables this is the subject of Section 3.

2.5 Ontology

In exposition of the ideas detailed in the preceding SubSections, I turn now briefly to a criticism of mathematical biolinguistics levelled by Behme (2015), which does not yet seem to have been directly addressed in the public domain[21]. Behme's article constitutes a response to Watumull (2013), which in turn is a response to Postal's (2009) critique of Chomsky's biolinguistic ontology — I concentrate here on the objection to Watumull, leaving the broader attack on Chomsky aside for lack of space (cf. Martins & Boeckx, 2016)[22]. Fundamentally, Behme's criticism is based upon a misunderstanding of the essence of Watumull's argument: Behme repeatedly makes reference to the idea that physical objects cannot have abstract properties, but this is incoherent under Watumull's assumption of mathematical realism and the MUH. Behme and Postal are unable to reconcile the abstract and the physical because they overlook the significance of computability theory (Turing, 1936): the infinite can be compressed into the finite, in the form of an effective procedure. To recite the quintessential example given by Turing (1936), one can design, and physically implement, a procedure that can (given infinite time and memory) compute the irrational (infinite) number pi. Pi has been compressed into a finitary procedure, one which can be physically implemented in a finite machine. In the exact same way, the LTM compresses the infinity of language into a finite, physically implementable machine. The infinite can be represented physically. To force the point: one can imagine a very simple program which writes the number '1' indefinitely[23]. Voila, infinity captured in a finite procedure, nullifying 'Behme's repeated denial of d-infinity, asserting that '[t]he knowledge [of language — LVS] of any speaker is finite because their brain and their lifetime are finite' (Behme, 2015, p. 39). Behme and Postal are implicitly adhering themselves to the connectionist dogma, that functions must be implemented using a lookup table architecture — the architecture adopted by neural network models. The assumption that neural networks accurately model what is happening in the brain leads to the conclusion that effective procedures, i.e., Turing machines, cannot be physically realised in the brain. What Behme and Postal do not appear to realise is that this approach is inherently anti-representational, as

[21] Martins and Boeckx (2016) briefly mention the article as part of a wider point.

[22] The criticism of Behme here is not to be taken as a glorification of Watumull's article, which is not infallible. For instance, Watumull's persistent use of short, even single-word, inline quotes taken out of context is not, as Behme notes, conducive to his argument. Further, Behme and Postal's overall critique of Chomsky's biolinguistics is provocative even in spite of their repeated misapprehensions of the programme, but a full discussion thereof lies beyond the scope of this dissertation (cf. Martins & Boeckx, 2016). For the record, I believe that there is considerably more scope for agreement between the two sides than is apparent from the rhetoric.

[23] i.e., with the pseudocode 'print '1' while true'. Another simple function which could be implemented in an effective (finite) procedure is the successor function *succ*(*x*) (cf. Roberts et al., in press).

Gallistel and King (2010) elaborate, since there is no way for the network to read its current state, without at the same time changing said state. In other words, only a god outside the machine can read off values from a neural network, making these representations meaningless within the network, and by extension within the brain[24]. In sum, '[c]onnectionists draw their computational conclusions from architectural commitments, whereas computationalists draw their architectural conclusions from their computational commitments' (Gallistel & King, 2010, p. ix). To be clear, Behme and Postal do not explicitly represent a connectionist position; however, Behme's implicit denial of the neurobiological relevance computational principles lying behind Watumull's (2013) mathematical biolinguistic ontology on the basis of their 'lack[ing] any biological content' (Behme, 2015, p. 37) is reminiscent of 'draw[ing] computational conclusions from [] architectural commitments'[25].

Further to this, Behme appears to mistake Watumull's claim of isomorphism between the physically encoded symbols in the brain and the symbols of mathematics to mean that the brain must interact with abstract Platonic forms in a kind of mangled Cartesian dualism. The rebuttal to this again follows from a point which is almost obvious from the perspective of computational cognitive science, assumed in the Gallistelian paradigm in which Watumull is working: isomorphism between representations does not mean that they are identical (see Gallistel, 2001). Rather, it is a defining property of representations that they are isomorphic to what they represent — otherwise they would be of no use at all. To analogise to an example of Gallistel and King (2010), if direction vectors were represented in the mind of an ant in a form such that the only operation available to them (say, +) merged the two vectors into a set (say, the form given to syntactic objects), the representation would be completely useless since it has no functional basis in the ant's representation of the world. In the context of language, Watumull's claim that the representations in his LTM are isomorphic to the representations encoded by the brain is an empirical hypothesis. Indeed, Behme (2015, p. 37) seems to overlook this important point, dismissing Watumull's perfectly acceptable claim (in another article, though repeatedly quoted by Behme) that 'as yet undiscovered neurobiological primitives' (Watumull, 2012, p. 227) are the basis for the kind of linguistic computation Watumull espouses. This is certainly in the spirit of Poeppel and Embick (2005) in that it sets the stage for potential neurobiological study, which may reveal the kind of sub-neural computation suggested by Gallistel and King (2010), to which Watumull is clearly implicitly alluding. As such, this contributes to a quintessential aspect of the biolinguistic enterprise, that of taking Poeppel's concerns regarding the potential for interdisciplinary work between linguistics and neurobiology seriously (Gallistel & King, 2010; cf. Boeckx, 2013; Martins & Boeckx, 2016).

Behme also appears to have the causal link between the investigation of biology and linguistics within this research programme backwards. The plausibility of the 'abstract' approach to language as a mathematical object derives from the theory surrounding the Minimalist Program. In short, in order for language to be evolvable and acquirable, it seems to be a reasonable starting point for a hypothesis to say that language is an 'optimal' solution[26]. What this hypothesis comes to mean under the Watumullian program is that language is isomorphic to the LTM. Correspondingly, as in the Minimalist Program generally, a large amount of effort goes towards understanding the 'interface systems', where 'imperfections' are hypothesised to arise. Biological constraints naturally arise—arguably nowhere is this more prominent than in the phonological and phonetic domains, where physical constraints like the need to

[24] The most commonly used mechanism to update the weightings of neurons in a neural network, back-propagation, relies on a god outside the machine in order to read and adjust these values, a matter reluctantly accepted by neuroscientists but one which makes their models ontologically incoherent (Gallistel & King, 2010).
[25] Note that 'architecture' in this quote is being used in a sense entirely separate from the topic of Section 3.
[26] Precisely what 'optimality' means here is a major theme of Section 4; cf. also Section 2.2 and Mobbs (2015).

externalise language in a chrono-linear fashion cause the output of the LTM to be manipulated in increasingly well-understood ways (see Section 3; as well as Kayne, 1994; Uriagereka, 1999; Chomsky, 2008). Pace Behme's interpretation, then, the goal is not to fine-tune the theory to be as bio logically appropriate as possible, but rather to assume that language is 'perfect', and then to discover the biological constraints that impose limits on the output of a physical implementation of such a system, and to determine whether these constraints can reliably be considered part of the 'interface' or whether they require high-level adjustment to the LTM. To this end, I submit that a proper hypothesis regarding the architecture of the linguistic system is needed, in order to clarify the relations between the LTM and other linguistic and non-linguistic cognitive domains. This is the topic of the following Section. Such a hypothesis allows inferences to the structure of the high-level structure of the LTM to be made, namely in the elaboration of what exactly 'optimal' means. Section 4 takes this a step further, suggesting that inferences can be made towards a refinement of the structure of the LTM via application of the minimax theorem upon this architectural basis[27].

3 Architecture

3.1 The lexicon and projection

The traditional generative view of the architecture of the grammar is captured by the 'feed-forward Y-model', as depicted in Figure 1.

As described in Section 1, the lexicon in this model is a list of exceptions composed of bundles of 'features', which are circularly defined as the primitives which make up lexical items. I shall argue that this identification of features proves to be wholly vague and reductionist, and as a result unfalsifiable and thus untenable as a theory. More acutely, it is difficult to find a coherent theory of features throughout the Chomskyan programme — often they are taken as axiomatic, leaving their ontology in the dark, a problem which is widely acknowledged in the literature (Adger & Svenonius, 2011). Furthermore, the definition of features given by Chomsky (1995b) can be argued to violate strict modularity, particularly in the case of 'interpretability' of features, as argued by Zeijlstra (2014). This approach implicitly assumes that computational symbols (interpretable features) must be shared between computational modules, when in fact the exact opposite is the case, as made explicit in the condition of domain-specificity. Even if phonological, grammatical and semantic features are clearly demarcated sets, strict modularity entails that they are completely illegible to the modules they do not belong to, unless they are translated at an interface. In the architecture adopted here, features which appear to be 'interpretable' are merely those features which are translated in a certain special way at the interface, but whose meaning is not dictated by the syntactic feature itself. This also gives room for manipulation of how features are interpreted to take part in the acquisition process — since substance-free features are underspecified in UG. Evidently, the issue of finding an appropriate definition of 'features' is inherently intertwined with architectural considerations — an issue already clear from the phonological literature (see Hale & Reiss, 2008).

[27] Indeed, Section 4 goes some way to addressing Braine's (1992, p. 79) related point of there not being a clear theory of the syntactic primitives which need to be somehow biologically (genetically) encoded, and subsequently employed to 'bootstrap' acquisition (cf. Pinker, 1984)—a citation which Behme (2015) also invokes in her criticism of Watumull (2013).

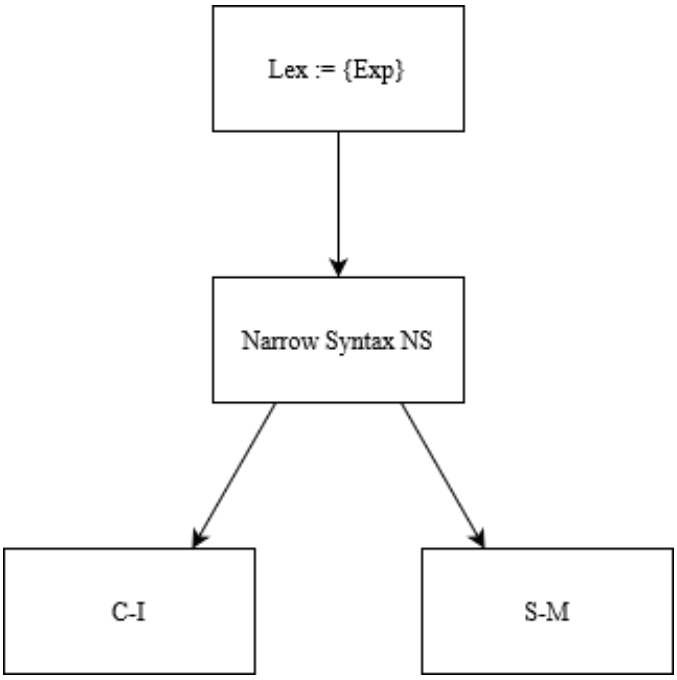


Figure 1: *The Y-model.*

In the Y-model, therefore, the so-called 'projection principle' is foundational. The principle was explicitly formulated in the Government & Binding era, predating modern Minimalism, as 'the idea that all syntactic structure is projected from the lexicon' (Haegeman, 1994, p. xx). In Bare Phrase Structure (Chomsky, 1995a) the principle in its original formulation is incompatible, due to the elimination of X-bar structure; however, Boeckx (2014, p. 6) argues that it persists so long as the lexicon remains the base of the derivation, as in Figure 1 'minimalism has so far failed to distinguish itself from previous transformational accounts, which relied on [the projection principle]'. This is clear from the way Chomsky (1995b) describes the process of lexical 'selection' into the 'numeration', the first step of the derivation. As such, lexical items and their constituent features drive the derivation: the syntax begins with a subset of the lexicon (call this the 'numeration' or 'lexical array' or whatever) and the derivation is constructed from this. Furthermore, the principle persists in conjunction with the Borer-Chomsky Conjecture (BCC) (Baker, 2008, p. 354), such that 'derivations [are] driven by morphological properties [i.e., features — LVS] to which syntactic variation of languages is restricted' (Chomsky, 1993, p. 44). This is the root of what Boeckx (2014) calls 'lexicocentrism'. In the following Sub-Sections, I make the case for eliminating lexicocentrism by eliminating the lexicon as the root of the derivation, in a manner similar to Distributed Morphology (DM) (Marantz, 1997). I integrate this architecture with the representations used in the cartographic, one-feature-per-head approach adopted by Nanosyntax and their interpretation at the interfaces (Baunaz & Lander, 2018; Caha, 2020).

3.2 The semantic spine

When defining an architecture, it is vital to decide on the division of labour between the different modules. Since each module is afforded its own encapsulated computational symbols and procedures, this decision can have empirical consequences. It is interesting, therefore, that in a framework like Nanosyntax, as

Baunaz and Lander (2018, p. 5) openly admit, 'syntax is assumed to be the vehicle for expressing grammatical semantics, and it does so by means of abstract syntactico-semantic features that are arranged by syntax'.

In discussion of 'syntactico-semantic features', it may prove a useful analogy to explore the perpetual focus on features in phonological theory, where defining the inventory and properties of features has been part and parcel of any theory dating back to Jakobson et al. (1928; 2002). It is likely that this is the result of the more apparent 'surfacy' effects of (traditional) phonological features, which have predominantly been linked directly to acoustic and/or articulatory properties, e.g., as in Chomsky and Halle (1968). The same cannot be said of many proposed syntactic features, such as at the extreme the elusive 'EPP-feature' (Chomsky, 2001, p. 4) and its variants, which do not have clear physical correlates. Scheer (2020) refers to this as the degree of 'faithfulness', which makes sense only at the phonology-phonetics interface, but not the syntax phonology one. Nevertheless, the 'substance-free' approach to phonological features entails that this faithfulness is illusory (Hale & Reiss, 2008). The central thesis of this research programme is that phonological features are doubly arbitrary computational symbols with no universal correlates to syntactic or phonetic features beyond language-specific realisation rules. Phonological features are in fact abstract symbols of computation, just as 'formal' syntactic features are[28].

Returning to syntax, cartographic approaches make no attempt to hide the radically sub stance*ful* nature of the functional hierarchy and its features. Blending cartographic insights with a substance-free theory is thus a fairly radical step; however, it is not at all paradoxical in the search for a truly explanatory theory. The directive that syntax and semantics ought to be separated in a modular sense holds, leading to the incentive to separate what is syntactic from what is conceivably some kind of semantic 'spine', corresponding to that of e.g., Cinque (1999), Wiltschko (2014), Ramchand and Svenonius (2014), or some combination.

This leads to the question now of where this 'spine' is stored — is it integrated into the syntactic inventory, as in cartography? Or is it somehow a phenomenon merging from the syntax semantics interface? It is of note here, as noted by Chomsky (2019a) and others, that much of what is typically called 'semantics' in linguistics is actually just a form of syntax. Grammars augmented with rules that are imbued with lambda calculus representations of compositional meaning, are inherently 'syntactic' — technically speaking, a 'semantics' only arises when this representation is linked to the world. This, then, makes sense of the cartographic fusion of 'syntactico-semantic' features which lead to the equation of the syntactic module in Nanosyntax to 'SMS [, which] stands for syntax, morphology, and semantics, which in nanosyntax are seen as one and the same module' (Baunaz & Lander, 2018, p. 11). This is further equivalent to a 'language of thought' as conceived of by Chomsky (2017, p. 200) and the similar idea expressed by Sigurdsson (2020, p. 7). The 'linking to the world', or more precisely to extralinguistic components in the mind/brain which may then transduce their contents into physical phenomena, is performed by C-I, which I have also been referring to as the semantic module.

My inclination is to say that these restrictions on syntactico-semantic structure are indeed interface phenomena. What is deemed a 'feature' in the cartographic sense is (a part of) an entry in the syntactic/semantic-pragmatic interface lookup table. The 'spine' is similarly a 'lexical' concept. Whilst I have thus far assumed for simplicity that all interface translation rules are generated in the process of language acquisition, it is plausible for some to be innate, or further for them to emerge in minimax fashion

[28] Terminologically, I avoid the conventional use of 'formal' to distinguish syntactic/grammatical features, since as symbols they are equally as formal as any other computational symbol. The question remains of what, if anything, substantively distinguishes the different sets of features.

out of constraints imposed within the syntactic or semantic-pragmatic modules themselves. The formalisation of acquisition lies beyond the scope of this work, but it is clearly at least plausible.

3.3 A modular architecture for FL

Each stage in the computation has access to a particular inventory of symbols and an inventory of computations. The module takes a representation formulated in said symbols and from this generates an output determined by the input and the computational inventory, formulated in the same set of symbols.

In between each module must be a translator which converts representations legible in one component to representations legible in another[29]. This is effectively a function between two mutually exclusive sets — compare the function(s) within a module, which are defined by closure over the set of symbols. There are two hypotheses for the nature of this transducer function, known as 'lexical translation and computational translation', as elaborated by Scheer (2020). Lexical translation takes the form of a lookup table — a weak computational architecture, representing the bare minimum needed to represent a function, with maximum specification and no compression. On the other hand, computational translation, as the name implies, applies a computational procedure to the input, transforming it into an entirely different symbolic space. In other words, computational translation is an unhappy hybrid of lexical translation with the fully-fledged computational properties of a module. A computational translator is thus a module with a symbol inventory $U := M_1 \cup M_2$, the union of the inventories of two other modules. This contradicts one of the key motivations for modules in the first place: domain-specificity. If a module can 'understand' two symbolic inventories, there's no need for a 'translation' to take place at all — all computation can take place within the single module, so 'interface devices [are] pointless: if modules can parse the vocabulary of their neighbours, no translation is needed in the first place' (Scheer, 2020, p. 183). Therefore, lexical translation is adopted in this architecture as the model for the interfaces. Furthermore, this corresponds with Boeckx's position (directly following from Marantz, 1997) that the lexicon should not be generative. In line with DM, in which the object that is 'distributed' is the lexicon itself (Marantz, 1997), it seems to be a natural conclusion that the lexicon is in fact an interface phenomenon—in the literal sense that the interfaces constitute the lexicon. Indeed, this is compatible with Scheer's lexical translation, unifying the two approaches. The distributed approach is incompatible with Nanosyntax, however, which avoids lexicocentrism in the sense that it reserves the syntax as the driving force of derivations, rather than projection from the lexicon, but which treats the lexicon as a unified object. Lexical entries in the nanosyntactic lexicon consist of three representations (SO, PF, LF), which can be modelled with the lookup architecture.

In sum, we have modules, and we have interfaces between modules, as depicted in Figure 2[30].

[29] This is in fact distinct from Pylyshyn's (1984) transducer (as adopted by Hale & Reiss, 2008), which converts between symbolic and physical representations, e.g., from phones to sound waves. The phonological system has its own computational module intervening between the output of syntax and SM, and thus its own rules and representations, which may be in the form of a 'phonological Turing machine' (see Vaux & Watumull, 2012).

[30] The increasingly technical specification of modules in the mind should not set off neurological alarm bells— the idea is not phrenological, literally dictating that there is a direct physical correlate of the 'interface-as-list' somewhere in the brain. Rather it constitutes a model for the operations of the mind, with plausible but undefined for each interface (corresponding to DM's Lists 2 & 3).

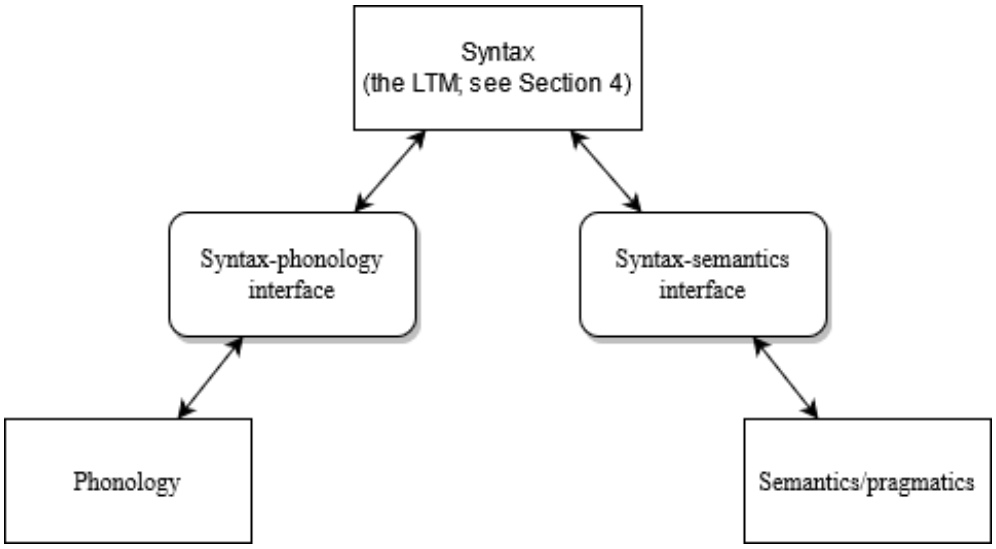


Figure 2: *The architecture.*

The LTM takes the role of Nanosyntax's SMS, incorporating a feature inventory with a Merge operation (the topic of Section 4). A computational module is equivalent to the function $f : X \rightarrow X$; an interface is equivalent to the function $g : X \rightarrow Y$; $X \cup Y = \emptyset$. Different interfaces use different sets of symbols corresponding to X and Y ; however, the general form of the representation is isomorphic between all instantiations (cf. Section 2). In line with the desideratum of mathematical biolinguistics, apparently complex concepts have been reduced to mathematical structures.

This architecture of grammar yields both methodological and conceptual advantages. Considering methodology: due to the flexibility of the interfaces and relative independence of the components, there arises little conflict between individual theories of each of the components, even when radically different philosophical standpoints are adopted. The phonological and semantic components are completely independent of the syntactic component and are permitted to use any form of input as appropriate, assuming that a computable function can be defined to transduce syntactic output into legible phonological/semantic representations — be they the linear representations of Government Phonology (Scheer, 2004) or the arboreal structures of Feature Geometry (McCarthy, 1988; Halle et al., 2000), or similar alternatives on the semantics side. The function g is equivalent to what is usually denoted the 'lexicon', although there is a lexicon neural implementation. The plausibility derives from my adoption of Gallistel and King's (2010) well-argued hypothesis that the brain effectively implements a computational model equivalent to the Turing machine (see Section 2).

With this architecture established, the next task is to establish the inner workings of the LTM, which is the subject of the next Section.

4 Minimax Feature Merge

4.1 Motivations

There are two key motivations for the reformulation of the LTM which is the focus of this Section. The first is deduced from the architecture presented in Section 3, which renders untenable Watumull's (2015) original conception of the LTM, with LIs as the irreducible (from the perspective of the LTM) urelements. In particular, since there is no pre-syntactic lexicon, but rather only a pre-syntactic set of features, the lexical item cannot be a primitive of computation. The procedure encoded by the LTM, f_{MERGE}, must directly manipulate features, or rather constituent structures composed of features — syntactic objects (SOs). Note that even complex SOs are not LIs as defined in Section 1, neither are the translation rules constituting the interface equivalent to LIs; the latter I denote 'lexical entries' to emphasise the distinction.

The second motivation derives from the 'minimax' idea introduced in Section 2.3, 'minimise the maximum loss and maximise the minimum gain'. Relying on LIs as primitives requires considerably more overhead, since LIs take up more space than individual features, and are thus less economical in terms of space complexity (cf. Mobbs, 2015, p. 141). Additionally, LIs are opaque: they are, for the most part, treated by Watumull's (2015) LTM as atomic — the urelements of Merge, which do not need to be 'seen into' for the LTM to function. Indeed, this atomism fails in practice: in the sample derivation provided by Watumull (2015, p. 97) complex featural computation is included in the LTM by way of another (undefined) function, f_{SELECT}, which putatively performs minimal search to detect the LI with an appropriate feature fingerprint, consequently providing it to his f_{MERGE}. The problem with this is immediate from the strictly modular approach: separate operations within the same module appear to be dealing with different computational primitives (symbols) — there is confusion between *syntactic* objects and features, the latter of which are not valid SOs without a parent LI[31]. The 'configuration' is not uniform — rather, two configurations are cascaded: the configuration of LIs and the configuration of features. This problem is entirely eliminated if features are inherently valid SOs, as in the present model, where the problematic intermediary of the LI is eliminated. This is achieved by redefining the LTM to use features, and syntactic objects composed of constituent features by Minimax Feature Merge, as the objects of computation, as is the subject of Section 4.2.

'Minimax Feature Merge' is inductively defined in Section 4.4; however, some preliminary notes are in order. Watumull (2015, p. 27) denotes f_{MERGE} as encoded in the LTM as 'Minimax Merge'. The function is 'minimax' for a conceptually simple, yet vital reason: 'the optimal function for generating syntactic objects is minimax: an 'effectively unary binary function'' Watumull (2015, p. 27, emphasis original). How can a binary function be 'effectively unary'? The answer is natural in the context of the Turing machine, since it corresponds directly to the concept of the 'tape': one of the inputs to the next operation of Minimax Merge is the output of the previous stage in the computation. The tape — the Turing machine's memory — persists between stages in the computation (this is one of the key properties which distinguishes the Turing machine from less powerful automata, see Gallistel & King, 2010). Thus, the first argument in Minimax Merge is always satiated by the current contents of the tape. Minimax Feature Merge inherits this property and is hence a minimax-optimised function, creating the maximal possible output from the minimal possible input. The idea of minimax can be further extended in the present case by looking at features in particular, and how they themselves may be optimised to provide the most information with minimal overhead. This is the topic of Section 4.3.

[31] This is not to say that an LI cannot *contain* a single feature. But if Merge manipulates individual features by incorporating e.g., Chomsky's (2015, p. 240) 'Move F', then, if Merge is also minimax (effectively unary), you arrive at the nanosyntactic view (and the one adopted in the present proposal) of the structure of syntactic representations for free.

It is of further interest to compare this notion of memory to the notion of the workspace (WS), as used by Chomsky et al. (2019). In this configuration, the Merge function is actually ternary (cf. the definition given by Chomsky, 2019a), needing three inputs: the two objects to be merged *and* the workspace as a whole, which may not be the same as the other two SOs. This conception of Merge is thus not minimax in the sense of being effectively unary. It requires far greater complexity, since two objects must be selected to be merged, rather than the single object needed by Minimax (Feature) Merge. Effective unarity also directly relates to the Extension Condition, whereby 'Merge always applies in the simplest possible form: at the root' (Chomsky, 2015, p. 227). This is a given, since one argument of Minimax (Feature) Merge is always satiated by memory. A non-minimax Merge requires the Extension Condition to be a stipulation, rather than something which emerges from the mathematical properties of the function itself. In the present proposal, WS is assumed to be isomorphic to the tape[32].

4.2 Reformulating the LTM

With this preliminary discussion in place, it is possible to provide a formal definition of the featural LTM. As compared to the LTM in Example 5, the featural LTM requires an entirely different vocabulary than that represented by Γ, since Γ contains 'lexical items' which have no status in my theory. Instead, the urelements will be denoted by F, and the vocabulary of the featural LTM by S:

(1) F is the set of syntactic primitives such that $|F| \geq 1$.
(2) The members of F comprise the urelements in S.
(3) S is defined as the union of F with the result of the inductive application of f_{MERGE} to F.
(4) $f_{MERGE} : S \times S \rightarrow S$

f_{MERGE} is a function that takes two members of S as input and returns a member of S as output.

$f_{MERGE}(X, Y) = \{X, Y\}$ for $X, Y \in S$

The featural LTM, as with the LTM in Example 5, is isomorphic to a free magma. A magma is the most basic mathematical structure in universal algebra, consisting of a set S and a binary operation closed under S. A magma is labelled 'free' if no other restrictions are applied (cf. Watumull, 2015, p. 7).

(5) The syntactic component L is isomorphic to the free magma (S, f_{MERGE}), which in turn is isomorphic to the featural LTM.

I return to some interesting consequences of this isomorphism in Section 4.5.

4.3 Features

[32] The question of how adjunction should work in Minimalist Syntax, since it appears to be incompatible with the Extension Condition, is still a topic of debate and there is not room for discussion here.

Despite the pervasive use of the term ‘features’, there remains no generally accepted adequate formal definition thereof. Kibort and Corbett (2010) summarise the issue well:

> ‘Despite their ubiquity and centrality in linguistic description, much remains to be discovered about [features]: there is, for example, no readily available inventory showing which features are found in which of the world’s languages; there is no consensus about how they operate across different components of language; and there is no certainty about how they interact.’ (Kibort & Corbett, 2010, online abstract)

The substance of features in this architecture was briefly discussed in Section 3.2 in relation to the notion of a cartographic ‘spine’. As defined in Section 4.2, however, the featural LTM in principle requires only a single feature to satisfy isomorphism with a free magma. As such, it also does not specify any kind of ‘content’ needed within a feature — to f_{MERGE}, all features are internally uniform (unlike LIs were), although f_{MERGE} is able to distinguish different features externally if $F > 1$. Note that increasing the cardinality of F adds a layer of complexity which takes f_{MERGE} beyond what Chomsky et al. (2019) denote Simplest Merge and note further that this would require corresponding adjustment of the transition function δ of the LTM in order to read and write the new symbol(s) directly.

It is worth then discussing what effects varying the contents of F has on the overall syntax, and the characters of the free magmas it generates. If $F = 1$, then we arrive effectively at Boeckx’s (2014) model, which has only a single feature, denoted the ‘edge feature’ EF. As the only feature, the edge feature is necessarily syntactically inert, serving merely as a remnant of the derivation. This leads to highly uniform trees, which seem perhaps overly simplistic for a language model. Boeckx (2014) resolves this uniformity by appealing to phase theory — the idea of cyclic transfer to the interfaces. If every other phrase is a phase, then phrases can be distinguished, and different properties emerge. Boeckx (2014, p. 44) argues that this gives transitivity: the transitive phase head and the intransitive phasal complement. Syntax is thus ‘regulated’ by the cycle, which prevents unconstrained iteration. A more complex approach is taken by Sigurdsson (2020), who allows for multiple kinds of EFs to be distinguished in the syntax.

The featural LTM with $F = 1$ can be denoted ‘interface-driven’ — in other words, Merge here is free, non-linear, and automatic, as in Boeckx’s (2014) ‘Merge-alpha’ and Chomsky et al.’s (2019) ‘Simplest Merge’. It stands in direct opposition to ‘crash-free’ syntax as perhaps best espoused in the work of Samuel David Epstein and colleagues (Epstein et al., 1998). From the lexicalist perspective, I would argue that this approach makes most theoretical sense, in line with Frampton and Gutmann (2002) and ‘pace’ (Brody, 2002). Any representation constructed by f_{MERGE}, i.e., any member of S, can simply be decomposed into the derivational sequences which constructed the representation. With such a barebones formulation of the function, it is difficult to see how ‘opacity effects’ which obscure the steps of the derivation can be introduced into a representation. This even applies at the interfaces, which need not allow for the ability to take complex, holistic representations as input rather, this data can be ‘serialised’, much as multiple Spell-Out entails in phase theory. Nevertheless, in the mathematical biolinguistic framework, Watumull (2015) argues that much of the derivation-representation issue is moot. The intuitive perspective of computational processes is that they proceed sequentially in time, but ‘the passage of time’ is a purely subjective experience. Objectively, spacetime as a mathematical structure just ‘is’ Tegmark (2014). A computation, like that described by the LTM, is just an alternative description of a mathematical structure. Thus Watumull’s (2015, p. 14) conclusion: ‘[] the derivation/representation debate in syntactic theory is meaningless: subjectively, linguistic expressions are derived in time; objectively, they are simply derived

— represented — in the abstract.' A derivation using the LTM is analogous, indeed, isomorphic, to a proof: the steps taken in the proof are the proof, representationally; analogously, the steps taken in the syntactic derivation using Minimax (Feature) Merge are evident in the final representation. It seems to me, however, that this actually lines up more or less entirely with the position taken by Epstein et al. (1998, p. 12), where 'the post-syntactic computational systems must examine not 'output representations' to determine [syntactic] relations, but, rather, the derivation'. In a free magma, there is by definition nothing more to a representation than the derivation – everything is there, nothing is opaque; the LTM is deterministic.

4.4 f_{MERGE}

Minimax Feature Merge is a recursive function in the sense provided by computability theory as detailed by Watumull et al (2014) and as is essential to its qualification as minimax. A recursive function has three properties: the first is that of Turing computability, the second and third those of induction — definition by induction and mathematical induction, respectively. The fact that f_{MERGE} is Turing computable is evident from its encoding in the featural LTM — namely, it is an effective (finite) procedure, as defined in Section 2.3. Its inductive properties have been implicit in the definition given of the free magma L — the set S is defined by the inductive application of f_{MERGE}. This set, consisting of the syntactic objects, is strongly generated, meaning that a structure given as the argument for f_{MERGE} is one that has already been generated by f_{MERGE}.

The second sense of induction, mathematical induction, extends strong generation to be unbounded – the set S is unbounded. Definition by induction is also a property of the successor function, which derives the natural numbers. Indeed, fMERGE further reduces to the successor function as shown in Section 4.5. Returning to a theme of Section 2: 'finite brains running in finite time literally do generate infinite sets' (Watumull et al., 2014, p. 3, emphasis original).

4.5 Mathematical language

In this final SubSection, I return to the more philosophical matters of Section 2, linking the featural LTM to the desideratum of explanatory adequacy.

The fact that language, assuming my argument to hold, can be 'reduced' to mathematics — that language is isomorphic to a mathematical structure — is not only a matter of beauty, but has concrete consequences for linguistic theory. Mathematics is absolute, so the search for a falsifiable theory of language 'as it is in people's mind/brains' is not only plausible but perhaps inevitable, given enough resources. This is only possible if 'strong generativity' is adopted as the goal, in itself a mathematical notion, associated with the theory of computation and the Turing machine, and thus with the TPLT as presented in Section 2.4. The evidence in favour of the MLH comes from many sources as is characteristic of the Minimalist Program, and indeed biolinguistics generally. Some are in the form of third factors (Chomsky, 2005): for instance, computational efficiency — which crucially has ties to the explanatory requirement of evolvability. As detailed above, f_{MERGE} minimises computational cost and maximises generative capacity, exploiting the nature of the free magma, the simplest mathematical structure but one with the potential for infinite (unbounded) generation, as demonstrated by the isomorphic LTM. There is also empirical evidence in favour of this, adapted from a famous experiment by computer scientist Marvin Minsky and one of his students, Daniel Bobrow. The set of all Turing machines is 'countably infinite' — countable, since they can be listed (unlike, say, the real numbers R), but infinite, since this list would go on forever (cf. Sipser,

2012, p. 206 for a proof). Minsky and Bobrow, in an 'unpublishable' 1961 paper (Minsky, 1967, p. 281) simulated the first few thousand of a subset of Turing machines, finding that, whilst the majority of machines did something entirely uninteresting (e.g., halt immediately, erase their input or get stuck in a loop), the small set which were interesting were all 'essentially the same' and performed a counting operation which 'increased by one the length of a string of symbols' (Minsky, 1985, p. 120). Formally, this is essentially the same as Merge, as Chomsky (2008) claims and as is evident from the definition of the natural numbers by induction using the successor function, shown in Example 12 (from Goldrei, 1996, p. 38):

(6) Given a set x, the successor of x, written as x^+, is the set
$x^+ = x \cup \{x\}$
So
$0 = \emptyset$
$1 = \emptyset^+ = \emptyset \cup \{\emptyset\} = \{\emptyset\}$
$2 = (\emptyset^+)^+ = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\}$
$3 = (\emptyset^{++})^+ = \{\emptyset, \{\emptyset\}\} \cup \{\{\emptyset, \{\emptyset\}\}\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$
. . .

Replacing $\cup$ with f_{MERGE}, then, allows the natural numbers to be derived by the LTM and, further, shows that language can be arithmetised. This is shown by Watumull (2015, p. 86) for his version of the LTM; however, the proof using his Minimax Merge is not directly comparable to one using Minimax Feature Merge because of the use of a 'dummy' LI to represent 1. Example 13 shows how the natural numbers can be defined as being derived using a single feature coupled with the empty set:

(13) $0 = \emptyset$
$1 = F_1$
$2 = f_{MERGE}(F_1, \emptyset) = \{F_1, \emptyset\}$
$3 = f_{MERGE}(\{F_1, \emptyset\}, \emptyset) = \{\{F_1, \emptyset\}, \emptyset\}$

Indeed, this formulation arguably makes maximal use of *more minimal* means, by only requiring a single feature, whereas Watumull's (2015, p. 86) appears to require multiple LIs (LI_1, LI_2, . . .).

Notably, this example also shows that a free magma can be generated using only a single feature. This implies that any additional features in F, be their origin in UG (evolution) or in acquisition, are there due to external necessity. Such necessity could be in the form of the semantic spine noted in Section 3.

Returning now to the Minsky & Bobrow experiment: the findings show that the 'useful' areas of computational space consist mostly, perhaps entirely, of functions which reduce to the successor function (and thus to f_{MERGE} as in Example 10). From an *evolutionary* perspective, then, the odds that the useful computation chanced upon by random genetic mutation is f_{MERGE} are relatively high. This affords independent support, then, to the SMT, and the devolution of as much as possible to the interfaces, as described in Section 3.

5 Conclusion

Despite the continual advances of generative grammar over the past more than half a century, 'virtually every aspect of (I-)language remains a problem' (Chomsky et al., 2019, p. 253). This dissertation has attempted to reframe the problem in terms of mathematical biolinguistics, tackling architectural concerns alongside the formulation of syntactic primitives and procedures.

Returning to the title, *{*Minimax *{*Feature Merge*}}* / *{{*Minimax Feature*}* Merge*}* captures the two central aspects of the dissertation: the formulation of features and their role in the architecture of FL (minimax features) and the minimax augmentation of the nanosyntactic disSection of the syntactic component (merging features, not lexical items). The focus has been on Merge in particular, leaving aside other operations like Agree and Label, which may 'fall out' from more general, unavoidable considerations regarding the interactions between cognitive modules[33].

There has not been room to discuss empirical consequences of the theory, and as such this dissertation potentially raises more questions than it answers, as is the nature of such a theoretical work. As I have sought to demonstrate, however, it remains well-founded upon the biolinguistic methodology (Di Sciullo & Boeckx, 2011), embracing interdisciplinarity both within and beyond narrowly syntactic theory.

6 References

Adger, D. & Svenonius, P. (2011). Features in minimalist syntax. In C. Boeckx (Ed.), *The Oxford handbook of linguistic minimalism* (pp. 27–51). Oxford University Press.

Baker, M. C. (2008). The macroparameter in a microparametric world. In T. Biberauer (Ed.), *The limits of syntactic variation* (pp. 351–374). John Benjamins.

Balari, S. & Lorenzo, G. (2012). *Computational phenotypes: Towards an evolutionary develop mental biolinguistics*. Oxford University Press.

Baunaz, L., Haegeman, L., De Clercq, K. & Lander, E. (Eds.). (2018). *Exploring nanosyntax*. Oxford University Press.

Baunaz, L. & Lander, E. (2018). Nanosyntax: The basics. In L. Baunaz, L. Haegeman, K. De Clercq & E. Lander (Eds.), *Exploring nanosyntax* (pp. 3–56). Oxford University Press.

Behme, C. (2015). Is the ontology of biolinguistics coherent? *Language Sciences*, *47*, 32–42.

Boeckx, C. (2013). Biolinguistics: Facts, fiction, and forecast. *Biolinguistics*, *7*, 316–328.

Boeckx, C. (2014). *Elementary syntactic structures: Prospects of a feature-free syntax*. Cam bridge University Press.

Braine, M. D. S. (1992). What sort of innate structure is needed to "bootstrap" into syntax? *Cognition*, *45*(1), 77–100.

Brody, M. (2002). On the status of representations and derivations. In S. D. Epstein & T. D. Seely (Eds.), *Derivation and Explanation in the Minimalist Program* (pp. 19–41). Blackwell.

Caha, P. (2020). *Nanosyntax: Some key features*.

Chomsky, N. (1975). *The logical structure of linguistic theory*. Plenum.

Chomsky, N. (1993). A minimalist program for linguistic theory. In K. Hale & S. J. Keyser (Eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 1–52). MIT Press.

[33] per Watumull (2015, f.n. 117): 'I set aside, or rather subsume within the concept of f_{MERGE}, other functions that enter into the computation to define relations. One such function is Agree . . . I need not posit it 'in addition to Merge' [(Chomsky, 2004, p. 555)] if I can assume the possibility that f_{MERGE} somehow subsumes Agree'.

Chomsky, N. (1995a). Bare Phrase Structure. In H. Campos & P. M. Kempchinsky (Eds.), *Evolution and revolution in linguistic theory: Studies in honor of Carlos P. Otero* (pp. 51– 109). Georgetown University Press.

Chomsky, N. (1995b). *The Minimalist Program*. MIT Press.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–156). MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. J. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). MIT Press.

Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *Structures and beyond* (pp. 104–131). Oxford University Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, *36*(1), 1–22. Chomsky, N. (2008). On phases. In R. Freidin, C. P. Otero & M. L. Zubizarrete (Eds.), *Founda tional issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*. MIT Press.

Chomsky, N. (2015). *The Minimalist Program* (20th Anniversary Edition). MIT Press.

Chomsky, N. (2017). The language capacity: Architecture and evolution. *Psychonomic Bulletin & Review*, *24*(1), 200–203.

Chomsky, N. (2019a). *Fundamental issues in linguistics* (Lecture series). MIT. Chomsky, N. (2019b). *The UCLA lectures* (Presentation recording).

Chomsky, N., Gallego, A. J. & Ott, D. (2019). Generative grammar and the faculty of language: ´ Insights, questions, and challenges. *Catalan Journal of Linguistics*, 229–261.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. MIT Press. Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford Univer sity Press.

Cohen, M. (2008). *On reality as a mathematical structure* (PhD Dissertation). Ben Gurion University of the Negev. Israel.

Deutsch, D. (2011). *The beginning of infinity: Explanations that transform the world*. Allen Lane.

Di Sciullo, A.-M. & Boeckx, C. (2011). Introduction: Contours of the biolinguistic research agenda. In C. Boeckx & A.-M. Di Sciullo (Eds.), *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty* (pp. 1–16). Oxford University Press.

Epstein, S. D., Groat, E. M., Kawashima, R. & Hisatsugu, K. (1998). *A derivational approach to syntactic relations*. Oxford University Press.

Frampton, J. & Gutmann, S. (2002). Crash-proof syntax. In S. D. Epstein & T. D. Seely (Eds.). D. W. Lightfoot (Ed.), *Derivation and explanation in the Minimalist Program* (pp. 90– 105). Blackwell.

Gallistel, C. R. (2001). Mental representations, Psychology of. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 9691–9695). Pergamon.

Gallistel, C. R. & King, A. P. (2010). *Memory and the computational brain: Why cognitive science will transform neuroscience*. Wiley-Blackwell.

Goldrei, D. C. (1996). *Classic set theory: For guided independent study*. Chapman & Hall. Haegeman, L. (1994). *Introduction to Government and Binding Theory* (2nd ed.). Blackwell.

Hale, M. & Reiss, C. (2008). *The phonological enterprise*. Oxford University Press.

Halle, M., Vaux, B. & Wolfe, A. (2000). On feature spreading and the representation of place of articulation. *Linguistic Inquiry*, *31*(3), 387–444.

Jakobson, R., Karcevsky, S. & Trubetzkoy, N. (2002). Quelles sont les methodes les mieux ´ appropriees ´ a un expos ` e complet et pratique de la phonologie d'une langue quelconque? ´ *Selected writings: Vol. 1. phonological studies* (3rd ed., pp. 3–6). Mouton De Gruyter. (Original work published 1928)

Kayne, R. S. (1994). *The antisymmetry of syntax*. MIT Press.

Kibort, A. & Corbett, G. G. (Eds.). (2010). *Features: Perspectives on a key notion in linguistics*. Oxford University Press.

Lenneberg, E. H. (1967). *Biological foundations of language*. John Wiley & Sons. Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania working papers in linguistics*, *4*(2), 14.

Martins, P. T. & Boeckx, C. (2016). What we talk about when we talk about biolinguistics. *Linguistics Vanguard*, *2*(1), 1–15.

McCarthy, J. J. (1988). Feature geometry and dependency: A review. *Phonetica*, *45*(2–4), 84–108. Minsky, M. L. (1967). *Computation: Finite and infinite machines*. Prentice-Hall.

Mobbs, I. (2015). *Minimalism and the design of the language faculty* (PhD Dissertation). Univer sity of Cambridge. Cambridge.

Pinker, S. (1984). *Language learnability and language development*. Harvard University Press. 27

Poeppel, D. & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-First century psycholinguistics: Four cornerstones* (pp. 103–118). Taylor & Francis Group.

Postal, P. M. (2009). The incoherence of Chomsky's 'biolinguistic' ontology. *Biolinguistics*, *3*(1), 104–123.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. MIT Press.

Ramchand, G. & Svenonius, P. (2014). Deriving the functional hierarchy. *Language Sciences*, *46*, 152–174.

Roberts, I. & Watumull, J. (2015). Leibnizian linguistics. In A. J. Gallego & D. Ott (Eds.), ´ *50 years later: Reflections on Chomsky's Aspects*. MIT Working Papers in Linguistics.

Roberts, I., Watumull, J. & Chomsky, N. (in press). Universal Grammar. In D. Vakoch (Ed.), *Xenolinguistics: Toward a science of extraterrestrial language*. Oxford University Press. Scheer, T. (2004). *A lateral theory of phonology: Vol. 1. what is CVCV and why should it be?* De Gruyter Mouton.

Scheer, T. (2020). On the lexical character of intermodular communication. *Radical: A Journal of Phonology*, *1*, 183–239.

Sigurdsson, H. A. (2020). Universality and variation in language: The fundamental issues. ´ *Evolutionary Linguistic Theory*, *2*(1), 5–29.

Sipser, M. (2012). *Introduction to the theory of computation* (3rd ed.). Cengage Learning. Tegmark, M. (2014). *Our mathematical universe: My quest for the ultimate nature of reality*. Penguin Books.

Tomalin, M. (2006). *Linguistics and the formal sciences: The origins of generative grammar*. Cambridge University Press.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *2nd series 42*(1), 230–265.

Uriagereka, J. (1999). Multiple spellout. In S. D. Epstein & N. Hornstein (Eds.), *Working minimalism* (pp. 251–282). MIT Press.

Vaux, B. & Watumull, J. (2012). *The phonological Turing machine*.

von Neumann, J. (1928). Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, *100*(1), 295–320.

Watumull, J. (2012). A Turing program for linguistic theory. *Biolinguistics*, *6*(2), 222–245. Watumull, J. (2013). Biolinguistics and Platonism: Contradictory or consilient? *Biolinguistics*, *7*, 301–315.

Watumull, J. (2015). *The linguistic Turing Machine* (PhD Dissertation). University of Cambridge. Cambridge.

Watumull, J. & Chomsky, N. (2020). Rethinking universality. In A. Bar´ any, T. Biberauer, J. ´ Douglas & S. Vikner (Eds.), *Syntactic architecture and its consequences II: Between syntax and morphology* (pp. 3–24). Language Science Press.

Watumull, J., Hauser, M. D., Roberts, I. G. & Hornstein, N. (2014). On recursion. *Frontiers in Psychology*, *4*.

Wiltschko, M. (2014). *The universal structure of categories: Towards a formal typology*. Cam bridge University Press.

Zeijlstra, H. (2014). On the uninterpretabilitty of interpretable features. In P. Kosta, S. L. Franks, T. Radeva-Bork & L. Schurcks (Eds.), *Minimalism and beyond: Radicalizing the interfaces* (pp. 109–129). John Benjamins.

Memory and Neurolinguistic Function in the Deaf

Lydia Wiernik
University of Edinburgh

Abstract. The Deaf community's relationship with language provides a new angle from which to study the role of phonemic elements in the interaction between lexemes and memory. Departing from a hearing canon and exploring a broader spectrum of language perception and production, we are able to revise questions on and develop insight into cognitive processes that influence memory. This paper will not attempt to ascertain whether the Deaf or the hearing have better memories; neurocognitive discrepancies between the groups prevent certain abilities from being compared under the same criteria. But there is no comprehensive "better," only the space afforded by these discrepancies that allows for deeper understanding of neurolinguistic processes in conjunction with memory. To investigate the processes that influence Deaf memory, I compare studies on American Sign Language (ASL) and Japanese Sign Language (JSL). This comparison finds that when research on memory in the Deaf compared to the hearing is limited to English, the Deaf's supposed deficit is the fault of the English language rather than a lack of phonological information. Linguistic models allow for a detailed understanding of information retrieval and the role of memory within language production. They also can narrow down variables of language-specific stimulus and when and where in cognition they come into play. The current standardized models of language production do not account for an inability to perceive audition. Thus, this paper will conclude by proposing a new Deaf language production model.

Keywords: Deaf linguistics; ASL; neurolinguistics; sociolinguistics; language production; memory

---

1 Introduction

The role of memory in psycholinguistics is indisputable. It has been seen as the groundwork on which all fundamental linguistic processes are built; lexical entry, access, and retrieval all combine memory with neurological interaction. In coordination with memory, physical senses of sight and hearing grant orthography and phonology. Research has determined phonology one of the most important factors in the recognition of cognates and the creation of links between lexical and semantic items. But when the phonological component is absent, how does this impact memory, our essential hard drive? The Deaf community's relationship with language provides a new angle at which to study the facilitating role of phonological elements in the interaction between lexemes and memory. It is necessary to look outside the canon of hearing memory processes and instead consider how this restrictive standard affects an understanding of Deaf brains and memory, as well as its impact on the education of Deaf children. As research in Deaf memory and language production continues to emerge, revised questions on cognitive processes and neurological structures follow. Retracing the processes behind Deaf language production allows for a step-by-step understanding of Deaf memory function and its interaction with language use. Thus, to conclude this paper, a Deaf language production model will be proposed, considering both neurological and socialized contributions to memory in the Deaf.

2 Neurolinguistic processes behind memory

2.1 Functionality of memory in the Deaf and the hearing

Almost every linguistic process involves phonology. Increasing research on phonology in language acquisition and production reveals just how crucial sound is. The 1999 Dijkstra et al. article, 'Recognition of cognates and interlingual homographs: The neglected role of phonology,' tested lexical access processes of bilinguals. The article studied how phonological similarities between the stimulus and a non-target word affected or influenced stimulus performance. Dijkstra (1999) noted that phonological factors are often overlooked regarding recognition, despite phonology playing a mediating role when the orthography of words is different. Phonological similarity between interlingual words proved an inhibitory contribution to reaction time for bilinguals; when there was a phonological overlap, the reaction time was slower. Monolinguals, on the other hand, with only one lexicon to sort through, were able to recognise non-words faster and with more accuracy.

Despite the abundance of literature on phonological importance, research on the effects of the absence of sound is an almost untapped domain. Comprehensive linguistic concepts must be reconsidered when phonology becomes inapplicable. Some linguistic facets, such as those tested by Dijkstra (1999), are hindered by an inability to perceive audition. Even the most rudimentary are impacted, down to the first stages of speech perception. For example, how would the accepted Cohort model account for a lack of acoustic stimulus? Activation, the first step of Cohort, is influenced by solely acoustic properties. The next step, selection, sorts through the activated cohort, or group of phonetically similar words. But the Deaf would not perceive phonetically similar words; without phonemes, the inhibitory factors of homophones and false friends are eliminated. Form still plays a role in a Deaf person's language activation and retrieval, but manifests differently than that of a hearing person's. Orthography and signs represent form for the Deaf, the latter replacing a hearing person's phonology. The presence of both orthography and signs, known as Simultaneous Communication or SimCom, is likened to bilingualism because the Deaf also possess two combined, simultaneously activated lexicons. SimCom's dual encoding accounts for slower orthographic word recognition and processing speed in the Deaf. Additionally, both Deaf and hearing people experience a form of memory load, the amount of information the brain can perceive and remember at a single moment. When there's an overload, hesitation phenomena occurs: as 'memory load increases, performance decreases' (Dehn, 2011, p. 61). This can present in hearing people as tongue twisters, the same phenomenon called finger fumblers in the Deaf. Repeats and false starts are universal quirks of language, extending to slips of the hand for the Deaf, especially with formationally or locationally similar signs. However, the difference in how the brain processes information comes into play as the memory load increases. As a Deaf person's memory becomes more saturated, their language performance will begin to deteriorate sooner than a hearing person's would (Dehn, 2011, p. 65). Seemingly, it is the process of sequential recall that is formulated and conceptualized differently in hearing and Deaf brains.

An individual's memory capacity is directly linked to language comprehension and acquisition (Hamilton, 2011). Because critical aspects of language acquisition, such as attention and processing speed, can suffer from a lack of auditory perception, it is in these processes that the hearing seemingly perform better than the Deaf. When a Deaf child is unable to develop sequential processing skills early enough and apply them to the brain's processing of linguistic information, their understanding of syntax may suffer (Hamilton, 2011). The ability to confidently grasp syntactic order is crucial in language development. If a

Deaf child does not completely master this, they may suffer impediments in learning and comprehending either signed or printed linguistic information. Language delay is common in Deaf children as a result, with only '50% of Deaf high school students reading at the fourth-grade level or below upon graduation' (Hamilton, 2011, p. 408). This suggests early intervention for strengthening syntactic order can be critical. However, this deficit cannot solely be attributed to a Deaf person's brain processes. Lack of 'accessible linguistic interaction with other signers and less than sufficient working memory skills to assist Deaf children during language learning period' are external factors that impact the memory capabilities of the Deaf (Hamilton, 2011, p. 408). Factors of linguistic inequality and a desire for Deaf children to speak or lip read instead of sign also comes into play. Despite being tested on the same technical skills, neurocognitive discrepancies between the Deaf and hearing prevents those skills from being compared using the same criteria. Perhaps this supposed deficit appears not because the Deaf are deficient to the hearing, but because the hearing standard forced upon them attempts to stimulate a neural pathway that simply does not exist.

2.2 Linguistic strengths of Deaf memory

In other areas, however, a lack of auditory perception boosts memory. In both nonlinguistic and linguistic terms, the Deaf have performed better than the hearing in tasks where information is presented in 'static visuospatial format' (Hamilton, 2011, p. 407). Visuospatial memory is involved in 'recalling and manipulating images to remain oriented in space and keep track of the location of moving objects' (Dehn, 2011, p. 80). Examples of impairment in visuospatial processing include difficulty driving due to an incorrect judgement of distance or other objects in the environment. The boost in this type of memory of Deaf individuals could also be seen as a form of linguistic determinism; biologically, the Deaf brain has adapted for improved visuospatial memory to make up for a lack of acoustic memory and perception. A hearing person's auditory cortex helps them process and comprehend sound, but in a Deaf person's brain, it becomes part of working memory (Cardin, 2018). Instead of being allocated to auditory processing, the cortex is now involved in cognitive processing, which accounts for a Deaf person's heightened memory and comprehension of locative and visual movement.

Compared to the hearing, Deaf brains perform better on visuospatial linguistic tasks such as arrays of 'blocks on a table or objects in a grid' (Hamilton, 2011, p. 406). This was determined by the Corsi block-tapping test and the Knox cube test. In the former, the experimenter 'touched blocks randomly arranged on a board' in a sequence, which would then be replicated by the subject (Hamilton, 2011, p. 406). In the latter, the same sequential task was created, but instead of a random arrangement, the blocks were in a straight line. In both experiments, Deaf participants proved superior to hearing participants in recalling these visual patterns and successive figures. Just as some cognitive functions in the hearing allow them to perform better on certain tasks, other functions in Deaf brains grant capabilities the hearing lack. Because of the physicality of sign language and the sequential memory involved in structuring and understanding the syntax and grammar of rapid hand movements, it is more than reasonable that the Deaf would perform better on these tasks.

3 The role of the English language

The previous comparisons of linguistic and nonlinguistic memory have been between hearing English speakers and American Sign Language users. But as with the brain's adaptation to a lack of audition,

memory function can be somewhat attributed to linguistic determinism and dependent upon the language itself. To investigate the influence of language-specific stimulus, I will compare memory function in American Sign Language (ASL) and Japanese Sign Language (JSL). JSL users were found to have better short and long term memories than ASL users, but so were Japanese speakers compared to English speakers (Flaherty, 2000, p. 243). A language's alphabet often has an innate connection between orthography and phonology, but since Deaf people have never encoded auditory features of language, so they're unable to recognise multiple phonemes. One consequence of this is that there is no positive correlation between phonemic awareness and reading comprehension. Many languages rely on this phonological encoding to build stronger cognitive links, but in languages like Japanese that is not the case (Hamilton, 2011). With three alphabets, kanji, hiragana, and katakana, the logogens build upon each other, rather than rely solely on phonological association. Kanji characters are meaning-based rather than sound-based like the English alphabet. In essence, this is sign language. Kanji's 'finite set of radicals and forms are spatially and visually related to one another in writing,' eliminating the necessity of sound to glean meaning. When reading, the Japanese Deaf use visual properties of letters, so they don't suffer from a lack of auditory perception when there's already no orthographic-phonetic correspondence. In Japanese, there are 2,000 kanji, almost 80 times English's 26 letters. Readers of Japanese, hearing or Deaf, simply must encode, and thus remember, more visual shapes and sequences than English readers. This innately circumvents poor sequential memory, so when compared to English speakers, Japanese speakers will prove more successful at memory tasks involving short-term sequential skills.

Perhaps when research on memory in the Deaf compared to the hearing is limited to ASL, the Deaf's deficit is the fault of the English language rather than a lack of phonological information. Is memory then a result of linguistic determinism? In the same way Russians take longer to categorize blue because of the way their language lexically orders colours, do the Japanese simply engage with the world in a different, more retentive way because of unconscious factors of their language? Flaherty and Aidan's 2004 study of sequential memory in Deaf and hearing students using English or Japanese addressed this, highlighting a key difference in how the students went about a short-term memory task (Flaherty & Aidan, 2004). Their research advanced the theory that the Japanese are influenced by technical aspects of their language, and so their mindsets are structured to remember sequences more adeptly. Two experiments were conducted. The participants were tested first on a series of words, then on abstract lines and shapes. In both experiments, the Japanese participants outperformed the English-signing or speaking. Flaherty (2000) notes that the study confirms 'the superior visuo-spatial memory of the Japanese' (Flaherty, 2000, p. 241). Students using JSL said they used 'a visual gestalt memory strategy, seeing the sequence as a whole,' whereas students using ASL said they saw the words as just a sequence (Hamilton, 2011, p. 406). This raises another question. Do the Deaf really have an elevated capacity for visual memory when verbal memory is eliminated? Or is 'a deficit in recall of English linguistic material more a function of the idiosyncrasies of written English rather than a deficit in recall of written material' (Flaherty & Aidan, 2004, p. 40)?

4 Constructing a new Deaf language production model

Research connecting Deaf memory deficits or surpluses to language production began to emerge in the 1970s. While many models of language production have been proposed, there is no one standardised form, much like the hearing bilingual model which continues to evolve as more research is realised. Deaf models by Fromkin (1971), Garrett (1975), Butterworth (1979), and the bilingual model by de Bot (2004) will serve

as groundwork for this paper's culminating model. Bilingual language production is likely the most similar to Deaf language production, given the likeness of SimCom to bilingualism. Grosjean (2008) collates the stages of these models to outline a common one. Deaf language production begins with non-linguistic message formulation such as an idea or a thought. It's then padded by pragmatic factors and grammatical structure assignment, and prosodic features are chosen. The next stage is termed 'lexical look-up,' which Grosjean defines as when 'the signer enters her lexicon and chooses the appropriate lexical items' (Grosjean, 1979, p. 324). Many factors involved in lexical activation are not wholly conscious, so an updated definition of lexical look-up would benefit Grosjean's explanation. These items are then assembled and a command is sent to the motor control of the brain. If motor commands were replaced with overt speech, this model would look similar to Levelt's. Levelt's model notes that form comes before phonology in the formulating stage, meaning the application of audition and overt speech are the last steps of producing language. Among minor phonetic changes in the formulator, the articulating stage is the only phase that will be significantly altered when accounting for differing neurological function.

The proposed model combines the general structure of Levelt's with the steps outlined by Grosjean (Grosjean, 1979). Elements of de Bot's bilingual updates to Levelt's model are also incorporated, including a combined lexicon. Figure 1 is a sketch of the model. Instead of the articulation and audition stages of Levelt's, the proposed model has a motor control centre. Here, the content of the lexical plan is sent to the motor cortex, a Section of the brain's frontal lobe in charge of the body's movements, such as facial expressions and hand signs ('The Motor Cortex').

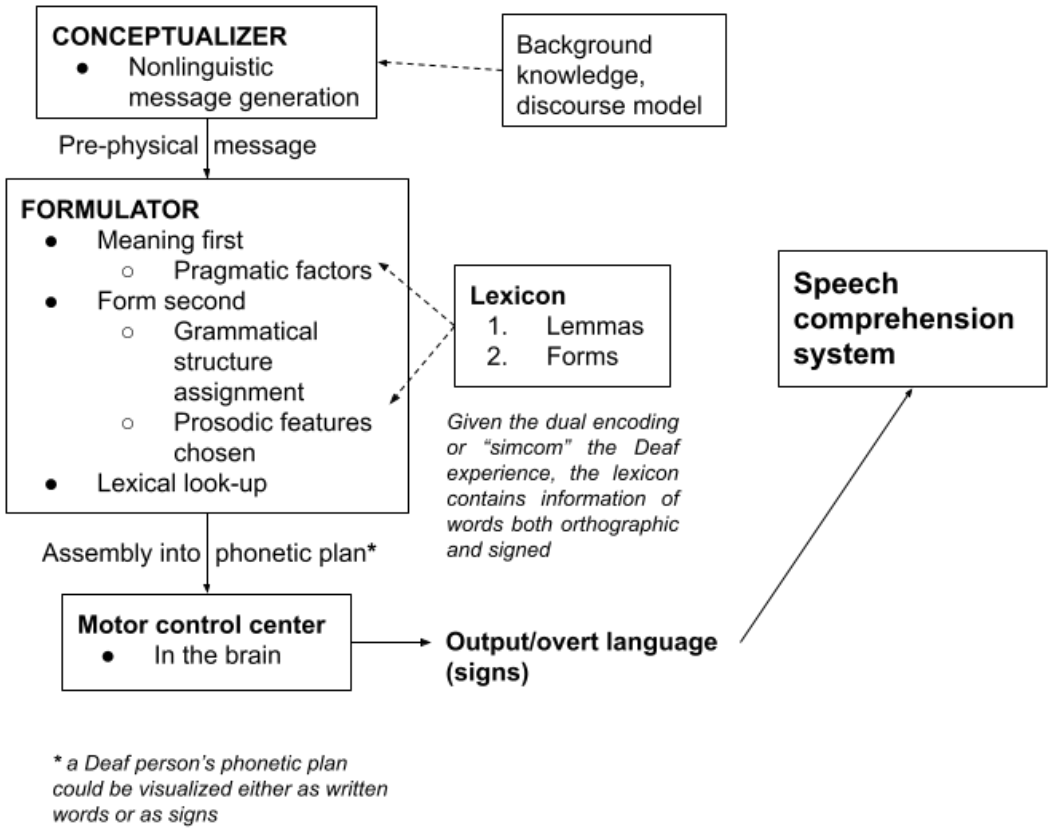


Figure 1: *Redeveloped Deaf language production model.*

Though de Bot's amendments to Levelt's model provide the groundwork for other models with dual encoding, not all of his suggestions would make sense in this application. Instead of having two separate languages with separate grammars and signifiers, as a bilingual would, a Deaf person's dual encoding is derived from the same base language, and thus it's unlikely there would be unequal command. De Bot adopted Green's (1986) idea of separate formulators for each language, but this is omitted in the Deaf model. De Bot's formulator generates meaning important to phonological processing and is largely language-specific, and in the Deaf model, phonology has been eliminated. A hearing person's perception of intonational meaning and metrical structure are different than a Deaf person's. Regarding the studies on ASL and JSL, is it possible to create a comprehensive Deaf model when there are such discrepancies in how language is processed and sorted language to language? In that respect, how vague must any model dealing with dual encoding be to account for linguistic determinism? The model proposed here is not exhaustive. It instead provides a simplified overview of Deaf language production based on reinterpretations of current models and the modern research aggregated for this paper.

5 Conclusion

The importance of studying Deaf memory and language production is an almost untapped resource in the field of psycholinguistics. The Deaf interact with language in a unique way. Their memories and perception of the world are influenced by a lack of audition, which no longer calls for an answer of better or worse but rather how and why. The hearing person's canon of language can be informed by the Deaf's; studying alternative memory function brings to light alternative learning strategies and competencies for all people. Many of the supposed deficiencies in Deaf memories can be remedied by engaging Deaf instructors to teach Deaf children, promoting communication through a Deaf lens rather than trying to conform to a hearing one. Creating a Deaf language production model helps to understand not only the many components that must align to facilitate language production but also the intricacies of the brain. But linguistic testing can only take us so far. It must be supplemented with brain scans and neurological tests for a view of the brain itself, not only how linguistic functions present verbally or nonverbally. Questions posed within this paper invite specialization on Deaf memory in fields such as neurolinguistics, linguistic anthropology, or sociolinguistics. Each study is a step closer to unravelling intricacies that will broaden perspectives on memory and neurolinguistic function. The realm of research on Deaf memory is ripe and ever-expanding, and its fruits are profoundly important to the Deaf community and linguistics as a whole.

6 References

Butterworth, B. (1979). Hesitation and the Production of Verbal Paraphasias and Neologisms in Jargon Aphasia. *Brain and Language*, 8(2), 133–161.

Cardin, V. (18 March 2018). How does the brain change in people who are deaf?. *Action on Hearing Loss*, Royal National Institute for Deaf People, www.actiononhearingloss.org.uk/live-well/our-community/our-blog/how-does-the-brain-change-in-people-who-are-Deaf/.

Dehn, M. J. (2011). *Working memory and academic learning: assessment and intervention*. John Wiley & Sons, Inc.

De Bot, K. (2004). The Multilingual Lexicon: Modelling Selection and Control. *International Journal of Multilingualism*, 1(1), 17–32.

Dijkstra, T., et al. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *Journal of Memory and Language*, 41(4), 496–518.

Flaherty, M. (2000). Memory in the deaf: a cross-cultural study in English and Japanese. *American Annals of the Deaf*, 145(3), 237–44.

Flaherty, M. & Aidan M. (2004). Deaf signers who know Japanese remember words and numbers more effectively than deaf signers who know English. *American Annals of the Deaf*, 149(1), 39–45.

Fromkin, V.A. (1971). The Non-Anomalous Nature of Anomalous Utterances. *Language*, 47(1), 27.

Garrett, M.F. (1975). The Analysis of Sentence Production. *Psychology of Learning and Motivation*, 133–177.

Grosjean, F. (1979). The production of sign language: psycholinguistic perspectives. *Sign Language Studies*, 25, 317–29.

Grosjean, F. (2008). *Studying Bilinguals*. Oxford University Press.

Hamilton, H. (2011). Memory Skills of Deaf Learners. *American Annals of the Deaf*, 156 (4), 402–23..

McGill University (n.d.) "The Motor Cortex." *The Brain From Top to Bottom*, Retrieved 01/02/2020, from https://thebrain.mcgill.ca/flash/a/a_06/a_06_cr/a_06_cr_mou/a_06_cr_mou.html

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Massachusetts Institute of Technology.

Linguistic Relativity and Grammatical Number: A Comparison between Native Slovenian and English Speakers

Ajda Krišelj
University of Edinburgh

Abstract. The theory of linguistic relativism suggests that language influences the way we think (Reinez & Prinz, 2009). The present study aims to extend the evidence into another domain, grammatical number. The methodology and idea stem from research by Phillips & Boroditsky (2003). My study investigated the differences in cognition between native speakers of a 2-way number system language (English) and native speakers of a 3-way number system language (Slovenian). The primary hypothesis was that Slovenians group pictures of two versus three by number, whereas the English group them by type. Dual and nondual dialects of Slovenian are also compared to exclude the possible cultural differences between the English and the Slovenian participants. The experiment was carried out online using Qualtrics survey software. A grouping task was used to explore whether there is a difference between Slovenian and English native speakers' mental categories. The design for ensuring balanced conditions was Latin square, and factorial ANOVA was used for the data analysis. There was no significant difference in grouping choice between the English and Slovenian group. There was a slight effect of dialect, which needs to be explored further. A big limitation was a significant effect of the device used for solving the survey on the groupings. There was no evidence found in favour of linguistic relativism in the domain of the grammatical number. The findings offer some compelling grounds for further research.

Keywords: language relativity; Slovenian dual; language cognition interaction; Slovenian dialects

---

1 Introduction

This dissertation investigates whether having a grammatical category of the dual number in your native language affects the way you think, or more specifically, the way you choose to group objects. The theory behind this study is called linguistic relativism, which states that the language we speak influences the way we think. It has been a source of controversy ever since it had been proposed by Whorf in 1956. The strong Whorfian view, or linguistic determinism, which says that thought and action are entirely determined by language, has been abandoned in the field. A more moderate question of whether language shapes or influences thought is still being actively researched. It has proven to be difficult to definitively answer, as there is evidence for both sides of the debate. Much recent evidence in favour of linguistic relativism comes from the work of Lera Boroditsky. Phillips and Boroditsky (2003) were testing whether the grammatical gender of inanimate objects leads people to think of them as having a gender. They found the effects of grammatical gender on people's perceptions of objects in a variety of settings. Her other experiments have also found effects of linguistic relativity in representations of space, time, and substances. This study investigates whether grammatical number also shows effects of linguistic relativity by exploring the differences in cognition between native speakers of a language with a 2-way number system (English) and native speakers of a language with a 3-way number system (Slovenian).

2 Linguistic Relativism

Languages differ immensely in how they describe the world. However, does having a different way of describing the world also means that speakers of different languages think about the world in different ways? This question is addressed by the theory of linguistic relativity or Sapir-Whorf hypothesis (Casasanto, 2008). Reinez and Prinz (2009) summed up the evidence of linguistic relativity according to four domains: grammatical gender, frames of reference, spatial categories, and noun types. The present study aims to extend the evidence into another domain; the grammatical number.

Hunt and Agnoli (1991) state that linguistic relativism occurs because different languages pose different challenges and provide differential support to cognition. They discuss the arguments that oppose and support the Sapir-Whorf theory from a cognitive psychology point of view. According to them, one of the major counterarguments of the theory is the intertranslatability issue. This issue asks whether a statement in one language can be translated into a statement in another language. That usually is the case, which disproves the strong version of the theory, linguistic determinism. However, the statements in different languages might be of different length and therefore easier or harder to create. The issue is then whether the naturalness of the translation between the languages is the same. Naturalness can be quantified in terms of the computational burden, which would make longer statements more expensive than single-word ones. Hunt and Agnoli (1991) compare this complexity argument to Whorf's idea, which says that the distinctions made by the grammar of a language are far more important determinants of thought than the explicit categorisations of the lexicon (Whorf, 1956). This connection between grammar and language is one notion that the present experiment is going to test.

Linguistic determinism goes against the view held by Noam Chomsky (2006), among others, which states that human cognition is largely universal and independent of language and culture. He says that all languages share the same deep structure of thought and that thought has a universal quality separate from language. The cognitive linguistics as a discipline rejects the notion of linguistic relativity. They view language as just another cognitive function operating under the same rules as other cognitive processes. Boroditsky's critics say that her findings, for example, that differences in English and Mandarin speaker's conceptions of time affect the temporal cognition (Boroditsky, 2001), are highly dependent on the context in which the experiments take place. They say that her findings neglect the reality that language is much less complex than our thoughts and as such, cannot determine them (January & Kako, 2007).

There is a lot of conflicting arguments and views in the field of linguistic relativism. In his influential book, 'The Language Instinct', Steven Pinker (1994) describes linguistic relativism as a 'conventional absurdity'. He claims that language cannot equal thought as it 'is hopelessly unsuited to serve as our internal medium of computation' (Pinker, 1994, p. 76). He lists five properties of the English language, which show problems for the notion that language equals thought. 'Ambiguity', in his opinion, shows that if there can be two thoughts corresponding to one word, thoughts cannot be equal to words. 'Lack of logical explicitness' shows that we need something else other than language to think, namely common sense, which is not available in the sentences alone. 'Co-reference' shows that we know who we are talking about, even if we refer to them with different expressions. 'Deixis' shows that some information can only be extracted from the context of the utterance, not words themselves. 'Synonymy' shows that we understand many different arrangements of words to have the same message, and we cannot get that simply by processing the words. These five properties of language are why he thinks that people think in the language of thought, 'mentalese', rather than in any particular natural language. Knowing a language then means knowing how to translate mentalese into strings of words and vice versa. However, language relativism is not claiming

that language equals thought. Casasanto (2008) says that a lot of the controversy in the field is due to conflating two distinct questions — Do we think in language? and Does language shape thought?. What Pinker is arguing is not the same as what Whorf initially proposed, which is that differences among languages cause differences in the thoughts of their speakers. Language may shape the way people think even if they do not think in a language (Casasanto, 2008).

The type of linguistic relativity that is considered in this study is structural relativism. This concerns how speaking one or more particular languages may influence thinking as a result of differences in the morphosyntactic configurations (Lucy, 2000). One of the differences between the world's languages is how they mark the number on different word classes. The languages chosen for comparison in this study are Slovenian and English, as they differ in their grammatical number systems. English distinguishes between being one, singular, and being two or more, plural. Slovenian adds a category for being exactly two, dual. The dual is a form of the plural and is the most marked category (Jakop, 2012).

3 The Slovenian Dual

Slovenian (also called Slovene) is a South Slavic language spoken primarily in the Republic of Slovenia. It is the native language of about 2 million people living in Slovenia and its global diaspora. It is one of the few Indo-European languages that still have the grammatical number category of dual, next to Upper and Lower Sorbian, which are minority West Slavic languages in eastern Germany. Some sources also mention sub-national language variety Kashubian as having dual (Jakop, 2010), but others claim that it only had dual in its earlier historical development (Slobodchikoff, 2013). Dual is otherwise present in around 200 non-Indo-European languages, the biggest one being Arabic (Jakop, 2010).

Dual is present in all paradigms in Contemporary Standard Slovenian (CSS), which is the language used in the literature and official writings. In addition to CSS, the Slovenian language has 36 dialects and 12 subdialects, which are separated into seven dialect groups. There is also a colloquial standard which is used in less formal speech settings such as television, radio, and theatre. It has formed between CSS and local dialects and therefore differs across different regions of the country. Dual is one of the more interesting areas of difference between the distinct codes of Slovenian, as it is used to various extents across the different varieties (Jakop, 2012). Some Slovenian dialects are witnessing a gradual disappearance of the dual and its replacement by plural forms. Dual is the most stable in nominative and accusative masculine case. In feminine nouns, dual forms have been mixed with plural forms since the 16th century (Jakop, 2008). In the colloquial language of central Slovenia, dual is still used for masculine nouns. However, the dual forms for feminine nouns have been replaced by the plural: ‘dva fanta’ [two boys] (masc. dual), ‘dve punce’ [two girls] (fem. dual/pl.). The CSS equivalent for the later has two different forms: ‘dve punci’ [two girls] (fem. dual) and ‘dve punce’ [two girls] (fem.pl.). The situation in the central Slovenian dialects is the same, while the dialects in the north-west and north-east still preserve the feminine dual. Masculine and feminine verbal forms exist in most Slovenian dialects, and in some of them, the distinction between genders occurs only in dual, which is not the case in CSS (Jakop, 2012).

Slovenian is an ideal language for investigating differences in mental representations between native speakers of a language with a two-way grammatical number system and native speakers of a language with a three-way grammatical number system. Not only because it uses the grammatical number category of dual, but also because the usage of dual is not equally distributed in the different dialects. This distribution provides a unique opportunity to eliminate possible other influences on different mental representations between Slovenian and English native speakers, such as cultural differences. Speakers of different dialects

of Slovenian have (roughly) the same culture, but they differ in the extent to which they use the dual grammatical number (Marušič et al., 2016). The present study will compare whether there is a difference between Styrian and Littoral dialects, which use dual more, and Upper Carniolan and Ljubljana dialects, which use dual less. This grouping of dialects was selected based on the description of the differential use of the dual by Jakop (2008).

4 The Present Study

An important point when researching linguistic relativity is not to mistake language relativity with language diversity. Opponents of the theory do not deny the fact that languages differ immensely between each other, not only in their lexicon but also in the way they form their descriptions of the world and mark the distinctions on the different word classes. When testing for language relativism, it is important not to make a circular question and test a function that is dependent on language, but to test other non-linguistic cognitive implications. In order to have evidence for linguistic relativity, there must be an assessment of cognition that is independent of immediate verbal production (Lucy, 2000). In consideration of this argument, a grouping task was chosen as a test of linguistic relativity in the present study.

A suggestion of the theory of linguistic relativity is that language structure cues attention. For example, if a language has a grammatical distinction between words for groups of two versus groups of three, its speakers will pay more attention to those cues, and they will form a category in their mind. The notion that language cues attention is an assumption of the present study.

The main research question is: Does having the dual grammatical number category in a language influence the way the native speakers of that language choose to group pictures? The research question will be tested with the following hypotheses:

(H0) Slovenian and English native speakers will group the same pictures, or the grouping difference will not be statistically significant.

(H1) Slovenian speakers will group pictures with two items by the number of items significantly more often than English native speakers.

(a) Same Picture Same Number (SPSN) item will be grouped with Distinct Picture (DISP) in the Plural Dual (PD) and Dual Plural (DP) conditions significantly more in the Slovenian native speaker group.

b) English and Slovenian participants will be as likely to group SPSN with DISP and Distinct Number (DISN) in the Singular Dual (SD), Singular Plural (SP), Plural Singular (PS), and Dual Singular (DS) conditions.

c) Nondual dialect Slovenian native speakers will respond similarly to English native speaker group, therefore grouping by the entity in DP and PD conditions.

d) Because the use of dual in different dialects contracts in the feminine cases most strongly, a comparison between dual and nondual dialects of Slovenian with only items of feminine grammatical gender will be made. Participants from dual dialects will group DISN and SPSN significantly more often than participants from nondual dialects.

Some other exploratory research questions will be tested:

(A) Does the item type of the picture – human, animal, object – influence the grouping?
(B) Does the device the survey is completed on influence the grouping?
(C) Does the sex of the participants influence the grouping?

To summarise, this study will explore whether native speakers of languages with different systems for marking number – Slovenian and English – show any effects of linguistic relativism, therefore if their native language's way of marking number affects their mental representations. This will be tested through a grouping task, where differences in grouping choices of pictures with two entities versus pictures with three entities will be observed. Different dialects of Slovenian which contrast in their usage of dual will be compared as well, especially the difference between masculine and feminine items. Grouping of different types of items (human, animal, object) will be investigated exploratorily. The influence of participants' sex and the device used on the grouping will be tested as well.

5 Methods

5.1 Participants

Participants were recruited online with convenience sampling, using a Qualtrics (Qualtrics, Provo, UT) anonymised link and posts on Facebook. 160 Slovenian native speakers and 50 English native speakers completed more than 90% of the survey, which was the cut-off for inclusion in further analysis. All participants in respective groups regarded Slovenian or English as their only native language. The average age of all participants was 23.4 years (SD = 6.86). Slovenian participants ranged in age from 18 to 57 years (M = 22.9, SD = 6.18). English participants ranged in age from 19 to 67 years (M = 23.0, SD = 8.60). Data from participants above the age of 30 were excluded before the analysis to make the sample more balanced and comparable. The new number of Slovenian native speakers was 151 (mean age = 21.49, SD = 2.10), and the new number of English native speakers 46 (mean age = 22.95, SD = 2.24). Out of the participants that answered the question about their sex, there were 35 males and 110 females in the Slovenian group, and 23 males and 21 females in the English group (see Figure 1).

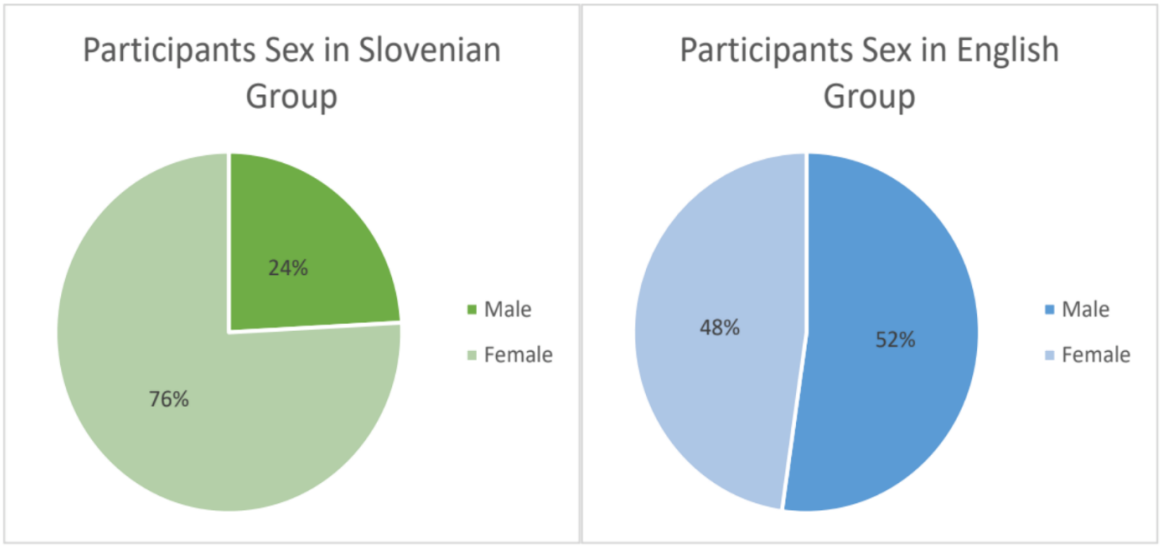


Figure 1: *Participant sex distribution in Slovenian and English groups.*

The Slovenian native speakers had, on average, 12.85 years of experience with English (*SD* = 3.02). The Slovenian group spoke on average 2.58 foreign languages (range 1 to 7, *SD* = 1.04), the most common being English (*n* = 143, 9 did not answer), German (*n* = 83), and Serbo-Croatian (*n* = 54). The English group spoke on average .89 foreign languages (range 0 to 5, *SD* = 1.30), the most common being French (*n* =15). The groups differed in education levels achieved by the participants (see Figure 2).

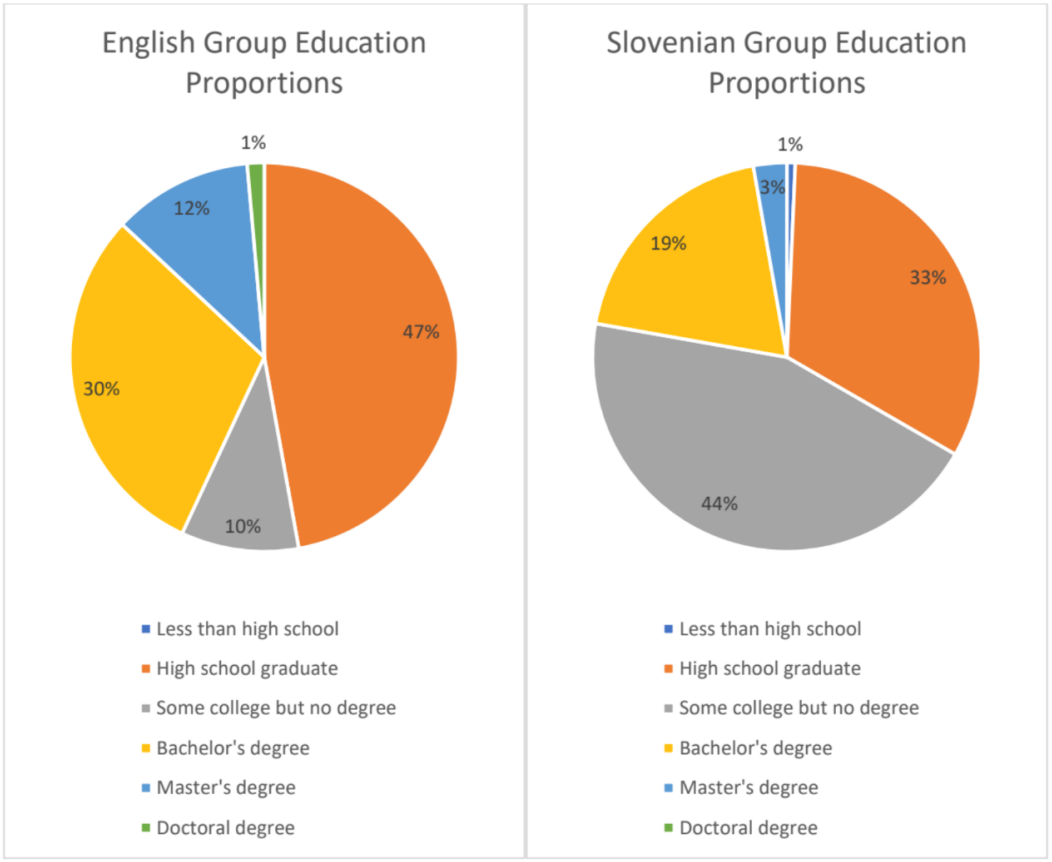


Figure 2: *Slovenian and English group participants' highest achieved education in proportions.*

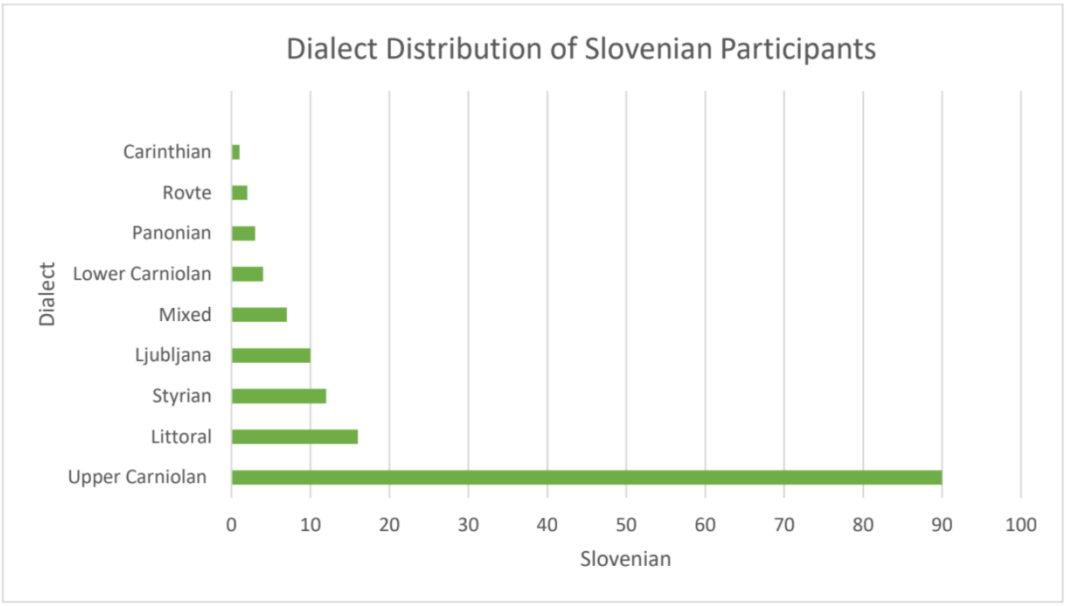


Figure 3: *The dialect distribution of the Slovenian participants.*

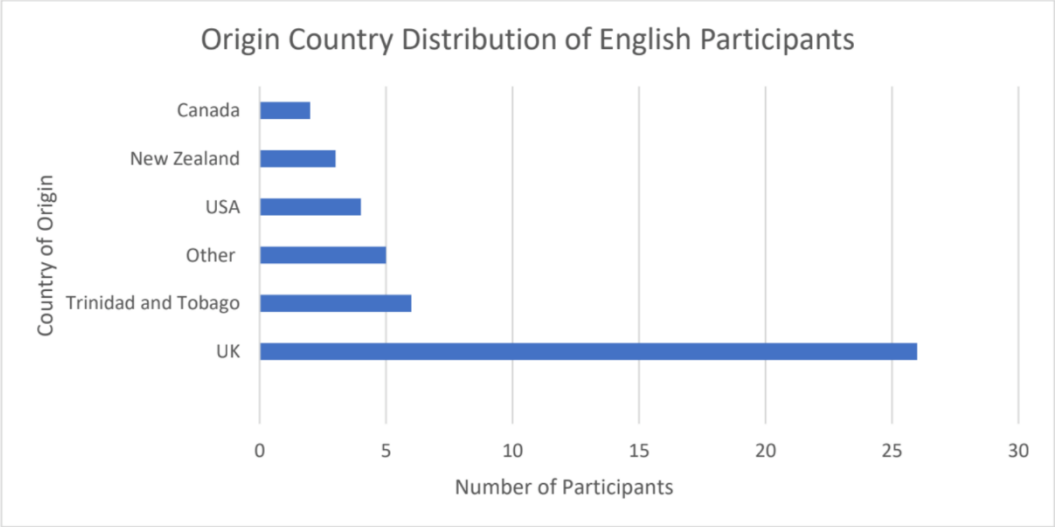


Figure 4: *Countries of origin for the English group.*

5.2 Materials

Materials comprised of 36 pictures. There were six categories of pictures, with six pictures in each category: 1 – 'human male', 2 – 'human female', 3 – 'animal male', 4 – 'animal female', 5 – 'object masculine' (the word for that object has masculine grammatical gender in Slovenian), 6 – 'object feminine' (the word for that object has feminine grammatical gender in Slovenian). The pictures were chosen by the researcher from Google clipart. For the full list of pictures, see Appendix Two.

The study had six conditions (SD, SP, DS, DP, PS, and PD). Every condition had a different combination of three pictures. For example, a question in SD (Singular Dual) condition had two pictures with a single item (singular), and one picture with two items (dual). One single picture was of the same type as the dual picture (e.g., the type of both was 'boy'). Different pictures of *boys* were used to avoid grouping due to the pictures being identical. The other singular item was a distinct picture (e.g., 'girl'). Therefore, the SD version of the first item (1a = 'boy') pictured 'one girl' (Distinct Picture – DISP), 'one boy' (Same Picture Same Number – SPSN), and 'two boys' (Distinct Number – DISN). For other conditions of this item, see Figure 5. A question in SP condition in the 'animal female' category had for example pictures of three cows, one cow, and one bull.

In three questions of each category (e.g., 'human male'), the DISP was of the opposite grammatical gender than DISN and SPSN (like in the example in Figure 5 – 'girl' is of the opposite gender to the 'boy'). In the other three questions of each category, the DISP was of the same gender (e.g., the PD condition in category 'animal male' had three bulls, two bulls, and three rams). This design served as a counterbalancing precaution to prevent grouping by gender of pictures.

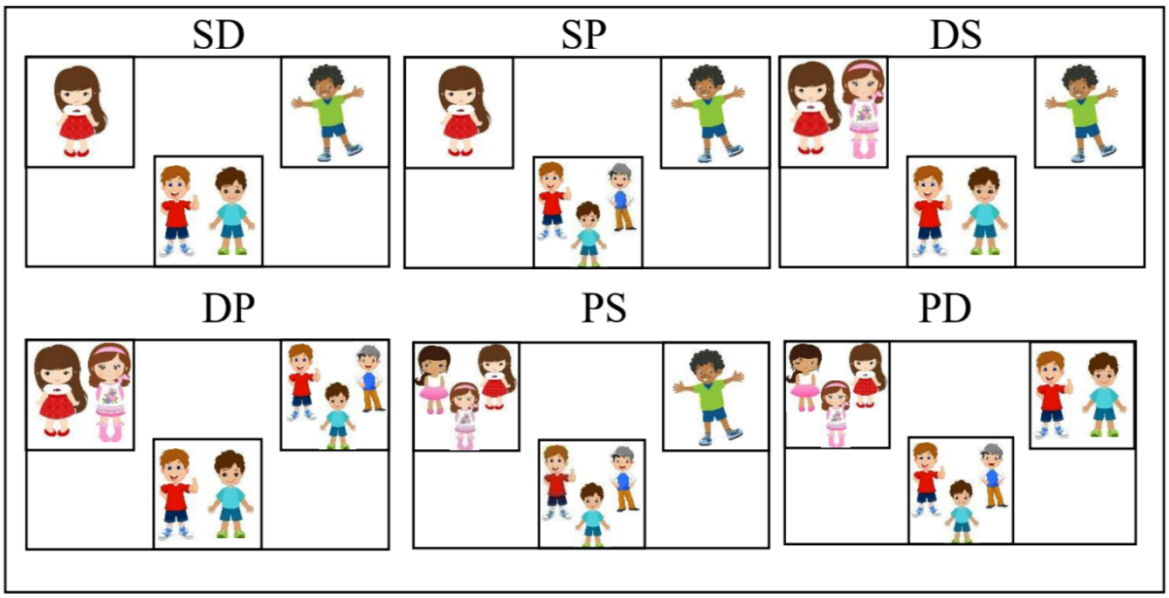


Figure 5: *Example of all conditions for the picture 1a – boy. There were 36 complete sets like this one in the experiment.*

5.3 Design

The design of the experiment was the Latin square, which means each participant saw only one version of each picture, and an equal number of pictures in each condition. For example, participants in group A saw the item *boy* in SD version, *man* in SP version, *ballet dancer* in DS version, *king* in DP version, *teacher* in PS version, *clown* in PD version. The order of answers in each question and the order of questions was randomised, to avoid any ordering effects. Participants were randomly presented with questions from one of six groups (A, B, C, D, E, F), with each group being presented to an equal number of people. Because not all participants finished all the questions, and some participant data was discarded in order to make the age groups more balanced, the end numbers of people in each group were not equal: A = 38, B = 34, C = 31, D = 34, E = 29, F = 31 (M = 32.83, SD = 2.91). To see the full list of questions for each group, see Appendix Three.

5.4 Procedure

The participants were first screened for their native language, and those who answered that their native language was only English or only Slovenian were able to proceed to the experiment. Bilinguals and people with other native languages were excluded. Participants were provided with the necessary information about their rights and asked for their consent. If they chose to proceed to the experiment, they were then provided with these instructions: 'Click on the two pictures you think are the most similar.' (Slovenian version: 'Izberi dve sliki, ki sta si najbolj podobni.'). First, they completed three practice questions to get used to the method. The practice questions included pictures that were not used in the main experiment. After the practice part, participants were notified that they are now going to start the main experiment. They had to respond to 36 questions (six for each condition, each one from a different category, in random order)

with the same instructions as above. When they finished, they were asked to provide some demographic data: age, sex, their level of education, rating of their proficiency in their native language, whether they speak another language and which language that is. Participants in the Slovenian group were also asked to specify which dialect they consider to be their primary dialect. They had to choose all that apply out of seven main dialect groups. The instructions for the demographic part of the survey told them that they could leave blank any question that they do not feel comfortable answering. They were asked the experimental questions first and demographic questions later so they would not be primed to think the experiment is about language, and so that more participants would answer the experimental questions. After they finished, participants were asked if they have any comments about the experiment. They were then debriefed about what the study was researching, the research hypotheses, and how to get in touch with the researchers if they have any questions or want to find out the results. In the end, they were asked for comments again and thanked for their participation. To see the survey flow from Qualtrics, see Appendix Five.

5.5 Plan of Data Analysis

The plan is to code the participant responses as '1' if they grouped pictures by number (e.g., 'two boys' and 'two girls' together), and as '0' if they grouped the pictures by type (e.g., 'two boys' and 'one boy' together). Responses that grouped by neither number or entity (e.g., 'three boys' with 'two girls') will be coded as '2' for the descriptive statistics and discarded before statistical analyses. There will be nine mixed factorial Analyses of Variance (ANOVA) tests conducted. Levene test will be used to check for the homogeneity of variance assumption (Navarro, 2015).

The Slovenian and English group responses will be compared using a two (grouping proportions) by two (native language) factorial mixed analysis of variance (ANOVA). The design will be mixed as within-subject variables (grouping proportions in different conditions), and between-subject variables (native language) will be compared. Three separate ANOVAs like the one above will be conducted: the first one with DP/PD as conditions, second with SD/DS as conditions, and the third one as PS/SP as conditions. There will be a general ANOVA testing the interaction between the means of the critical versus the means of control conditions and Slovenian versus English group.

Another 2x2 ANOVA will be conducted to test whether the dialect of the Slovenian participants influenced grouping in the critical conditions (dual dialect/nondual dialect vs DP/PD). A separate test will be conducted on only feminine items to test whether the grammatical gender has a significant effect on the grouping choice.

A 3x2 ANOVA will be conducted to test whether the item type of pictures influences grouping. The two factors will be the item type (human/animal/object) versus the native language (Slovenian/English).

Additional two 2x2 ANOVAs will investigate whether there was an effect of the device used for solving the survey (phone vs computer) on the groupings; and whether there was an effect of participant's sex on the groupings.

All data analysis will be done in R (Version 3.4.4., R Core Team, 2018). The coding and some participant statistics will be done in Excel.

5.6 Ethics

The study has been approved by the School of PPLS Ethics Committee, reference number 143–1920/2. For participant information sheet, consent form, and participant debriefing sheet, see Appendix Four.

6 Results

Data were collected anonymously from 210 people who finished the online experiment through Qualtrics. There were 151 Slovenian native speakers and 46 English native speakers who fit the criteria for the analysis.

Data were coded according to the grouping of pictures. Participants grouped pictures either by the entity (e.g., 'one cow' with 'two cows'), by number (e.g., 'one cow' and 'one sheep'), or by neither of those (e.g., 'one cow' with 'two bulls'). Grouping by the entity was coded as '0', grouping by number as '1' and grouping by neither was coded as '2'. Because the hypotheses concerned only grouping by number or by the entity, the responses that grouped by neither ('2') were discarded (1.1% of all responses). Therefore, all the means for the analyses were values between 0.0 and 1.0.

6.1 Interaction of the Native Language and the Grouping Choice

Firstly, the main hypothesis that there is a difference in the grouping of pictures between the Slovenian and the English groups was addressed. There were six questions in each condition in each of the six groups; therefore, 72 questions for each pair of conditions analysed. Response means of the conditions were calculated by averaging all responses in a condition from each participant.

The effect of the native language of grouping pictures in the critical conditions DP (two pictures in dual, and one picture in plural) and PD (two pictures in plural and one picture in dual) was analysed with a mixed factorial analysis of variance (ANOVA). The ANOVA was conducted with two factors, between-subject (language – English vs Slovenian) and within-subject (response proportions in DP and PD conditions). The Levene test of homogeneity of variance (Navarro, 2015) was not significant ($F(3,392) = .26$, $p = .85$), which shows that the homogeneity assumption of the ANOVA was not violated. There was no main effect of condition ($F(1, 390) = .02$, $p = .88$) or language ($F(1, 390) = .90$, $p = .34$). There was no interaction between condition or language ($F(1, 390) = .05$, $p = .83$). These results suggest that the native language of participants does not have a statistically significant effect on the grouping of pictures in the DP and PD conditions.

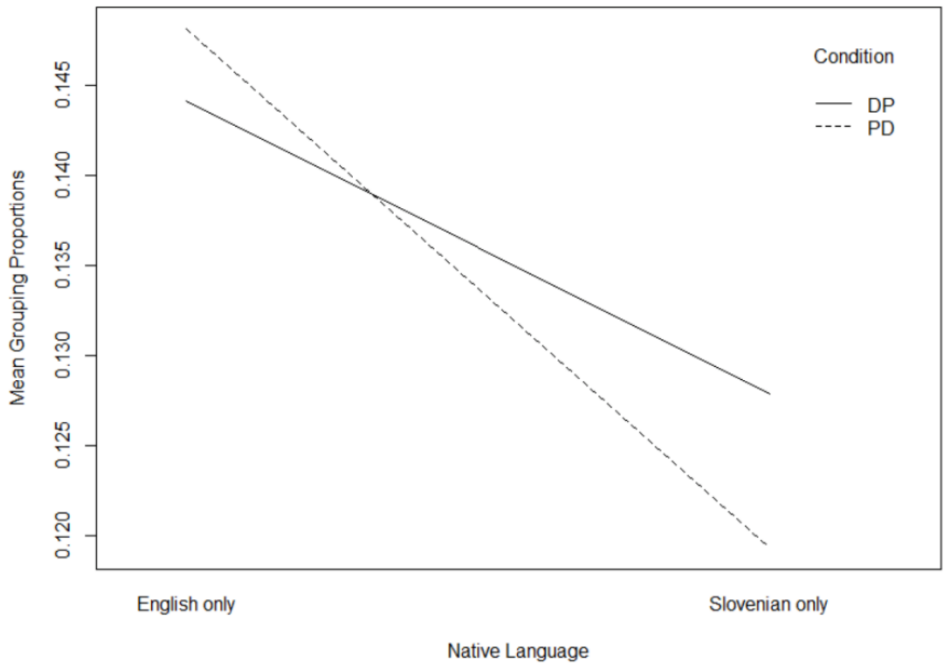


Figure 6: *An interaction plot showing how the mean grouping proportions of DP and PD conditions interact with the native language of participants.*

Next, the control conditions were analysed to see whether there is an effect of native language on grouping there. A mixed factorial ANOVA was conducted to investigate whether the native language of the participants (Slovenian vs English) interacted with the grouping of pictures in the SD and DS conditions. There was no main effect of condition ($F(1, 390) = 1.5$, $p = .22$). There was a main effect of language ($F(1, 390) = 4.49$, $p = .04$), with English participants grouping more by number than Slovenian participants (see Figure 7). There was no interaction between condition and language ($F(1, 390) = .59$, $p = .44$). The Levene test of homogeneity of variance was significant ($F(3,392) = 4.02$, $p < .001$), which shows that this assumption was violated. The main effect of language therefore needs to be interpreted with caution (Navarro, 2015).

Another factorial mixed ANOVA was conducted to investigate whether the second control conditions SP and PS interacted with the native language of participants. The Levene test of homogeneity of variance was significant ($F(3,392) = 2.95$, $p < .01$), which shows that this assumption was violated. There was a main effect of condition ($F(1, 390) = 5.64$, $p < .01$), with participants grouping more pictures by number is the PS condition (see Figure 7). There was no main effect of language ($F(1, 390) = .18$, $p = .67$). There was no interaction between condition and language ($F(1, 390) = .01$, $p = .95$). The main effect of the condition needs to be treated with caution because the Levene test was significant, which means the variances of the two groups are not homogenous.

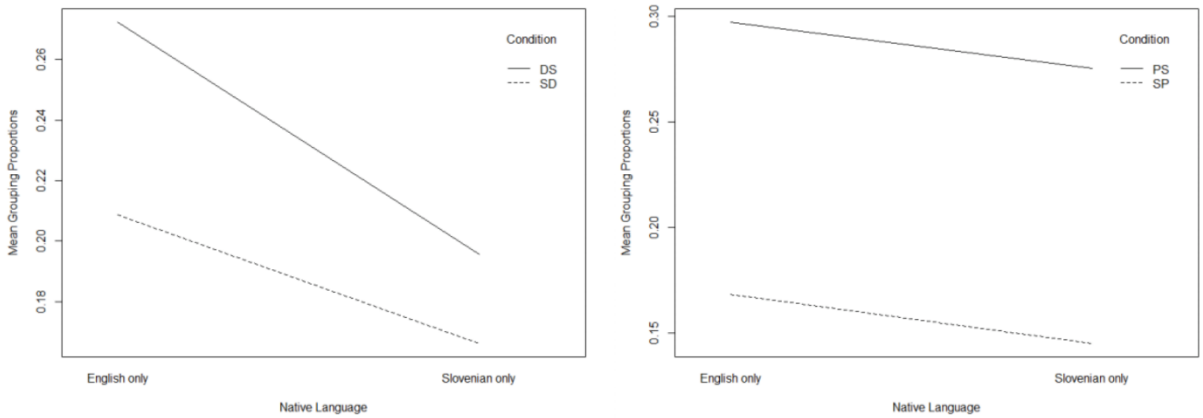


Figure 7: *Interaction plots showing how the mean grouping proportions of DP and PD (top), DS and SD (bottom left), SP and PS (bottom right) conditions interact with the native language of participants.*

A mixed factorial ANOVA was conducted between the mean proportions of the critical condition groupings (DP and PD) and control conditions groupings (SD, DS, SP, PS). The Levene test of homogeneity of variance was not significant $F(3,390) = .45$, $p = .72$). There was no main effect of condition ($F(1, 388) = 1.22$, $p = .27$) or language ($F(1, 388) = 1.32$, $p = .25$). There was no interaction between condition and language ($F(1, 388) = .02$, $p = .90$). These results suggest that the native language of participants does not statistically significantly affect the grouping of pictures in the critical and noncritical conditions.

These analyses uncovered that there might be a difference in the grouping proportions between conditions that have a 'lower' grammatical number (i.e., singular is 'lower' than dual, dual is 'lower' than plural) presented twice in a question (SD, DP, SP) and the conditions that have a 'higher' grammatical number (DS, PD, PS) presented twice in a question. The mean proportions in PD, DS, and PS were all higher than their 'lower' number twice counterparts (see Figure 8). A mixed two ('lower' number twice vs 'higher' number twice) by two (Slovenian vs English) factorial ANOVA was conducted to see if there is an interaction between the native language and mean proportions of critical and control conditions. The Levene test of homogeneity of variance was not significant ($F(3,390) = 1.67$, $p = .17$). There was no main effect of condition ($F(1, 388) = 1.82$, $p = .18$) or language ($F(1, 388) = 1.44$, $p = .23$). There was no interaction between condition or language ($F(1, 388) = .04$, $p = .84$). These results suggest that the native language of participants does not statistically significantly affect the grouping of pictures in the 'lower' number twice and 'higher' number twice conditions.

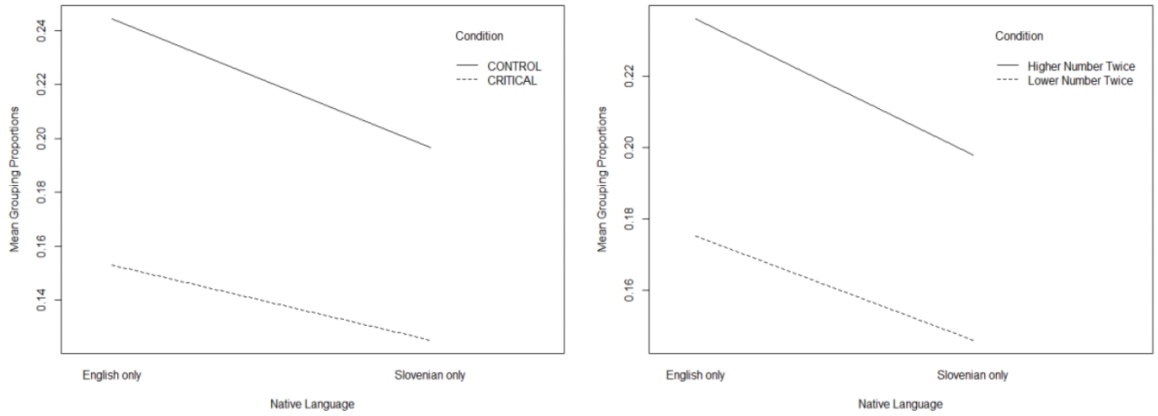


Figure 8: *Interaction plots of the critical and control conditions (left), and 'lower' number twice and 'higher' number twice conditions' (right) mean grouping proportions and their interaction with the native language of the participants.*

6.2 Interaction of the Slovenian Dialect and the Grouping Choice

The second part of the analysis looked at whether there is an interaction between dual vs nondual dialects of Slovenian participants and their grouping choices. Upper Carniolan and Ljubljana dialects were classified as the dual dialects, and Styrian and Littoral dialects as nondual dialects (Jakop, 2012). The rest of the Slovenian and all English data was not used for these analyses. First, a mixed factorial analysis of variance was conducted to investigate whether the dialect of the Slovenian participants (dual vs nondual) influenced their grouping in the critical conditions (DP and PD). There was no main effect of condition ($F(1, 256) = .00$, $p = .95$) or dialect ($F(1, 256) = 2.69$, $p = .10$). There was no interaction between condition or dialect ($F(1, 256) = .02$, $p = .90$). The Levene test of homogeneity of variance was not significant ($F(3,258) = .67$, $p = .57$). These results suggest that the dialect of participants did not affect the grouping of pictures in the critical conditions (DP and PD).

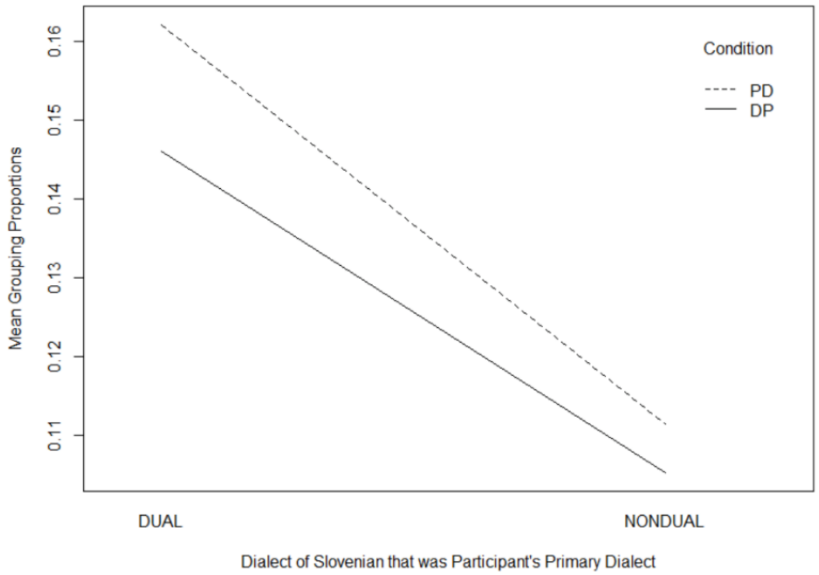


Figure 9: *An interaction plot showing how the mean grouping proportions and the dialect of Slovenian that was participant's primary dialect interact in critical conditions.*

Since the biggest difference in dual usage between dialects of Slovenian is the presence of the feminine dual, responses of dual vs nondual speakers were tested for interactions with responses on masculine vs feminine items. A two (dual vs nondual) by two (masculine vs feminine items) mixed factorial analysis of variance was conducted to investigate whether there is an interaction between the dialect of the Slovenian participants and their groupings of feminine and masculine items. The Levene test of homogeneity of variance was not significant ($F(3, 248) = .31$, $p = .82$), which suggests that this assumption was not violated. There was no main effect of grammatical gender ($F(1, 246) = .00$, $p = .99$) or dialect ($F(1, 246) = .94$, $p = .33$). There was no interaction between condition and dialect ($F(1, 246) = .01$, $p = .91$). These results suggest that the dialect of participants did not affect the grouping of masculine and feminine items. These findings were the same also when comparing groupings of feminine and masculine items between English and Slovenian native speakers. There was no effect of grammatical gender ($F(1, 388) = .03$, $p = .87$) or language ($F(1, 388) = 1.47$, $p = .23$). There was no interaction between condition and language ($F(1, 388) = .11$, $p = .75$). The Levene test of homogeneity of variance was not significant ($F(3, 390) = .49$, $p = .69$).

7 Exploratory Research Questions Analysis

7.1 Interaction of Item Type (Human, Animal, and Object) and Grouping Choice

Exploratory analysis of the influence of the item type on the grouping choice was carried out to see if pictures depicting humans, animals or objects were grouped differently. A three (human, animal, object) by two (English vs Slovenian groupings) mixed factorial ANOVA was conducted. The Levene test of homogeneity of variance was significant ($F(5, 585) = 5.28$, $p < .001$). There was a main effect of item type

($F(2, 582) = .53$, $p = < .001$), which indicated that the human items were grouped more by number than animal and object (see Figure 10), if unaffected by the results from the Levene test. The was no main effect of language ($F(1, 582) = 2.13$, $p = .15$) which indicates that there was no significant difference in grouping between the Slovenian and English group. There was no interaction between item type and language ($F(2, 582) = .10$, $p = .91$) which indicates that there was no difference in the grouping of different item types between the English and Slovenian group.

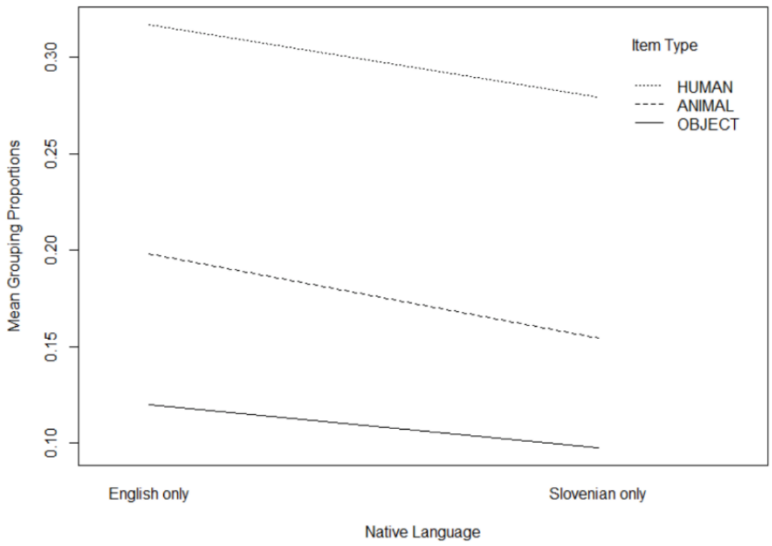


Figure 10: *An interaction plot of how the item type and native language groupings interact.*

7.2 Interaction of Device Used and Grouping Choice

A two (phone vs computer) by two (Slovenian vs English) mixed factorial analysis of variance was conducted to check for the interaction between the device used for solving the survey and the grouping proportions in the Slovenian and English groups. Data from all conditions were used for this analysis. Out of the participants who specified the device they used for the survey, there were 139 (74%) participants who used their phone and 48 (26%) participants who used their computer. The Levene test of homogeneity of variance was not significant ($F(3, 370) = 2.46$, $p = .06$). There was no main effect of condition ($F(1, 368) = .10$, $p = .18$). However, there was a main effect of the device ($F(1, 368) = 4.05$ $p < .05$), which suggests that the participants tended to group more by number when solving the experiment on the phone rather than on the computer (see Figure 11). There was no interaction between language and device ($F(1, 368) = .25$, $p = .62$), which suggests that there was no significant difference in groupings between Slovenian and English group.

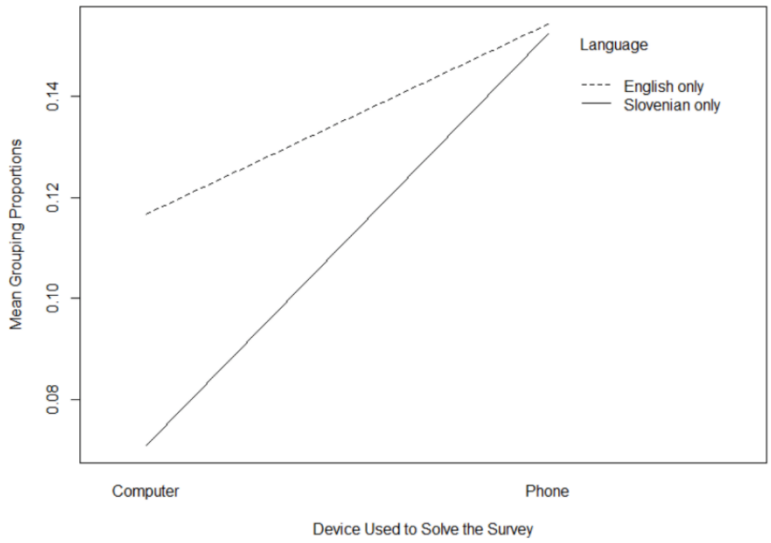


Figure 11: *An interaction plot showing how the device used to solve the survey, mean grouping proportions, and the native language of the participants interact.*

7.2.1 Interaction of Participant Sex and Grouping Choice

A two (female vs male) by two (critical condition proportion vs control condition proportion) mixed factorial analysis of variance has been conducted to check for the interaction between the sex of the participants and groupings in the critical and noncritical conditions. Data from all conditions were used for this analysis. Out of the participants who specified their sex, there were 131 (69%) females and 58 (31%) males in both groups. The Levene test of homogeneity of variance was not significant ($F(3, 374) = .29$, $p = .83$). There was no main effect of condition ($F(1, 372) = 1.14$, $p = .29$), or sex ($F(1, 372) = .93$, $p = .37$). There was no interaction between the sex of participants and grouping ($F(1, 372) = .00$, $p = .97$). These results suggest that the sex of the participants did not influence the grouping of pictures in a statistically significant way.

8 Discussion

8.1 Interaction of the Native Language and the Grouping Choice

A hypothesis test of the main hypothesis that grouping of pictures will differ between the Slovenian and English group showed a non-significant interaction between the grouping proportion and native language in the critical conditions. Therefore, we failed to reject the null hypothesis. An ANOVA comparing the control conditions SD vs DS showed a main effect of language, with English participants grouping more by number than Slovenian participants (see Figure 7). This effect is the opposite of the hypothesised effect and against the hypothesis H1 (b), which stated that English and Slovenian participants would be as likely to group SPSN with DISP and DISN in the control conditions. This main effect needs to be treated with

caution because the Levene test was significant. That shows that the homogeneity of variance assumption was violated, which means that the variances between groups were not homogenous. This outcome causes a problem when interpreting the results as robust. If we disregard the violation of the ANOVA assumption, this result shows that the English native speakers grouped the cases where there were either two single items and a dual item, or two dual items and a singular item, by number significantly more than the Slovenian group. Both groups still grouped more by the entity. The effect found could be a result of a considerable smaller number of participants in the English group.

There was a main effect of condition in the PS vs SP analysis, which suggests that participants grouped significantly more pictures by number is the PS condition than in the SP condition (see Figure 7). The Levene test of homogeneity of variance was significant, so this effect needs to be treated with caution. Nevertheless, this result could point to the fact that people tend to group pictures by number in conditions with 'higher' number twice (PD, DS, PS) more than in conditions with 'lower' number twice (DP, SD, SP). To test this, another mixed factorial ANOVA was conducted, this time to see if there is an interaction between the native language and mean proportions of the 'lower' number twice and 'higher' number twice conditions. There were no significant effects found, although there is a general trend that all 'higher' number twice cases have higher mean proportion. More data is needed to explore this effect.

A general ANOVA between the means of the critical (DP, PD) and control conditions (SD, DS, PS, SP) was done to investigate if there is an interaction with the native language. No significant interaction or main effect was found.

There are two possible conclusions from these results, either the theory of linguistic relativism does not apply to grammatical number (at least in the case of Slovenian versus English), or this experiment was flawed. Both possibilities will be discussed below.

8.2 Interaction of the Slovenian Dialect and the Grouping Choice

The hypothesis, which stated that Slovenian nondual dialect speakers would respond similarly to the English native speaker group (grouping by the entity in the critical conditions), was not significant. This result suggests that there was no difference between groupings of participants that regarded dual dialects of Slovenian as their primary dialects and participants that regarded nondual dialects of Slovenian as their primary dialect. However, the *p*-value for the main effect of dialect was 0.1, which was deemed significant by cognitive scientists in the past when discussing the variance for a variety of linguistic performance (Hunt & Agnoli, 1991). Therefore, the effects found could indicate some difference between dual and nondual dialect speakers. The mean of the dual dialect group was higher than the mean of the nondual dialect group (see Figure 9). That suggests that dual dialect participants group more by number than nondual dialect participants, which supports the hypothesis H1 c). This finding could indicate that mental representations depend on the spoken language (dialect) more than on the standard written language (CSS Slovenian).

A possible limitation is that the dialects cannot be separated into two clear-cut categories of whether they use dual or they do not. The ANOVA that tested responses on feminine vs masculine items between the dialects did not yield any significant interaction. That is a problem for the previous findings, as the dialects contrast the most in the use of feminine dual. Also, no speaker speaks in their dialect all the time because of the influence of education and media in CSS. The number of participants by the group was quite different; there were 29 dual, and 102 nondual dialect participants. However, the Levene test did not show a violation of the assumption of homogeneity of variance.

8.3 Exploratory Research Questions Analysis

8.3.1 Interaction of Item Type (Human, Animal, and Object) and Grouping Choice

The exploratory research question – Does the item type of the picture – human, animal, object – influence the grouping? – was analysed with three (item type) by two (grouping in the critical vs noncritical conditions) mixed factorial ANOVA. The analysis showed that participants grouped pictures of humans significantly more often by number than by the entity. There was no significant difference between the languages, so the results do not concern the main hypothesis. However, this indicates some interesting characteristics of the experimental setup that could be relevant for further research. Perhaps different humans seem more like each other than different objects (e.g., 'two men' and 'two kings' look more like each other, than 'two chairs' and 'two laptops'), and that is why participants grouped humans by number more. Another option is that the number of people bears more meaning than the number of, for example, erasers. A distinction between a single person, a pair of people, and a group of three people is more important to our everyday lives than a distinction between one, two, or three chairs. We even think about pairs of people (parents, couples, twins), and very rarely in pairs of objects.

Another possible influence is that, in some cases, an identical picture was used for the singular version and the dual or plural version for animals and objects, due to not finding many different appropriate pictures. The human items were consistently not identical between conditions. That could lead to grouping items that are identical in appearance, with no linguistic processing involved. The choice of pictures was a major flaw of the experiment, which needs to be improved in further research.

Another reason for differences in grouping between items of a different type was highlighted by participant feedback. It was reported that the colours of the animals were distracting (especially the eagles and rabbits) and that they chose to group the same colours, irrespective of item type or number.

8.3.2 Interaction of Device Used and Grouping Choice

The Qualtrics survey could be solved on the computer or the phone. Most participants used their phone (74%) rather than a computer (26%) to solve the experiment. A hypothesis test showed that people who solved the survey on their phone grouped pictures by number more. This result pointed out a major flaw of the experiment execution. Possible reasons for differences between solving on a phone and a computer are that participants had to scroll down to see all pictures in one question in the phone version of the survey, which could have prolonged the response time. Additionally, participant's comments pointed out that the phone version of the Qualtrics survey sometimes had delays in loading all pictures of one question on the screen. Since the participants were supposed to answer based on their first intuition, the different exposure times to certain pictures in a question might have influenced their answer. This finding could mean that when people have more time to think about the similarities, they choose the number more often. A suggestion for further research is to specify that the survey can only be solved on the computer.

8.3.3 Interaction of Participant Sex and Grouping Choice

The participant sample was quite heterogeneous, with a large difference in the number of participants between the two groups (77% of participants were Slovenian, 23% English). This proportion was justified by the English speaker group being the control, so there were fewer participants required. However, it

nevertheless could have influenced the difference between the groups not yielding a significant result. There was a big difference between the number of males and females between groups as well; there was 24% Slovenian males and 52% English males. Some studies (e.g., Payne & Lynn, 2011) have found that females have a stronger module for second language processing than males, which could affect the results of this experiment. English as their second language could have influenced females in the Slovenian group when responding to the grouping questions. However, there was no statistically significant interaction found between the participant's sex and their grouping choice.

8.4 Relation of the Results to the Theory of Linguistic Relativism

The results of this study failed to support the theory of linguistic relativism and are, therefore, not in line with the findings of Phillips and Boroditsky (2013). The only effect was found between the dual and nondual dialect of Slovenian. However, it is not robust as the *p* value is 0.1, there is a big difference between groups, and it is hard to make a sharp distinction between dialects. Nevertheless, comparing dialects of the same language is somehow better than comparing two separate languages, as there are no confounding cultural effects. Perhaps the lack of effect we found between languages is due to other factors, such as culture. Furthermore, the results can be explained by a finding of Athanasopoulos (2006) that the second language acquisition may alter cognitive dispositions established by the first language. Since all the Slovenian participants were fluent in English, the English number system could have affected their cognitive representations of numbers. That effect has been controlled for in the dialect comparison. These results might be possible evidence in favour of linguistic relativism.

Nonetheless, more research is needed to establish whether the grammatical number is one of the domains in which the effects of linguistic relativism can be observed. The experimental design of this study was not completely appropriate for testing this question. The task used would have to be more implicit, perhaps including some distractor questions, so participants do not figure out what the experiment is about, to test how people think about groups of two versus groups of three and one better. Only telling people to group the more similar pictures is too vague and too easy. People just chose one characteristic of the pictures to focus on, be it the colour, the type, the arrangement, or the number, and then grouped them quite consistently according to that.

8.5 Other Limitations and Outlook

The Qualtrics survey has been translated from English to Slovenian by the researcher, and English was the original language of the survey. A considerable amount (n = 27) of Slovenian participants completed it with English instructions, even though they were specifically asked by the researcher and by text in the survey itself to change the language. That could have led to a priming effect of English, and so the Slovenian participants who were all fluent in English could have subconsciously employed more of an English number system. This suggestion is only a stipulation, as a statistical comparison was not made. Further research would need to address this question to ensure no effects of English on the Slovenian participants.

The survey has been conducted online, using convenience sampling of participants over social media. That attracted a heterogeneous sample in terms of sex and native language, but a homogenous sample in terms of education level (see Figure 2), at least to the extent of almost everyone having a high school education, and the majority being students. To a certain extent, that characteristic shows that a similar socio-economic background (to each other, and the researcher) can be inferred. That could be an indicator of

response bias, which occurs when only highly motivated people return a survey. Out of all participants that attempted the survey (n = 306), 67 of them completed less than 90% of the survey, which was the cut-off point for inclusion in further analysis. These participants were therefore not included in the study, which means that a part of the population (less motivated one) was not represented, which leads to a potential low external validity of the study. Further research would benefit from recruiting a more representative sample of the population.

Other cognitive implications of having dual in the native language could be investigated as well, for example, memory. If the assumption that language cues attention is correct, the retention of the number of items could be affected by different number systems of languages. This phenomenon could be compared between the Slovenian language and a language without the dual number, or the between the different dialects of Slovenian.

9 Conclusion

This study attempted to investigate whether the participants' native language affects their mental representations and categorisations. A grouping task was used to compare Slovenian and English native speakers. The relevant difference between these two languages is their different grammatical number systems. The Slovenian language has three grammatical numbers – singular, dual, and plural, whereas the English language only uses two grammatical numbers – singular and plural. The experimental hypothesis was that Slovenian native speakers would group items in the Dual Plural and Plural Dual conditions significantly more by number. In contrast, English native speakers would group them by the entity. This hypothesis was based on the theory of linguistic relativity, which would suggest that since Slovenian language has three distinctive categories for number, that distinction in the language would reflect in the mental representations. The mental representations were tested by grouping pictures of a different kind (human, animal, object) and different grammatical gender (masculine and feminine) in an arrangement of either one, two, or three items combined according to different conditions (DP, PD, DS, SD, PS, SP). The results failed to reject the primary null hypothesis. There was a difference found between dual and nondual dialects of Slovenian, but further research needs to confirm it with a more robust experimental design. There was a significant difference found between results from participants solving the survey on a phone versus a computer. This flaw in the design needs to be considered in further research. The theory of linguistic relativity was not supported by the results.

10 References

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. Bilingualism: Language and Cognition, 9(1), 89–96.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43(*1), 1–22.

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. Language Learning, 58, 63–79.

Chomsky, N. (2006). Language and mind (3rd ed.). Cambridge: Cambridge University Press.

Fox, J. & Weisberg, S. (2011). An {R} companion to applied regression (2nd ed.). Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Hunt, E. & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review, 98*, 377–389.

Jakop, T. (2008). *The dual in Slovene dialects*. Brockmeyer.

Jakop, T. (2010). Dvojinske in pluralizirane oblike v slovenskih narečjih. *Slavia Centralis, 2010*(1*)*, 74–87.

Jakop, T. (2012). Use of dual in standard Slovene, colloquial Slovene and Slovene dialects. *Linguistica, 52*(1), 349–362.

January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition, 104*(2), 417–426.

Lucy, J. A. (2000). Introductory comments. In Niemeier, S., Dirven, R. Eds. Evidence for linguistic relativity. (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory; v.198). Amsterdam: J. Benjamins.

Marušič, F., Plesničar, V., Razboršek, T., Sullivan, J., & Barner, D. (2016). Does grammatical structure accelerate number word learning? Evidence from learners of dual and nondual dialects of Slovenian. *PloS one, 11*(8).

Navarro, D. J. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners (Version 0.5).* The University of Adelaide, Adelaide.

Payne, T., & Lynn, R. (2011). Sex differences in second language comprehension. Personality and Individual Differences, 50(3), 434–436.

Pederson, E., E. Danziger, D. Wilkins, S. C. Levinson, S. Kita, and G. Senft. (1998). Frames of reference and language concepts. *Language, 74,* 557–89.

Phillips, W., & Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 25.

Pinker, S. (1994). *The language instinct.* New York: William Morrow.

R Core Team (2018). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Version 3.4.4. (2018-03-15). Vienna, Austria. URL: https://www.R-project.org/.

Reinez, M., & Prinz, J. (2009). Reviving Whorf: The return of linguistic relativity. *Philosophy Compass, 4*(6), 1022–1032.

Rudzinska, I. (2013). Are females better than males in communication in second language? *The Journal of Education, Culture, and Society, (2),* 148–154.

Slobodchikoff, T. G. (2013). The Slavic dual: Morphosyntactic feature economy as a factor in language change (Doctoral dissertation). University of Arizona, Arizona.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf.* New York: Wiley.

Wickham, H., & Henry, L. (2019). Tidyr: Easily tidy data with 'spread()' and 'gather()' functions. R package version 0.8.3. Retrieved from https://CRAN.R-project.org/package=tidyr.

Wickham, H., Francois, R., Henry, L., & Muller, K. (2019). Dplyr: A grammar of data manipulation. R package version 0.8.0.1. Retrieved from https://CRAN.R project.org/package=dplyr.

Packages used for the data analysis in R (R Core Team, 2018) were 'car' (Fox & Weisberg, 2011), 'tidry' (Wickham & Henry, 2019), and 'dplyr' (Wickham et al., 2019).

11 Appendices

11.1 Appendix One: Acronyms Used

CSS	**Central Standard Slovene**
SD	Singular Dual
SP	Singular Plural
DS	Dual Singular
DP	Dual Plural
PS	Plural Singular
PD	Plural Dual
DISP	Distinct Picture
DISN	Distinct Number
SPSN	Same Picture Same Number
d	Distinct Picture (DISP) is of distinct grammatical gender to the Same Picture
s	Distinct Picture (DISP) is of the same grammatical gender to the Same Picture

11.2 Appendix Two: Full List of Pictures Used

Slovenian translations are in the brackets.

1. Human masculine:
    a. boy (fant)
    b. man (moški)
    c. ballet dancer male (baletnik)
    d. king (kralj)
    e. male teacher (učitelj)
    f. clown (klovn)
2. Human feminine:
    a. girl (punca)
    b. woman (ženska)
    c. queen (kraljica)
    d. female farmer (kmetica)
    e. ballerina (baletka)
    f. female teacher (učiteljica)
3. Animal masculine:
    a. dog (pes)
    b. bull (bik)
    c. rabbit (zajec)
    d. eagle (orel)

e. ram (oven)
f. rooster (petelin)

4. Animal feminine:
   a. cow (krava)
   b. cat (mačka)
   c. bird (ptica)
   d. hen (kura)
   e. sheep (ovca)
   f. squirrel (veverica)
5. Object masculine:
   a. pen (svinčnik)
   b. mobile phone (telefon, mobitel)
   c. laptop (računalnik)
   d. notebook (zvezek)
   e. chair (stol)
   f. pot (lonec)
6. Object feminine:
   a. table (miza)
   b. bottle (steklenica)
   c. spoon (žlica)
   d. eraser (radirka)
   e. folder (mapa)
   f. handbag (torbica)

11.3 Appendix Three: List of Questions and Conditions that Each Group of Participants Saw

Participants in group A saw:

1. 1aSD d	2. 1bSP s	3. 1cDS d	4. 1dDP d	5. 1ePS d	6. 1fPD s
7. 2aSD d	8. 2bSP d	9. 2cDS d	10. 2dDP s	11. 2ePS s	12. 2fPD s
13. 3aSD d	14. 3bSP d	15. 3cDS s	16. 3dDP s	17. 3ePS s	18. 3fPD d
19. 4aSD d	20. 4bSP d	21. 4cDS s	22. 4dDP d	23. 4ePS s	24. 4fPD s
25. 5aSD d	26. 5bSP d	27. 5cDS s	28. 5dDP s	29. 5ePS s	30. 5fPD d
31. 6aSD d	32. 6bSP s	33. 6cDS d	34. 6dDP s	35. 6ePS d	36. 6fPD s

Participants in group B saw:

1. 1aSP d	2. 1bDS d	3. 1cDP s	4. 1dPS s	5. 1ePD d	6. 1fSD s
7. 2aSP d	8. 2bDS d	9. 2cDP s	10. 2dPS s	11. 2ePD s	12. 2fSD d
13. 3aSP d	14. 3bDS d	15. 3cDP s	16. 3dPS s	17. 3ePD s	18. 3fSD d
19. 4aSP d	20. 4bDS s	21. 4cDP d	22. 4dPS d	23. 4ePD s	24. 4fSD s
25. 5aSP d	26. 5bDS s	27. 5cDP s	28. 5dPS d	29. 5ePD d	30. 5fSD s
31. 6aSP d	32. 6bDS d	33. 6cDP s	34. 6dPS d	35. 6ePD s	36. 6fSD s

Participants in group C saw:

1. 1aDS d	2. 1bDP d	3. 1cPS d	4. 1dPD s	5. 1eSD s	6. 1fSP s
7. 2aDS d	8. 2bDP d	9. 2cPS d	10. 2dPD s	11. 2eSD s	12. 2fSP s
13. 3aDS d	14. 3bDP d	15. 3cPS d	16. 3dPD s	17. 3eSD s	18. 3fSP s
19. 4aDS d	20. 4bDP d	21. 4cPS d	22. 4dPD s	23. 4eSD s	24. 4fSP s
25. 5aDS d	26. 5bDP d	27. 5cPS d	28. 5dPD s	29. 5eSD s	30. 5fSP s
31. 6aDS d	32. 6bDP d	33. 6cPS d	34. 6dPD s	35. 6eSD s	36. 6fSP s

Participants in group D saw:

1. 1aDP d	2. 1bPS d	3. 1cPD d	4. 1dSD s	5. 1eSP s	6. 1fDS s
7. 2aDP d	8. 2bPS d	9. 2cPD d	10. 2dSD s	11. 2eSP s	12. 2fDS s
13. 3aDP d	14. 3bPS d	15. 3cPD d	16. 3dSD s	17. 3eSP s	18. 3fDS s
19. 4aDP d	20. 4bPS d	21. 4cPD d	22. 4dSD s	23. 4eSP s	24. 4fDS s
25. 5aDP d	26. 5bPS d	27. 5cPD d	28. 5dSD s	29. 5eSP s	30. 5fDS s
31. 6aDP d	32. 6bPS d	33. 6cPD d	34. 6dSD s	35. 6eSP s	36. 6fDS s

Participants in group E saw:

1. 1aPS s	2. 1bPDs	3. 1cSD s	4. 1dSP d	5. 1eDSd	6. 1fDP d
7. 2aPS s	8. 2bPDs	9. 2cSD s	10. 2dSP d	11. 2eDSd	12. 2fDP d
13. 3aPS s	14. 3bPDs	15. 3cSD s	16. 3dSP d	17. 3eDSd	18. 3fDP d
19. 4aPS s	20. 4bPDs	21. 4cSD s	22. 4dSP d	23. 4eDSd	24. 4fDP d
25. 5aPS s	26. 5bPDs	27. 5cSD s	28. 5dSP d	29. 5eDSd	30. 5fDP d
31. 6aPS s	32. 6bPDs	33. 6cSD s	34. 6dSP d	35. 6eDSd	36. 6fDP d

Participants in group F saw:

1. 1aPD s	2. 1bSD s	3. 1cSP s	4. 1dDS d	5. 1eDP d	6. 1dPS d
7. 2aPD s	8. 2bSD s	9. 2cSP s	10. 2dDS d	11. 2eDP d	12. 2fPS d
13. 3aPD s	14. 3bSD s	15. 3cSP s	16. 3dDS d	17. 3eDP d	18. 3fPS d
19. 4aPD s	20. 4bSD s	21. 4cSP s	22. 4dDS d	23. 4eDP d	24. 4fPS d
25. 5aPD s	26. 5bSD s	27. 5cSP s	28. 5dDS d	29. 5eDP d	30. 5fPS d
31. 6aPD s	32. 6bSD s	33. 6cSP s	34. 6dDS d	35. 6eDP d	36. 6fPS d

11.4 Appendix Four: Participant Information Form, Consent Form, and Debriefing Form

PARTICIPANT INFORMATION SHEET – ENGLISH:

You are being asked to take part in a research study on the connection between language and cognition. This study aims to investigate whether there is a difference in mental representations between Slovenian and English native speakers.

This research is being conducted as part of Honours Dissertation in Psychology by [redacted], 4th-year student of Cognitive Science (Humanities). It is supervised by Professor Martin Pickering, Department of Psychology, University of Edinburgh. For more information, please contact me on [redacted]@sms.ed.ac.uk or the supervisor on mpickering@ed.ac.uk.

You will be asked to group certain pictures and to complete a few demographic questions at the end.
The whole survey will not last more than 10 minutes. Please respond to the questions as fast as you can because we are interested in your immediate opinions.

You can decide to stop being a part of the research study at any time without explanation. You have the right to ask that any data you have supplied to that point be withdrawn/destroyed.

You have the right to omit or refuse to answer or respond to any question that is asked of you.

You have the right to have your questions about the procedures answered (unless answering these questions would interfere with the study's outcome). If you have any questions as a result of reading this information sheet, you should ask the researcher before the study begins.

There are no known benefits or risks for you in this study.

Your participation in this study is voluntary. I immensely appreciate your help. The data we collect does not contain any other personal information about you. No one will link the data you provided to the identifying information you supplied (e.g., email, age, area of domicile). Your data will be anonymised and stored on a password-protected computer. From that time, there will be no record that links the data collected from you with any personal data from which you could be identified. Up until the point at which your data have been anonymised, you can decide not to consent to have your data included in further analyses.

CONSENT FORM – ENGLISH:

By clicking to proceed, you are confirming that:

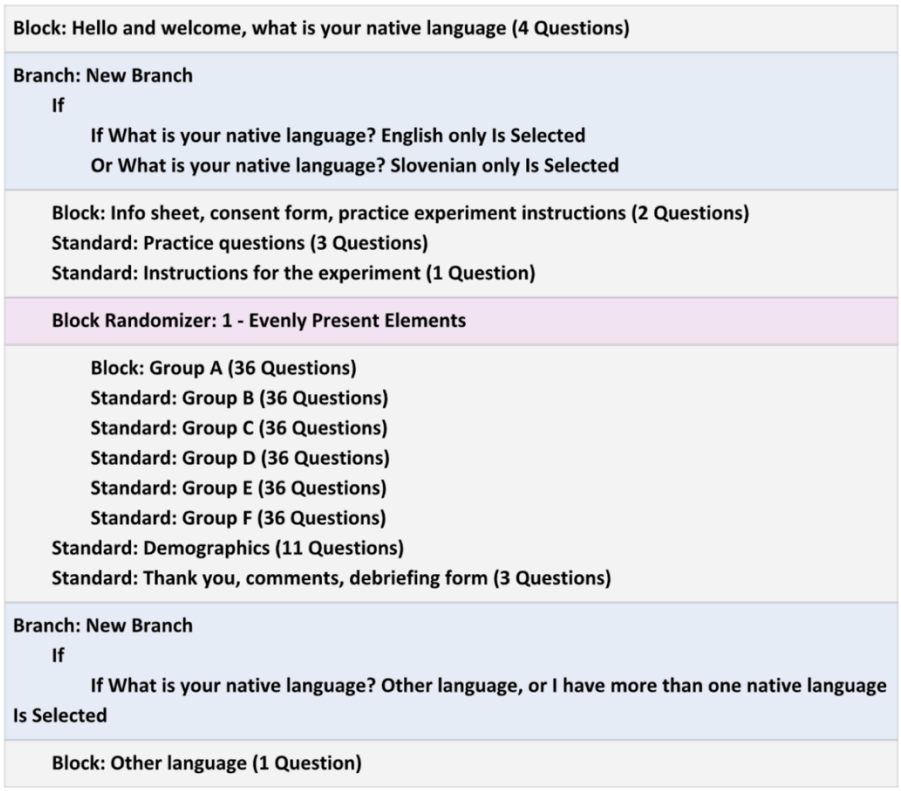
(1) you have read and understood the above information,
(2) questions about your participation in this study have been answered satisfactorily,
(3) you are aware of the potential risks (if any),
(4) you have the right to withdraw your participation at any point without giving a reason,
(5) you are taking part in this research study voluntarily (without coercion), and
(6) your anonymised data may be used for the purposes of this research.

DEBRIEFING SHEET – ENGLISH:

Here is a summary of what I was researching. If you have any further questions, please feel free to contact me ([redacted]) on [redacted]@sms.ed.ac.uk, or my supervisor (Prof. Martin Pickering, Department of Psychology, University of Edinburgh) on mpickering@ed.ac.uk.

This research is following the logic of Phillips and Boroditsky (2003) on how quirks of grammar can affect the way you think. They were testing whether the grammatical gender of inanimate objects leads people to think of them as having a gender. They found the effects of grammatical gender on people's perceptions of objects in a variety of settings. Instead of looking at gender, the present research will be exploring the effects of the dual grammatical number on people's perceptual inclination towards groups of two versus groups of three. Slovenian, a South Slavic Indo-European language, still has the dual number in the standard language and many of its spoken dialects. This part of grammar died out in other Indo-European languages (except in Sorbian) centuries ago. The possible difference in cognitive representations is investigated by comparing Slovenian and English native speakers' responses to a grouping task.

Hypotheses:
Null hypothesis: If grammar does not guide thought, then Slovenian and English native speakers will not differ in the way that they group items.



Experimental hypothesis: If grammar guides thought, then Slovenian native speakers will be less likely to put groups of two and three people/animals/objects together than English native speakers.

11.5 Appendix Five: Survey Flow from Qualtrics

Showing the design of the experiment.
Survey Flow

The Seneca Language and Bilingual Road Signs: A Study in the Sociology of an Indigenous Language

Anna Taylor
The University of Ohio

Abstract. One of the fundamental types of human rights concerns collective-developmental rights which allow minorities to use heritage languages and practices without external interference (Vašák, 1977). The protected status of minority language rights is a critical part of language revitalisation in which speakers of heritage languages, faced with the encroachment of more socially, politically, and economically dominant languages, embark on vigorous programs to ensure the survival and continued usage of their language. The Five Nations Iroquoian language, Seneca, has just a few remaining speech communities and a variety of ongoing language revitalisation initiatives (Mithun, 2012). To revitalise their traditional language, community classes through the Seneca Language Department and the Faithkeepers Montessori School Seneca Language Nest for young speakers have concentrated their efforts on preserving Onöndowa'ga:' Gawë:nö' the indigenous name for the Seneca language (Bowen, 2020; Murray, 2015). In the public sphere, a push by the Seneca Nation of Indians Department of Transportation fulfilling the intent of the federal Native American Tourism and Improving Visitor Experience (NATIVE) Act enacted in 2016, specifically included bilingual signs for state roads running through indigenous land in addition to other significant components (Figura, 2016). In an area whose geographic names are strongly connected to Iroquoian languages including Seneca, these bilingual signs represent more public and visible Seneca language presence and stand as symbols of language revitalisation. The place names and information that appear on the signs have considerable significance for community identity as well as linguistic and economic impacts, among others. Through oral histories collected from Seneca Nation members and language advocates in addition to a representative from the New York State Department of Transportation, this study pursues an analysis of the Seneca public usage of their heritage language and the various language revitalisation efforts occurring among indigenous and minority communities internationally. As the COVID-19 pandemic threatens already vulnerable populations, heritage languages that have been historically oppressed face a global language crisis that disproportionately harms and disadvantages speakers of heritage and minority languages (Roche, 2020). While the language of road signs may seem mundane, this study reveals how the Seneca bilingual signs play a significant role in awareness of indigenous territory and consequently stimulation of the local economy as well as supporting language learning, revitalisation, and de stigmatization. Primarily through the efforts of the Seneca community, the bilingual signs represent the expression of language rights in the public sphere and one part of the ongoing language revitalisation.

Keywords: language revitalization, bilingualism, indigenous languages, sociolinguistics

---

1 Introduction

1.1 Positionality

Before I moved to Columbus, Ohio to pursue my undergraduate degree in linguistics, I lived in Tonawanda, New York for my entire life. Although I had some awareness of the etymologies of the place names that dominated this area of western New York, it took three years of linguistics education at The Ohio State University before I set out to research and translate local names like Tonawanda, Niagara, Scajaqueda, Allegany, and others from their Iroquoian origins. It was at that point that I realised that my hometown of Tonawanda is also home to the Tonawanda Band of Seneca, which is one of just a few remaining speech communities for the Seneca language (Delgado N.d.) [34].

The Seneca language, or the Language of the People of the Great Hills, is the traditional language of the Seneca Nation of Indians and the Tonawanda Band of Seneca, though it currently faces a dwindling fluent speaker population (Delgado N.d.). This has heightened the urgency of the revitalisation programs for Seneca as part of the global language crisis currently facing minority languages around the world that are in similar danger of disappearing if not for active intervention on the part of community and non-community members alike (Roche, 2020). In recognition of my status as a non-community member studying the Seneca language, the focus of my research is on public displays of 'local definitions of language such as 'how a community connects to each other and how they express … themselves and their culture to each other'' in following with Wesley Leonard's 2019 'Musings on Native American Language Reclamation and Sociolinguistics' which credits this positionality to a Miami language teacher (Leonard, 2019).

There is considerable linguistic knowledge to be gained from studying all human languages, though I believe that the views and goals of the traditional speakers should be considered the priority for researchers and should be recognised by non-community members as a human right inherent to minority language speakers. While the death of any language is a significant loss in terms of linguistic diversity, the decline in the Seneca speech community represents much more considering the indigenous knowledge systems, culture, and history that are at stake[35]. My perspective as a sociolinguistic researcher has been significantly influenced by the insights of my thesis advisor, Dr. Brian Joseph, and my Global Arts + Humanities Discovery Theme Fellowship advisor, Puja Batra-Wells, as well as the Seneca language advocates and community organisers, Jody Clark and Flip White, who have kindly spoken with me about their traditional language, its history, and the ongoing language revitalisation programs. Finally, I would be remiss if I did not acknowledge the contributions of the other members of my thesis defence committee, OSU professors Andrea Sims and John Low, who provided valuable insights on previous drafts of this paper.

In light of this positionality, this study pursues an examination of language rights in the public sphere through Seneca bilingual signs as well as some points of comparison of bilingual signs and language revitalisation initiatives in other minority communities around the world. Specifically, I posit the following research questions: to what extent does Seneca bilingual signage, as well as such signage in general, encompass collective-developmental human rights and language rights in the public sphere and how do the public usage of the Seneca language and the Seneca revitalisation programs compare to minority speech communities internationally?

[34] There is limited data available at the moment concerning where the Seneca language is currently used among the various land under Seneca jurisdiction or where speakers of Seneca live.

[35] It would be academically irresponsible not to mention the failings of using a biological framework to analyse language endangerment considering that language 'death' does not necessarily mean that the language can no longer be used. For one thing, languages can survive in specialized niches, such as the use of Latin in the Catholic Mass into the mid-twentieth century (see also Section 3.2). In addition, though linguists had once considered the *myaamia* language to be 'extinct,' the Miami Tribe of Oklahoma has revitalized the sleeping language through archival records (Leonard, 2019).

1.2 Background

The Seneca language, which is natively referred to as Onöndowa'ga:' Gawë:nö', is considered a Haudenosaunee or Iroquoian language (Delgado N.d.). Based on Marianne Mithun's 2012 work on Iroquoian languages, the Seneca language is part of the Five Nations Iroquoian language family, which can be seen in the chart in Figure 1 below (Mithun, 2012). The Seneca are one of the original five nations of the Haudenosaunee (Iroquoian) Confederacy along with the Cayuga, the Mohawk, the Oneida, and the Onondaga, with the eventual addition of the Tuscarora in 1722 (Delgado N.d.). One source writes that the confederacy, also known as the Great League, was created around 1450 as part of an effort to create an 'ethnic confederacy among the Iroquois' (Wallace, 1969, p. 42). The six languages of the League along with Wendat and Wyandot make up the Northern Iroquoian languages, all of which are said to be 'highly polysynthetic, with templatic morphological structure…[which] are largely the same across the languages' (Mithun, 2012, p. 247). However, the Northern Iroquoian languages are classified as mutually unintelligible, while the Five Nations Iroquoian languages have had mixed reports of mutual comprehension based on the intensity of exposure (Mithun, 2012).

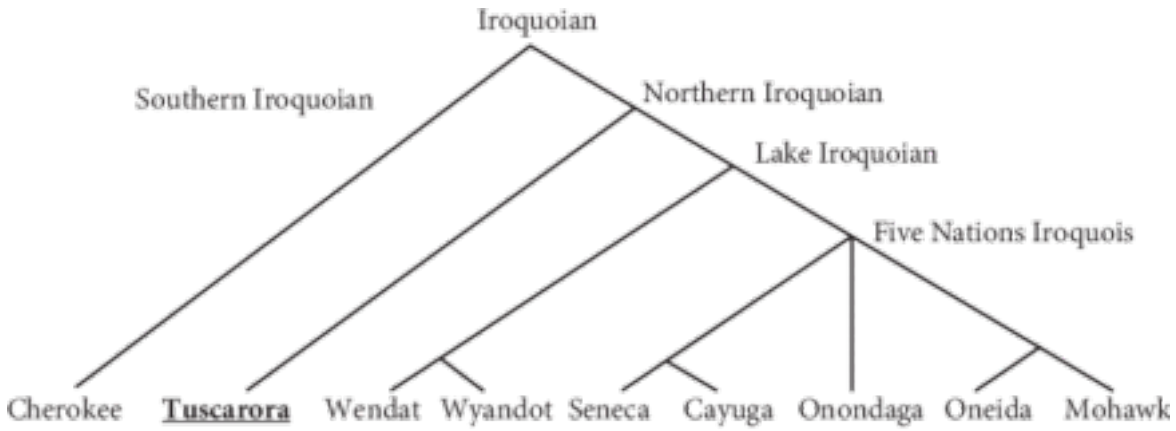


Figure 1: *Iroquoian language family chart reproduced with permission from Dr Mithun (2012).*

In more recent times, Northern Iroquoian languages are spoken primarily in parts of New York State and Canada (Mithun, 2006). While 'the Seneca Nation holds land titles for the Cattaraugus Reservation, the Allegany Reservation, the Oil Springs Reservation, and at reservations at Niagara and Buffalo' as well as the reservation of the Tonawanda Band of Seneca which spans three counties in western New York (Delgado N.d.), Mithun writes that 'Seneca is now spoken in three communities in western New York: Cattaraugus, Allegany, and Tonawanda' (Mithun, 2006, p. 32)[36]. Estimates of the current size of the speaker population are difficult to ascertain for a variety of reasons which include the inaccessibility of the speech community and the subjectivity of fluent speaker status among other considerations. The matter of determining a speaker's proficiency in their heritage language is a complex topic that often leads to

[36] It is worth mentioning that the United States government besieged and flooded 9,000 acres of land or about a third of the Seneca territory at the Allegany Reservation during the Kinzua Dam project in 1964, which had a profound impact on the Seneca community (Borgia-Askey, 2010).

questions of what it means to be a speaker and how to qualify. This is an especially sensitive issue with a language spoken only by a relatively small number of speakers in a dwindling speech community. In a personal communication in 2007 with the author of a dissertation on language preservation at Ohi:yo' or the Seneca Allegany territory, Wallace Chafe is quoted saying 'my guess is that there are less than 50 speakers altogether' (Borgia-Askey, 2010). Yet, based on the homepage for the Faithkeepers School Montessori Seneca Language Nest, a Channel 2 WGRZ news story featuring the school remarks that 'there are less than 30 people who can speak the language fluently' (Faithkeepers, 2021). These may be the most recent statistics generally available regarding the Seneca speaker population since the historical stigmatization of the language through governmental suppression, residential schools, and other injustices has made the linguistic situation into a complex and sensitive topic.

Regardless of the exact number of Seneca speakers, the population is dwindling, and most currently proficient speakers tend to be older in age and relatively isolated from other Seneca speakers and learners, especially considering the recent circumstances brought about by the COVID-19 pandemic. In 2007, Chafe classified the Seneca language as Stage 7 out of 8 according to Fishman's 1991 'Graded Intergenerational Disruption Scale' for measuring language endangerment (Borgia-Askey, 2010). This classification means that 'most speakers are beyond childbearing age, and a language in that position is seriously endangered' (Borgia-Askey, 2010). While the terminology of language endangerment — as opposed to language oppression — will be discussed further in later Sections of this study, the purpose of this classification is to show the urgency of language revitalisation programs.

Despite this pessimistic observation, there is hope within the Seneca community for the language as well as the cultural traditions and teachings that accompany it, to be practiced and passed on in the present and into the future. Various language revitalisation efforts are ongoing throughout Seneca territories with financial support from the Seneca community. Through the website for the Seneca Nation Language Department, there are community classes offered for both beginner and advanced students as well as a wide variety of language-learning resources publicly available (Bowen, 2020). In addition, the Faithkeepers Montessori School Seneca Language Nest provides a full-time, comprehensive learning environment for young children from three to twelve years of age to mindfully engage with traditional Seneca teachings and express themselves in the Seneca language (Faithkeepers, 2021). The school also offers immersion programs for dedicated adult learners on the Allegany territory. Other publicly available resources include the Learning the Seneca Indian Language website assembled by Dr. Quinata Delgado with funding from Humanities New York (Delgado N.d.). Yet, the most publicly accessible and visible indicator of Seneca language revitalisation are the bilingual Seneca-English road signs on the Allegany and Cattaraugus territories.

In the context of Seneca language revitalisation, the issue of bilingual signage becomes particularly worthy of study. While the exact dates of the construction of bilingual signage on Seneca territory may not be publicly available, their support is officially recognised in the federal Native American Tourism and Improving Visitor Experience (NATIVE) act of 2016 (Public Law No: 114–221) introduced by Senator Brian Schatz of Hawai'i (114 Cong. Rec., 2016). Although the act did not include financial support for the bilingual signs, it acknowledged their significance and increased awareness of bilingual signage on indigenous territories within the United States. In an online news story from the same year, bilingual signs in various Iroquoian languages across New York State are acknowledged as public monuments of indigenous cultural heritage (Figura, 2016). In the present study, I examine the impact of bilingual signage and focus primarily on community efforts at Seneca language revitalisation.

1.3 Theoretical Framework

It was Michael Goodhart's 2008 work, 'Human Rights and Global Democracy' that proposed that human rights represent normative claims, or reflections of how the state of affairs *should* be (Goodhart, 2008). Therefore, a certain partisan point of view of human rights is advanced that adheres to liberties and values at the individual level rather than a national or even global scale. While this conceptualization may be more sensitive in terms of cultural relativism in allowing all individuals to determine their own perspective on the rights of humans based on their own value system, this system is in contrast with the idea that certain human rights are universal and guaranteed.

In support of the Universal Declaration of Human Rights, Karl Vašák distinguished three broad categories of human rights norms in his 1977 commentary. Vašák divided these norms into 'civil-political' which encompass physical and civil securities such as freedom of religion and freedom from enslavement, 'socio-economic', which include social and economic needs like the right to fair wages and adequate shelter, and 'collective-developmental', which involve the rights of minority groups in matters such as self-determination, the use of heritage languages, and more (Vašák, 1977). In terms of language revitalisation, this third and final human rights norm is essential. While language has considerable cultural significance for identity and as a framework through which speakers interact with the world, this conceptualization of language use as a human right and a developmental norm is especially significant for endangered and oppressed languages.

Clarifying Vašák's view that language rights represent a distinct category of human rights, the United Nations Educational, Scientific, and Cultural Organization published the viewpoint of Fernand de Varennes that language rights are an integral part of human rights (De Varennes, 2001). De Varennes explains the imprecise nature of conceptualizing language rights as a separate entity from human rights: 'language rights of minorities are an integral part of well-established, basic human rights widely recognised in international law' (De Varennes, 2001). Since language rights are inherent to human rights which are already included in international law, any additional recognition of specific minority language rights would be superfluous and suggest language rights are not guaranteed to all humans. Rather than advancing the cause of heritage language preservation, De Varennes argues that the exclusion of minority language rights from a general conceptualization of human rights serves to deny the inherent rights of minorities.

The terminology used to describe languages with dwindling speech communities has been shifting in recent years to more accurately encapsulate the circumstances and varying reasons for this language shift. The current global language crisis examined in Gerald Roche's 2020 work for UNESCO raises awareness that 'at least half of the languages used today will most likely no longer be used by the end of the century' (Roche, 2020). This crisis is not entirely due to natural language evolution or shift; rather it is inherently tied to language rights, and more broadly, collective-developmental human rights. While dominant languages have historical connections to social and economic mobility, the speakers of under-represented and under studied languages have faced many challenges from outright persecution to stigmatization, which have consistently been accompanied by inequitable allocation of resources, respect, power, and more (Roche, 2020). Since these languages are under-represented in the population at large, governmental institutions including the education and justice systems have also played significant roles in historically disadvantaging speakers of minority languages.

Yet, the terminology of 'language endangerment' has failed to convey these dire conditions and substitutes more nuanced terms in favour of one borrowed from evolutionary biology. The use of 'endangerment' to describe language shift implies that other languages have more speakers and more

prestige due to their superior 'fit' for their environment, reminiscent of Darwinian 'survival of the fittest.' In reality, those attributes are bestowed on dominant languages by institutions of authority that have historically oppressed the speakers of what may be considered 'endangered' languages. While terms like 'language oppression' may be a closer approximation of the reasons that the Seneca language is no longer widely spoken, some may find fault in this description since the revitalisation programs for Seneca are currently thriving and strongly supported. In this study, where possible, more descriptive phrases will be used that merely point out the size of the speech communities rather than allude to the various causes of or factors relating to language endangerment, oppression, and death.

In an attempt to clarify a potentially fraught linguistic situation, language revitalisation refers to efforts through which the learning and use of heritage languages with dwindling speaker populations are expanded and passed on to future generations of a particular culture. This contrasts Wesley Leonard's 2019 framework for language reclamation, which 'describe[s] and theorize[s] efforts by Indigenous communities to claim their right to speak their heritage languages and to set associated goals in response to community needs and perspectives' (Leonard, 2019). While revitalisation focuses on the continual expansion of speech communities, reclamation focuses on the right of minority groups to use their traditional language. Leonard draws attention specifically to the goals of the speech communities for whom the language represents part of their heritage and culture, rather than academics who seek to obtain language documentation for intellectual pursuits that may be inaccessible to the traditional speakers of the language in question[37]. Although the focus of this study deals primarily with the sociology of the Seneca bilingual road signs as part of a larger, more complex linguistic situation, there is an element of reclamation occurring through the bilingual signage as markers claiming the right of the Seneca to use their heritage language on their territories. These public displays of the Seneca language serve as indicators that the indigenous population is still there and retains their sovereignty evident through the use of their heritage language on their own land.

Furthermore, in recognition of the diversity of the indigenous populations in North America, I acknowledge that these group names — American Indian, Native American, First Nation, indigenous, etc. — are not universally accepted. In order to be sensitive to these issues, I will follow the customary terminology in the field of linguistics and the most widely accepted terminology. Since terms such as 'Indian' have been reclaimed by certain groups including the Seneca Nation of Indians, that term will be used when referring to official names or quotations.

2 Methodology

2.1 Oral Histories

The role of the bilingual signs in western New York first became apparent to me through a local news story by David Figura in 2016 where Jessica Robinson, the former deputy director of the Seneca Nation of Indians Department of Transportation, is quoted explaining the multifaceted value of the signs. She notes that 'language is integral to Native culture, history, and future. Signage is one facet or tool in preserving language as well as to educate the public and acknowledge [the] Tribe's connection to the land as well as

[37] In recognition of the goals of minority speech communities, Ladefoged's 1992 opinion-piece in *Language* acknowledges that linguists and speech communities may have contradictory views concerning the revitalisation of traditional languages (Ladefoged, 1992). However, that discussion is beyond the scope of this study and has limited relevance to Seneca language revitalisation.

their sovereignty as nations across the country' (Figura, 2016). Based on the information provided in the news story, I contacted the named representative who referred me to her former supervisor, Jody Clark, who was said to have been deeply involved in efforts to introduce bilingual signage to Seneca territory. In actuality, she is the former director of the Seneca Nation of Indians Department of Transportation and a member of several national committees promoting indigenous bilingual signage across the country. She was generous enough to provide an oral history of the process of securing signage in the Seneca language as well as other matters of Seneca history, language, and culture. In a phone call that lasted several hours, she spoke at length about her experiences on national committees advocating for indigenous language representation across the United States. She also supplied me with the contact information for other language advocates to consult for more information, for which I am extremely grateful.

Based on the information provided by Jody Clark, I reached out to Seneca community organiser, Flip White, who is currently the Project Director for the Seneca Language Nest educational program for young Seneca language learners. The Seneca Faithkeepers School was initially started in 1998 by Lehman 'Dar' Dowdy and his wife, Sandy Dowdy, on the Allegany Reservation in order to combat the disappearance of the Seneca language and the accompanying loss of cultural heritage (Herbeck, 2004). In several extensive phone calls, I interviewed Flip White about the Seneca language revitalisation efforts for young speakers and their families on Ohi:yo', the Allegany territory. He was very encouraging of the ongoing revitalisation and shared with me that the program has recently received a federal grant to support the expansion of its services to a wider range of age groups and other modalities of language learning. The federal funds also support the development of a Seneca language-learning app *Memrise* that would make the information more accessible to parents of children at the Language Nest school, other interested parties, and the Seneca community in general. He emphasized the stark contrast between the past residential schools in the area that punished Seneca speakers for the use of their heritage language, and the current efforts in the community to bring the Seneca language back through a kind and loving communal process. His candid expression of language ideologies and attitudes towards the various forms of Seneca language revitalisation has been invaluable to this research and is also greatly appreciated.

Since both the Seneca Nation and New York State are considered sovereign entities within overlapping territory, the question of who has jurisdiction over the bilingual signs on state roads crossing the Seneca territories is complex. While the public outreach for the New York State Department of Transportation (NYSDOT) declined to comment on the topic of the Seneca bilingual signs, an Assistant Regional Design Engineer for Region 5 of the NYSDOT who was involved in the process, Sanjay Singh, provided further information from the state government on the collaborative process of introducing the bilingual signs to Seneca territory and state roads. Although the 'decision to include bilingual signs and other SNI [Seneca Nation of Indians] cultural enhancements resulted from collaborative consultations between NYSDOT and SNI officials,' the NYSDOT takes responsibility for future maintenance of signage whereas the SNI is responsible for the maintenance of artwork[38]. He stated that funding for the projects on Seneca land are provided in part by Federal Aid and the Nation itself. Singh also shared references for cultural artwork displayed on the territories as well as an image of a bilingual sign commemorating Chief Cornplanter Memorial Bridge on the Allegany territory, which can be seen in Figure 2 below[39].

[38] This information is taken directly from personal communication via email with Sanjay Singh on March 22 and 29, 2021.
[39] Chief Cornplanter was a notable Seneca warrior and leader who advocated for Iroquoian neutrality during the American Revolution and later acted as a mediator between the Seneca and the United States' government. Part of the territory near the Allegany River that he received in 1795 as a land grant has since been flooded by the Kinzua Dam Project ('Cornplanter' 2021).

Figure 2*: Bilingual sign on the Allegany territory shared by Sanjay Singh via email on March 22, 2021.*

2.2 Fieldwork

While the bilingual road signs may not be the most well-known or longest-lasting aspect of Seneca language revitalisation, they are the most publicly visible. It is this accessibility to the public that appealed to me as a resident of Tonawanda. Through public records, maps, blogs, and other information, I located several bilingual signs on the Cattaraugus territory in the Irving area. I photographed bilingual signs for Cattaraugus Creek, one of which appears on each side of the bridge over the creek, seen in Figure 3 below. In addition, I took a photograph of the Seneca language on a highway overpass entering the Cattaraugus territory which can be seen below in Figure 4. I also noted several signs in the Seneca language marking some local businesses and residences, though I was asked not to photograph or share images of the private signs.

Figure 3: *My photograph of a bilingual sign on the Cattaraugus territory on January 10, 2021.*

Figure 4: *My photograph of a highway overpass entering the Cattaraugus territory on January 10, 2021.*

Examination and comparison of the Seneca signs from various areas reveal differences in the orthography used to represent the Seneca language. In addition to varying usage of diacritics symbolizing nasal vowels, glottal stops, and vowel accents, the capitalization also changes based on the location and information conveyed, which is evident in the publicly posted images in Figures 5 and 6 below. Compared to American English road signs which historically used exclusively capital letters up until a recent change by the Federal Highway Administration expected to be implemented across the country by 2015 (Copeland, 2010), the Seneca orthography only uses capital letters at the beginning of certain words and never uses all-caps even for road signs. These orthographic details reflect regional differences in the language and the variety of written expression that exists for the Seneca language. Although Seneca speakers had historically recorded their laws and history through wampum strings consisting of various patterns of shells and sinew, a Presbyterian missionary developed a Roman-based alphabet and typographic system for the Seneca language in the late 19th century (Borgia-Askey, 2010, pp. 38–55). However, this written record for the language has not been universally accepted by all current Seneca speakers, leading to the diversity in orthography seen today. The attention to detail visible in the bilingual signs shows the consideration and care that went into the translation processes for the Seneca signs with input from language experts in each region, which cannot be said for all bilingual signs that feature minority languages.

Figure 5: *Photograph by Roger Green of a bilingual sign entering the Allegany territory.*

Figure 6: *Photograph by Todd Smith of a bilingual sign on the Allegany territory.*

3 Conclusions

3.1 Impacts of Bilingual Signage

While I initially hypothesized that the bilingual Seneca signs supported language learning efforts by providing public displays of the heritage language, the interviews that I have had with members of the Seneca community have demonstrated that there are a variety of linguistic, cultural, and economic impacts. More than a mere acknowledgment of Seneca sovereignty on their land, the signs represent a physical, publicly accessible, and visible marker of Seneca territory that has numerous intangible benefits. Speaking as a non-Native resident of Tonawanda, it can be difficult to know the exact boundaries of the current Seneca territory or ancestral homeland. The signs provide a marker in the public sphere for community and non-community members to increase awareness of the local indigenous population.

As a result of the visible markers of Seneca territory, Jody Clark noted that travellers through the area make more deliberate efforts to support the local economy when they are made aware of the indigenous presence. This contributes to significant stimulation of the tribal economy since the area relies in part on funds from tourism. Even though bilingual road signs may not be typical examples of tourist attractions, their public presence leads to intentional support of other tourism initiatives and local businesses.

Since the Seneca bilingual signage has come about largely through the efforts and fundraising of the Seneca community and community organisers such as Jody Clark, they are a visible source of pride for the local community. For a language that has been systematically stigmatized, the public display of Onöndowa'ga:' Gawë:nö' represents the cultural heritage of the area and its indigenous inhabitants as well as a visible symbol of the ongoing language revitalisation. The signs themselves, of course, do not create new speakers or further language revitalisation in the traditional sense, but they do offer support for existing representations of the Seneca language as well as increasing awareness of the continued presence of the language. Considering the views of language rights discussed previously, the usage of the Seneca heritage language in the public sphere shows how collective-developmental human rights are actually realised in modern times. Although the NYSDOT maintains that the signs have been a collaborative effort between the state and tribal governments, the experiences of advocates for indigenous signage reveal how non-Native governmental forces have tried to limit the kinds of signage that are allowed to display indigenous languages. Since road signs can be divided into informative (place names, hospital, parking, etc.), regulatory (one-way, yield, stop, etc.), and warning (merging lane, roadway conditions, road work, etc.) signs, indigenous language advocates have met resistance to any forms besides informational signs, which represents a challenge to collective-developmental rights in the public sphere. In spite of this challenge, there are still regulatory signs available in the Seneca language, such as the bilingual stop sign on the Allegany Reservation shown in Figure 7 below with a capital letter only at the beginning of the word as well as markings for nasal and accented vowels and glottal stops.

Figure 7: *Photograph of Seneca stop sign on Allegany territory.*

3.2 Comparison with Other Bilingual Signage

There are a variety of other language revitalisation programs occurring around the world which are too numerous and diverse for the scope of this study. Including but not limited to the other nations of the Haudenosaunee or Iroquois Confederacy, the Potawatomi Nation, the Wampanoag, the Diné or Navajo Nation, the Cherokee Nation, the Three Affiliated Tribes of North Dakota, the Pawnee Nation, Native Hawaiians, and Maoris (i.e., indigenous New Zealanders), the depth and breadth of indigenous language revitalisation is on the rise on a global scale (Borgia-Askey, 2010). While further discussion of these revitalisation programs must be relegated to future studies, it is important to note that only certain groups have attained signage in their heritage language regardless of their interest in having bilingual signage.

Historically, the United States does not have an official language, though the English Only Movement has gained significant attention as a topic of recent debate. In spite of constitutional issues regarding freedom of expression, 'to date 18 states have enacted laws designating English as the official state language' (Padilla et al., 1991). In addition to xenophobic motivations for such a policy, the intended effect of a national language policy infringes on the human rights of the many sovereign indigenous nations within the United States. For the Seneca community to display their heritage language in the public sphere as well as on the signs of private businesses indicates their continued presence in the region and as such it represents a challenge to those who would prefer that they communicate only in English or cater to those who are unfamiliar with the local language. Bilingual public signage acts as a literal and figurative signpost, a tangible and observable focal point for language rights and conceivably for language tension as well.

As a public record of the community reactions to the Seneca bilingual signs, the respect for these signs is demonstrated through a total lack of vandalism or theft. The same cannot be said for monolingual signs on the territories as well as bilingual signs that feature minority languages elsewhere. Through the past fieldwork of my thesis advisor, Dr. Brian Joseph, I have been made aware of the bilingual Albanian-Greek signs in Southern Albania where signs in the minority language of the region, Greek, have been defaced, as seen in Figure 8 below. Although the linguistic circumstances of Greek in Southern Albania

and Seneca on the Allegany and Cattaraugus territories in New York State differ in many ways, the contrast between the community reactions to the bilingual signs is representative of drastically different language ideologies concerning the use of minority languages in the public sphere.

Figure 8: *Photograph of a vandalized bilingual sign in Southern Albania reproduced with permission from Dr Joseph.*

Additionally, the regional differences in the Seneca language reflected in the bilingual signs show a level of consideration and mindful engagement with the Seneca community that has not been common across all bilingual signage. A variety of publicly accessible bilingual signs show evidence of translation errors or entirely inappropriate attempts to represent the minority language of the region. Specifically, the English-Welsh bilingual sign shown in Figure 9 below has gained popularity recently for the improper translation into Welsh displayed. Rather than a Welsh translation of, 'No entry for heavy goods vehicles. Residential site only,' the Welsh exhibited on the sign reads 'I am not in the office at the moment. Send any work to be translated.'. (Cycling Wales, 1996).

Figure 9: *Photograph of an improperly translated Welsh bilingual sign in Swansea.*

While there is an unfortunate lack of quantitative evidence to directly support the impact of Seneca bilingual signs on the revitalisation of the language, there is a clear symbolic value to the signs for Seneca community members in addition to many intangible benefits. Since the language is used in specific contexts such as the bilingual road signs and the Seneca Faithkeepers School, one could compare the linguistic situation to that of liturgical Latin used exclusively in Catholic church services. Just as the Seneca Faithkeepers School initially promoted the use of Seneca for the Longhouse religious ceremonies (Herbeck, 2004), the Catholic Church utilized Latin exclusively for religious services for many years. When the Second Vatican Council officially allowed masses to be conducted in vernacular languages in the 1960s, the move was met with criticism from some who mourned the loss of liturgical Latin to express their faith (Kennedy, 2011). If the language did not have significant value in that context, then its loss would not have incited attention, criticism, and other commentary. Although the Seneca community's intentions are to expand the use of their heritage language beyond ceremonial and educational settings, which differs notably from the intentions of those worshipping in Latin, the Seneca language as it is used in specific contexts has a clear value even if it may not be apparent to those outside of the community.

3.3 Further Areas for Research

I am extremely grateful to the language advocates and community organisers who generously shared their expertise and experiences with me, though if time and circumstances permitted, I had hoped to interview other active Seneca community members about the bilingual signs. A community survey of attitudes towards the Seneca bilingual signage would also help elucidate the language ideologies of the area concerning the revitalisation of the heritage language. For future studies, I would like to focus more on the varying language revitalisation programs occurring throughout the world and the similarities/differences in their goals, methodology, and outcomes.

4 References

114 Cong. Rec. (2016). Public Law 114–221. Retrieved 06/04/2021, from https://www.congress.gov/114/plaws/publ221/PLAW-114publ221.pdf.

Borgia-Askey, M. (2010). An Overview of Language Preservation at Ohi:yo', the Seneca Allegany Territory. Ph.D. dissertation, Indiana University of Pennsylvania.

Bowen, J. (2020). Onöndowa'ga:' Gawë:nö'. Retrieved from 03/03/2021 from https://senecalanguage.com/community-classes.

Copeland, L. (2010). ALL CAPS? Not OK on road signs, federal government says. *USA Today.* Retrieved 02/05/2021, from https://usatoday30.usatoday.com/news/nation/2010-10-21-road-signs-all-caps lowercase_N.htm.

'Cornplanter'. (2021). Encyclopedia of World Biography. *Encyclopedia.com.* Retrieved 03/04/2021, from https://www.encyclopedia.com/history/encyclopedias-almanacs-transcripts-and maps/cornplanter

Cycling Wales. (1996). Lost in Translation.Retrieved 06/04/2021, from https://cyclingwales.co.uk/badlytranslated.html.

Delgado, Q. (N.d.) Learning the Seneca Indian Language. Retrieved 03/03/2021, from https://www.learningthesenecalanguage.com/.

De Varennes, F. 2001. Language Rights as an Integral Part of Human Rights. *IJMS: International Journal on Multicultural Societies*, 3(1), 15–25.

Faithkeepers Montessori School. (2021). Faithkeepers School Montessori Seneca Language Nest – Ganöhsesge:kha:' Hënödeyë:stha' Onödowa'ga:' Gawë:nö' Wadehsäyë'. Retrieved 04/04/2021, from https://faithkeepermontessori.com/.

Figura, D. (2016). Bilingual Road Signs: Growing trend on state roads crossing Indian lands. *Advance Local Media LLC.* Retrieved 04/04/2021, from https://www.newyorkupstate.com/news/2016/10/bilingual_road_signs_growing_trend_on _state_roads_crossing_indian_land.html

Goodhart, M. (2008). Human Rights and Global Democracy. *Carnegie Council for Ethics in International Affairs*, 22(4): 395–420.

Herbeck, D. (2004). Seneca Faithkeepers School Tries to Keep Alive the Tribe's Traditional Ways, Language. *Canku Ota 114.* Retrieved 03/04/2021, from https://web.archive.org/web/20150402120836/http://www.turtletrack.org/Issues04/Co060 52004/CO_06052004_Seneca_Faithkeepers.htm.

Kennedy, P. (2011). Christianity: An Introduction. London: I.B. Tauris.

Ladefoged, P. (1992). Another view of endangered languages. *Language*, 68(4): 809–811.

Leonard, W. (2019). Musings on Native American Language Reclamation and Sociolinguistics. *Social Science Research Council.* Retrieved 03/03/2021, from https://items.ssrc.org/sociolinguistic-frontiers/musings on-native-american-language-reclamation-and sociolinguistics/?_cldee=am9zZXBoLjFAb3N1LmVkdQ%3d%3d&recipientid=contact 0a6f7b0eac02dd11bef0001cc477ec70-4da6265f8f5746c6975e8530c0406a7b&esid=583f41dd-dade-e911-a97f-000d3a34afa9.

Mithun, M. (2012). Challenges and Benefits of Contact among Relatives: Morphological Copying. *Journal of Language Contact*, 6(2):243–270.

Mithun, M. (2006). Iroquoian Languages. *Encyclopedia of Language and Linguistics*, 6:31–34.

Murray, A. (2015). Reclaiming Traditional Seneca Culture. *The Allegheny Front* http://archive.alleghenyfront.org/story/reclaiming-traditional-seneca-culture.html Accessed on March 30, 2021.

Padilla, A, et al. (1991). The English-Only Movement Myths, Reality, and Implications for Psychology. *American Psychological Association.* Retrieved 03/04/2021, from https://www.apa.org/pi/oema/resources/english-only.

Roche, G. (2020). Towards a New Language of the Global Language Crisis. *UNESCO.* Retrieved 03/03/2021, from https://catedra-unesco.espais.iec.cat/en/2020/11/23/50-towards-a-new-language-of-the global-language-crisis/.

Vašák, K. (1977). Human Rights: A Thirty-Year Struggle: the Sustained Efforts to give Force of law to the Universal Declaration of Human Rights. *UNESCO Courier* 11:29–32.

Wallace, A. (1969). *The Death and Rebirth of the Seneca*. Alfred A. Knopf, Inc.

Links to publicly posted photographs:

Figure 5: https://www.rogerogreen.com/2016/08/08/western-new-york-seneca-nation

Figure 6: https://www.pinterest.com/pin/432275264234751490/

Figure 7:
https://upload.wikimedia.org/wikipedia/commons/0/0b/Sáë%27he%27t_stop_sign_in_Seneca.jpg

Figure 9: https://cyclingwales.co.uk/badlytranslated.html

Voice Onset Time in English Voiceless Initial Stops in Long Read and Spontaneous Monologue Speech of Thai Students with English as a Second Language

Chanakan Wittayasakpan
Chulalongkorn University

Abstract. Proximity of L1 and L2 VOT (Voice Onset Time) values illustrates that L1 stop categories are used to acquire L2 ones. While a number of studies have found that VOT is very sensitive to many factors, how speech time and VOT correlate is still understudied. This study tests whether L1 transfer in terms of VOT would enhance as speech time elapses. Six university students with Thai as L1 and English as L2 were selected to produce long spontaneous monologue speech and long read monologue speech produced. VOT values in English initial voiceless stops were then segmented and analysed using the mix-effects model. The result reveals that raw VOT in spontaneous speech is significantly shorter than in read speech and no correlation between adjusted VOT and elapsed speech time is found. does not correlate with elapsed speech time. Implications of the result are discussed in terms of stylistic variation and second language acquisition.

Keywords: voice onset time; second language acquisition; L1 transfer; Thai; English

---

1 Introduction

Voice onset time, hence shortened as VOT, is the temporal lag of vocal fold vibration following a stop consonant release, or, in other words, the interval between the burst of a stop consonant and the onset of voicing. VOT serves as a phonetic cue to distinguish voicing categories and is generally classified into three ranges: voice lead, short lag, and long lag. Across languages, VOT values could differ even in the same category, and this is the case for English and Thai. In both languages, stops are produced in three places of articulation: bilabial, alveolar, and velar. However, English has two types of stops: voiced and voiceless, whereas Thai has three types: voiced, voiceless unaspirated, and voiceless aspirated. English voiceless stops and Thai voiceless aspirated stops fall into the long lag VOT category (Kessinger & Blumstein, 1997) but have different ranges of VOT values. VOT values in monolingual English voiceless stops range from minimally greater than 30 milliseconds to approximately 90 milliseconds (Lisker & Abramson, 1964). On the other hand, those in monolingual Thai voiceless aspirated stops range from approximately 40 milliseconds to 120 milliseconds (Lisker & Abramson, 1964; Shimizu, 1996; Shimizu, 2011). In summary, despite the overlap, Thai voiceless aspirated stops have been found to have higher range as well as higher mean VOT values than English voiceless stops.

VOT in stops in English produced by Thai ESL speakers was examined by Shimizu (2011). In the study, Thai ESL participants were asked to produce stops in their native tongue, Thai, and their L2, English. As a result, English voiceless stops produced by the participants have VOT values close to Thai voiceless aspirated stops. This illustrates that Thai speakers use their native stop categories to acquire English ones.

Whether this L1 transfer will enhance with time is thus worth exploring for it could lead to better understandings of L1 transfer and bilingual phonetic realization.

Besides L1, plenty of studies (e.g., Lisker & Abramson, 1964; Yao, 2009; Smith et al., 2015) have illustrated that VOT is sensitive to many factors, such as speech rate, place of articulation, follow vowel height, phonetic context, utterance position, and speaker styles. One of the understudied factors that could affect VOT is speech time. The reason time is an interesting factor is because speech variability tends to intensify during a period of long speech, resulting in reduced accuracy of speech recognition (Frankish et al., 1992). Thus, a question arises if VOT duration values will increase or decrease as a form of this variability.

So far studies regarding the correlation of VOT and time have mostly dealt with code switching. Balukas and Koops (2015) focused on English VOT and Spanish VOT of Mexican bilinguals in spontaneous code-switching and found that English VOT values rose in the first few seconds after code-switching and then stabilised. However, the same effect was not found in Spanish stops which were the first language of the participants. A similar result was found in the study by Piccinini and Arvaniti (2015). As time progressed from a code-switch point, Spanish VOT values remained steady whereas English VOT values became higher. In such studies, the focus was on dyadic speech and time elapsed from code-switching apparently lasted up to 30 seconds. Though there are studies into VOT in longer speech without code-switching, they tend to focus on analysing mean VOT values (e.g., Hillman & Gilbert, 1977; Grosjean & Miller, 1994) and not its correlation with time.

When examining the correlation between VOT and time, it is worth including different speech styles. Previous studies have found that VOT values in isolated words tend to be higher than those in words read in sentences (e.g., Baran et al., 1977; Chodroff & Wilson, 2017). This is in line with the study by Nakamura et al. (2008), which found that, comparing with read speech, the spectral distribution was significantly reduced, and phonemes varied more in spontaneous speech. Considering that, a greater degree of variability in spontaneous speech could result in more substantial change than or even a different trend from that of read speech.

The present study has two aims. The first aim is to test if L1 transfer in terms of voice onset time varies with time in long monologue speech. L1 transfer here refers to VOT values in English voiceless stops produced by Thais, which are close to those of Thai voiceless aspirated stops. The second aim is to test if the presence or absence of such variation is the same in different running speech styles, that is, in spontaneous speech and in read speech. Thus, I have two hypotheses. First, VOT values in should significantly increase as speech time elapses because speech production should become more accented, i.e., more similar to Thai, resulting in higher VOT values. Second, VOT values in spontaneous speech should be lower and vary more greatly than those in read speech, resulting in a steeper slope.

2 Methods

2.1 Participants

Six university students, aging from 18 years old to 23 years old, participated in this study. All participants reported that they had Thai as their first language and English as their second language and used only Thai at home. All had studied in an international school or English programme during their primary education and none of them had lived outside Thailand for more than six consecutive months. To ensure the ability

to fulfil the tasks, all of the selected participants also had been trained to debate and were capable of making a 7-minute speech without interruption.

2.2 Tasks and Recording Procedures

For spontaneous speech, a debate in Asian parliamentary format was hosted on Mixidea.org, a website for online debating. Each team consisted of three speakers and each speaker was assigned to give a 7-minute speech. The motion was released 30 minutes before the debate, so each team had half an hour to prepare its case. This was to ensure that speech was spontaneous as the given preparation time would not suffice to write an entire script and would compel all speakers to improvise. The motion on the floor was 'This House would punish natural or legal persons who are accused of cultural appropriation'. The motion and the wordings were chosen with the aim to ensure a sufficient number and skewed distribution of the stops throughout each speaker's speech since the discussion must revolve around 'culture' and 'punish'. Interruption, point of information, and clapping were not permitted during speech so as to ensure continuity of long speech as well as to minimise noise. Recordings were conducted in two means. The first means was the recording function coming with mixidea.com. Each participant was also required to co-record using their own phone. The sound to be analysed was chosen based on minimum noise and minimum missing signals. Eventually, two files from mixidea.com and four files from participants' phones were used.

For read speech, each participant was asked to read the article 'Kept Women' by Marina Benjamin (2013). The article was excerpted and rearranged so as to evenly distribute the stops and make the speech last approximately seven minutes, which was the expected time of spontaneous speech. Practice before recording was allowed and participants were asked to finish the entire speech in one recording so as to ensure the continuity of speech. All read speech files were recorded using participants' phones.

2.3 Analysis

Spontaneous speech was manually transcribed. All speech was then auto-segmented with WebMAUS Basic service (Kisler et al., 2017) provided by Bavarian Archive for Speech Signals (BAS). VOT in initial voiceless stops was then manually segmented in Praat (Boersma & Weenink, 2017) and measured in milliseconds. Each VOT value was coded along with the time at which the stop was produced, its place of articulation, its following vowel duration, its following vowel height, the participant who produced the token, and the word containing the stop. The speech time here is defined as the time from the onset of the first word. Based on the transcription, approximately 1401 were expected, approximately 903 tokens from spontaneous speech and 498 tokens from read speech. Mean overall time is 7.21 minutes for spontaneous speech and 7.36 minutes for read speech.

2.4 Exclusion Criteria

The analysis excluded tokens which underwent problems. The first is stops without clear burst signals. This was most likely caused by technical problems, such as movement of participants' headphones/microphones or background noise in participants' settings. The second is stops which underwent deletion. As a consequence of these problems, identifying the point at which a release occurred became inaccurate, if not impossible, since no clear burst signal could be detected. The third and the most common is stops which

underwent affrication or frication. Participants, especially P1 and P3, tended to produce /t/ as fricative and /p/ as affricate. Apparently, this was due to regional dialects and/or free variants of the participants. Though VOT values could be measured, they were excluded since they were not from stops and could create noise for the analysis. The fourth is stops whose following vowels underwent devoicing, as it prevented accurate identification of the onset of vowels and resulted in drastically higher VOT values. And the last is stops which underwent voicing, as they would yield negative VOT and fall out of the scope of this study. After the step, the final number of tokens is 890 tokens, 523 tokens from spontaneous speech and 367 tokens from read speech.

3 Results

Firstly, let us summarize the distribution of raw VOT values. Figure 1 illustrates central tendencies and the variability of VOT values from both spontaneous and read speech. The symbol '+' signifies the mean value. The VOT means and standard deviations for both speech styles are given in Table 1.

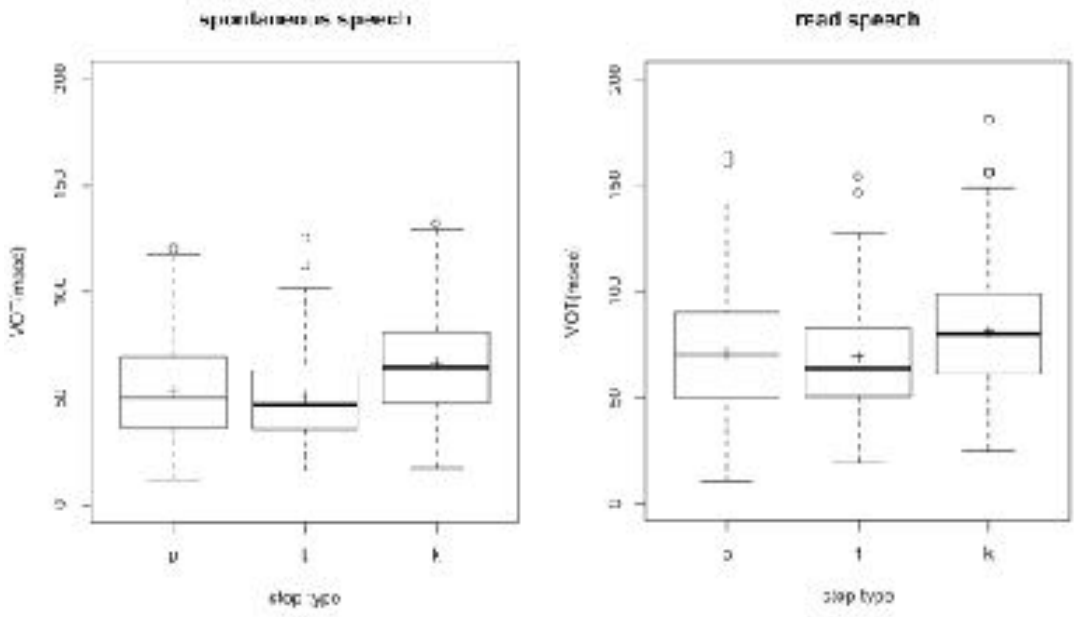


Figure 1: *Spontaneous and read speech distributions of raw VOT values.*

Table 1: *VOT means (ms), standard deviations (ms), amount of tokens.*

	Spontaneous speech			Read speech		
	mean	SD		mean	SD	
/p/	53.705	24.301	n = 116	70.601	31.952	n = 119
/t/	50.469	21.379	n = 113	68.888	26.590	n = 85
/k/	66.541	23.754	n = 294	81.080	28.243	n = 163
grand mean	56.905	23.144	n = 523	73.523	28.928	n = 367

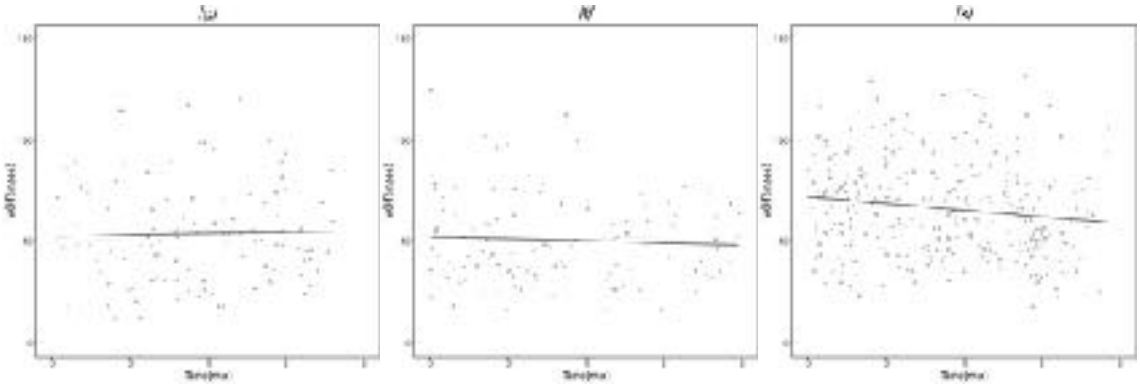


Figure 2: *Spontaneous speech VOT values by time.*

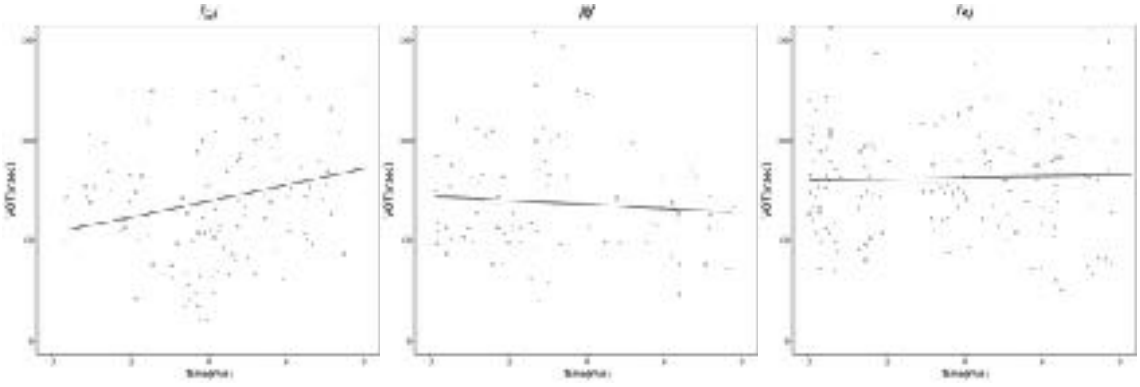


Figure 3: *Read speech VOT values by time.*

From Figure 1 and Table 1, in all stop types VOT values in read speech are on average 16.6 milliseconds higher than those in spontaneous speech. The standard deviations in read speech are also higher, indicating that the range of VOT values in the style is both higher and wider. In both styles, /p/ has slightly higher

VOT values than /t/, while /k/ has the highest VOT values. A major overlap between /p/ and /t/ could be seen in both speech styles. The mean VOT values and standard deviations in read speech are close to those of English voiceless stops produced by Thai speakers in Shimizu (2011).

Next, let us turn to the relationship between raw VOT and values speech time. As mentioned, if English voiceless stops became more similar to Thai voiceless aspirated stops, VOT values would be expected to increase. In the scatterplots in Figure 2 and Figure 3, linear regression lines are imposed to illustrate trends. In spontaneous speech, the /p/ line shows a slightly upward trend whereas the /t/ line shows a slightly downward trend. However, the slopes are not steep, indicating that neither /p/ nor /t/ undergoes any significant change. Only one stop type, /k/, shows a visible decrease in VOT values, signifying that the temporal gap between the burst of /k/ and the onset of voicing becomes shorter as time progresses in spontaneous speech.

The trends are distinct in read speech. While /k/ is the only stop that undergoes a clear change in spontaneous speech, it is the most stagnant stop in read speech. In contrast, /p/ here has an upward trend with a steeper slope than that in spontaneous speech, while /t/ still has the same slightly downward trend. It should be reasonable to conclude here that no systemic correlation between raw VOT values and speech time is found in either speech style.

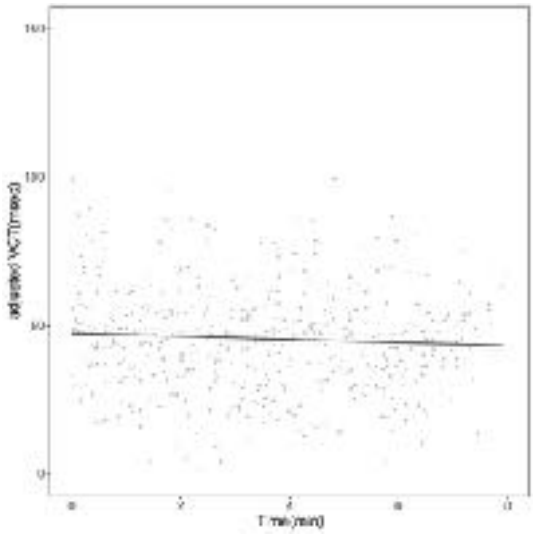


Figure 4: *Regression line showing slight decrease in spontaneous speech (slope=-0.52).*

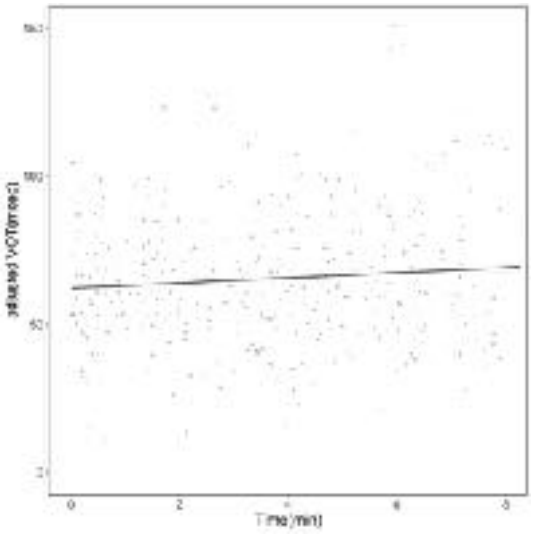


Figure 5: *Regression line showing slight increase in read speech (slope=0.85).*

As stops in running speech are greatly influenced by their environment (Yao, 2009; Smith et al., 2015), a linear mixed-effect model using random intercepts was constructed using lme4 package (Bates et al., 2015) in the statistical software R. This is to minimise and control variables which may affect VOT values. The model was adapted from Balukas and Koops (2015) by using the same coded fixed and random effects but without a logistic transformation. This is due to the fact that their study focused on code switching and the phonetic convergence was present only in the earlier part of speech, so the relation between VOT durations and the time from a code-switch point was non-linear. On the other hand, the present study aims to establish a linear relationship between VOT durations and the time from the onset of the first word since the effect should persist throughout speech. On that account, a logistic transformation is not included. Here, coded fixed effects are a) speech time and b) following vowel duration. The following vowel duration is used as an indirect measurement of speech rate as well as stress. Coded random effects are a) place of articulation, b) vowel height, c) word containing the stop, and d) participant who produced the token.

Figure 4 and Figure 5 show the trends of VOT values after mixed effects are calculated. All the grey dots in the scatterplots have been intercept-adjusted. In spontaneous speech, VOT values slightly drop as time elapses. In read speech, VOT increases slightly but with a greater slope than that from spontaneous speech. Here, VOT values in both speech styles do vary but very much slightly. Each minute that goes by, VOT durations become shorter by 0.52 milliseconds in spontaneous speech and longer by 0.85 milliseconds in read speech. So, the change in neither speech style tends to be greater than seven milliseconds throughout seven minutes of speech time. Recognising that there is much room for VOT durations to extend up to 120 milliseconds, which is the upper bound of Thai voiceless aspirated stops, the resulted change is barely significant. An F-test via Kenward-Roger approximation also affirms that elapsed time is not a statistically significant predictor of VOT in either speech style. (p-value = 0.24 in read speech, 0.20 in spontaneous speech). In conclusion, there is no clear relationship between speech time and adjusted VOT values.

4 Discussion

Both hypotheses fail. VOT values do not significantly increase as speech time elapses. Though lower, VOT values in spontaneous speech do not vary more significantly that those in read speech. Thus, with the present study, it should be reasonable to conclude that Thai ESL speakers tend to produce English initial voiceless stops without the VOT duration values becoming closer to those of Thai voiceless aspirated stops. In other words, they tend to adhere to a narrow range of VOT values throughout long speech in both spontaneous and read speech. Thus, L1 transfer does not vary with time.

Let us begin with discussing L1 transfer. English voiceless stops in read speech in the present study have the mean VOT values and standard deviations rather close to English voiceless stops produced by Thai ESL speakers in Shimizu's (2011) study. The same could not be said for spontaneous speech for Shimizu's study analysed VOT in read isolated words. Though more information is needed for comparison, it would not be inconsistent with the previous study to conclude that Thai ESL speakers produce English voiceless stops in read speech, whether in citation forms or running speech, with VOT values close to those in Thai.

The marked contrast of the raw VOT values between the two styles is in line with previous studies (Baran et al., 1977; Chodroff & Wilson, 2017), which found that VOT in read speech was longer than that in spontaneous speech. It also strongly supports stylistic variation, that is, attention and awareness affect stop articulation. When participants read, they tend to be more aware of their speech, resulting in more articulation rate and thus higher VOT values. In contrast, in spontaneous speech like a debate speech, they tend to be less aware of their production since they have to constantly engage themselves with the content at hand, resulting in less articulation rate and thus lower VOT values. Also, long spontaneous monologue usually contains many linguistic/pragmatic constraints and repetition. Many times, words either are repeated as a means to buy time to think mid-speech or could be predicted based on the syntactic structure and the context. This greater degree of redundancy and less amount of information permit VOT reduction while still maintaining intelligibility of speech (Coker & Umeda, 1975; Baran et al., 1977).

A possible explanation for the stability of VOT values is that even if L1 transfer is present, ESL learners still try to maintain VOT values of English voiceless stops in a particular range so that they will not overlap with those of Thai stops. Basically, it could be considered a way in which a bilingual attempt to keep a set of phonetic properties of stops separated and exclusive to each language. A study into Thai stops in long speech produced by ESL Thai learners is needed to compare and prove the hypothesis.

Though minimal, the difference between the resulting trends, i.e., that VOT values in spontaneous speech tend to decrease while those in read speech tend to increase, is worth discussing. The downward trend in spontaneous speech could stem from vocal fatigue after long continuous speech production. Since the effect of vocal fatigue on stops was minimal (Caraty & Montacié, 2010), the downward trend turned out to be only slight. Regarding the upward trend in read speech, I offer two explanations. First, the upward trend itself could be the result of L1 transfer which was minimised, if not neutralised, by the attempt to separate stop categories for each language. It is apparent in only read speech since the effect was overshadowed by neither cognitive load nor redundancy. Second, the increase could also be a form of compensation. Since participants tended to be more aware of their speech when reading, it could follow that they were aware of fatigue, which should result in less articulation or greater imprecision. Consequently, they tried to compensate it by carefully articulating, thus resulting in higher VOT values than the earlier part of reading in which they did not try to compensate fatigue.

5 Conclusion

The present study examines VOT in long spontaneous speech and long read speech elicited from Thai speakers with English as a second language. The results show that VOT values in spontaneous speech are lower than those in read speech. The marked contrast between VOT values in the two speech styles support stylistic variation and constraints in spontaneous speech. Also, no correlation between VOT values and speech time is found in either speech style. One possible explanation is that a bilingual tries to maintain the phonetic exclusivity for each language. Though insignificant, the downward trend in spontaneous speech could stem from vocal fatigue while the upward trend in read speech could be due to minimised L1 transfer or compensation.

Further studies may use tasks without a topic to obtain spontaneous speech in order to reaffirm the result and may also group participants according to levels of English fluency to test if the result would differ among people from different levels. Since this study includes only English stops, future studies should also examine Thai stops in long speech produced by native Thai speakers to investigate whether the trends would be similar when there is no L1 transfer effect.

6 References

Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, 19(4), 423–443.

Baran, J. A., Laufer, M. Z., & Daniloff, R. (1977). Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics*, 5(4), 339–350.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Benjamin, M. (2013). Kept Women. [Online] Retrieved from *Aeon*: <https://aeon.co/essays/why-do young-rural-women-in-china-become-mistresses>.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program, version 6.0.33]. Retrieved 13/01/2020, from *Praat*: <http://www.praat.org/>.

Caraty, M. J., & Montacié, C. (2010). Multivariate analysis of vocal fatigue in continuous reading. INTERSPEECH, 2010.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, 61, 30-47.

Coker, C. H., & Umeda, N. (1975). The importance of spectral detail in initial-final contrasts of voiced stops. *Journal of Phonetics*, 3, 63–8.

Frankish, C., Jones, D., & Hapeshi, K. (1992). Decline in accuracy of automatic speech recognition as a function of time on task: fatigue or voice drift?. *International Journal of Man-Machine Studies*, 36(6), 797–816.

Grosjean, F. (1980). Temporal variables within and between languages. In H. W. Dechert, & M. Raupach (Eds.), *Towards a cross linguistic assessment of speech production* (pp. 39–53). Frankfurt: Lang.

Grosjean, F., & Miller, J. L. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science*, 5(4), 201–206.

Hillman, R. E., & Gilbert, H. R. (1977). Voice onset time for voiceless stop consonants in the fluent reading of stutterers and nonstutterers. *The Journal of the Acoustical Society of America*, 61(2), 610-611.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25(2), 143–168.

Kisler, T., Reichel U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326–347.

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, 20(3), 384–422.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, 10, 1–28.

Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language*, 22(2), 171–184.

Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish–English spontaneous code switching. *Journal of Phonetics*, 52, 121–137.

Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech. *International Congress of Phonetic Sciences*, 607–610.

Shimizu, K. (1996). *A Cross-Language Study of Voicing Contrasts of Stop Consonants in Asian Languages*. Seibido Publishing Co.

Shimizu, K. (2011). A Study on VOT of Initial Stops in English Produced by Korean, Thai and Chinese Speakers as L2 Learners. *International Congress of Phonetic Sciences*, 1818–1821.

Stuart-Smith, J., Sonderegger, M., Rathcke, T. & Macdonald, R. (2015). The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology*, 6(3-4), 505–549.

Yao, Y. (2009). Understanding VOT Variation in Spontaneous Speech. *UC Berkeley Phonology Lab Annual Report,* 5(5), 29–43.

7 Appendices

7.1 Appendix One

Table A1: *Speech time.*

	Spontaneous speech	Read speech
P1	6.57	7.50
P2	7.24	8.22
P3	6.34	6.44
P4	7.24	7.41
P5	7.56	6.23
P6	7.01	6.50
Mean	7.21	7.36

Table A2: *Fixed effects summary.*

	Spontaneous speech			Read speech		
	estimate	Std. Error	t-value	estimate	Std. Error	t-value
intercept	47.259	7.308	6.467	62.194	7.5672	8.219
time	-0.523	0.403	-1.296	0.8516	0.7147	1.192
vowel duration	179.896	27.804	6.47	111.2686	33.5296	3.319

Table A3: *Random effects summary.*

	Spontaneous speech		Read speech	
Group	Variance	SD	Variance	SD
stop type	36.03	6.003	16.22	4.028
word	78.91	8.883	141.37	11.89
vowel height	22.31	4.723	40.63	6.374
participant	144.03	12.001	109.67	10.473
residual	335.77	18.324	564.31	23.755

Table A4: *Random intercept values for the variable 'stop type'.*

	Spontaneous speech	Read speech
/p/	-2.892	-3.193
/t/	-3.499	0.752
/k/	6.391	2.441

Table A5: *Random intercept values for the variable 'following vowel height'.*

	Spontaneous speech	Read speech
high	-4.700	-4.812
low	2.105	6.083
mid	2.595	-1.271

Table A6: *Random intercept values for the variable 'participant'.*

	Spontaneous speech	Read speech
P1	-2.725	-3.347
P2	5.008	14.368
P3	15.852	2.181
P4	6.489	6.210
P5	-6.531	-4.502
P6	-18.092	-14.910

Acknowledgements

First of all, I am extremely grateful to Assistant Professor Sujinat Jitwiriyanont for inspiring me to pursue phonetics, encouraging me to send my paper to conferences, and advising this paper from a term project of the Acoustic Phonetics subject to even after the course ended. I'm deeply indebted to all my dearest linguistics seniors who have helped me understand linguistics, conferences, and researching inexpressibly deeper: P' Ponds Vatcharit, P'Tuner Pongbodin, and P'Tongla Ponrawee. This paper would not have been possible without P'Beam Asawa, my debate senior who helped me survive R and obtain the very result. Thanks also to P'Sun Suntisuk, P'Pao, P'Est, Book, U-Center friends, and my family for accompanying me

throughout this project, which unfortunately coincided with the pandemic, and making the progress a lot less hard for me.

Fight the virus, stick with the rules and reduce the peak: An Analysis of the Metaphors Used by Boris Johnson and Nicola Sturgeon to Discuss the COVID-19 Pandemic in Their Official Press Conferences from March to October 2020

Mitchell McKee
University of Glasgow

Abstract. The field of metaphor research has shown that metaphor can shape thought (Lakoff & Johnson, 1980) and affect reasoning (Thibodeau & Boroditsky, 2011), emotion (Hendricks et al., 2018) and behaviour (Hauser & Schwarz, 2015). Recently, this power of metaphor has been applied to investigate how the metaphors used by prominent political leaders frame the COVID-19 pandemic (Semino 2020b). There have been few in-depth studies on the metaphors used by UK politicians, specifically politicians from the Scottish Government, like Nicola Sturgeon, and those from the UK Government, like Boris Johnson. This study investigates the discourse of these politicians and asks what metaphors they used to discuss the pandemic and if they differed in any way. I extracted the metaphors from a sample corpus of daily press conferences between March and October 2020 using the Metaphor Identification Procedure (Pragglejaz Group, 2007). The results gathered from this process reveal two main conceptual metaphors which are shared by both speakers: THE COVID-19 VIRUS IS A POWERFUL AGENT, which instils the virus with weight and power, and THE PANDEMIC IS A JOURNEY – society is 'moving towards' normality and 'follows' scientific data as a 'guide' to get there. However, there is nuance between the two speakers: Johnson uses much more negatively valanced violence metaphors, such as 'fight' and 'battle', compared to Sturgeon who avoids these violent metaphors and foregrounds defence in words like 'protect'. I discuss the potential effect of these metaphors on public perception of the crisis and avenues for possible future research.

Keywords: metaphor; discourse analysis; cognitive linguistics

---

1 Introduction

The COVID-19 pandemic is one of the most calamitous events of the 21st century. The virus originated in China in late 2019 and spread throughout Europe in early 2020 and by March it had spread worldwide. The initial 'first wave' saw thousands die, millions of businesses close and billions confined to their homes under government enforced 'lockdowns'. The COVID-19 pandemic has influenced every part of daily life and, as a result, many researchers from different fields have studied it and all its facets. From a linguistics perspective, the COVID-19 pandemic has greatly impacted everyday language use. The Oxford English Dictionary found that the most frequent words in their corpora between April and June 2020 were virus related with 'PPE', 'lockdown', 'pandemic', and 'ventilator' being the most popular (OED Editorial, 2020). Researchers have noted a shift towards 'coronaspeak', which refers to coronavirus related vocabulary that has dominated everyday conversation (Thorne, 2020).

One area of language that has received a lot of focus in particular is metaphor. This is not only because metaphors are ubiquitous in speech, accounting for around 20% of natural discourse (Thibodeau et al., 2017, p. 854), but also because discussion surrounding the pandemic has been largely metaphorical. The pandemic has been described as a 'war' which needs to be 'fought' (Heffernan, 2020) and a natural disaster where humanity is caught in the 'perfect storm' with a 'tsunami' of cases which throws society into a 'meltdown' (Nerlich, 2020a).

A focus of metaphor research specifically examines politicians' use of metaphor as this can affect how the public forms opinions about issues (Charteris-Black, 2011, p. 32). An issue as catastrophic and frightening as the COVID-19 pandemic meant that politicians had to choose their metaphors carefully. This was the case for the UK as the primary source of communication the public received about how to live and act during the pandemic came from the leading politicians in the nation. Research conducted on what metaphors were used by these politicians can illuminate how the pandemic and its many facets was discussed and delivered.

As of yet, there has been no large-scale study conducted on the metaphors used by UK politicians to discuss the pandemic. This essay aims to examine the metaphors used by two major political leaders in the UK, Nicola Sturgeon, the leader of the Scottish Government, and Boris Johnson, the leader of the UK Government, in their public press conferences delivered during the critical period of the pandemic between March and October 2020. I ask the following questions:

- What are the metaphors used by both speakers to discuss a set of salient target domains related to the COVID-19 pandemic?
- How do these metaphors frame aspects of the pandemic?
- Are there similarities and/or differences in the metaphors used between the speakers?
- What is the potential impact of these metaphors on public perception of the crisis?

In the next Sections, I will discuss the power and significance of metaphor, previous research on metaphor and COVID-19 and the Scottish and UK political dimension which will further illuminate these questions.

1.1 The Importance of Metaphor

An approach to metaphor research developed in the 1980s by Lakoff and Johnson found that some metaphors can exist in systems where they express very similar ideas. For example, there are several metaphors which express emotions in terms of heat, e.g., 'he made my blood boil', 'she's cold hearted', 'a lukewarm response' etc. (Lakoff, 2016, p. 270). Systematic metaphors like these exist in everyday speech and can pattern our discourse. Lakoff and Johnson tried to account for this in their Conceptual Metaphor Theory (CMT) by arguing that there exists a range of 'conceptual metaphors' in the mind of speakers which generate these systematic metaphors found in speech (Lakoff & Johnson, 1980, p. 7). Conceptual metaphors are created by the mapping between two concepts in the mind: the 'source domain', which is the concept speakers take knowledge and vocabulary from, is mapped onto a 'target domain', which is the concept that is being described (Lakoff & Johnson, 1980, p. 252). In the metaphors where emotions are discussed in terms of heat, the source domain HEAT is mapped onto the target domain EMOTIONS to form the conceptual metaphor: EMOTIONS ARE HEAT (following the conventions of CMT, the source and target domains are capitalised). This conceptual metaphor can then generate the linguistic metaphors found in

discourse like ‘he made my blood boil’ etc. It is important to note that conceptual metaphors are a property of thought and not of discourse. The metaphors found in discourse are taken as evidence to suggest the existence of an underlying conceptual metaphor.

Due to this cognitive dimension of metaphor, Lakoff and Johnson argue that the metaphors we use can reflect our conceptual structure and reveal how we perceive and think about concepts in the world. This was demonstrated in an experimental study where two groups of participants read different passages about crime. Each passage was written according to one of two common conceptual metaphors used to discuss crime: CRIME IS A VIRUS or CRIME IS A BEAST. These conceptual metaphors describe crime differently through systematic metaphors, e.g., in CRIME IS A VIRUS, crime is ‘spreading’ and ‘plaguing’ cities, and in CRIME IS A BEAST, crime is ‘ravaging’ cities and perpetrators need to be ‘hunted’. Participants were then asked how they think the crime problem should be solved. A major result found was that the type of metaphors the participants read affected their answers. Participants who read the CRIME IS A VIRUS passage suggested that the crime problem should be solved by rooting out the cause of the issue and creating more social reform. On the other hand, participants who read the CRIME IS A BEAST passage suggested that more officers should be hired, and more jails should be built to ‘catch and cage the criminals’ (Thibodeau & Boroditsky, 2011, p. 5). Participants thought about the issue of crime in terms of the metaphor they were exposed to, which suggests that metaphor can influence a reader’s thought process. A potential explanation for this, the researchers argue, is that the source domains which are used (VIRUS/BEAST) contain different knowledge structures. When the mapping between one of the source domains to the target domain (CRIME) takes place, the stored knowledge specific to the source domain highlights some aspects of the target domain, therefore creating a different understanding of the target (Thibodeau et al., 2017, p. 852).

This study shows that conceptual metaphors can have real world effects on the thoughts and reasoning abilities of listeners and readers. If the participants in the study were politicians, the metaphors used to describe crime could have influenced how they, for example, voted on legislation. Lakoff further discusses this effect of conceptual metaphor by arguing that some conceptual metaphors were used to ‘justify’ the Gulf War in 1990. He argues that one popular conceptual metaphor which was developed was THE STATE IS A PERSON. This conceptual metaphor is indicative of the politicians’ thoughts: they believed that the state needs resources to ‘survive’ and oil was its ‘lifeline’. If it were cut off then it would ‘die’, so they went to war to effectively ‘save’ themselves (Lakoff, 1991, p. 26).

A criticism of CMT is that it is unclear how many conceptual metaphors exist and how much systematicity of metaphors in discourse is needed in order to suggest an underlying conceptual metaphor (Gibbs, 2011, p. 533). Conceptual metaphors can also be shared by a culture or a group of people (Lakoff & Johnson, 1980, p. 8), so if only one instance of a metaphor was used by a speaker, this could still suggest a conceptual metaphor which is shared generally by speakers in the same culture. On the other hand, one instance of metaphor may not indicate an underlying conceptual metaphor; it could be a possible novel use. This means that it is challenging to ‘find’ conceptual metaphors in the minds of speakers which makes it difficult to discuss their potential impact on thought. CMT is still a common theory, however, and has shaped metaphor research for decades. Studies on metaphor commonly use the terms ‘source’ and ‘target domain’ which I will continue to adopt. Also, in my analysis for metaphors for the COVID-19 crisis used by Johnson and Sturgeon, the existence of possible underlying metaphors in the minds of the speakers will only be suggested when a set of very similar metaphors occur frequently in their speech.

Looking more generally at metaphor in discourse rather than in thought, metaphor is still a very powerful linguistic device due to its ability to ‘frame’ concepts. This is a process where speakers select

'aspects of a perceived reality' to make them more noticeable and important against a background of other aspects (Entman, 1993, p. 52). Framing is inherent in metaphor because speakers need to choose what concept, and the vocabulary from that concept, they will use to describe another concept. For example, in the metaphor 'read the road', the choice of 'read' frames the road as a book, which foregrounds its ability to be understood and processed like a text. If the metaphor were 'decipher the road', then the framing would change: the road is now a code. This foregrounds its complexity as it now must be decoded to understand it. Crucially, both of these metaphors offer a different interpretation of what a road is. The former describes it as something approachable and easy to understand, whereas the latter makes it appear as something difficult. This means that the framing ability of metaphor can create a particular interpretation of a concept which affects the way people understand it (Boeynaems et al., 2017, p. 119; Semino et al., 2018, p. 626).

This framing ability of metaphor has multiple effects on listeners. Experimental evidence suggests that metaphorical framing can influence behaviour and reasoning. A study was conducted on participants who were exposed to different framings of cancer. After reading excerpts of text where cancer was either framed as an 'enemy' or as a 'balance', e.g., an 'unbalance of cells', participants were asked questions about how likely they were to engage in certain behaviours. Those who read cancer as an 'enemy' were much less likely to engage in behaviours which would reduce their risk of cancer, like reducing red meat consumption, engaging in more exercise and so on, compared to participants who read cancer as a 'balance' (Hauser & Schwarz, 2015, p. 71).

As well as influencing behaviour and reasoning, metaphor can also influence the way we emotionally evaluate a situation. Hendricks et al. (2018) conducted a study where they asked participants to read excerpts about someone suffering from cancer and answer questions about the protagonist's experience. The protagonist's experience was either framed as a 'journey' or a 'battle'. One finding was that participants who read the 'journey' passage felt that the protagonist was more likely to make peace with their situation compared to those who read the 'battle' passage who stated that the protagonist would be more likely to feel guilty about their illness (Hendricks et al., 2018, p. 271). A similar study on patients who received involuntary outpatient treatment for mental health illnesses found that the metaphors they used reflected whether they felt positively or negatively towards their treatment. Some patients viewed it positively as a 'safety net', but others were more negative and framed it as a tool for punishment and control, e.g., like a 'tranquiliser' or as if they were 'locked up' (Lawn et al., 2016, pp. 5–10).

Similar studies to these have been conducted on a broad range of target domains, such as financial markets (Morris et al., 2007), obesity (Atanasova, 2018), democracy (Nasirci & Sadik, 2018), climate science (Deignan et al., 2019) and so on. These studies all compound the same suggestions: metaphors can frame aspects of real life in certain ways which can influence our interpretations and opinions, as well as affect our reasoning, our emotions, our behaviour, and, potentially, how we think about a situation or concept. As a result of the power of metaphor, it is the central concern of this essay which will particularly focus on the framing effects of metaphors for COVID-19 and how groups of similar metaphors can potentially suggest the existence of underlying conceptual metaphors.

1.2 Metaphor, Infectious Diseases and COVID-19

Science communication to the public typically uses metaphor to discuss complicated ideas in simpler and more understandable terms (Deignan et al., 2019, p. 379). Infectious diseases are a particularly relevant area to the public as deadly viruses can disrupt daily life. This makes the metaphorical framing of diseases important as it can help the public avoid catching any viruses. Work on the framing of SARS in British

newspapers found that the most common framing technique of the disease used by journalists was to describe the virus as a 'lethal' and 'deadly threat' and a 'killer bug' (Washer, 2004, p. 2565). Other metaphors, such as 'slams', 'hurt', 'hammered' and 'gripped', frame the virus as a powerful and forceful danger in society (Wallis & Nerlich, 2005, p. 2635). This was the most dominant framing constructed for the virus, and since it was delivered by the media, it managed to pervade the lives of the public.

COVID-19, which is like the SARS virus, has been framed in a very similar way. It has been described as an 'invisible killer' and a 'threat' (BBC News, 2020a). A number of studies have criticised the use of military metaphors for COVID-19, which describes doctors as being on the 'frontline' in the 'battle' 'against' the disease, by arguing that they negatively frame the pandemic as chaotic and filled with unnecessary deaths and suffering (Wise, 2020). This type of framing can also cause doctors to be viewed as brave 'heroes' who are 'fighting' for the people, which puts even further pressure on healthcare workers (Hannan, 2020).

As a result of the prolific use of war metaphors, researchers, namely Semino, Koller, Sobrino, and others, started the #ReframeCOVID project (the hashtag indicates that it originally started on Twitter) in order to find other ways of framing the pandemic from different languages and cultures across the world (Semino, 2020b; Nancy, 2020). They found that the crisis and its consequences can be framed and understood in many different ways, such as framing the virus itself as a 'fire' which needs to be 'put out' (Semino, 2020a) or framing the pandemic as a 'journey' which everyone is 'on' (Cox, 2020).

The #ReframeCOVID project shows the vast extent to which the COVID-19 pandemic has been discussed metaphorically. It also highlights the complexity of the metaphors used: aspects of the pandemic can be framed differently through a range of metaphors used by different people across the world. Each metaphor then has its own effects on people's understanding and construal of events. The next Section will specifically focus on the UK and Scotland's approach to COVID-19 and how this could have impacted the metaphors used.

1.3 UK and Scotland's Approach to COVID-19

The UK's approach to the COVID-19 pandemic was unique as alongside the overarching UK Government, there are also devolved governments in Scotland, Wales, and Northern Ireland which have their own leaders. This means that the people from these countries were not only asked to listen to their devolved leader but also to the UK's leader, Boris Johnson. In Scotland, the Scottish people followed the advice of First Minister Nicola Sturgeon about how to live and act during the pandemic, but so too were they asked to follow Boris Johnston's advice. The two governments have aimed to be 'united' in their approach throughout the crisis (Macnab, 2020) but have taken drastically different decisions. For example, the leaders 'eased' the lockdown restrictions at different stages throughout Summer 2020 with Johnson making decisions weeks before Sturgeon. Scotland also has a different 'tier system' from the rest of the UK and enacted it incongruently with England in October and November 2020 where large swathes of Scotland saw pub and restaurant closures, whereas England did not (Scottish Parliament Information Centre, 2021). The different approaches from both Johnson and Sturgeon diverged to such an extent that a poll conducted on a group of people from Scotland found that three quarters preferred Sturgeon's handling of the pandemic compared to Johnson's (Flockhart, 2020).

Johnson and Sturgeon approached the pandemic differently in their policy decisions, but of interest here is whether they also approached it differently in their metaphor use. Finding a potential difference, or similarity, in how the leaders framed aspects of the pandemic through their metaphor choices can further

reveal insight into their reasoning and show how they delivered crucial information to millions of people throughout the nation.

1.4 Summary

In summary, metaphors are powerful tools: when used on listeners and readers, metaphors can affect the way they reason about and interpret concepts which can influence their behaviour and emotions. They can also suggest how speakers think about concepts when evidence suggests the existence of an underlying conceptual metaphor. The following Sections of this essay will now apply these previous findings about the power of metaphor to speech data from Sturgeon and Johnson. Section 2 describes how the metaphors used to discuss aspects of the COVID-19 pandemic were extracted and analysed. In Section 3, I will discuss how the metaphors frame each main aspect of the crisis and if they indicate possible underlying conceptual metaphors. Section 4 offers a wider discussion which pulls the metaphors together to suggest how the speakers generally framed the pandemic. Section 5 will conclude on the main findings and lead into Section 6 which offer avenues for potential future research.

2 Methodology

2.1 Corpora Construction

The primary texts for this study are the public press conference speeches delivered by Boris Johnson and Nicola Sturgeon. There are many reasons why these texts have been chosen. Firstly, they were important to the public as they contained the key information about how to live and act during the pandemic and were viewed by millions of people at the height of the crisis (Boyle, 2020). Secondly, each briefing was of the same genre as the other, i.e., an address to the public and were all in the same format where a set speech was delivered followed by a question-and-answer session with journalists. Language differs between different genres and contexts so the similarity between the briefings allowed for a robust comparison of metaphor use (Semino et al., 2018, p. 8). Finally, the speeches are not ephemeral and hard to access compared to quotes in newspapers or discussions with journalists in interviews. Instead, they are fully scripted and easily accessible on the websites of the UK and Scottish Government.

Although the press conferences are robust texts to analyse, collating them into corpora posed some difficulties. Both speakers did not deliver the same number of briefings throughout the pandemic, nor did they deliver them on the same date. This not only creates inconsistency but also a vast amount of data. For the purposes of this study, only the first briefing of the month delivered by the two speakers was analysed. Defining the parameters of the corpus in this way still creates bias and does not account for all the data. However, the first briefing of the month was taken as a milestone for the development of the pandemic and allows for comparison between the speakers.

The time period that was chosen was between March and October 2020. The reason for this is that mid-March was when the pandemic began in the UK which subsequently started the daily press briefings on the crisis. This was also the start of ‘lockdown’ which was gradually eased throughout the summer then reinstated in many forms and complexities around September/October. Case numbers also rose in March then fell until September before rising again around October (BBC News, 2020b). This means that March-

October contained a large portion of the events and developments of the pandemic making this time period a relevant source for linguistic data.

The speeches from Johnson and Sturgeon were compiled into two separate corpora:

Table 1: *The public press conference speeches which constitute each corpus.*

Month	Date of Delivery	
	Boris Johnson (BJ)	**Nicola Sturgeon (NS)**
March	16th	20th
April	30th	2nd
May	11th	1st
June	3rd	1st
July	3rd	1st
August	31st July	3rd
September	9th	3rd
October	12th	2nd
TOTAL SPEECHES	**8**	**8**
TOTAL WORD COUNT	**9,618**	**15,239**

The first briefing by Johnson was not on March 16th as there were other smaller briefings before that. However, these were not of the same scale and intensity as the 16th, and also after the 16th is when the briefings started to be delivered daily which marks this date as a significant turning point in the UK Government's approach. Sturgeon also delivered a statement on the 17th of March, but this was to the Scottish Parliament and was not a press conference. She followed the UK Government's approach and started daily updates starting from the 20th of March. For August, the 31st of July was chosen for Johnson because he did not deliver a briefing in August, but he did deliver one on the final day in July, so this is taken to represent August. The total word count is also different; however, the raw metaphor counts were normalised per thousand words as to allow for comparison. The small size of these corpora also proves beneficial as an analysis of smaller corpora can be very detailed and fine-grained, adding to what Deignan describes as the 'richness of interpretation' (Deignan, 2005, p. 93).

These corpora were uploaded to eMargin, an online webpage which allows researchers to annotate parts of texts and write comments (https://emargin.bcu.ac.uk/). This webpage was used in previous research to facilitate the extraction of metaphors (Semino et al., 2018, p. 63).

2.2 Identifying Target Domains

In line with Semino et al.'s work in 2018, which analysed metaphors for cancer, a set of target domains were identified before the corpus data was analysed (Semino et al., 2018, p. 61). A target domain, which is a term from CMT, refers to the concept which is being discussed metaphorically (Lakoff & Johnson, 1980, p. 252). The reason they are defined first is because metaphor is ubiquitous in language and many conventional subjects are discussed metaphorically, e.g., time is usually discussed metaphorically in terms

of space (Boroditsky, 2000, p. 3). Conventional metaphors such as these are not of interest to this study because they do not have any ties with the research focus. Defining a set of target domains directs the analysis to the specific metaphors which are relevant to the topic of the research. In this study, the focus is metaphors for the COVID-19 pandemic. This is a very broad area with multiple different aspects.

Therefore, I defined the target domains of interest as the following:

- The COVID-19 virus
- The restrictions
- The pandemic
- Action by the governments
- COVID-19 infection rates
- Scientific data

If these target domains were discussed using metaphorical language, that linguistic data would be extracted. Moreover, identifying the target domains allows for the identification of underlying conceptual metaphors as they align with Lakoff and Johnson's source and target domain structure, i.e., THE TARGET DOMAIN IS THE SOURCE DOMAIN, like the ARGUMENT IS WAR conceptual metaphor (Lakoff & Johnson, 1980, p. 4). In this case, the target domains take the first part of this construction, e.g., THE COVID-19 VIRUS IS X.

2.3 A Method for Identifying Metaphor

There are currently no fully computational method researchers can use to extract all metaphorical language from a corpus. This means that this process must be carried out manually (Deignan, 2005, p. 92). Manual analysis is problematic because it is time consuming and does not offer completely reliable data as judging metaphorical language is difficult and subjective (Charteris-Black, 2004, p. 35; Semino, 2008, p. 14). Multiple methods have been devised to assist researchers in this manual analysis (Charteris-Black, 2004, p. 35; López & Llopis, 2010, p. 3301; Neuman et al., 2013; Sun, 2020), but one of the most popular is the Metaphor Identification Procedure (MIP) created by a group of researchers known as the Pragglejaz Group (Pragglejaz Group, 2007). The MIP offers a step-by-step process to limit the number of decisions and subjective interpretations made by the researcher(s):

- Step 1: read the text through to understand what it means.
- Step 2: identify the lexical units in the text. These can be individual words, but can also be phrasal verbs, compounds, proper nouns, polywords etc.
- Step 3: for each lexical unit, establish:
    - A – its meaning in the context of the sentence.
    - B – its most general and basic meaning in other contexts as outlined by a dictionary. The sense that should be chosen should relate to concreteness/tangibility and can be imagined and related to bodily action and lived experience. The sense chosen should also be the oldest where possible but should still be in use in modern discourse.

    - C – whether the contextual meaning (A) contrasts with the basic meaning (B) but can be understood in terms of it.
- Step 4: if there is this contrast outlined by C, mark the lexical unit as metaphorical. (Adapted from Pragglejaz Group, 2007, p. 3)

This method was applied to the corpus data. An example of this is outlined in the following extract from Boris Johnson's speech from June:

> 'Now that the rate of transmission in the UK has significantly fallen from its peak, we need to take steps to manage the flare-ups and stop the virus re-emerging in the UK.' (Johnson, 2020d)

Following Step 1, this extract concerns COVID-19 and its rate of infection in the UK. Step 2 requires that the lexical units be identified:

> Now / that / the / rate / of / transmission / in / the / UK / has / significantly / fallen / from / its / peak, / we / need / to / take / steps / to / manage / the / flare ups / and / stop / the / virus / re-emerging / in / the / UK.

The MIP hinges its decision about the metaphoricity of a lexical unit entirely on the contrast between the basic and contextual meaning of the lexical unit. This is why Step 3 and 4 are the most crucial part of the process as it separates the metaphorical language from the literal. To do this, I used the Oxford English Dictionary Online (OED, 2020a) to identify the most basic and oldest meaning still in use and contrasted this with the contextual meaning. In this extract, I mark 'fallen', 'peak', 'steps', 'flare-ups', and 're-emerging' as metaphorical but not the other words, most notable are 'transmission' and 'stop'. Table 2 illustrates the reasons for these decisions:

Table 2: *Example of steps three and four in the MIP with an added column referring to the target domains of this study.*

Lexical Unit	Basic Meaning	Contextual Meaning	Metaphor?	Target domain
Fallen	To move from a high position to a lower one by force of gravity (OED 2020b).	The number of infections is reducing from the maximum number.	Yes	COVID-19 infection rates
Peak	The highest point on a mountain (OED 2020c).	The maximum number of COVID-19 cases.	Yes	COVID-19 infection rates

Steps	An act of bodily motion of moving the feet to facilitate walking, climbing etc. (OED 2020d).	Action taken to prevent COVID-19 outbreaks and reduce its effects.	Yes	Action by the governments
Flare-ups	Sudden burst into flames (OED 2020e)	The sudden increase of COVID-19 cases.	Yes	The COVID-19 virus
Re-emerging	To rise out of a liquid (OED 2020f).	The ability for COVID-19 to spread again in the UK.	Yes	The COVID-19 virus
Transmission	To cause something to be passed/transferred from one person to another (OED 2020g).	The number of times COVID-19 is spread between people.	No	-
Stop	To cease to move or act (OED 2020h).	To end COVID-19 effects.	No	-

The final step in marking the lexical unit as metaphorical depends on whether the contextual meaning is discussed *in terms of* the basic meaning. For example, in the instance of 'steps', action against COVID-19 is discussed in terms of progressive movement with the feet. Conversely, in 'transmission', the basic meaning is the same as the contextual meaning as they both refer to senses of transferring something from person to person.

2.4 The MIP: Problems and Solutions

The MIP provides a systematic way for identifying metaphor, but one of its biggest shortcomings is that it still requires researchers to make decisions which can potentially lead to biased and subjective data. The first decision made is in Step 2 as it requires researchers to divide the data into lexical units. Due to complexities inherent to morphology in trying to determine word boundaries, it is not clear what should be treated as a word, separate words, compounds, and so on. Researchers who created the MIP refined the procedure into the Metaphor Identification Technique Vrije Universiteit (MIPVU) and produced more

guidance about how lexical units should be identified[40]. Compounds and novel words should be treated as their own lexical units, whereas the constituents making up multi-word expressions and polywords should be divided into individual parts (Steen et al., 2010, p. 186). This means that 'flare-ups' is treated as a singular word expression, but in 'Test and Protect' (Sturgeon, 2020c), both 'test' and 'protect' are treated separately. More difficult cases arise from this refinement, for example from Boris Johnson's speech in April:

> 'We have come through the peak, or rather we've come under what could have been a vast peak as though we've been going through some huge alpine tunnel, and we can now see the sunlight and pasture ahead of us. And so it is vital that we do not now lose control and run slap into a second and even bigger mountain.' (Johnson, 2020b)

This is an extended metaphor which primarily discusses the epidemic in the UK. However, dividing each lexical unit loses the sense of the unified metaphor. For example, 'sunlight' and 'pasture' are dependent on previous lexical units 'going' and 'tunnel'. A researcher could argue that this statement is one metaphor, but the requirements of the MIP state that each word must be divided and judged individually for metaphoricity. Both perspectives lose elements of granularity in the metaphor: viewing it as one metaphor means that 'come' and 'going' are not counted as two separate movement metaphors which could have implications for the overall data analysis, but dividing each lexical unit loses the context of the unifying metaphor. A solution to this, as discussed below, is to provide a real example for each lexical unit so its wider context is not lost.

Another decision that must be made by researchers is the division between basic and contextual meaning. There are many instances where this division is not so clear, for example in the word 'stop' as shown in Table 2. The first sense in the OED refers to blocking up a passageway. This implies that COVID-19 is moving forward and action is taken to 'block' this movement. However, an even more basic meaning than this which all the senses relate to is the idea of stopping general action of the virus, which the UK Government is trying to do. In instances like this where most basic meaning of a word is unclear; the MIP suggests using the oldest meaning which is still in use. The issue with this is that the OED offers multiple literal meanings and for some words it is impossible to determine which came first. Also, determining what is the most 'basic' meaning can be affected by researchers' personal opinions and knowledge of the world. In this case, it was decided that the most basic meaning of 'stop' refers to the cessation of all action of an entity. This does not contrast with the contextual meaning, so it was not marked metaphorically. On the other hand, another researcher may have decided that the most basic meaning of 'stop' refers to the action of blocking movement, and they would have marked the lexical unit as metaphorical.

Difficulties such as these show that metaphorical language can be inherently difficult to notice and judge (Semino et al., 2004, p. 1272). Another example of this is the word 'sacrifice'. This appears multiple times in the data, e.g., '…the public who continue to make very hard sacrifices right now' (Sturgeon 2020f). The literal meaning of 'sacrifice' is an 'offering' (OED, 2020i). The contextual meaning refers to the act of giving something up, e.g., socialising, in order to end the pandemic and reduce its effects. Similar to 'stop', it is difficult to judge whether this is a metaphor — the ultimate decision lies with the researchers.

[40] The MIPVU is more complex than the MIP as it contains multiple different types of metaphor and a more difficult identification process. For the purpose of this study, I will use the original MIP, but I will use the guidance from the MIPVU discussed here.

The Pragglejaz Group's solution for resolving subjective decisions such as these is to have several researchers make multiple 'passes' at the texts and convene to discuss differences (Pragglejaz Group 2007, p. 17). In a study such as this, this process is impossible. However, in order to move forward with the procedure, I applied the MIP to both corpora three times with at least two weeks apart. Each 'pass' allowed me to be more critical about the choices I made. I also applied the 'When-In-Doubt-Leave-It-In' approach, which has been used by other researchers (Semino et al., 2018, p. 63), as this ensures that no possible metaphor is missed.

The final shortcoming of the MIP is that it depends on contextual information to determine metaphoricity, but the data created at the end of the procedure is a selection of lexical units removed from context. This makes it difficult to understand how the words are actually metaphorical. This has further implications for data visualisation as shown by Table 3.

Table 3: *The raw and normalised frequencies for a selection of metaphors to describe the COVID-19 virus without examples.*

Metaphor	**Total (raw)**		**Total (per thousand words)**	
	BJ	**NS**	**BJ**	**NS**
campaign	2	0	0.21	0.00
fight	11	1	1.14	0.07
protect	10	47	1.04	3.08

It can be inferred that 'fight' is metaphorical by assuming the context, such as 'fight the virus', but for 'campaign' it is much more difficult. A solution to this is to provide a real stereotypical example for each instance of metaphor so the context can be understood.

Table 4: *The raw and normalised frequencies for a selection of metaphors to describe the COVID-19 virus with examples.*

Metaphor	**Examples**		**Total (raw)**		**Total (per thousand words)**	
	BJ	**NS**	**BJ**	**NS**	**BJ**	**NS**
campaign	We're leading a campaign to fight back against this disease (March)		2	0	0.21	0.00

fight	our fight against coronavirus (June)	as we fight this battle against coronavirus (April)	11	1	1.14	0.07
protect	protect the NHS (April)	people working in Test and Protect and our local public health teams (August)	10	47	1.04	3.08

Despite its shortcomings, the MIP is still a very robust and common technique used in metaphor research. It is easy to use and produces quantitative and qualitative datasets for fruitful discussion, hence its continuous application across a range of different metaphor studies such as on illness (Beck, 2016, p. 78; Chircop & Scerri, 2018, p. 2625; Semino et al., 2018, pp. 57–59), finance (Cheng & Ho, 2017, p. 264); music (Pérez-Sobrino, 2014, p. 302), tourism (Jaworska, 2017, p. 166), and even previous work on metaphors and COVID-19 (Semino, 2020a).

2.5 Extracting and Compiling the Metaphors

The MIP was applied to both corpora on eMargin so that every lexical unit was read and its metaphoricity was determined. If a lexical unit was considered metaphorical, it was highlighted and underlined.

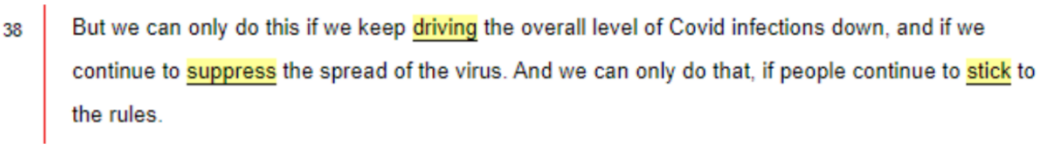


Figure 1: *Extract from Sturgeon's speech in June on eMargin demonstrating how the extraction process works. (Taken from https://emargin.bcu.ac.uk/)*

Each highlighted metaphor then had to be grouped into a target domain (cf. Section 2.2). Doing this posed some difficulties. For example, is 'flare-ups' describing the COVID-19 infection rates or is it describing the COVID-19 virus? A case can be made for each of these. In terms of infection rates, 'flare-ups' describe how the numbers are rising rapidly like a fire. In terms of the COVID-19 virus, the virus itself is a fire which is re-alighting or 'flaring up' again. In each 'fuzzy' instance such as this, a decision had to be made about which target domain the metaphor was grouped into. In this case, I concluded that 'flare-ups' best fits into the COVID-19 virus target domain as it is more describing the virus itself as a 'fire'. The multiple 'passes' I made on the data weeks apart elucidated the decisions made over time, so the target domains became clearer and more refined.

An Excel Spreadsheet (Microsoft Corporation, 2016) was made for each target domain and the metaphors which related to them were compiled into it:

	Examples		COUNTS							
			TOTAL (raw)		TOTAL (per tousand words)		MARCH		APRIL	
Metaphor	BJ	NS	BJ use	NS use	BJ use	NS use	BJ use	NS use	BJ use	NS use
campaign	We're leading a campaign to fight back against this disease (March)		2	0	0.21	0.00	2			
fight	our fight against coronavirus (June)	as we fight this battle against coronavirus (April)	11	1	1.14	0.07	2			1
protect	protect the NHS (April)	people working in test and protect and our local public health teams (August)	10	47	1.04	3.08	1	7	3	5
face	challenge that we face (March)	to reduce the risks that we face (September)	1	4	0.10	0.26	1	1		
shield	we would pause shielding nationally (August)	our advice for some children who are shielding (July)	5	1	0.52	0.07	2		1	
defeat	defeat this virus (April)		1	0	0.10	0.00			1	
beat	the British people have the resolve to beat this virus (October)		5	0	0.52	0.00			1	
threat	the level of threat posed by the cirus (May)		1	0	0.10	0.00				
front	progress on three fronts (June)	the supply and distribution of PPE to frontline staff (April)	1	3	0.10	0.20				3

Figure 2: *An image of the Spreadsheet for the COVID-19 virus target domain which shows the metaphor in each month with examples, the total raw instances of the metaphor and the normalised frequency per thousand words. Only instances for March and April are shown but this Spreadsheet extends to October and the total contains all instances from March-October.*

The next step was to further group together the metaphors within each target domain to understand their patterns of use. There have been many approaches in linguistics as to how concepts should be organised, e.g., into domains, frames, cognitive models, and so on (Croft & Cruse, 2004, p. 8). In studies on metaphor, metaphors are normally grouped together by their shared meaning. An approach adopted by Semino et al. (2018, p. 64) was to use the UCREL Semantic Analysis System (USAS). This gave every metaphor a semantic tag which the researchers used to group the metaphors together by their shared meaning. This approach is systematic and offers an objective process for semantic organisation, but the tags are often inconsistent and can be inaccurate and vague, for example a possible tag can be 'the universe' (Semino et al., 2018, p. 67).

The USAS is useful for larger corpora, but a more granular process can be used on smaller corpora. Chircop and Scerri's approach was to group metaphors according to the shared 'semantic field' of which the words were a part (Chircop & Scerri, 2018, p. 2625). It is unclear how they define a 'semantic field', but it is similar to a lexical field. As described by Croft and Cruse, a lexical field 'groups together words that are associated in experience… words are defined relative to other words in the same lexical field' (Croft & Cruse, 2004, p. 10).

Table 5: E*xample of the lexical fields into which the metaphors for the COVID-19 virus are grouped.*

Lexical field	Metaphor	Examples	
		BJ	**NS**
VIOLENCE	campaign	We're leading a campaign to fight back against this disease (March)	
VIOLENCE	fight	our fight against coronavirus (June)	as we fight this battle against coronavirus (April)
FORCE	against	inoculating ourselves against this disease (April)	our collective efforts against Covid (October)
FORCE	strengthen	But as we grieve we are strengthened in our resolve (April)	
FORCE	engulf	the tragedy that engulfed other parts of the world (April)	
MANAGEMENT	control	Stay Alert, Control the Virus and Save Lives (May)	they make the job of everybody working to try to control this virus that much harder (August)
MANAGEMENT	contain	we are continually exploring smarter means of containing the virus (July)	give them the best chance of containing outbreaks (August)
MANAGEMENT	under control	we're getting the virus under control in the UK (June)	as we go into the winter months in keeping this virus under control (October)

The words were grouped into their lexical fields through the analysis of their meaning in context. This is because some words can differ in meaning from one context to another (Saeed, 2016, p. 57). For example, 'engulf' has the general sense of swallowed up powerfully (OED, 2020j). However, in Table 5, 'engulf' in this context refers to the strong, powerful nature of the COVID-19 virus. This links with other words with similar senses of force and power, like 'strengthen' and 'against'. These similar words have shared semantic meaning which constitutes the construction of the lexical field, FORCE. In some cases, it is difficult to

group metaphors together into a lexical field as there are not enough of them in the data and they show no similarity with other metaphors. Where this is the case, their lexical field is marked as 'MISC'.

This method is not perfect as it still requires intuitive decisions to be made about the lexical fields to which the metaphors belong; however, a purely systematic and 'objective' approach like the USAS tagger also requires making decisions because it can provide multiple tags for one word where only one can be chosen. This is because, as Saeed states, 'word meaning is slippery' (Saeed, 2016, p. 56). No solution for grouping will offer a completely objective result, but this approach allows for the general patterns of metaphor use to be identified for discussion.

2.6 Summary

This Section has outlined how the corpora were constructed, how the metaphors were extracted using the Metaphor Identification Procedure and how the metaphor data was synthesised. In each of these steps, criticisms of the method and procedure have been discussed, but so too have the solutions that were found in order to provide robust results which work as the foundation for the next Section: analysis and discussion of the metaphors for the COVID-19 crisis.

3 Results and Discussion

Section 3 is further divided into Sections which correspond to each of the target domains discussed in Section 2.2. In each Section, I will identify and compare the main metaphors used by both speakers to discuss each target domain. I discuss how these metaphors frame each target domain and if they can signal the existence of potential underlying conceptual metaphors. Section 4 offers a more detailed discussion of these metaphors and brings them together to suggest how the pandemic in general was framed by both speakers, whether there are similarities or differences in this framing and the potential impact.

3.1 Metaphors for the COVID-19 Virus

Table 6: *Metaphors used to describe and discuss the COVID-19 virus by Boris Johnson and Nicola Sturgeon. This shows the raw number and normalised frequency per thousand words of each metaphor as well as the lexical field with examples.*

Lexical field	Metaphor	Examples		Total (raw)		Total (per thousand words)	
		BJ	NS	BJ	NS	BJ use	NS use
VIOLENCE	campaign	We're leading a campaign to fight back against this		2	0	0.21	0.00

		disease (March)					
VIOLENCE	fight	our fight against coronavirus (June)	as we fight this battle against coronavirus (April)	11	1	1.14	0.07
VIOLENCE	protect	protect the NHS (April)	people working in test and protect and our local public health teams (August)	10	47	1.04	3.08
VIOLENCE	face	challenge that we face (March)	to reduce the risks that we face (September)	1	4	0.10	0.26
VIOLENCE	shield	we would pause shielding nationally (August)	our advice for some children who are shielding (July)	5	1	0.52	0.07
VIOLENCE	defeat	defeat this virus (April)		1	0	0.10	0.00
VIOLENCE	beat	the British people have the resolve to beat this virus (October)		5	0	0.52	0.00
VIOLENCE	threat	the level of threat posed by the virus (May)		1	0	0.10	0.00
VIOLENCE	front	progress on three fronts (June)	the supply and distribution of PPE to	1	3	0.10	0.20

			frontline staff (April)				
VIOLENCE	defence		we are the first line of defence (September)	0	3	0.00	0.20
VIOLENCE	enemy	We're fighting an invisible enemy (June)		1	0	0.10	0.00
VIOLENCE	vicious	against this vicious disease (July)		1	0	0.10	0.00
VIOLENCE	secure	continuing to work from home or attending a COVID Secure workplace (August)		10	0	1.04	0.00
VIOLENCE	cruel		It is a particularly cruel virus (April)	0	1	0.00	0.07
VIOLENCE	battle		this battle against coronavirus (April)	0	1	0.00	0.07
FORCE	cope	give our NHS the chance to cope (March)		2	0	0.21	0.00
FORCE	against	inoculating ourselves against this disease (April)	our collective efforts against COVID (October)	12	6	1.25	0.39

FORCE	strengthen	But as we grieve, we are strengthened in us resolve (April)		1	0	0.10	0.00
FORCE	engulf	the tragedy that engulfed other parts of the world (April)		1	0	0.10	0.00
FORCE	impact	reduce the impact of the virus across the globe (June)	reduce the impact of this virus (March)	1	2	0.10	0.13
FORCE	pressure	the NHS will swiftly be under intolerable pressure (October)		2	0	0.21	0.00
FORCE	overwhelm	at no stage has our NHS been overwhelmed (April)		4	0	0.42	0.00
FORCE	take on		take on the challenge of this virus (March)	0	1	0.00	0.07
FORCE	suppress	measures taken in Leicester and Luton have suppressed the virus (August)	we continue to suppress the spread of the virus (June)	2	7	0.21	0.46
FORCE	struggle	struggling to keep the virus under control	the struggle against COVID (October)	1	1	0.10	0.07

		(August)					
FORCE	squash	we can squash this virus wherever it appears (October)		1	0	0.10	0.00
FORCE	tackle	to tackle the resurgence of the virus (October)	our efforts to tackle this virus (May)	2	5	0.21	0.33
MANAGEMENT	control	Stay Alert, Control the Virus and Save Lives (May)	they make the job of everybody working to try to control this virus that much harder (August)	13	1	1.35	0.07
MANAGEMENT	contain	we are continually exploring smarter means of containing the virus (July)	give them the best chance of containing outbreaks (August)	4	6	0.42	0.39
MANAGEMENT	under control	we're getting the virus under control in the UK (June)	as we go into the winter months in keeping this virus under control (October)	7	4	0.73	0.26
MANAGEMENT	get a grip	in order to get a grip on emerging outbreaks (July)		1	0	0.10	0.00
MANAGEMENT	out of control	If it starts running out	if this virus gets out of control	1	1	0.10	0.07

		of control again (July)	(September)				
FIRE	flare-ups	we need to take steps to manage the flare-ups (June)	we’re seeing significant flare- ups (August)	1	1	0.10	0.07
FIRE	hotspot	we will restrict access to places which become hotspots for the virus (July)		1	0	0.10	0.00
FIRE	firefighting		But it is like fighting forest fires (August)	0	2	0.00	0.13
MOVEMENT	come back	I have also consistently warned that this virus could come back (August)		1	0	0.10	0.00
MOVEMENT	emerge	emerging outbreaks (July)		2	0	0.21	0.00
MOVEMENT	keep at bay		we're going to be able to keep this virus at bay (July)	0	1	0.00	0.07
MOVEMENT	past us		in these occasions when the virus gets past us (September)	0	1	0.00	0.07
MOVEMENT	through		the more we let the virus through	0	1	0.00	0.07

			(September)				
MOVEMENT	import	ensure we don't reimport the virus from abroad (June)		2	0	0.21	0.00
MOVEMENT	run	If it starts running out of control again (July)		1	0	0.10	0.00
MOVEMENT	foothold	an attempt to stop the virus from gaining a foothold in the UK (June)		1	0	0.10	0.00
MOVEMENT	race		It takes enormous efforts and enormous resource and it's always a race against time (August)	0	1	0.00	0.07
BROKENNESS	fix		This is not a quick fix (April)	0	2	0.00	0.13
BROKENNESS	fragile		Incidents like the one in Aberdeen remind us of how fragile things are (October)	0	3	0.00	0.20
BROKENNESS	tool	so many more tools at our disposal to deal with it (August)	it is a really important additional tool in tackling	1	1	0.10	0.07

			COVID (October)				
OPPORTUNITY	give chance to		give the virus a chance to spread (September)	0	5	0.00	0.33
OPPORTUNITY	deny chance to		continue to reduce the risk of transmission – by denying the virus opportunities to spread (August)	0	1	0.00	0.07
MISC	trouble	We must be willing to react to the first signs of trouble (August)		1	0	0.10	0.00
MISC	growing	We will have to beat this disease by our growing resolve (April)		1	0	0.10	0.00
MISC	detect		they cannot reliably detect the virus during the incubation period (April)	0	1	0.00	0.07
MISC	solution		I've said before that the app is not a magic solution (October)	0	1	0.00	0.07
MISC	address		the progress we have	0	1	0.00	0.07

			made in addressing COVID (July)				

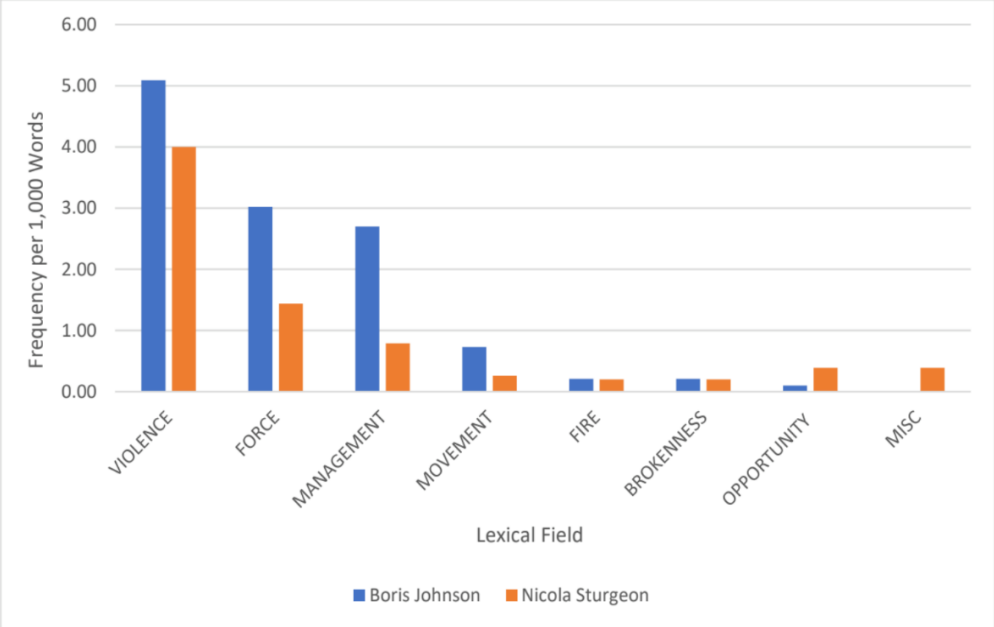


Figure 3: *The lexical fields of the metaphors in the COVID-19 virus target domain and their frequency of occurrence. The frequency is calculated by dividing the total number of metaphors in each lexical field by the word count of each corpus and normalising by 1,000 words. Following figures similar to this in each target domain contain the same calculation.*

The most salient metaphors in this lexical field all suggest that the virus is a powerful agent which moves independently in society and can make its own choices. FORCE metaphors are used by both leaders to frame the virus as something which has weight: it can 'overwhelm' the NHS, so it needs to be 'tackled' and 'suppressed' and we need to 'strengthen' ourselves against it. BROKENNESS metaphors are also similar to this which suggests that the virus is forcefully breaking society, so a 'tool' is needed to 'fix' it. MOVEMENT metaphors frame the virus as moving independently in society. It can 'come back' if it is not controlled, and it must be 'kept at bay' so it does not get 'past us'. The virus is actually 'moved' by people through contact, coughing, sneezing, etc., but this framing implies that it has its own independent movement.

As a result of this, the virus needs to be managed in some way. The leaders state that it must be 'controlled' and 'contained', suggesting that it is moving throughout society and the country needs to 'get a grip' of it. Sturgeon also goes as far as to use OPPORTUNITY related metaphors such as 'give' or 'deny chance to' the virus which frames it as a free agent who is looking for ways to cause more harm — it is just waiting for its chance.

The systematicity of these metaphors signal a possible underlying conceptual metaphor: THE COVID-19 VIRUS IS A POWERFUL AGENT. This suggests that the leaders view the virus as something

that can move freely, and when it does it is weighty and unruly. It therefore must be 'controlled' and stopped from moving. These metaphors instil the virus with goals and a plan.

Furthermore, this conceptual metaphor also allows insight into how the leaders believe the virus should be stopped. The dominant lexical field in the COVID-19 virus target domain is VIOLENCE. These metaphors can be further subdivided into OFFENSIVE, DEFENSIVE and NEUTRAL violence metaphors:

Table 7: *Further division of the metaphors in the lexical field of VIOLENCE.*

Metaphor	**Description**	**Total (raw)**		**Total (per thousand words)**	
		BJ	**NS**	**BJ**	**NS**
campaign	Offensive	2	0	0.21	0
fight		11	1	1.14	0.07
defeat		1	0	0.10	0
beat		5	0	0.52	0
enemy		1	0	0.10	0
battle		0	1	0	0.70
threat		1	0	0.10	0
face		1	4	0.10	0.26
	TOTAL	**22**	**6**	**2.29**	**0.39**
protect	Defensive	10	47	1.04	3.08
shield		5	1	0.52	0.07
front		1	3	0.10	0.20
defence		0	3	0	0.20
secure		10	0	1.04	0
	TOTAL	**26**	**54**	**2.70**	**3.54**
vicious	Neutral	1	0	0.10	0
cruel		0	1	0	0.07
	TOTAL	**1**	**1**	**0.10**	**0.07**

Table 7 suggests that Johnson discusses the COVID-19 virus using many more aggressive metaphors and views it as something which needs to be attacked. On the other hand, Sturgeon uses considerably fewer offensive metaphors. She said 'battle' once, but this was in April, and she did not use such aggressive fight metaphors after that. She still uses VIOLENCE metaphors, but she instead foregrounds defensive elements, e.g., she states that nurses and doctors are on the 'frontline' but are not attacking the virus — they are acting as a line of 'defence' against it. A large portion of the defensive metaphors used by Sturgeon is taken up by the word 'protect' which appears in the phrase 'Test and Protect', which is the name of the track and trace service in Scotland. This service is integral to stopping the effects of the virus, so foregrounding defensive elements in such a ubiquitous name further reinforces the idea of defence and protection against the virus and backgrounds aggression.

A novel use of metaphor describes the virus as a 'fire' and stopping 'flare-ups' is like 'firefighting'. This group of metaphors is supported by Semino because they accurately characterise the virus: in a fire, removing flammable material, like removing people by reducing contact and forcing them into isolation, can stop it spreading (Semino, 2020a). There are only a few of these metaphors used in this sample (Johnson = 2, Sturgeon = 3) which were adopted a few months into the pandemic. Although the sample size is small, these suggests different ways of framing the virus and how it can be stopped. However, the overriding metaphors frame the virus as a powerful aggressor which is moving through society with agency. Johnson describes it a violent agent which needs violent counter-measures; Sturgeon, on the other hand, views it as an aggressor but prioritises defensive tactics in her discourse. The next Section discusses metaphors for the restrictions which were put in place to stop the virus which will further illuminate the framing of the crisis.

3.2 Metaphors for the Restrictions

Table 8: *Metaphors used to describe and discuss the restrictions involved in the COVID-19 crisis*

Lexical field	**Metaphor**	**Examples**		**Total (raw)**		**Total (per thousand words)**	
		BJ	**NS**	**BJ use**	**NS use**	**BJ use**	**NS use**
OBJECT	ease	As lockdown eases (July)		3	0	0.31	0.00
OBJECT	relax	We relaxed the rules on meeting outside for a very specific reason (June)	And if all of that happens restrictions will have to be reimposed, rather than being relaxed (June)	3	1	0.31	0.07
OBJECT	strengthen	we are simplifying and strengthening the rules (September)		1	0	0.10	0.00
OBJECT	toughen	we are today simplifying, standardising and in some places toughening		1	0	0.10	0.00

		local rules (October)					
OBJECT	tough	That is why it is so important that we take these tough measures now (September)	I know how tough this is, but please stick with it (April)	1	5	0.10	0.33
OBJECT	hard		Remember to physically distance, I know it's really hard (August)	0	1	0.00	0.07
OBJECT	break	Breaking these rules now could undermine and reverse all the progress (July)		3	0	0.31	0.00
OBJECT	stick		And we can only do that, if people continue to stick to the rules (June)	0	11	0.00	0.72
OBJECT	calibrate	Local lockdowns will be carefully calibrated depending on the scientific and specific circumstances (July)		1	0	0.10	0.00
OBJECT	disrupt	The answer is that we are asking people to do	There is disruption for businesses	3	1	0.31	0.07

		something that is difficult and disruptive of their lives (March)	right now, across the economy (March)				
OBJECT	impact	lockdown has saved many hundreds of thousands of lives – but it has also had a devastating impact (July)	As we gather more hard data on how the virus is spreading in Scotland and on what impact these lockdown measures are having (April)	2	3	0.21	0.20
OBJECT	blow	I know that the steps we are taking will be a heavy blow to many people (August)		1	0	0.10	0.00
OBJECT	erode	No one, least of all me, wants to impose these kinds of erosions of our personal liberty (October)		1	0	0.10	0.00
OBJECT	lift	lockdown has not yet been lifted entirely (July)	All of us want to be able to lift more restrictions (June)	3	1	0.31	0.07
OBJECT	alleviate		our approach to changing, and hopefully alleviating, the lockdown	0	1	0.00	0.07

			measures (May)				
OBJECT	bear down	By bearing down on social contact and improving enforcement, we can keep schools and businesses open (September)		1	0	0.10	0.00
PLACE	into	At the same time, you'll remember that international travel plummeted as countries around the world went into lockdown (June)		2	0	0.21	0.00
PLACE	out of		we can continue to move in the right direction, out of lockdown (September)	0	6	0.00	0.39
PLACE	emerge from		approach will play a part in helping us emerge gradually from lockdown (May)	0	1	0.00	0.07
PLACE	exit		as we exit lockdown (August)	0	1	0.00	0.07

PLACE	return to	This is how we will avoid a return to full national lockdown (August)		1	0	0.10	0.00
LOCKDOWN	lockdown	Throughout the period of lockdown which started on March 23rd (May)	what impacts these lockdown measures are having (April)	19	13	1.98	0.85
MISC	follow	Because if everyone stays alert and follows the rules, we can control coronavirus (May)	it is so important to remind everyone of the advice we are asking you to follow (March)	5	6	0.52	0.39
MISC	travel corridors	we will explore the possibility of international travel corridors with countries that have low rates of infection (June)		1	0	0.10	0.00
MISC	targeted	Fourth, targeted restrictions (July)	The question then became what was the most targeted action we could take (September)	5	3	0.52	0.20
MISC	bubble	For example, if a single household or support bubble is larger than 6, they can still gather (September)		2	0	0.21	0.00

MISC	throw up	this rule of six will of course throw up difficult cases (September)		1	0	0.10	0.00

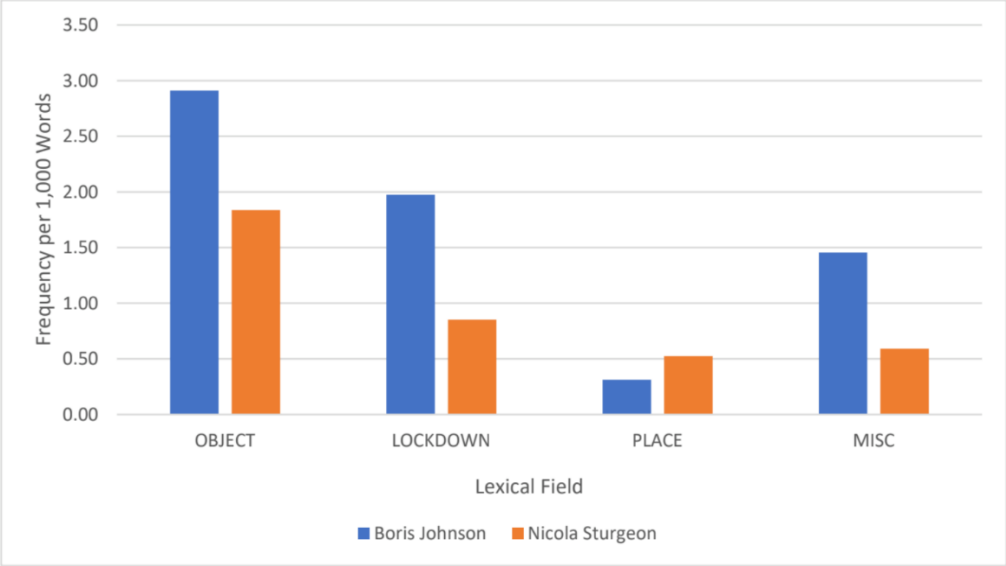


Figure 4: *The lexical fields of the metaphors in the restrictions target domain and their frequency of occurrence.*

The restrictions are what were used by the leaders to stop the effects of the virus. OBJECT metaphors are the most salient in this target domain. The metaphors within these lexical fields are all semantically similar and suggest an underlying conceptual metaphor: RESTRICTIONS ARE AN OBJECT. This conceptual metaphor is sophisticatedly developed by both speakers as they instil the metaphorical object with different features such as tangibility, weight, and volume:

Table 9: *Different metaphors in the OBJECT lexical field used to describe object features.*

Metaphor	Object Feature	Total (raw)		Total (per thousand words)	
		BJ	NS	BJ	NS
ease	Tangibility	3	0	0.31	0
relax		3	1	0.31	0.07
strengthen		1	0	0.10	0
toughen		1	0	0.10	0
tough		1	5	0.10	0.33
hard		0	1	0	0.07
break		3	0	0.31	0
flexible		0	1	0	0.07
extend		1	0	0.1	0
stick		0	11	0	0.72
blanket		1	1	0.10	0.07
	TOTAL	**15**	**29**	**1.56**	**1.25**
full	Volume	1	0	0.10	0
within		0	1	0	0.07
	TOTAL	**1**	**1**	**0.10**	**0.07**
disrupt	Weight	3	1	0.31	0.07
impact		2	3	0.21	0.20
blow		1	0	0.10	0
lift		3	1	0.31	0.07
alleviate		0	1	0	0.07
bear down		1	0	0.10	0
under		1	1	0.10	0.07
	TOTAL	**10**	**6**	**1.04**	**0.39**

The restrictions have tangibility, i.e., they can be ‘relaxed’ or ‘eased’ if society is reducing the effects of the virus, or they can be ‘strengthened’ to be made ‘hard’, so they are not ‘broken’. Sturgeon also views the restrictions as ‘sticky’, asking people to ‘stick to the rules’. The restrictions also have weight as they can ‘bear down’ on society and can be ‘lifted’; they can have an ‘impact’ and can be a hard ‘blow’. Finally, they can also have volume, e.g., lockdown can be ‘full’.

This developed frame allows both speakers to discuss the restrictions, which are abstract and complex and can affect different parts of the country in different ways, as a singular concrete object. A possible motivation behind this framing could be a reaction to the COVID-19 virus. As the virus is framed and understood as a powerful aggressor by both speakers, then something heavy, large, and strong is needed to

counteract it. Therefore, the two conceptual metaphors, COVID-19 IS A POWERFUL AGENT and RESTRICTIONS ARE AN OBJECT, interact.

Similarly, one main restriction was to place the country into 'lockdown'. This metaphor was placed its own lexical field due to its uniqueness as it did not semantically relate to other metaphors. It was still frequent in the data meaning that it could not be simply placed in MISC. It is a term that is originally referred to the enforced confinement in prison for security purposes (Poole, 2020). This framing further emphasises the COVID-19 IS A POWERFUL AGENT conceptual metaphor because society needs to be 'locked' away from the virus. This reinforces the danger of COVID-19 and its ability to overcome simple restrictions; it takes the maximum level of confinement possible to stop its effects on society.

Another interesting metaphor used by both speakers frames 'lockdown' as a place which we can go 'into', come 'out of', 'exit' or 'emerge from'. This suggests that society is moving forward, perhaps on a 'journey', and 'lockdown' is a place on that journey which society is trying to move away from but is forced to 'return to' because of the virus. There are few of these metaphors for the restrictions which are overshadowed by the OBJECT metaphors; however, they do link with the next Section which covers metaphors for the pandemic and further explores the 'journey' frame.

3.3 Metaphors for the Pandemic

Table 10: *Metaphors used to describe and discuss the COVID-19 pandemic target domain.*

Lexical field	**Metaphor**	**Examples**		**Total (raw)**		**Total (per thousand words)**	
		BJ	**NS**	**BJ use**	**NS use**	**BJ use**	**NS use**
JOURNEY	entering	We are entering a new and crucial phase (October)	phase 1, which we entered on Friday (June)	1	1	0.10	0.07
JOURNEY	move	As we move to the next stage of our fight against coronavirus (June)	we can continue to move in the right direction (August)	2	3	0.21	0.20
JOURNEY	return	allow life to return closer to normality (September)	allow a return – a very careful return – to greater	3	2	0.31	0.13

			normality (July)				
JOURNEY	close	closer to normal before Christmas (September)		2	0	0.21	0.00
JOURNEY	brake	putting on the brakes and re imposing restrictions (July)		3	0	0.31	0.00
JOURNEY	long haul		we are in this for the long haul (April)	0	1	0.00	0.07
JOURNEY	tunnel		see some light at the end of the tunnel (June)	0	1	0.00	0.07
JOURNEY	ease up		But that progress is fragile and if we ease up now (June)	0	2	0.00	0.13
JOURNEY	setback		potential setbacks in the weeks ahead (July)	0	1	0.00	0.07
DIVISION	stage	the next stage of our fight against coronavirus (June)	an early stage in this pandemic (September)	3	4	0.31	0.26
DIVISION	phase	in phase one of this disease (May)	It is important now and will be important in the next phase (May)	2	5	0.21	0.33
CONTAINER	in		we all have to play our	0	3	0.00	0.20

			part in this (September)				
MISC	hard		we know times are hard (March)	0	1	0.00	0.07

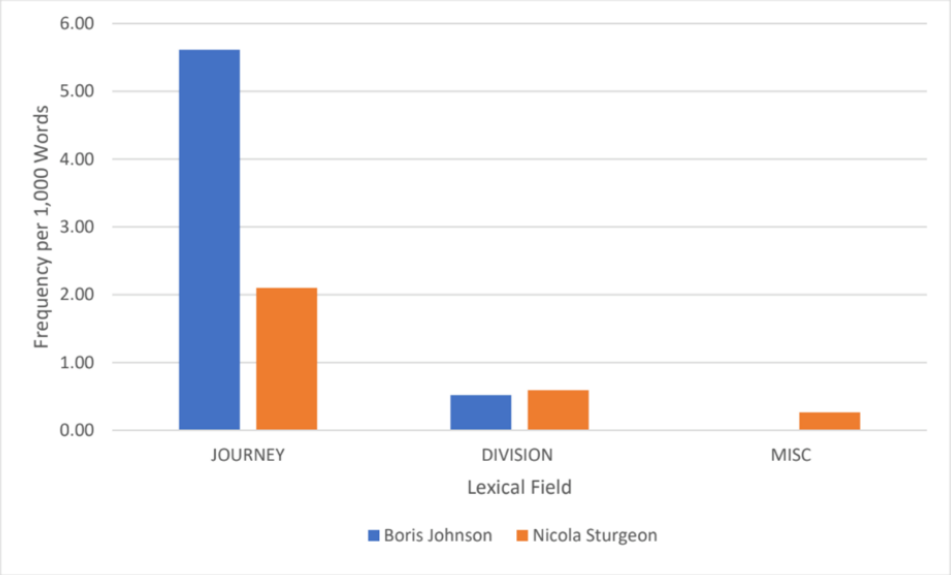


Figure 5: *The lexical fields of the metaphors in the pandemic target domain and their frequency of occurrence.*

This target domain describes the pandemic in general as a singular concept. Figure 5 shows that the pandemic is described by both speakers overwhelmingly as a 'journey'. Similar to other target domains, it is sophisticated and has many features which can be further broken down and analysed:

Table 11: *Further breakdown of the JOURNEY metaphors.*

Metaphor	Feature of Journey	Total (raw)		Total (per thousand words)	
		BJ	NS	BJ	NS
entering	Places	1	1	0.10	0.07
return		3	2	0.31	0.13
close		2	0	0.21	0
woods		1	0	0.10	0
there		4	0	0.42	0
(light at the end of the) tunnel		0	1	0	0.07
approaching		1	0	0.10	0

past		2	0	0.21	0
ahead		1	0	0.1	0
coming		1	0	0.10	0
	TOTAL	**16**	**4**	**1.66**	**0.26**
move	Movement	2	3	0.21	0.20
ease up		0	2	0	0.13
progress		10	10	1.04	0.66
through		3	1	0.31	0.07
	TOTAL	**15**	**16**	**1.56**	**1.05**
brake	Mode of Transportation	3	0	0.31	0
reverse		1	2	0.10	0.13
steps		16	3	1.66	0.2
strides		0	1	0	0.07
lose control		1	0	0.10	0
run		1	0	0.10	0
	TOTAL	**22**	**6**	**2.29**	**0.39**
road	Terrain	1	0	0.10	0
long haul		0	1	0	0.07
slog		0	3	0.00	0.20
	TOTAL	**1**	**4**	**0.10**	**0.26**
setback	Obstacles	0	1	0	0.07
	TOTAL	**0**	**1**	**0**	**0.07**
route map	Navigation	0	1	0	0.07
	TOTAL	**0**	**1**	**0**	**0.07**

From this data, it suggests that both speakers have the conceptual metaphor THE PANDEMIC IS A JOURNEY. This conceptual metaphor also contains elements of the Invariance Hypothesis of conceptual metaphor, which was developed by Lakoff as part of CMT. This is when the mapping between the source and the target domain preserves the fine structural details of the source domain (Lakoff, 1990, p. 54). In this case, the detailed features of a JOURNEY, like places seen on a journey, the mode of transport taken, the obstacles encountered, the type of the path taken, etc., are mapped onto the PANDEMIC. The effect of this is that it allows the pandemic to be almost entirely discussed in terms of a journey.

A journey involves movement in some way and both speakers understand that society is 'moving' which is expressed by equal instances of 'progress'. They are moving along a terrain which is 'long' and is a difficult 'slog'. But society is not just moving aimlessly. Like on a journey, a final destination is envisioned which is being moved towards. In this case, the final destination is 'normality' which is where society started but is venturing away from due to the virus. This is suggested by metaphors like 'return to'

normality, aiming to get ‘there’, being ‘close’ and trying to ‘get out of the woods’ and reach the ‘light at the end of the tunnel’.

There is some suggestion that Johnson develops this conceptual metaphor much more than Sturgeon as he uses more journey related metaphors in discourse surrounding the pandemic. Johnson describes different places along the journey, understanding the infection rate curve as a place which can be ‘coming’ at us. He also states that London is ‘ahead’ which suggests that different cities are on the journey and have different speeds depending on their level of infection. Additionally, he discusses modes of transport in the journey focusing on walking as shown in several instances of ‘steps’. Sturgeon still views the pandemic as a journey, but uses considerably fewer metaphors in this case. She understands the country as ‘moving’, but does not foreground elements such as place and modes of transportation. She does discuss elements of the terrain of the journey, obstacles on it and how to navigate it with a ‘route map’. The difference in the number of metaphors used is discussed in the following Section.

This conceptual metaphor also interacts with metaphors which describe restrictions as a ‘place’ (cf. Table 8). ‘Lockdown’ is frequently described as a place which people can ‘emerge from’ or go ‘into’. This suggests that on this journey, ‘lockdown’ is one of the destinations on the road to normality which people are trying to get away from but are forced to ‘return to’ because of the virus. The next Section discusses the action taken by both governments to try and ‘get back to normality’ and make the ‘journey’ as easy as possible.

3.4 Metaphors for the Action by the Governments

Table 12: *Metaphors used to describe and discuss the action by the governments during the COVID-19 crisis.*

Lexical field	**Metaphor**	**Examples**		**Total (raw)**		**Total (per thousand words)**	
		BJ	**NS**	**BJ use**	**NS use**	**BJ use**	**NS use**
MOVEMENT	bring forward	bringing forward the right measures at the right time (March)		1	0	0.10	0.00
MOVEMENT	bring in	Why bring in this very draconian measure? (March)		1	0	0.10	0.00
MOVEMENT	go	we may need to go	we are going ahead with	5	1	0.52	0.07

		further (August)	many of the Phase 2 changes (July)				
MOVEMENT	explore	we will explore the possibility of internationa l travel corridors with countries (June)		3	0	0.31	0.00
MOVEMENT	move	We will of course study the data carefully and move forward with our intention to open up (August)	the expansion of testing that I set out today is separate and distinct from our move to establish a Test, Trace, Isolate system (May)	3	1	0.31	0.07
MOVEMENT	step in		if we don't do our job, Test & Protect steps in (September)	0	2	0.00	0.13
MOVEMENT	expand		I can confirm that we will now expand that approach (May)	0	13	0.00	0.85

MOVEMENT	extend	we should extend this advice to mass gatherings as well (March)	that means we can further extend eligibility for testing (May)	3	3	0.31	0.20
MOVEMENT	tackle	everyone responsible for tackling these problems (April)		1	0	0.10	0.00
MOVEMENT	throwing	We are throwing everything at it, heart and soul, night and day (April)		1	0	0.10	0.00
MOVEMENT	push		We welcome the support the UK Governmen t has announced so far but we are pushing them to do more (March)	0	1	0.00	0.07
WEAPONS	aim	as long as the data allows, we aim to allow (May)	we are aiming to reach 8,000 by the middle	2	2	0.21	0.13

			of this month (May)				
WEAPONS	target	we have the ability to target that capacity at local areas (July)	Test and Protect enables us to be much more targeted and proportionate (September)	2	6	0.21	0.39
THEATRE	play a part	I know we are going to beat this – if each and every one of us plays our part (August)	Fundamentally we all have to play our part in this if we are going to succeed (September)	4	6	0.42	0.39
THEATRE	play		try to keep transmission under control - and what part Test & Protect plays in that (September)	0	2	0.00	0.13
THEATRE	perform		The really important role Test & Protect has to perform for us (September)	0	1	0.00	0.07

MISC	build		we are building towards being able to carry out 3,500 tests per day (April)	0	3	0.00	0.20
MISC	flow		We will also use our procuremen t systems and government contracts to keep financial support flowing (March)	0	1	0.00	0.07
MISC	face	the logistical problems we have faced in getting the right protective gear (April)		1	0	0.10	0.00
MISC	response	I will firsthand over to Chris to take us through the latest data before I set out how we are responding to it (September)	I want to provide you with a further update on Scotland's response to the Covid-19 epidemic (April)	3	2	0.31	0.13

MISC	open up	our intention to open up as soon as we possibly can (August)		1	0	0.10	0.00
MISC	scrap	That doesn't mean we're going to scrap the programme entirely (September)		1	0	0.10	0.00
MISC	put		the government takes action to put money into people's pockets (March)	0	1	0.00	0.07
MISC	moonshot	Our plan – this moonshot that I am describing (September)		1	0	0.10	0.00

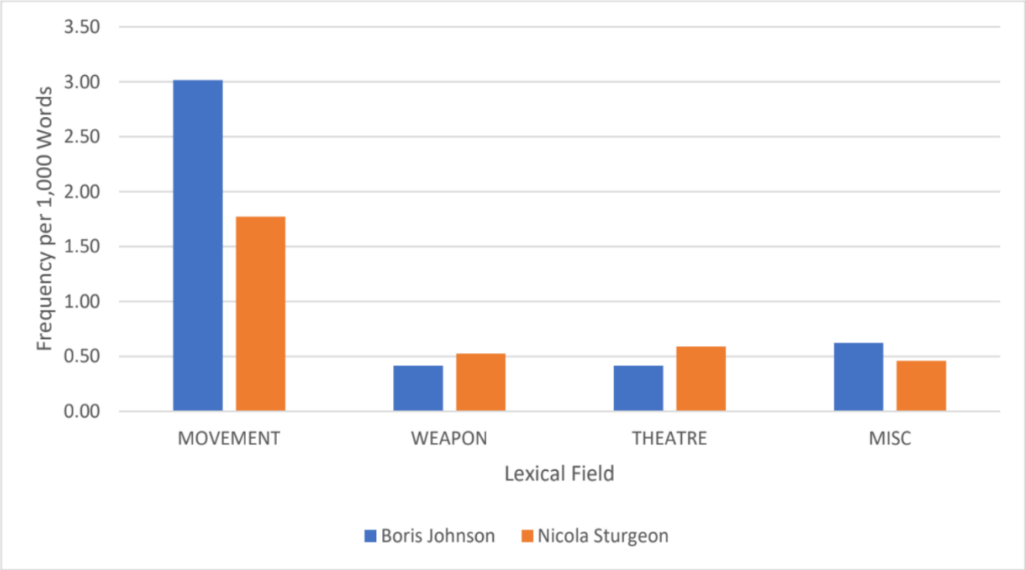


Figure 6: *The lexical fields of the metaphors in the action by the government's target domain and their frequency of occurrence.*

The largest proportion of metaphors used to discuss the action by the governments during the COVID-19 pandemic is taken up by MOVEMENT metaphors. Many of these are similar to the JOURNEY metaphors discussed in the previous Section, but they also refer to the general animated movement of the speakers:

Table 13: *Features of the MOVEMENT metaphors used by both leaders.*

Metaphor	Types of Movement	Total (raw)		Total (per thousand words)	
		BJ	**NS**	**BJ**	**NS**
go	Movement on a Journey	5	1	0.52	0.07
explore		3	0	0.31	0
move		3	1	0.31	0.07
lead		3	0	0.31	0
overcome		2	0	0.21	0
approach		5	6	0.52	0.39
step in		0	2	0	0.13
	TOTAL	**21**	**10**	**2.18**	**0.66**
work though	General Movement	1	0	0.1	0
expand		0	13	0	0.85
extend		3	3	0.31	0.20
bring forward		1	0	0.10	0
bring in		1	0	0.10	0
tackle		1	0	0.10	0
throwing		1	0	0.10	0
push		0	1	0	0.07
	TOTAL	**8**	**17**	**0.83**	**1.12**

Both leaders further describe the pandemic as some kind of 'journey' suggested in metaphors such as 'going forward' with decisions, having different 'approaches' to problems, 'moving' to make choices, and so on. There are also more general movement metaphors which do not suggest a journey but suggest some kind of animated action by both governments. They 'bring in' different measures which they then 'expand' and 'extend', suggesting physical movement and also the physicality of the measures themselves (cf. Table 9). The high frequency of these metaphors suggest that the governments are physically exerting themselves to end the crisis. The pandemic is complex and changing; therefore, physical, strenuous action is needed. This further develops the COVID-19 VIRUS IS A POWERFUL AGENT conceptual metaphor as if the virus is moving fast in society with its own freewill, the UK and Scottish Government so too have to 'move' quickly. This is also echoed in FORCE metaphors like 'tackle' and 'throwing everything at it'.

A novel use of metaphor are metaphors which relate to the theatre, such as the governments 'playing a part' and health services 'performing a role'. There is not enough data to suggest an underlying conceptual metaphor such as THE PANDEMIC IS A PLAY; however, these metaphors illuminate the metaphorical repertoire available to discuss aspects of the pandemic and strategy towards it. Weapon metaphors are also somewhat frequent, but are more general and do not suggest aggression, such as 'our aim' and 'we target'. The overriding metaphors used to describe this target domain are movement metaphors which further reinforces the conceptual metaphors THE PANDEMIC IS A JOURNEY and the COVID-19 VIRUS IS A

POWERFUL AGENT. Similar to these conceptual metaphors is the COVID-19 infection rates target domain discussed in the next Section to reveal further framing of aspects of the crisis.

3.5 Metaphors for COVID-19 Infection Rates

Table 14: *Metaphors used to describe and discuss the COVID-19 infection rates.*

Lexical field	Metaphor	Examples		Total (raw)		Total (per thousand words)	
		BJ	NS	BJ use	NS use	BJ use	NS use
NATURAL WORLD	peak	the rate of transmission in the UK has significantly fallen from its peak (June)	the peak of the outbreak (June)	12	1	1.25	0.07
NATURAL WORLD	through	We have come through the peak (April)		2	0	0.21	0.00
NATURAL WORLD	under	we've come under what could have been a vast peak (April)		1	0	0.10	0.00
NATURAL WORLD	slope	we are on the downward slope (April)		1	0	0.10	0.00
NATURAL WORLD	mountain	run slap into a second and even bigger mountain (April)		1	0	0.10	0.00

NATURAL WORLD	wave	prevent a second wave of the virus reaching the UK (June)		3	0	0.31	0.00
NATURAL WORLD	spike	the spike in Leicester (July)		2	0	0.21	0.00
NATURAL WORLD	growth	the fast growth part of the upward curve (March)		1	0	0.10	0.00
MOVEMENT	slow	slowing the spread of the disease (March)	slow down the spread of this virus (March)	4	3	0.42	0.20
MOVEMENT	creeping up	those numbers creeping up (July)		1	0	0.10	0.00
MOVEMENT	pace	the virus is now gathering pace (July)		1	0	0.10	0.00
MOVEMENT	other direction		gone in the other direction today (July)	1	0	0.00	0.07
MOVEMENT	spring up		the number of cases and the number of clusters and the number of outbreaks that will spring up	0	1	0.00	0.07

			(September)				
HEIGHT	fall	infection rate is falling (April)		3	0	0.31	0.00
HEIGHT	lift	lift R or the reproduction rate of that disease back above one (April)		1	0	0.10	0.00
HEIGHT	drive down	drive this virus down (September)	we keep driving the overall level of COVID infections down (June)	1	1	0.10	0.07
HEIGHT	keep down	that keeping the R down is going to be absolutely vital to us recovery (April)		4	0	0.42	0.00
HEIGHT	bring down	bring the R level down (May)		3	0	0.31	0.00
MISC	wake-up call		the situation… I really think should be a wake-up call for all of us (September)	0	1	0.00	0.07

MISC	flashing	These tables are flashing at us like dashboard warnings in a passenger jet (October)		1	0	0.10	0.00

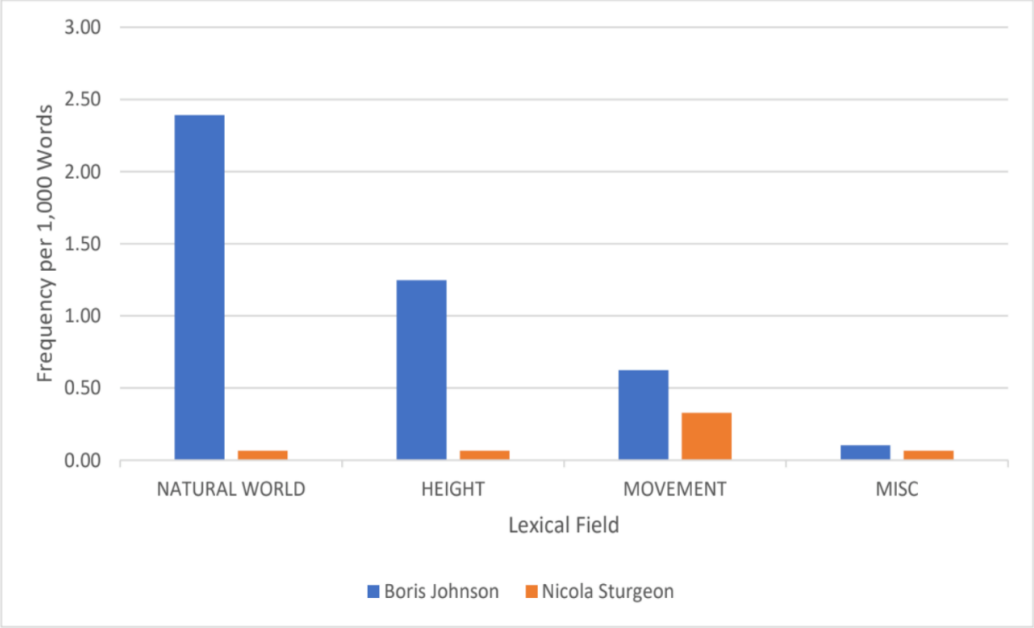


Figure 7: *The lexical fields of the metaphors in the COVID-19 infection rates target domain and their frequency of occurrence.*

COVID-19 Infection Rates is its own target domain as it is different from the biological virus itself and instead refers to the numbers which indicate its presence in society. Both speakers converge is their understanding of the infection rates as 'moving', e.g., they can be 'slow', they can have 'pace', they can go in the 'other direction' and so on. This is similar to the COVID-19 IS A POWERFUL AGENT conceptual metaphor as it frames COVID-19 as having agency in terms of movement through space. This is understandable as the infection rates appear like they are 'moving' because of reductions and increases in number; however, this particular framing implies the numbers have physical movement which suggests that the COVID-19 virus itself is moving.

Sturgeon is noticeably absent in this target domain, whereas Johnson uses many metaphors to discuss the infection rates. The most prominent are NATURAL WORLD metaphors, such as 'peak' and 'waves' of infections; 'slopes', 'mountains' and 'spikes'; and the ability for numbers to 'grow'. Johnson also understands the infection rates as having height, i.e., the need to be 'kept down' and be made to 'fall'. This

suggests that Johnson is metaphorically framing literal graphs and curves of the pandemic as natural objects like mountains, whereas Sturgeon is does not discuss these aspects metaphorically and presents the infection rates much more literally.

The final target domain under discussion here is the scientific data target domain which will provide the last example of framing of the crisis from both speakers[41].

3.6 Metaphors for the Scientific Data

Table 15: *Metaphors used to describe and discuss the scientific data target domain.*

Lexical field	Metaphor	Examples		Total (raw)		Total (per thousand words)	
		BJ	NS	BJ use	NS use	BJ use	NS use
MOVEMENT	guide	we are being guided by the science (April)	the information coming through Test & Protect guided us in the decisions we took (September)	1	2	0.10	0.13
MOVEMENT	guidance	that guidance remains unchanged (August)		4	14	0.42	0.92
MOVEMENT	led		that analysis and intelligence led us (September)	0	2	0.00	0.13
MOVEMENT	follow	close down premises and cancel events which are not following COVID Secure guidance (August)	as we go into this weekend and beyond, need to follow the guidance (March)	1	11	0.10	0.72

[41] In this target domain and the following, tables which further break down the metaphors into their different qualities, like in previous Sections, are not provided due to the small number of the metaphors in each target domain.

MOVEMENT	driven		they are driven by the evidence (September)	0	3	0.00	0.20
OBJECT	based on	everything we do is based scrupulously on the best scientific advice (March)		2	0	0.21	0.00
OBJECT	hard		As we gather more hard data (April)	0	1	0.00	0.07
COMMUNICATION	says	only if the data says it safe (May)		1	0	0.10	0.00
COMMUNICATION	tells	the data tells us is driving the current increase in cases (August)	this week's statistics again tell of the real and sustained progress that we are making (July)	1	2	0.10	0.13
MISC	in light of	look again at the measures we have in place nationally in light of the data (August)		1	0	0.10	0.00

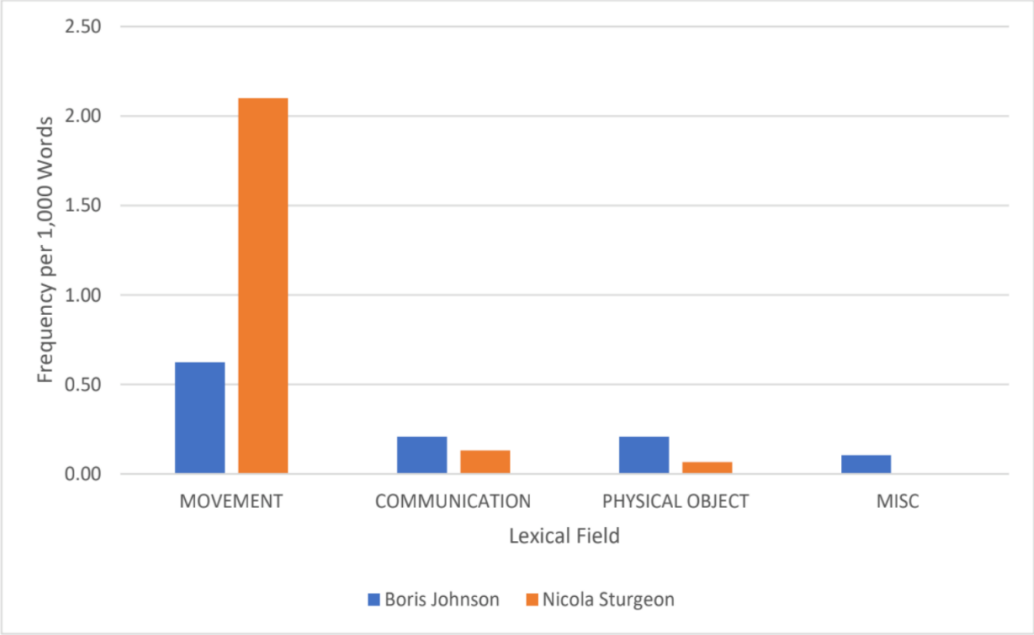


Figure 8: *The lexical fields of the metaphors in the scientific data target domain and their frequency of occurrence.*

The main metaphors used by both speakers understand scientific data as 'moving'. Specifically, it is described as 'guidance' which people should 'follow'. It also 'guides', 'drives', and 'leads' both governments' decision making. It also has elements of communication, it can 'say' what is safe and it 'tells' policy makers what to do. These metaphors personify scientific data as a wise guide who is leading us through the pandemic. In conceptual metaphor terms, this can be written as SCIENTIFIC DATA IS A GUIDE. This is fundamentally linked to THE PANDEMIC IS A JOURNEY conceptual metaphor: on this journey the 'data' is a guide which we are 'following'.

The potential impact of these metaphors, as well as overall similarities and differences, are discussed in the following Section.

4 Shared Metaphors

The systematicity and frequency of similar metaphors used to discuss each target domain suggests that there are several underlying conceptual metaphors shared by both speakers:

- THE COVID-19 VIRUS IS A POWERFUL AGENT
- RESTRICTIONS ARE AN OBJECT
- THE PANDEMIC IS A JOURNEY
- SCIENTIFIC DATA IS A GUIDE

These conceptual metaphors shared by both speakers generate the metaphors found in their speech and frame aspects of the pandemic in similar ways. Both speakers understand the pandemic as a 'journey' where everyone is trying to get back to the final destination of 'normality'. They have been taken away from

normality due to the crisis, but they struggle to get back there and follow scientific data as a 'guide' to try and 'return'. However, negatively affecting this move 'forward towards' normality is the COVID-19 virus. It is described as a powerful agent with freewill which is physically harming both nations. In order to stop the COVID-19 virus' negative effects, each government has weighty and powerful restrictions which have an 'impact'. The effect of this overall framing causes the complex and frightening experience of the crisis to be grounded in everyday lived experience. Something like a virus is difficult to see and understand, whereas concepts like weight and power, instilled in the virus and the restrictions, can be more easily understood. Moreover, describing the pandemic as a 'journey' gives it a destination, i.e., 'normality', and a way to get there, such as by 'moving forward'. The lived experience of a 'journey' allows the crisis, which seems to be continuous and without end, to have a general goal in mind to 'keep people going' in a sense.

This essay has not conducted any experimental study to examine the potential impact of these metaphors on public perception. However, previous research has revealed some findings which can be applied to this data to provide a wider discussion about the impact of these metaphors in society. Research on journey metaphors in the context of a disease like cancer suggests that they can have positive effects on listeners and readers. One study found that there was a correlation between 'journey' metaphors and someone being more likely to make peace with their illness (Hendricks et al., 2018, p. 271). Journey metaphors also contain the idea sharing, as in 'sharing the journey' with someone (Semino et al., 2017, p. 64). When applied to a difficult situation, like a debilitating illness or a global pandemic, journey metaphors foreground the aspect that we are on the road 'together' in a shared experience. This is especially the case considering that scientific data is 'on the journey' with us and leads us along. As a result of these positive effects, journey metaphors have been praised by some metaphor researchers (Nerlich, 2020b).

4.1 Divergent Metaphors

There is a range of subtle differences between the speakers. These can be divided into differences in the number of metaphors and differences in the type of metaphors. In terms of the number of metaphors between the two speakers, Johnson generally uses more metaphors for each lexical field across all target domains (except from when scientific guidance is described as a 'guide' as Sturgeon is much more frequent in this area). This variability could be attributed to the overall difference in metaphor use between the speakers as Johnson uses many more metaphors than Sturgeon (35.1 per 1,000 words compared with 21.5 per 1,000 words). Also, it has been noted that Johnson's language is noticeably more elaborate than other politicians and contains higher use of metaphor (Hayward, 2019). It is therefore difficult to discuss the potential impact of metaphor frequency due to this general variability and difference in speech style.

What is more significant and clearer to discuss is the differences in the type of metaphors chosen by both speakers. One such difference is that Johnson choses to use metaphors to describe the COVID-19 infection rates, such as 'peak', 'mountain', slope' etc., whereas Sturgeon chooses not to metaphorically frame this target domain. COVID-19 infection rates are sometimes complex and can be difficult to understand. Metaphor is frequently used to explain scientific concepts like this to a lay audience (Kampourakis, 2016, p. 947). Johnson, although not a scientist, is following this convention by describing the infection rates in terms of natural objects. However, it has been noted that some metaphors for scientific concepts can potentially mislead listeners. Kampourakis (2016, p. 947) notes that because of the ability of metaphor to foreground some information and background others, the public may 'overlook' aspects of a concept not included in the metaphor. This was demonstrated in a study on metaphors for the climate where

researchers found that high school pupils sometimes came to wrong conclusions about aspects of the climate because the metaphors they read did not include all the necessary information (Deignan et al., 2019, p. 399).

This research can be applied to Johnson's use of natural world metaphors. For example, the 'peak' of a mountain is often considered a goal because it is the limit and anything beyond it is smaller and easier in comparison. The coronavirus, on the other hand, has shown that even past the 'peak' of infections, the pandemic still poses numerous threats, such as economic challenges and the ability to find a vaccine (and, more recently, the effect of new mutations). Moreover, natural peaks of a mountain cannot rise again. Passing the 'peak' is a milestone and only downhill remains. As Johnson states, we are on the 'downward slope' (cf. Table 14). However, in the case of COVID-19, it has been shown that there can be multiple 'peaks' and numbers can start to rise again. Johnson's framing of the infection rates as a 'mountain' could have potentially negative effects on public understanding of the crisis as it overlooks these crucial aspects of virus. Sturgeon's notable avoidance of these type of metaphors suggests she discusses infection rates in more literal terms and avoids this potential impact.

Another key difference between the speakers is how they use violence metaphors: Johnson foregrounds violence when framing the COVID-19 virus as a powerful agent, whereas Sturgeon focuses on defensive elements. Johnson's metaphors frame the virus as an aggressor and the action taken to defeat it as violent. The effect of this is that it generally frames the pandemic as a 'war' between the virus and the people, as shown in the numerous instances of 'fight'.

War metaphors are used commonly in public discourse surrounding viruses and disease (Flusberg et al., 2018, p. 1). These violent metaphors have potential harmful effects as evidence has shown that they can negatively affect people's emotions. For example, they can make people feel guilty if they catch a disease as they have not 'fought hard enough' (Flusberg et al., 2018, p. 9). Experimental work on cancer has shown that violent metaphors can make cancer treatment seem more difficult which caused people to become fatalistic about the disease. This negatively impacted people's behaviour as it made them less likely to take steps to avoid developing the disease (Hauser & Schwarz, 2020, p. 1703). Extending this to the pandemic, constant aggressive framing of the pandemic as a violent 'fight' could mean that people take fewer steps to prevent catching and spreading the virus. This evidence highlights the potential negative effects of war metaphors and could explain why Johnson has been repeatedly criticised for his use of aggressive violence metaphors by political commentators (Tisdall, 2020; Clark, 2020), metaphor researchers in the #ReframeCOVID project (Semino, 2020b; Nancy, 2020) and health professionals (Marron et al., 2020, p. 625).

On the other hand, war metaphors can 'motivate people to pay attention' and take action (Flusberg et al., 2018, p. 6). In the pandemic, everyday actions had to be changed immediately. Violence metaphors could have stimulated this by emulating a violent situation which heightens the senses. Moreover, experimental work on cancer has found that violence metaphors can sometimes be empowering, e.g., people can be 'fighters' against the disease which emphasises feelings of solidarity (Semino et al., 2017, p. 63).

This shows that violence metaphors are not entirely negative, nor positive. However, the widespread criticism of violent metaphors which have been used by Johnson do suggest that their impact has been negative. Sturgeon's choice not to use aggressive violent metaphors could suggest that she recognised these negative effects, hence her decision to foreground defensive elements.

4.2 Summary

Both speakers generally frame aspects of the pandemic in similar ways, such that we are on a 'journey' towards 'normality' and 'follow' scientific data to get there but are disrupted by the powerful COVID-19 virus and must use restrictions to stop it.

Despite some subtle differences in the types of metaphors chosen, such as Johnson's more violent framing of the virus and natural world metaphors to describe infection rates, this similar framing subsists despite the two leaders having different policy approaches to the pandemic.

Previous research on these metaphors suggest that journey metaphors have a positive impact as they promote features such as togetherness and provide goals. This can be reassuring and can create a sense of peacefulness and direction. On the other hand, war and violence metaphors, although having potentially positive implications, have an overall more negative effect on the public perception as they have been shown to harm emotion and behaviour. Natural world metaphors may also mislead as they do not cover all major aspects about the virus and its rate of infection.

5 Conclusion

This essay has extracted the metaphors used by Nicola Sturgeon and Boris Johnson to discuss aspects of the COVID-19 pandemic between March and October 2020. From these quantitative results, I have discussed how the speakers frame aspects of the crisis: they both understand the pandemic as a journey towards normality where society follows scientific data and uses robust restrictions as a counter against the powerful COVID-19 virus which is instilled with force and weight. A divergence in metaphor use was also found between the two speakers: Johnson uses more violent metaphors which foreground aggression against the virus, whereas Sturgeon foregrounds defence. Johnson also frames the infection rates using natural world metaphors, whereas Sturgeon uses few metaphors to discuss this concept. Finally, I grounded the discussion of these framing effects in previous metaphors research to suggest their potential impact on public perception of the crisis.

6 Further Research

As mentioned in 4.1, determining the effect of metaphors on reasoning and thought of listeners is difficult without an experimental study. This thesis has quantitatively identified the metaphors from both speakers and found the general framing of aspects of the crisis as well as some subtle differences. Future research can then use this evidence to construct experimental studies, similar to Thibodeau and Boroditsky's (2011) study on CRIME IS A VIRUS/BEAST, which examines the effects of the framing on participants. Of particular interest is the difference between describing the COVID-19 virus as something that needs to be 'fought' and something which needs to be 'protected' against. The significance of the journey frame also needs further investigation. It has been argued that the journey frame is beneficial, however, it may also negatively affect people's behaviour. For example, viewing the pandemic as a 'journey' may elicit different 'paths/routes' to get back to normality. Also, constant framing of restrictions as 'tough' and 'hard' which 'bear down' on society could have negative impact on the emotions of people if they are used continuously throughout time.

A further avenue of research created by this project is to compare this data from Scotland and the UK to other counties to understand different framing techniques around the world. This could illuminate cultural differences in understanding the pandemic and its major aspects. For example, other countries have

pushed the war framing, such as the US executive who declared 'war' on the virus (Bates 2020: 5), whereas other countries, like Germany, avoid this framing entirely (Paulus 2020).

Finally, there was no space in this study to discuss the effect of time on these metaphors. However, the data does allow for this analysis as it charts the number of metaphors used across each month (cf. Figure 2). A potential future study could examine if metaphors develop and change as the pandemic progresses. This could suggest a shifting reasoning. There is already evidence of this happening, for example the 'journey' has now become a 'race' for a vaccine (BBC News 2021).

7 References

Atanasova, D. (2018). "Keep moving forward. LEFT RIGHT LEFT": A critical metaphor analysis and addressivity analysis of personal and professional obesity blogs. *Discourse, Context & Media,* 25, 5-12.

Bates, B. R. (2020). The (In)Appropriateness of the WAR Metaphor in Response to SARS-CoV-2: A Rapid Analysis of Donald J. Trump's Rhetoric. *Frontiers in Communication,* 50(5), 1-12.

BBC News. (2020a). Coronavirus: Boris Johnson's address to the nation in full. https://www.bbc.co.uk/news/uk-52011928 (9 February 2021).

BBC News. (2020b). Covid-19: UK could face 50,000 cases a day by October without action – Vallance. https://www.bbc.co.uk/news/uk-54234084 (9 February 2021).

BBC News. (2021). Sturgeon: Vaccination programme needs to win the race. https://www.bbc.co.uk/news/av/uk-scotland-55544717 (18 February 2021).

Beck, C. T. (2016). Post-traumatic Stress Disorder After Birth: A Metaphor Analysis. *The American Journal of Maternal/Child Nursing,* 41(2), 76-83.

Boeynaems, A., Burgers, C., Konijn, E. A., & Steen, G. J. (2017). The Effects of Metaphorical Framing on Political Persuasion: A Systematic Literature Review. *Metaphor and Symbol*, 32(2), 118-134.

Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition*, 75, 1-28.

Boyle, D. (2020). How viewing tables for No10's coronavirus briefings have wavered from record highs of 27.5m to lows of 2m viewers as ministers scrap Saturday and Sunday addresses. Retrieved 09/02/2021 from <https://www.dailymail.co.uk/news/article-8393215/How-Number-10s-Covid-weekend-briefings-lost-25-million-viewers.html>

Charteris-Black, J. (2004). *Corpus Approaches to Critical Metaphor Analysis*. Basingstoke: Palgrave Macmillan.

Charteris-Black, J. (2011). *Politicians and Rhetoric: The Persuasive Power of Metaphor,* 2nd Ed. Basingstoke: Palgrave Macmillan.

Cheng, W. & Ho, J. (2017). A Corpus Study of Bank Financial Analyst Reports: Semantic Fields and Metaphors. *International Journal of Business Communication,* 54(3), 258-282.

Chircop, D. & Scerri, J. (2018). The use of metaphors in non-Hodgkin's lymphoma patients undergoing chemotherapy. *Journal of Advanced Nursing,* 74(11), 2622-2629.

Clark, E. (2020). Muddled Metaphors. Retrieved 01/02/2021 from <https://fabians.org.uk/muddled-metaphors/>

Cox, P. (2020). Fires, orchestras, parachutes. Some other ways to describe coronavirus — besides war. Retrieved 09/02/2021 from <https://www.pri.org/stories/2020-04-28/fires-orchestras-parachutes-some-other-ways-describe-coronavirus-besides-war>

Croft, W. & Alan Cruse, D. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Deignan, A. (2005). *Metaphor and Corpus Linguistics*. Amsterdam: John Benjamins.

Deignan, A., Semino, E. & Paul, S. (2019). Metaphors of Climate Science in Three Genres: Research Articles, Educational Texts, and Secondary School Student Talk. *Applied Linguistics,* 40(2). 379-403.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication,* 43(4). 51–58.

Flockhart, G. (2020). BBC poll suggests Nicola Sturgeon has handled Covid-19 pandemic better than Boris Johnson. Retrieved 09/02/2021 from <https://www.scotsman.com/health/coronavirus/bbc-poll-suggests-nicola-sturgeon-has-handled-covid-19-pandemic-better-boris-johnson-3040619>

Flusberg, S. J., Matlock, T. & Thibodeau, P. H. (2018). War Metaphors in Public Discourse. *Metaphor and Symbol,* 33(1), 1-18.

Gibbs, R. (2011). Evaluating Conceptual Metaphor Theory. *Discourse Processes,* 48, 529-562.

Hannan, R. (2020). Why we shouldn't use war metaphors to talk about healthcare. Retrieved 09/02/2021 from <https://www.thersa.org/blog/2020/05/war-language-healthcare>

Hauser, D. J. & Schwarz, N. (2015). The War on Prevention: Bellicose Cancer Metaphors Hurt (Some) Prevention Intentions. *Personality and Social Psychology Bulletin,* 41(1), 66-77.

Hauser D. J. & Schwarz, N. (2020). The War on Prevention II: Battle Metaphors Undermine Cancer Treatment and Prevention and Do Not Increase Vigilance. *Health Communication,* 35(13), 1698-1704.

Hayward, F. (2019). Bojo: Orator, Etonian and Mountebank. Retrieved 01/02/2021 from <https://www.nytimes.com/reuters/2019/07/23/world/europe/23reuters-britain-eu-leader-johnson-language.html>

Heffernan, V. (2020). Metaphors Matter in a Time of Pandemic. Retrieved 15/01/2021 from <https://www.wired.com/story/metaphors-matter-in-pandemic-coronavirus/>

Hendricks, R. K., Demjén, Z., Semino, E. & Boroditsky, L. (2018). Emotional Implications of Metaphor: Consequences of Metaphor Framing for Mindset about Cancer. *Metaphor and Symbol,* 33(4), 267-279.

Jaworska, S. (2017). Metaphors We Travel by: A Corpus-Assisted Study of Metaphors in Promotional Tourism Discourse. *Metaphor and Symbol,* 32, 161-177.

Johnson, B. (2020a). Prime Minister's statement on coronavirus (COVID-19): 16 March 2020. Retrieved 09/02/2021 from <https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-16-march-2020>

2324650 91

Johnson, B. (2020b). Prime Minister's statement on coronavirus (COVID-19): 30 April 2020. Retrieved 09/02/2021 from <https://www.gov.uk/government/news/prime-ministers-statement-on-coronavirus-covid-19-30-april-2020>

Johnson, B. (2020c). Prime Minister's statement on coronavirus (COVID-19): 11 May 2020. Retrieved 09/02/2021 from <https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-11-may-2020>

Johnson, B. (2020d). Prime Minister's statement on coronavirus (COVID-19): 3 June 2020. Retrieved 20/12/2020 from <https://www.gov.uk/government/speeches/pm-statement-at-the-coronavirus-press-conference-3-june-2020>

Johnson, B. (2020e). Prime Minister's statement on coronavirus (COVID-19): 3 July 2020. Retrieved 09/02/2021 from <https://www.gov.uk/government/speeches/prime-ministers-statement-on-coronavirus-covid-19-3-july-2020>

Johnson, B. (2020f). Prime Minister's statement on coronavirus (COVID-19): 31 July 2020. Retrieved 09/02/2021 from <https://www.gov.uk/government/speeches/prime-ministers-statement-on-coronavirus-covid-19-31-july-2020>

Johnson, B. (2020g). Prime Minister's statement on coronavirus (COVID-19): 9 September. 2020. Retrieved 09/02/2021 from <https://www.gov.uk/government/speeches/pm-press-conference-statement-9-september-2020>

Johnson, B. (2020h). Prime Minister's statement on coronavirus (COVID-19): 12 October 2020. Retrieved 09/02/2021 from <https://www.gov.uk/government/speeches/prime-ministers-statement-on-coronavirus-covid-19-12-october-2020>

Lakoff, G. & Johnson, M. (1980). *Metaphors we Live By*. Chicago: University of Chicago Press.

Lakoff, G. (1990). The Invariance Hypothesis: is abstract reason based on image-schemas*? Cognitive Linguistics*, 1(1), 39-74.

Lakoff, G. (1991). Metaphor and War: The Metaphor System Used to Justify War in the Gulf. *Peace Research,* 23, 25-32.

Lakoff, G. (2016). Language and Emotion. *Emotion Review,* 8(3), 269-273.

Kampourakis, K. (2016). The Bad Use of Metaphors and the Use of Bad Metaphors. *Science & Education*, 25, 947-949.

Lawn, S., Delany, T., Pulvirenti, M., Smith, A. & McMillan, J. (2016). Examining the use of metaphors to understand the experience of community treatment orders for patients and mental health workers. *BMC Psychiatry,* 16(82), 1-16.

López, A.M.R. & Llopis, M.A.O. (2010). Metaphorical pattern analysis in financial texts: Framing the crisis in positive or negative metaphorical terms. *Journal of Pragmatics,* 42, 3300-3313.

Marron, J.M., Dizon, D.S., Symington B., Thompson, M.A. & Rosenberg, R.A. (2020). Waging War on War Metaphors in Cancer and COVID-19. *JCO Oncology Practice,* 16(10), 624-627.

Macnab, S. (2020). Boris Johnson and Nicola Sturgeon pledge 'united' approach after talks. Retrieved 09/02/2021 from <https://www.scotsman.com/news/politics/boris-johnson-and-nicola-sturgeon-pledge-united-approach-after-talks-2978873>

Microsoft Corporation. (2016). Microsoft Excel. [Computer Software] https://office.microsoft.com/excel

Morris, M.W., Sheldon, O.J., Ames, D.R. & Young, M.J. (2007). Metaphors and the market: Consequences and preconditions of agent and object metaphors in stock market commentary. *Organizational Behaviour and Human Decision Processes,* 102(2), 174-192.

Nancy, S.L. (2020). #ReframeCovid at RaAM 2020Virtual. Retrieved 09/02/2021 from <https://eng.inn.no/conferences/previously-held-conferences/raam-2020/keynotes-and-plenaries/reframecovid-at-raam-2020virtual>

Nasirci, H. & Sadik, F. (2018). Investigating primary school teachers' perception about democracy through metaphor analysis. *European Journal of Educational Research,* 7(1), 121-132.

Nerlich, B. (2020a). Metaphors in the time of coronavirus. Retrieved 19/01/2021 from <https://blogs.nottingham.ac.uk/makingsciencepublic/2020/03/17/metaphors-in-the-time-of-coronavirus/>

Nerlich, B. (2020b). Being on a journey while staying at home: More about corona metaphors. Retrieved 18/02/2021 from <https://blogs.nottingham.ac.uk/makingsciencepublic/2020/04/10/being-on-a-journey-while-staying-at-home-more-about-corona-metaphors/>

Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N. & Frieder, O. (2013). Metaphor Identification in Large Texts Corpora. *PLoS ONE,* 8(4).

Oxford English Dictionary Editorial. (2020). Using corpora to track the language of Covid-19: update 2. 15/01/2021 <https://public.oed.com/blog/using-corpora-to-track-the-language-of-covid-19-update-2/>

Oxford English Dictionary (OED). (2020a). Retrieved 17/12/2020 from <https://www.oed.com/>

Oxford English Dictionary (OED). (2020b). fall, v. Retrieved 17/12/2020 <https://www.oed.com/view/Entry/67829?rskey=H15WQd&result=3>

Oxford English Dictionary (OED). (2020c). peak, n.2 and adj. Retrieved 17/12/2020 from <https://www.oed.com/view/Entry/139282?rskey=D0DoY7&result=2&isAdvanced=false>

Oxford English Dictionary (OED). (2020d). step, n.1. Retrieved 17/12/2020 from <https://www.oed.com/view/Entry/189836?rskey=9EbALU&result=2&isAdvanced=false>

Oxford English Dictionary (OED). (2020e). flare, v. 17/12/2020 <https://www.oed.com/view/Entry/71121?rskey=NsoF7H&result=1&isAdvanced=false>

Oxford English Dictionary (OED). (2020f). emerge, v.1. Retrieved 17/12/2020 from <https://www.oed.com/view/Entry/61126?rskey=gTZJMQ&result=2>

Oxford English Dictionary (OED). (2020g). transmit, v. Retrieved 17/12/2020 from <https://www.oed.com/view/Entry/204911?rskey=ykOKzR&result=2>

Oxford English Dictionary (OED). (2020h). stop, v. Retrieved 17/12/2020 from <https://www.oed.com/viewdictionaryentry/Entry/190889>

Oxford English Dictionary (OED). (2020i). sacrifice, n. Retrieved 21/12/2020 from <https://www-oed-com.ezproxy.lib.gla.ac.uk/view/Entry/169571?rskey=Mlf0Zt&result=1>

Oxford English Dictionary (OED). (2020j). engulf | ingulf, v. Retrieved 11/01/2021 from <https://www-oed-com.ezproxy.lib.gla.ac.uk/view/Entry/62331?redirectedFrom=engulf>

Paulus, D. (2020). How politicians talk about coronavirus in Germany, where war metaphors are avoided. Retrieved 15/02/2021 from <https://theconversation.com/how-politicians-talk-about-coronavirus-in-germany-where-war-metaphors-are-avoided-137427>

Pérez-Sobrino, P. (2014). Let's Talk Music: A Corpus-Based Account of Musical Motion. *Metaphor and Symbol,* 29, 298-315.

Poole, S. (2020). From barges to barricades: the changing meaning of 'lockdown'. Retrieved 05/02/2021 from <https://www.theguardian.com/books/2020/apr/02/changing-meaning-of-lockdown>

Pragglejaz Group. (2007). MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, 22(1). 1-39.

Saeed, J. (2016). *Semantics.* 4th Ed. West Sussex: Blackwell.

Scottish Parliament Information Centre. (2021). Timeline of Coronavirus (COVID-19) in Scotland. Retrieved 23/02/2021 from <https://spice-spotlight.scot/2021/02/19/timeline-of-coronavirus-covid-19-in-scotland/>

Semino, E. (2008). *Metaphor in Discourse*. New York: Cambridge University Press.

Semino, E. (2020a). 'A fire raging': Why fire metaphors work well for Covid-19. Retrieved 12/01/2021 from <http://cass.lancs.ac.uk/a-fire-raging-why-fire-metaphors-work-well-for-covid-

19/?utm_source=Twitter&utm_medium=social&utm_campaign=SocialSignIn%20%20July%201%202020%20by%20Elena%20Semino%20(accessed%208th%20September%202020)>

Semino, E. (2020b). Beyond the battle, far from the frontline: a call for alternative ways of talking about Covid-19. Retrieved 15/01/2021 from <https://www.lancaster.ac.uk/linguistics/news/beyond-the-battle-far-from-the-frontline-a-call-for-alternative-ways-of-talking-about-covid-19>

Semino, E., Heywood, J. & Short, M. (2004). Methodological problems in the analysis of metaphors in a corpus of conversations about cancer. *Journal of Pragmatics,* 36, 1272-1294.

Semino, E., Demjén, Z. & Demmen, J., Koller, V., Payne, S., Hardie, A. & Rayson, P. (2017). The online use of Violence and Journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study. *BMJ Supportive & Palliative Care*, 7, 60–66.

Semino, E., Demjén, Z. & Demmen, J. (2018). An Integrated Approach to Metaphor and Framing in Cognition, Discourse, and Practice, with an Application to Metaphors for Cancer. Applied Linguistics 39(5). 625-645.

Semino, E., Demjén, Z., Hardie, A., Payne, S. & Rayson, P. (2018). *Metaphor, Cancer and the End of Life: A Corpus-Based Study*. New York: Routledge.

Steen, G. J., Dorst, A.G., J., Herrmann, B., Kaal, A., Krennmayr, T. & Pasma, T. (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. Amsterdam: John Benjamins.

Sturgeon, N. (2020a). Coronavirus (COVID-19) update: First Minister's speech 20 March 2020. Retrieved 09/02/2021 from <https://www.gov.scot/publications/first-minister-covid-19-update/>

Sturgeon, N. (2020b). Coronavirus (COVID-19) update: First Minister's speech 2 April 2020. Retrieved 09/02/2021 from <https://www.gov.scot/publications/first-minister-covid-19-update-10/>

Sturgeon, N. (2020c). Coronavirus (COVID-19) update: First Minister's speech 1 May 2020. Retrieved 09/02/2021 from <https://www.gov.scot/publications/coronavirus-covid-19-update-first-ministers-speech-1-2020/>

Sturgeon, N. (2020d). Coronavirus (COVID-19) update: First Minister's speech 1 June 2020. Retrieved 09/02/2021 from <https://www.gov.scot/publications/coronavirus-covid-19-update-first-ministers-speech-01-june-2020/>

Sturgeon, N. (2020e). Coronavirus (COVID-19) update: First Minister's speech 1 July 2020. Retrieved 09/02/2021 from <https://www.gov.scot/publications/coronavirus-covid-19-update-first-ministers-speech-1-july-2020/>

Sturgeon, N. (2020f). Coronavirus (COVID-19) update: First Minister's speech 3 August 2020. Retrieved 09/02/2021 from <https://www.gov.scot/publications/coronavirus-covid-19-update-first-ministers-speech-3-august-2020/>

Sturgeon, N. (2020g). Coronavirus (COVID-19) update: First Minister's speech 3 September 2020. Retrieved 09/02/2021 from <https://www.gov.scot/publications/coronavirus-covid-19-update-first-ministers-speech-3-september-2020/>

Sturgeon, N. (2020h). Coronavirus (COVID-19) update: First Minister's speech 2 October 2020. Retrieved 09/02/2021 from <https://www.gov.scot/publications/coronavirus-covid-19-update-first-ministers-speech-2-october-2020/>

Sun, S.A. (2020). A Corpus-based Analysis of Frames and Metaphors in the Media Reports of COVID-19 [Conference Presentation]. Retrieved 09/02/2021 from <https://corporadiscourse.com/covid-19-panel-videos/>

Tisdall, S. (2020). Lay off those war metaphors, world leaders. You could be the next casualty. Retrieved 01/02/2021 from <https://www.theguardian.com/commentisfree/2020/mar/21/donald-trump-boris-johnson-coronavirus>

Thibodeau, P.H. & Boroditsky, L. (2011). Metaphors We Think With: The Role of Metaphor in Reasoning. *PLoS ONE,* 6(2).

Thibodeau, P.H., Hendricks, R.K. & Boroditsky, L. (2017). How Linguistic Metaphor Scaffolds Reasoning. *Trends in Cognitive Sciences,* 21(11), 852-863.

Thorne, T. (2020). #CORONASPEAK – the language of Covid-19 goes viral – 2. Retrieved 15/01/2021 from <https://language-and-innovation.com/2020/04/15/coronaspeak-part-2-the-language-of-covid-19-goes-viral/>

Wallis, P. & Nerlich, B. (2005). Disease metaphors in new epidemics: the UK media framing of the 2003 SARS epidemic. *Social Science & Medicine,* 60, 2629-2639.

Washer, P. (2004). Representations of SARS in the British newspapers. *Social Science & Medicine,* 59, 2561-2571.

Wise, A. (2020). Military Metaphors Distort the Reality of COVID-19. Retrieved 09/02/2021 from <https://blogs.scientificamerican.com/observations/military-metaphors-distort-the-reality-of-covid-19/>

An Experimental Approach to the Perception of Empathy in Speech

Amanda McHugh
University of Cambridge

Abstract. With advances in the techniques and naturalness of speech synthesis, and the increasing commercial contexts in which it is used, the need for natural affective synthesis has grown. There is a need to understand the acoustic correlates of emotions in natural speech to optimise this affective synthesis, particularly regarding 'secondary emotions' like empathy. This original research seeks to shed more light on the nature of empathy in speech, using a parametric approach to synthesis. Although an older technique compared to machine learning synthesis, it is found that this approach allows for a greater degree of control over acoustic correlates, and allows for a more precise image of empathy to emerge. This research splits empathy into production and perception; the first experiment looks at the difference between non-empathetic and empathetic contexts in a dialogue context with 10 participants (5 male and 5 female, in pairs), which allows for the identification of empathetic correlates in natural speech production. Following this, an experiment using resynthesised versions of the non-empathetic productions explores which combinations and amounts of the correlates observed in the first experiment (including pitch, duration, and voice quality) must be used in order to elicit empathy. In doing so, it becomes clear that empathy is not a unified concept in speech, instead behaving differently within and between production and perception.

Keywords: phonetics; empathy; resynthesis; affective synthesis

---

1 Introduction

1.1 Empathy

Language does not exist in a vacuum in relation to other human and sociological phenomena. Any investigation into the manifestation of emotions in speech needs to begin with an understanding of how emotions relate to one another. Becker and Wachsmuth (2006) note that in order to model emotions into a socially-intelligent agent capable of communicating with and relating to others (in which speech clearly plays a part), one must first explore theories of emotion. One common starting point is to divide emotion into two types: primary and secondary. Primary emotions are typically defined as emotions that are innate, supporting reactive response behaviour and tending not to involve higher-level processing (Becker & Wachsmuth, 2006; James et al., 2018), whereas secondary emotions rely on this higher-level processing. Primary emotions can be 'understood as a prototypical emotion types [sic] which can already be ascribed to one year old children' (Becker-Asano & Wachsmuth, 2008, p. 3). As noted by Kemper (1987), there are large variations in what are considered to be primary, arguing for four (fear, anger, depression, satisfaction), where others argue for as few as two (Brenner, 1980) or as many as eleven (Emde, 1980). There appears to be a cross-cultural universality of primary emotions: Ekman (1973) found that fear, anger, sadness, happiness, disgust, and surprise were similarly realised across a number of unrelated languages and cultures.

This implies that the manner of primary emotions' realisation in speech is consistent or mostly-consistent across languages.

Due to being more sociologically-conditioned, secondary emotions such as empathy are more subjective. Conclusions about their realisation in speech are limited to the cultures and contexts in which they are observed. Secondary emotions require a higher level of cognitive processing (James et al., 2018), and build on the ability to evaluate preferences over outcomes and expectations (Becker-Asano & Wachsmuth, 2008).

Empathy is a secondary emotion, combining the influences of self/other state and inherent traits (Cuff et al., 2016). There is a divide, particularly in research on personality and developmental disorders, between cognitive and affective empathy, where affective empathy is not consciously elicited, but cognitive empathy can consciously manipulate these affective elements (Cuff et al., 2016). There is no clear means of dividing these two types of empathy when studying its manifestation in speech, but it can be assumed that in acted speech contexts, cognitive elements are being employed. However, because this distinction cannot be explicitly made, the term 'empathy' will be used in this study without subdivision, and further work will be needed to understand how these two types differ in speech. For the purposes of this study, empathy is defined as when the emotion felt by one person is felt by another person, allowing them to emotionally respond in a manner appropriate to the situation. As a result of its status as a secondary emotion, empathy may occur in different degrees (Boukricha et al., 2013), and may vary across languages.

Sympathy and compassion are two distinct concepts that are commonly confused with empathy (Cuff et al., 2016), but they are not the same as empathy. Singer and Lamm (2009) give the example of empathising with a person who is feeling sad, which produces a feeling of being sad in the self. Sympathy, on the other hand, will cause a feeling of pity for the other, and compassion will create a feeling of compassionate love for the other, but neither will cause the same feeling of sadness. Empathy is also not just mimicry or emotional contagion: while they are low-level contributors to empathy, they do not account for the whole experience of empathy (Singer & Lamm, 2009). Additionally, these processes seem to be largely automatic (Hatfield et al., 2014), which rules them out from being secondary in the sense that empathy is.

1.2 Empathy in Natural Speech

In studying the realisation of empathy in speech, the same correlates are explored as are used for studying other emotions like anger or joy. These studies of other emotions employ pitch, duration, and intensity in particular (Erickson, 2005), although voice quality (Gobl et al., 2002) and articulatory precision are also frequently included (Cahn, 1990; Murray & Arnott, 1995; Schröder, 2001). These same correlates are explored for empathy; the correlates at our disposal are typically considered to be pitch, duration, loudness, and voice quality (Laver, 1994). There is reason to assume that these are all manipulated in the production of empathy, and no reason to assume the existence of other major influencing correlates.

Pitch can be manipulated by a speaker to realise emotions through raising or lowering the entire pitch level, or they can reduce the pitch span by widening or flattening the pitch contour. Busso et al., (2009) found both of these features to be important in emotional communication, although with more focus on pitch level than contour. Their paper leaves open the question of whether there are particular parts of an utterance that are more affected by these changes and thus more salient for empathy.

As with pitch, there are two different types of duration changes that can be observed, these being global and temporal (or local) speech rate (Mozziconacci & Hermes, 2000). Global speech rate, commonly

referred to as just 'speech rate' in the literature (Burkhardt & Sendlmeier, 2000; McHenry et al., 2012), refers to the number of segments produced over time in a whole utterance. Local speech rate is a measure of segment length. Mozziconacci and Hermes (2000) explored a number of primary emotions, asking participants to rank durational deviations from neutral by how well they expressed a given emotion. They found that speakers ranked changes in global duration anywhere from a 9% deviation (for fear), all the way to a 31% deviation (for boredom) as optimal for different emotions. Deviations perceived as optimal ranged from +25% to -31% and differed in most cases from changes found in their production study. These results correspond with those found in other studies (Bezooijen, 2011), including in unrelated languages like Japanese (Kitahara & Tokhura, 1992). McHenry et al. (2012) looked at three secondary emotions: 'caring', 'sympathetic' and 'competent,' as classified by an oncologist, a psychiatrist, and a clinical psychologist. They found smaller deviations from neutral speech rate than Mozzioconacci and Hermes (2000). Speakers increased speech rate between 5% ('competent') and 8.6% ('sympathetic'). Therefore, if there is a change in duration for empathetic speech, it is likely to lie in the region of these secondary emotions, but there is scope for duration change to deviate further from neutral.

Studies of the perception of empathy in speech have investigated the same correlates as for other emotions. Xiao et al. (2015), for example, focus on speech rate as a cue for empathy in therapist communication in American English. McHenry et al. (2012) also focus on speech rate, finding it to be reduced in empathetic contexts alongside reduced pitch in their statistical study of empathetic communication (American English). Weiste and Peräkylä (2014) conducted a qualitative rather than statistical analysis of empathetic communication in a therapy context (Finnish), and found that mimicry of a patient's intonation and voice quality was a signal of empathy. Clearly, therefore, there are variations in the ways in which empathy is claimed to be realised in speech, but whether this is a case of language, cultural or situational context, or experimental method is unclear. Nevertheless, despite differences in results, it is clear that the correlates focused on are some (but not all) of those used in the study of other emotions in speech.

Empathy in speech is vital to understand due to its social and developmental role as a marker of self/other discrimination (Asada, 2015). This split allows for socially-intelligent agents (either human or robot) to assume a social role in relation to others, as well as allowing for perspective-taking, which has clear purpose in inter-personal relations. In understanding empathy in speech, we understand the manifestation of these crucial human-human interactions, and it is then possible to model these into human-robot interactions as well (Asada, 2015).

1.3 Parametric Synthesis

Speech synthesis, and the resynthesis of natural speech, are useful tools in the investigation of the correlates of empathy. Machine learning approaches to speech synthesis (and expressive synthesis in particular) can replicate natural speech with great ease and accuracy, although using a parametric synthesis technique is more effective for exploring the manipulation of correlates to understand emotions in speech.

The effectiveness of machine learning approaches is not to be ignored, however: Tacotron (Wang et al., 2017), an end-to-end synthesis approach that produces spectrograms from text inputs, was found to outperform production parametric systems, and could be trained to produce expressive output (based on richly expressive but noisy data) from scratch with a random initialisation (Wang et al., 2017). However, since the system is fully end-to-end, there is little that can be gleaned in terms of which features the system is identifying as salient in emotional speech, and how they are reproduced by the system. The only thing

that can be learned from this system is that it is possible to produce expressive speech outputs using a neural network, and nothing about the correlates of the emotions themselves.

Similarly using a voice-AI system, but adding features that focus on acoustic correlates, Cohn and Zellou (2019) demonstrated that with the Amazon Alexa 'speechcons' (Amazon, 2021), pitch and duration could be altered to make certain words sound more 'expressive'. Altering pitch and duration is a step towards understanding how these correlates factor into expressive speech, and it was demonstrated that these expressive speechcons aid human vocal alignment towards the voice-AI (Cohn & Zellou, 2019). However, there is no explanation of what 'more expressive' actually means in terms of emotional output, and additionally there is no room for understanding the gradience of emotions, since there are only the 'regular' and 'expressive' possibilities.

It is clear therefore that machine learning methods can provide the most effective techniques for replicating the human voice, which allows for more trustworthy and engaging social agents (Yalçın, 2019). This is a key step in achieving the manifestation of self/other distinctions and relations as realised in empathy. However, the amount that we can learn about the processes of emotional perception and production through synthesis is limited. Clearly, the end-to-end systems do not provide a means to investigate which correlates have been identified as salient. Even non-end-to-end synthesis systems that use statistical parametric synthesis cannot tell us much about which correlates are necessary for elicitations of emotions such as empathy. EMPHASIS (Li et al., 2018), which uses a cascade model, groups features that are relevant to different emotions in the input layer. However, the features that this system uses automatically to generate emotions such as empathy cannot be made explicit, and the correlates used to encode and decode emotional speech remain hidden. As such, using older techniques such as formant synthesis are more useful for advancing the theoretical understanding of how emotions like empathy are produced in speech.

Older models, based on formant synthesis and parametric approaches, may produce less natural speech, but grant a much greater degree of control over individual correlates, which will aid in understanding emotional speech. This is often achieved using formant synthesisers, which derive approximations to a speech waveform (Klatt, 1980). For example, Klatt's original (1980) synthesiser, based on the Liljencrants-Fant Model (Fant et al., 1985), offered 39 control parameters, which allows for specific alterations in acoustic features. This allows for greater understanding of encoding emotions in speech. Additionally, many formant synthesisers are commercially-available and easy to use (Bangayan et al., 1997), so although the output is far less realistic than the most up-to-date methods, they are far more practical for understanding emotional speech.

Parametric alteration of acoustic correlates for affective synthesis, as with the study of productions of natural affective speech, tends to focus on pitch, duration, voice quality, and precision of articulation in various combinations. Cahn's (1990) Affect Editor manipulates pitch, duration, and articulatory precision to produce various primary emotions. Murray and Arnott (1995) build on Cahn's model by adding Voice Quality, as do Schröder (2001) and Burkhardt and Sendlmeier (2000). These parametric syntheses can still be incorporated into models of emotions in socially-intelligent agents. Kismet (Breazeal, 2001) uses the AVS model (Arousal-Valence Space) to plot the relationships between different emotions in the system. One of the most common mappings of emotions is the PAD-space (Pleasure-Arousal-Dominance) (Park et al., 2011), which opts for a three-dimensional mapping of the emotional space for an agent. In order to realise this emotional framework in speech, some means of affective synthesis is required, and parametric synthesis techniques are perfectly sufficient to fulfil this role, although with the caveat above that naturalness may be slightly lessened compared to machine-learning techniques.

It is therefore a parametric approach that shall be used in this research, since the greater degree of control over correlates and greater transparency in producing affective synthesis renders it far superior for understanding and recreating the true correlates of emotions in speech. Although the methods are not the most up-to-date, a parametric approach to synthesis provides the best toolkit for the task at hand, as well as being more commercially-available and accessible than cutting-edge machine learning techniques. Additionally, the findings from this research can still be incorporated into artificial emotional frameworks like the AVS model or PAD-space.

1.4 Aims of this Study

This study explores the perception and production of empathy in speech in a British English and non-clinical context. Previous studies (such as McHenry et al., 2012; and Weiste & Peräkylä, 2014) focus on languages and dialects other than British English, and focus on empathy in clinical contexts such as therapy (Weiste & Peräkylä, 2014) or oncology (McHenry et al., 2012). The correlates that are being focused on will be the same as those studied in other empathetic (and general emotional) speech studies, these being pitch, duration, and voice quality. Due to the experimental methodology not guaranteeing fixed distance from the microphone during recording, as well as Praat's resynthesis tools normalising intensity, the correlate of intensity will not be explored, although evidence does suggest some salience in other kinds of emotional speech (Erickson, 2005), as well as in empathy (Alam et al., 2018). The case is the same for articulatory precision: although it is perceptually salient in some emotional productions (Murray & Arnott, 1995), it could not be feasibly produced or assessed in the present work.

The first study (Section 2) explores empathy in speech production, using acted speech dialogues between pairs of participants to determine which correlates explored are salient to empathy, as well as if other factors such as gender are relevant to the production of empathetic speech. The second study (Section 3) manipulates these same correlates for a perception test, where samples from the first experiment are resynthesised. The gradience of empathetic speech is unknown (Boukricha et al., 2013), and this is also explored in this study. The results are considered jointly in Section 4, and Section 5 concludes and suggests further directions for this work.

2 Production Study

2.1 Methodology

2.1.1 Design

Ten participants were selected for this experiment (five male and five female), all speakers of British English and aged between 20 and 26. Participation was voluntary and participants were not compensated. To elicit productions of empathy, the participants were paired and asked to read eight different dialogues, each preceded by a short scenario description (e.g., 'A and B are discussing difficulties in a relationship'). Each dialogue had a stimulus phrase that was placed into both a non-empathetic and empathetic context, and was one of the following: 'I know where you're coming from,' 'That's hard,' 'I understand,' and 'I hear what you're saying' (labelled from 1 to 4 accordingly). For example, 'I know where you're coming from' was placed into a dialogue where A asks B for directions, and also in a dialogue where A and B are

discussing work stress over a coffee (the full list of dialogues can be found in Appendix One). Half of the participants produced Stimuli 1 and 2, and the other half produced 3 and 4.

Participants were paired as follows:

Table 1: *Participant Pairings.*

Participant One (Stimuli 1 and 2)	**Participant Two (Stimuli 3 and 4)**
1F	1M
2F	2M
4M	3M
3F	4F
5M	5F

Participants were recorded in a recording booth using a Sennheiser ME-64 microphone, recorded onto a MixPre-6 recorder (256kbps, 48kHz) at a distance of around 50 centimetres (resulting from the room size and participants sharing a microphone). They were instructed to read through the dialogues twice, clapping once at the end of each dialogue, and twice at the end of the full set, as this created spikes in the waveform that made it easier to distinguish the recordings during analysis. One pair, 3F and 4F, had to re-record the full set, as they accidentally switched parts halfway through. For all recordings, the second set of dialogues was used for analysis unless there was disruption to the recording, on the basis that speakers were more comfortable with the dialogues after having read them through once already.

The recordings were trimmed using Audacity so that only the stimuli remained, and Praat was used for analysis. Different analyses were conducted for different correlates being explored. For pitch, there were two different strategies that could have been used by speakers. The first is changing the pitch contour, which is a well-attested correlate of emotional speech (Busso et al., 2009), as well as a raising or lowering of the overall pitch, which is noted as a correlate of empathy in bad news discussion by McHenry et al. (2012). To measure the variations both in overall pitch and contour, or 'level' and 'span' (Ladd, 2008), I employed a method modelled on Graham (2014), which, instead of comparing the entire pitch contour, uses Patterson's (2000) method to plot pitch onto tonal targets. This was chosen on the assumption that speakers are manipulating tonal targets to alter prosody, rather than abstractly changing the pitch contour, and so these targets are key in understanding how pitch was modified for empathy. Using the ToBI system and using Praat (with pitch set to the autocorrelation method), the pitch was recorded at the onset, H and L* for shorter stimuli (2 and 3), and for the longer stimuli (1 and 4), at the Onset, initial H* (iH*), L, H* and final low (FL). Span, as a means of measuring the width of the pitch contour, was determined by subtracting FL from iH* for longer stimuli and L* from H for shorter ones. The full list of pitch measurements is found in Appendix Two.

For duration, there were also two types of change measures, these being global and temporal (or local) duration, both different kinds of speech rate (Mozziconacci & Hermes, 2000). Global duration was measured by comparing the lengths of the stimuli between the non-empathetic and empathetic conditions, under the logic that the number of segments produced in the utterance would be the same for both conditions, and so measuring total length was sufficient for a measure of rate. Local speech rate was ascertained by measuring the length of the first accented vowel (iH* for Stimuli 1 and 4 and L* for Stimuli 2 and 3).

The only voice quality that appeared aside from modal voice was creaky voice, which typically arises from a low frequency vocal fold vibration and low F_0 (Laver, 1980), caused by slack, thick, and compressed vocal folds with a short vibrating length (Keating et al., 2015). The level of creak was measuring by subtracting the amplitude of the second harmonic (H2) from that of the first (H1), which is a common way to measure creaky voice (Keating et al., 2015). This was done using a Praat Script to extract the spectrum of the vowel being analysed, and manually finding H1 and H2 values.

2.1.2 COVID-19 Adjustments

The recordings in this experiment took place during the second national COVID-19 lockdown, and consequently measures were taken to ensure the utmost safety of participants. This included sanitisation of the recording area, social distancing and wearing masks where possible, and ensuring recordings were taken as soon after a negative asymptomatic swab test as possible. One male participant had to be replaced as a result of a positive swab test, and three female participants had to be replaced as a result of rapidly-changing travel plans in response to the pandemic.

Additionally, the building containing the recording facilities was locked for the duration of the Christmas vacation, which meant that the recording of 3M and 4M had to take place in an empty room, using a mobile phone with recording quality set to the maximum of 256kbps and 48kHz, since 4M had to be found last-minute as a replacement for the other male participant who contracted COVID-19, delaying the recording. There was echo interference in this recording, which made it impossible to measure voice quality, but analysis of duration in was still possible once the dynamic range was changed from 90dB to 70dB.

2.2 Results

2.2.1 Pitch

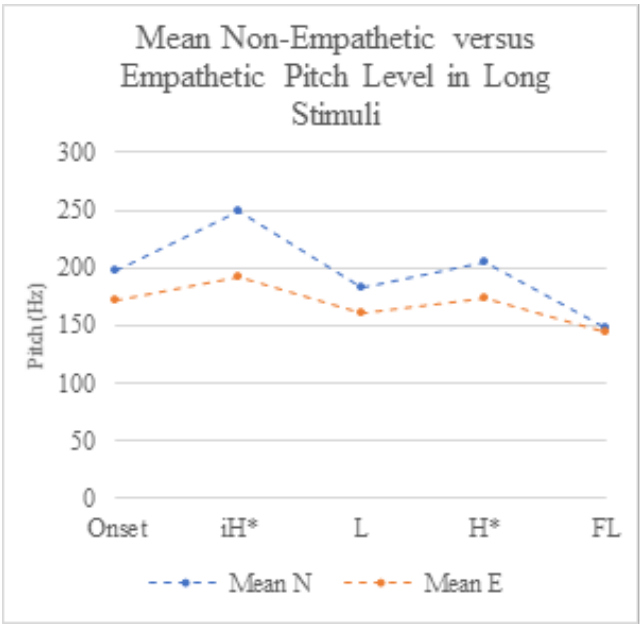


Figure 1: *Difference in Mean Pitch by Condition.*

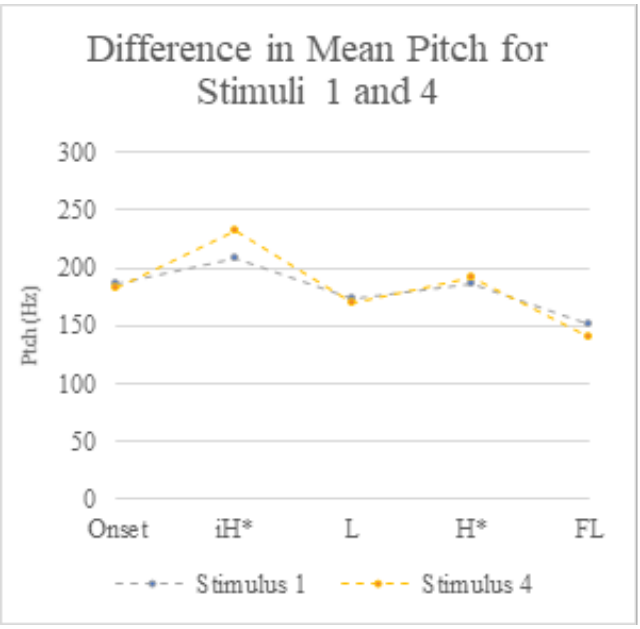


Figure 2: *Difference in Mean Pitch by Stimulus.*

There is a clean divide between stimulus length in relation to pitch level as a cue for empathy. For long utterances, a repeated measures ANOVA was conducted with five levels (onset, iH*, L, H*, FL), with condition (non-empathetic and empathetic), stimulus (4 or 1, these being the two longer utterances) and gender as between-subject factors. This ANOVA revealed strong significance between pitch level and condition ($F(4, 40)=8.932$, $p=<0.001$), being lower in the empathetic condition (as illustrated in Figure 1), as well as pitch level and stimulus produced ($F(4,40)=4.701$, $p=0.003$), where Stimulus 1 had a lower pitch level than Stimulus 4 (Figure 2). Three-way effects of pitch level, condition, and gender ($F(4,40)=2.414$, $p=0.065$) and pitch level, condition, and stimulus ($F(4,40)=2.434$, $p=0.063$) were found to be approaching significance, which is promising given the sample size. Gender was revealed to be insignificant in relation to pitch level ($F(4,40)=1.432$, $p=0.241$).

Individual ANOVAs were conducted for each pitch point, to see which pitch points were most salient for this difference in pitch level. It was found that only iH* was significant between conditions, but this was a strong significance ($F(1,16)=10.333$, $p=0.005$). This can be seen in Figure 1, as the difference in mean pitch level for non-empathetic and empathetic conditions is much greater for iH* than other points.

For shorter utterances, a repeated measures ANOVA was also conducted, with three levels (onset, H, and L*), and between subject factors as stimulus (2 or 3 this time), condition, and gender (both as before). Unlike for the longer utterances, this ANOVA revealed no significance between pitch level and any other factors. Individual ANOVAs for each pitch point also revealed no significant effects, which suggests that pitch level manipulations for empathy are limited to longer utterances.

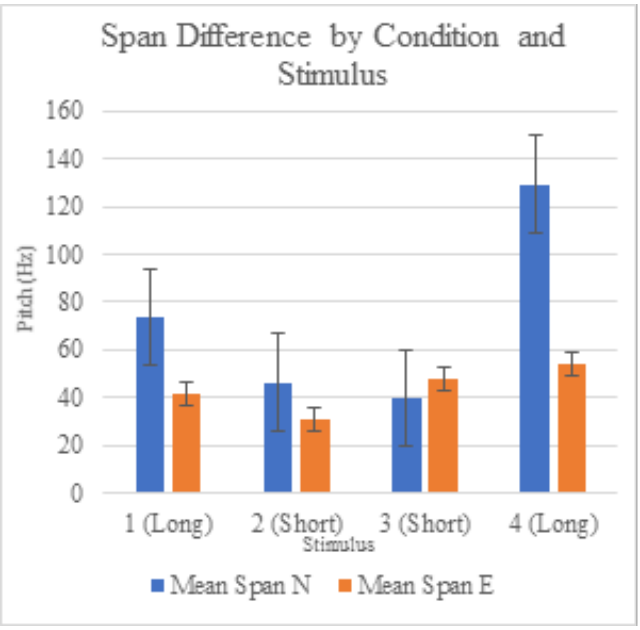


Figure 3: *Differences in Mean Pitch Span by Condition and Stimulus.*

Span was determined by subtracting the FL pitch from the iH* pitch for longer utterances, and L* pitch from the H pitch for shorter utterances. The results for difference in mean span between condition and stimulus are found in Figure 3. An ANOVA revealed that condition was strongly significant for longer utterances ($F(1,12)=12.102$, $p=0.001$) and showed a large decrease between non-empathetic and empathetic, but not significant at all for shorter ones ($F(1,12)=0.275$), $p=0.61$). Stimulus was also found to be significant for longer utterances ($F(1,12)=6.615$, $p=0.024$), where stimulus 4 showed a much larger span difference, and no other effects were found in the data. This suggests that span decrease between conditions and stimuli is again more salient for longer utterances.

2.2.2 Duration

Using a paired samples *t* test, it was found that global duration difference for Stimulus 1 was approaching significance ($t(4)=-2.44$, $p=0.071$), which is promising given the sample size, although still insignificant. No significance was found for Stimulus 2 ($t(4)=0.656$, $p=0.548$), Stimulus 3 ($t(4)=-0.453$, $p=0.674$), or Stimulus 4 ($t(4)=1.335$, $p=0.253$). The difference between the two conditions (non-empathetic minus empathetic), as well as the % change from the non-empathetic condition, can be found in Appendix Two. Figure 4 demonstrates the difference between condition for global duration by stimulus.

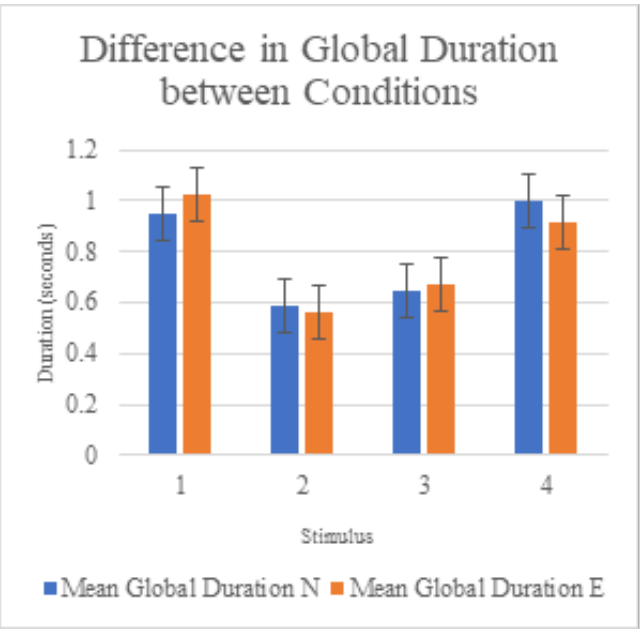


Figure 4: *Global Duration Difference by Condition and Stimulus.*

Local duration was found to be significant particularly in shorter stimuli according to Huggins' (1972) Just-Noticeable Difference (JND) for segment length in natural speech, where an increase of 1ms or decrease of 2ms was deemed to be significant. It was significant only in 50% of longer stimuli, whereas for shorter stimuli it was significant in 80% (Table 2; values over this JND are italicised). To ensure that this was not an effect of pre-boundary lengthening, since temporal expansion at prosodic phrase edges is common, (Fletcher, 2010), the final segment lengths of the longer stimuli were measured. The segment lengthening was only significant for 20% of cases, and only within Stimulus 1 (Table 3), and so the duration effect seen in shorter stimuli is an empathy cue rather than pre-boundary lengthening.

Table 2: *Duration Change in First Accented Syllable.*

Stimulus 1	
Participant	Change (ms)
1F	+0.2
2F	+0.5
3F	*-8.6*
4M	*+4*
5M	-1.4

Stimulus 2	
Participant	Change (ms)
1F	*-7.7*
2F	*+11.4*
3F	-1.1
4M	*-6.1*
5M	*+2.4*

Stimulus 3	
Participant	Change (ms)
1M	*+2.4*
2M	*-3.4*
4F	*-9.3*
3M	*+10.3*
5F	-0.2

Stimulus 4	
Participant	Change (ms)
1M	-1.5
2M	+0.7
4F	*-2.1*
3M	-6.5
5F	*+4.7*

Table 3: *Final Segment Lengthening.*

Stimulus 1	
Participant	Change (ms)
1F	0
2F	*+4.9*
3F	+0.5
4M	*+6.3*
5M	-0.4

Stimulus 2	
Participant	Change (ms)
1F	*-7.7*
2F	*+11.4*
3F	-1.1
4M	*-6.1*
5M	*+2.4*

Stimulus 3	
Participant	Change (ms)
1M	*+2.4*
2M	*-3.4*
4F	*-9.3*
3M	*+10.3*
5F	-0.2

Stimulus 4	
Participant	Change (ms)
1M	+0.2
2M	+0.3
4F	+0.2
3M	0
5F	+0.9

2.2.3 Voice Quality

The measurements of H1-H2 for each speaker's long and short stimuli are listed in Appendix Two (3M and 4M are omitted due to recording quality). An ANOVA was conducted, with H1-H2 as the dependent variable, and speaker gender, condition (non-empathetic or empathetic), and stimulus length (short or long) as the factors. This test revealed no significant effects for condition or stimulus length, but showed a strong significance for gender ($F(1, 24)=12.5$, $p=0.002$). Gender was not found to be significant when combined with other factors such as condition ($F(1,24)=0.183$, $p=0.673$) or length ($F(1, 24)=0.280$, $p=0.601$). Male speakers used creaky voice (indicated by a negative or small positive H1-H2 value) to a far greater degree than female speakers (as illustrated in Figure 5), so while creaky voice is a reliable cue for speaker gender, it is not a reliable cue for empathetic speech.

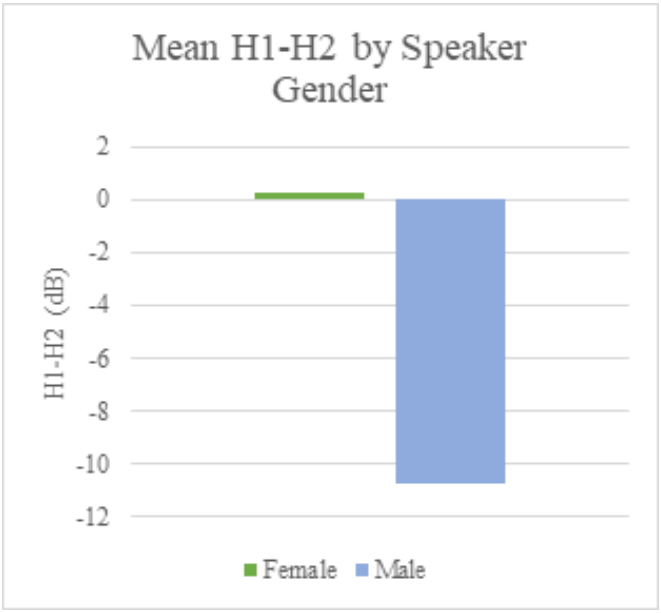


Figure 5: *Mean H1-H2 for Male and Female Speakers.*

2.3 Discussion

The results of the production study demonstrate that there are correlates responsible for the elicitation of empathy in speech, but their behaviour is not identical across all empathetic productions. The main factor deciding which correlates employed is the length of the production. Empathy in longer utterances was elicited by a significant decrease in pitch level, centred around the first accent (iH*), as well as significant reduction in pitch span. For shorter utterances, on the other hand, it was found that empathy was elicited by a significant increase in local duration. Global duration change and voice quality were not found to be markers of empathy.

A number of general points must be considered before the data presented above is analysed. Firstly, as a result of the methodology demanding acted speech, the phrases chosen for this experiment were fixed, which means that they may not have been natural for the participants and could have changed the realisation of empathy. After participating, 5M commented that he felt the phrases in the dialogues were not ones he would have chosen to be empathetic. The extent to which meaning factors into productions of empathy is unclear, and requires further work.

In terms of pitch, the difference between the results for longer and shorter stimuli indicate that pitch is a far better marker of empathy in longer utterances than shorter ones. The lowering of pitch in the empathetic condition supports McHenry et al.'s (2012) study in American English, which demonstrates the same decrease in pitch in empathetic scenarios. The significance of only the iH* in longer utterances as an individual pitch point suggests that this is also an extremely localised effect in terms of what is being manipulated, focused on the first accent in an utterance. For pitch level, it may be that no significance is realised in shorter utterances because the first (and only) accent is attached to an L rather than an H, but work on short utterances where the first accent is attached to an H would be needed to confirm this. The reduction in pitch level is also accompanied by a reduction in pitch span in longer utterances, something

again not seen in shorter utterances, which suggests that the two are connected in pitch-based productions of empathy.

The shorter stimuli showing no significance in pitch level or span between conditions complicates the image of McHenry et al.'s (2012) unified portrayal of empathetic speech, especially in light of the fact that local-level duration is far more salient for shorter utterances than longer utterances. This suggests a sensitivity to length that goes against the idea of empathy having a single realisation.

The significant stimulus difference for longer utterances, where stimulus 4 showed a larger span and higher level, can be explained by looking at the context of Dialogue 6 ('Speaker A is trying to do a video call with Speaker B'). Stimulus 4 (non-empathetic) was contained in this dialogue, and the context seemed to trigger the Lombard effect, where speakers commonly increase vocal intensity and fundamental frequency in noisy environments or when there are difficulties being heard (Garnier & Henrich, 2014). Therefore, had this not occurred, it is possible that the inter-stimulus difference would not be significant, as demonstrated by the fact that there was no inter-stimulus difference for shorter utterances, where the pragmatic content was different but there was no elicitation of the Lombard Effect.

The increased reliance on durational cues in shorter utterances, as shown by the results for local duration change, show that correlates of empathy change with utterance length. It is not simply the case that shorter utterances show no significant changes in empathetic contexts, as would be concluded with pitch alone. The insignificance for all stimuli of global duration, despite the mean percentage deviation for each stimulus being roughly that of changes found in previous studies, suggests that the effect observed is more specific than just speech rate. Even though speech rate tends to affect stressed syllables more than unstressed syllables (and vowels more than consonants) (Burkhardt & Sendlmeier, 2000), a change in speech rate would still affect the duration of the entire utterance, which does not happen to significance in any circumstance here.

Gender did not significantly contribute to any results apart from the use of creaky voice, where there was no significant difference between the non-empathetic and empathetic conditions, or gender and condition combined. Creaky voice is therefore a better signal of the gender of the speaker than a cue for empathy. This creak was much stronger for male speakers than it was for female speakers, which agrees with cross-dialectal studies that find that male speech usually contains more creak (c.f. Podesva, 2011).

2.3.1 Interim Conclusions and Implications for Resynthesis

These results do not fully correspond with previous studies on the production of empathy in speech. This difference between prediction and empirical results could come from a number of sources, such as the fact that the speech was acted rather than natural as it was in previous studies, or the use of British English as the speaker dialect, or the more everyday speech context, instead of being a clinical context. Nevertheless, the elicitation of empathy in these results seems to be much more guided by the length of the utterance, where longer utterances rely more on manipulation of pitch level and pitch span, and shorter ones more on local-level duration change. Creaky voice played no part in the signalling of empathy, instead serving to be a marker of gender rather than emotional state. Therefore, it is expected that pitch level, pitch span, and local-level duration will be the most salient correlates of empathy in the perception study.

3 Assessing Perceptions of Empathy through Analysis-by-Synthesis

In the production study, strong relationships were found between pitch level and span and empathy for longer utterances, and local duration and empathy for shorter ones. The perception study aims to see if these correlates can be reproduced through resynthesis, and if they still communicate empathy to the same significance as in the production study. This tests to see if listeners prefer a certain correlate or set of correlates in perceiving empathy, and if this corresponds to the correlates produced by the speakers. It additionally tests to see if empathetic speech is perceived as gradient.

3.1 Methodology

3.1.1 Production of Resynthesised Samples

One male and female production of a longer stimulus and a shorter stimulus, both non-empathetic, were chosen to be resynthesised using the correlates found in the production study. Stimulus 1 was chosen to avoid interference from the Lombard Effect, and speakers 5M and 2F were the samples chosen. The shorter utterance chosen was Stimulus 3, and 1M and 4F were chosen as samples.

It was decided that intonational manipulations (pitch level and widened/flattened contour) would be grouped (per Burkhardt & Sendlmeier (2000)) and applied to longer utterances, and local duration effects would be manipulated in shorter samples. This was to maximise the potential for significant results according to the production study, and also because combining all correlates in different quantities would have resulted in thousands of samples, which was not possible with the resources available. There was a chance that speakers would rely on correlates present in the non-empathetic samples that had not been deemed significant in production, but this was an accepted consequence of a resynthesis approach, although widening and flattening were applied to shorter utterances as well, to test if perception strategies relied on different correlates to production.

Pitch level was altered in Praat by increasing or decreasing the pitch of the entire sample every two semitones. It was increased and decreased until the level reached the upper and lower limits of male and female conversational speech pitch (50–250Hz for men and 120–480Hz for women (Laver, 1994)). The pitch contour was altered using a manipulation in Praat and removing all but 'turning points' in the pitch contour. These were then widened by increasing the distance from the mean (above or below) by one semitone until the upper or lower conversational limit was reached, and flattened by bringing each point towards the mean by one semitone until no longer possible. Certain contour change samples were also subjected to pitch level change, to see if there was an interaction of correlates.

Duration was manipulated using a Praat Script, increasing or decreasing the duration of the accented vowel by 2ms to ensure the change was greater than the Just-Noticeable Difference for segments in natural speech (Huggins, 1972). Duration was increased until it reached +20ms and decreased until the segment length was 0. Duration change was applied to a number of different widened or flattened contours as well to test for interaction.

Creaky voice was initially going to be resynthesised using an open-source version of the Klatt Synthesiser (d'Heureuse, 2021; Klatt, 1980), recreating the formants of the original vowel and increasing flutter as a means of synthesising jitter and shimmer. However, not only was flutter an inappropriate substitute for jitter and shimmer, this resynthesis was not at all natural. Given that creak was only a marker of gender in the production study, it was not important to perfect this, and any effects of creaky voice could be analysed using 4F's sample (H1-H2=-9.7).

A pilot study following the design of (Yiu et al., 2002) was conducted. All 344 samples were placed into a survey with each block of samples randomised. Ten volunteer participants had to determine if the sample sounded ‘normal’ or ‘abnormal’ to them. Using a binomial distribution table (Runyon et al., 2000) with a 95% confidence level, it was determined that 8 out of 10 participants needed to agree on a sample for it to pass as natural enough for the perception study. Of these samples, 85 of them were deemed normal.

3.1.2 Design

The samples that passed the pilot naturalness test were placed into a survey, with the order totally randomised to avoid task fatigue and shifting criteria. Participants were given a definition of empathy, as well as one of the dialogues from the production experiment that was not the context for any resynthesised sample (Dialogue 4). This was to minimise the possibility of confusing empathy for other emotions, since listeners seem to be at risk of confusing emotions in perception (Breazeal, 2001). The introduction to the survey as described is available in Appendix Three. Participants were asked to confirm their gender, as well as confirming they had not been involved in prior experiments, and also that they were either a native speaker of British English, or were a first language speaker of English who had spent two or more years in England, in order to control for the potential cultural variation across English-speaking areas.

Each sample was ranked according to a five-point Likert scale, since this is cited as a fairly common means of ranking empathy (Yalçın, 2019). Instead of ranking from 1 to 5, this scale went from 0 to 4, with the assumption that a sample could be seen as totally unempathetic, whilst also maintaining the hypothetical notion that empathy could be gradient.

In total, 33 participants completed the survey (20 female and 13 male), all naïve listeners. Participants were given the chance to enter a prize draw with a cash prize after completing the survey. A range of ages can be expected, since they were recruited online, and age was not asked for in this experiment. Given the means of recruitment (through personal connections), it can be assumed that there is a skew towards younger speakers, with a handful of older speakers.

3.1.3 Analysis

Too few samples passed as natural in the pilot experiment to form statistical conclusions about the interactions of correlates in empathetic perception, and so results focus on manipulations of single correlates, these being pitch level, pitch span, and local duration.

The Likert data from the experiment was analysed in SPSS. The first test was a Repeated Measures Ordinal Logistic Regression (RMOL) via a Generalised Estimating Equation function. This test was used as it is appropriate for ordinal data such as results from a Likert scale, as in this experiment. Subject ID and Level of Manipulation (either pitch level, widened or flattened pitch contour, or segment duration) were treated as the subject variables, and the gender of the respondent (the rater of samples) was a within-subject variable. The dependent variable in this test was the empathy rating. The test of model effects was the Type III Wald Chi-Square test, which determines whether the predictors are significant and contribute to the model.

The second statistical test was a Friedman test, which is a non-parametric K-related samples test, where the repeated samples were the different levels of manipulation. This provides a mean rank for each of the factors, allowing for comparison, which shows if there is a statistically significant difference between the means for each category.

3.2 Results

As regards pitch level change, the results for both 5M and 2F show no significance, despite their significance in the production study. The Repeated Measures Ordinal Logistic Regression (RMOL) confirmed that in 5M, there was no significant impact of rater gender ($\chi^2(1, N=33)=0.97$, $p=0.32$), level of manipulation ($\chi^2(4, N=33)=3.05$, $p=0.55$), or of level of manipulation and gender combined ($\chi^2(4, N=33)=3.93$, $p=0.42$), presented in Figure 6. Similarly for 2F, (Figure 7) no significant effect was found for level of manipulation ($\chi^2(2, N=33)=0.64$, $p=0.968$), and neither for gender and level of manipulation ($\chi^2(2, N=33)=1.434$, $p=0.488$). The gender of the rater alone was close to approaching significance ($\chi^2(1, N=33)=3.639$, $p=0.56$), although still insignificant.

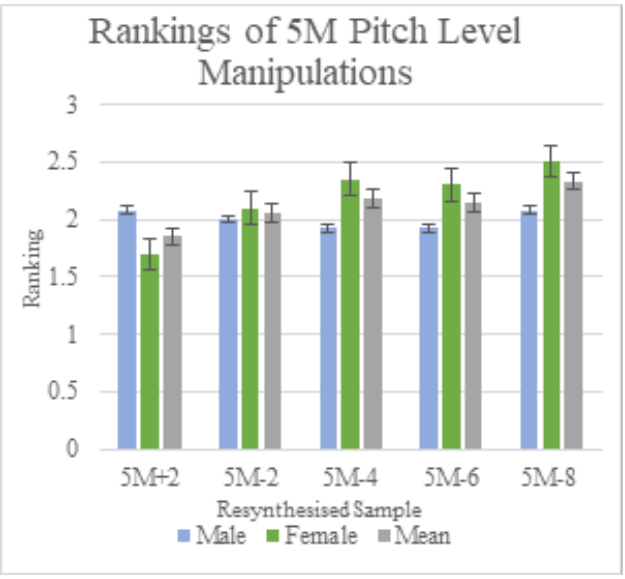


Figure 6: *Overall Mean, Male Mean, and Female Mean Rankings for 5M's Pitch Samples.*

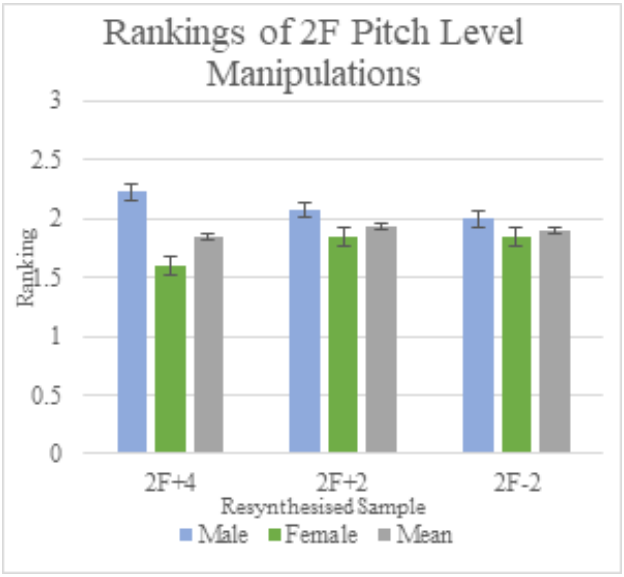


Figure 7: *Overall Mean, Male Mean, and Female Mean Rankings for 2F's Pitch Samples.*

The Friedman test confirmed that for both 5M ($\chi^2(4, N=33)=3.93$, $p=0.42$) and 2F ($\chi^2(2, N=32)=0.282$, $p=0.868$) that no level of manipulation was significant, and thus that listeners did not perceive any level of manipulation to be more empathetic than another at a statistically significant level.

For the widening and flattening of the pitch contour, the image was the same. The RMOL for 2F showed no significant effects for gender ($\chi^2(1, N=50)=0.829$, $p=0.591$), level of manipulation ($\chi^2(1, N=50)=0.004$, $p=0.952$), or the two combined ($\chi^2(1, N=50)=0.29$, $p=0.778$), confirmed by a Friedman test ($\chi^2(1, N=33)=0.474$, $p=0.491$). 5M also showed no significance in the RMOL for gender ($\chi^2(1, N=16.7)=0.879$, $p=0.349$), level of manipulation ($\chi^2(5, N=16.7)=6.425$, $p=0.267$) or the two combined ($\chi^2(5, N=16.7)=2.479$, $p=0.780$), with insignificance confirmed by a Friedman test ($\chi^2(5, N=33)=7.022$, $p=0.219$).

The shorter utterances were similarly insignificant for pitch widening or flattening: 4F was not significant for rater gender ($\chi^2(1, N=6.3)=2.494$, $p=0.114$), level of manipulation ($\chi^2(15, N=6.3)=8.619$, $p=0.897$), or the two combined ($\chi^2(15, N=6.3)=4.223$, $p=0.997$), verified with a Friedman test ($\chi^2(15, N=32)=9.526$, $p=0.848$). Gender ($\chi^2(2, N=33)=2.318$, $p=0.126$), level of manipulation ($\chi^2(2, N=33)=0.613$, $p=0.736$) and the two combined ($\chi^2(2,N=33)=0.415$, $p=0.813$) were not significant for 1M either, also verified by a Friedman test ($\chi^2(2, N=33)=2.333$, $p=0.311$).

Duration was only incorporated into (and thus investigated in) shorter stimuli, but both speakers' results show different significances. 1M showed no significance at all in the RMOL for level of manipulation ($\chi^2(2, N=33)=1.110$, $p=0.574$) or gender and level combined ($\chi^2(2, N=33)=0.618$, $p=0.734$), but showed a very strong significance for the gender of the rater ($\chi^2(1, N=33)=6.897$, $p=0.009$), as demonstrated in Figure 8. The Friedman test confirmed that no level of manipulation was significantly more empathetic to listeners as a whole ($\chi^2(2, N=33)=3.057$, $p=0.217$). Individual ANOVAs conducted on each increased duration showed that gender was particularly salient for a duration of +10ms, which was ranked significantly lower by female raters ($F(1, 31)=4.868$, $p=0.035$). This potentially indicates increased significance of gender when duration changes become more extreme, especially since Figure 8 seems to show this gender division widening with increased duration.

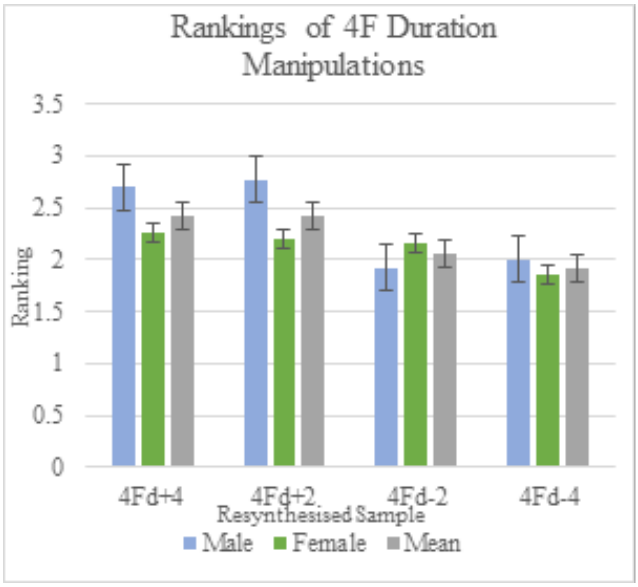


Figure 8: *Overall Mean, Male Mean, and Female Mean Rankings for 4F's Duration Samples.*

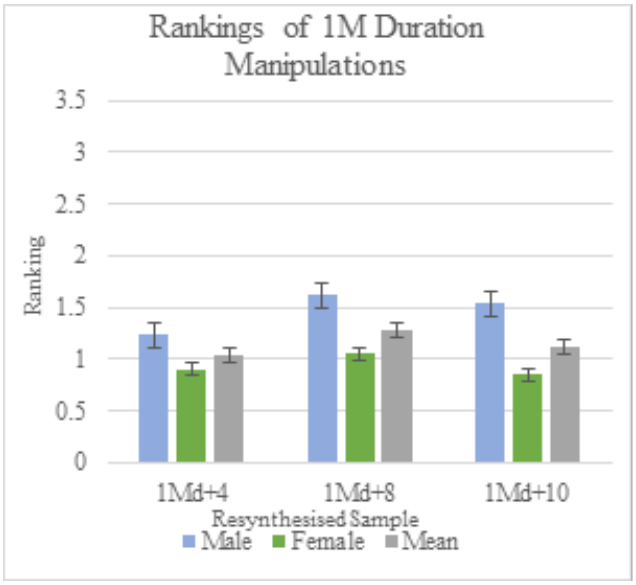


Figure 9: *Overall Mean, Male Mean, and Female Mean Rankings for 1M's Duration Samples.*

For 4F, however, the level of manipulation was confirmed by the RMOL to be statistically significant ($\chi^2(1, N=25)=9.187$, $p=0.027$), where gender of the rater ($\chi^2(1, N=25)=2.014$, $p=0.156$) and gender and level combined ($\chi^2(1, N=25)=4.148$, $p=0.246$) were not (Figure 9). This significance was re-confirmed by a Friedman test ($\chi^2(3, N=33)=9.128$, $p=0.028$). It was found that a duration of +4ms was the highest-ranked sample in the Friedman test, which means that for 4F, an increase in duration of 4ms on the accented vowel elicited maximal perceived empathy in this context. This result is drastically different from the 9.3 millisecond decrease between non-empathetic and empathetic conditions produced by 4F in stimulus 3 of the production study. A duration increase of +2ms approached significance in a Friedman test once the

+4ms sample was removed ($\chi^2(2, N=33)=5.763$, $p=0.056$), which indicates a potential element of gradience, but since this result is not actually significant, this conclusion can only be tentative.

3.3 Discussion

The results of the perception study are noticeably different from the production study. Where pitch level and span were key cues for empathy in longer utterances in the production study, they were insignificant in the perception study. Pitch contour change (which changes span) was, as expected, found to be insignificant for shorter utterances. Duration produced significant effects in shorter utterances, per the production study, but only for 4F did it show significance relating to the level of manipulation. For 1M, on the other hand, it was found that only the gender of the person rating the samples was significant, where females rated samples significantly lower than males, which cannot be definitively attributed to gender-based differences in empathy perception.

These results clearly paint a different picture to the results found for the production study, and this may be for several reasons. Firstly, the removal of the context for those listening may have altered how empathetic the speech was perceived to be, and the correlates that were salient in production were context-dependent. Alternatively, the resyntheses, although passing a naturalness pilot test, may still be in some way unnatural to listeners, which could cause interference with how empathetic they perceived the samples to be. This could potentially explain the significant result in 4F's sample that matches what was found in the production study, as the creaky voice appeared to minimise the naturalness problem by disguising changes in pitch or duration; 37 of the 85 samples that passed the pilot test were from 4F, implying that creaky voice can aid with natural resynthesis. On the other hand, it may simply be that correlates such as a lowering of pitch or reduction of span by flattening the pitch contour are far less salient in perception than they are in production, but given the strength of the significance in production, this seems unlikely.

Participants were not asked to offer feedback on the experiment, but several did. One participant (F;22) said that she found the shorter utterances harder to distinguish than the longer utterances. Another participant (M;22) said he found the decision process to be 'largely arbitrary' based on whether a speaker sounded 'nice or insincere', which indicates that the judgements were not based on conscious criteria, although this does not discount naturalness having a subconscious influence on decisions. Two participants (M;54 and F;30) independent of one another said that their judgements were based on what they thought the speaker was going to say next: the male participant said 'I found myself imagining what the next words coming out of their mouths would be — some of them I thought were more interested to talk about themselves, others would have followed with a 'but…''. The female speaker said that 'sometimes you could hear the unspoken 'but',' which again demonstrates that respondents were not basing judgements on conscious decisions of correlates or naturalness, but rather intuitive judgment, which does not rule out the influence of naturalness issues.

Another participant (F;21) said that on the whole, she found that she felt the female voices to be more empathetic across the board, which suggests possible socio-cultural influences. The significant difference in ranking based on the gender of the rater in 1M seems to support some kind of gender-based variation in empathetic perception as noted by the participant F;21. However, it seems limited in scope given that this was the only sample for which there was a difference, and it was only at extreme variations in duration that this occurred to a significant level. It could, therefore, be attributable to female respondents noticing less natural samples and rating them as less empathetic, rather than this being a concrete cue to gender-based differences in empathy.

Given the fact that 4F's samples appear to have been more perceptually natural to participants in the pilot experiment, the results demonstrating that the level of duration manipulation is significant for participants' rankings of empathy are particularly encouraging. Not only does this increase in stressed syllable length support the duration increases found in previous studies of empathetic speech such as Weiste and Peräkylä (2014), the approach to significance in a gradient fashion appears to suggest gradience in the perception of empathy, although clearly more work is needed to solidify this conclusion. The fact that these results correspond to Weiste and Peräkylä's (2014) Finnish study also indicates that there may be some cross-linguistic and cross-cultural consistency in the way empathy is realised in speech, which counters the idea of secondary emotions such as empathy being completely socio-culturally determined.

Although 4F's was the only sample where the results matched those of the production study, it does seem to point towards the idea that, with natural enough samples, the perception of empathetic correlates may match those that are most significant in the production of empathy.

3.3.1 Interim Summary

These results demonstrate the importance of naturalness in empathy perception: although the resynthesised samples passed the pilot naturalness test, the best explanation for the insignificance of these results is a subconscious sensitivity to the naturalness of samples. This could also provide an explanation for the gender divide in the perception of 1M's resynthesised samples, especially since this significance only appeared at the extremes of duration change. This naturalness issue is an accepted shortfall of parametric synthesis, although perhaps a total resynthesis would place respondents into a position where they accept the samples to be entirely artificial and are not subconsciously affected by otherwise perfectly human-sounding voices having elements of unnaturalness.

The results for 4F, where the creak seems to have disguised some of these naturalness issues, are indeed promising. The significance of increasing the accented segment duration in both the production and perception studies for shorter utterances indicates that it is the same correlates that are targeted in perception as in production, regardless of the gender of the person listening. This points towards a model of empathy governed by utterance length rather than gender of speaker or listener, production or perception, or a fixed set of correlates across the board.

4 Discussion

The results from both of these experiments show differences not just between empathy production and perception, although this could be a consequence of the naturalness of samples, but also that within production and perception, there is a divide in the treatment of longer and shorter empathetic productions. This helps to show that empathy is not unified in speech, as previous studies such as McHenry et al. (2012) or Weiste and Peräkylä (2014) convey. However, the fact that the results of this study are in agreement with some of the previous results, which use a different dialect of English and a different language respectively, seems to imply a cross-cultural consistency in at least some aspects of empathetic realisation.

Correlates, such as intensity and precision, that have been used in other emotional speech studies were not included in the analysis of the recorded samples. This could potentially alter the overarching conclusions drawn on the perception of empathy, but a clear image of empathy is created through analysing pitch and duration in particular. Empathetic speech appears to be divided between shorter and longer

utterances, where shorter utterances signal empathy more through local duration increase, and longer ones through a decrease in both pitch level and span.

The results from the perception experiment are hindered by their naturalness, but where this is overcome in 4F's sample, it seems that the results from the perception experiment match those of the production experiment, and there is potentially even some kind gradience, which provides evidence for Boukricha et al.'s (2013) underlying assumption of gradience in empathy. Of course, all conclusions drawn at this stage must be tentative, due to the insignificance of any other factors in other samples in relation to empathy, and further work must be done with more natural samples, or indeed samples that are further from the natural voice to avoid unwanted subconscious interferences from naturalness. However, the difference between longer and shorter empathetic utterances still holds, as the only results that showed significance were shorter empathetic utterances.

The discrepancy between theoretical predictions and the empirical production evidence can be explained by the fact that acted speech does not behave the same way as natural speech (Johnstone & Scherer, 1999), and as such it would be prudent to explore other contexts in which empathy is elicited alongside this acted speech context, to see if these discrepancies are methodological or part of the nature of empathy itself.

4.1 Limitations

The naturalness of a parametric approach proved to be an obstacle in obtaining results in the perception study. Although resynthesis was chosen over synthesis from scratch to increase the naturalness of samples, and although a pilot naturalness test was conducted, it seemed that naturalness was still an issue for the empathy perception samples. It could be the case that this resynthesis approach created a kind of acoustic 'uncanny valley' effect, where samples were close to sounding completely natural, but there were elements of unnatural speech sufficient to influence empathy rankings in participants. Experiments using resynthesis versus synthesis are necessary to verify this and ensure that the discrepancy between production and perception was methodological.

This study was seriously limited by the COVID-19 pandemic, and it is hoped that further work, free from pandemic-based restrictions, can help solve a large number of the issues encountered. A larger number of participants, recorded in the proper environment, would be possible, allowing for more data to be created and analysed. This could potentially include a larger number of stimuli, as the four stimuli in this experiment constitute a small sample size. In-person perception tests, not possible due to the pandemic, could also allow investigation of intensity as a potential correlate, since in-person tests allow experimenters to control audio output device as well as playback volume. Managing two experiments in a pandemic proved to be particularly difficult with ever-changing restrictions, and it is hoped that future work will not be seriously affected by external circumstances as this work was.

5 Conclusion

There is much that we do not yet know about the behaviour of empathy in speech, and this dissertation has aimed to outline its key correlates, both in production and perception, through a combination of dialogue recordings and analysis-by-(re)synthesis.

Empathetic speech manipulates pitch and duration, depending on the length of the utterance, both in production and perception. This is despite these experiments having different contexts, with the production experiment focusing on acted speech with contexts outlined, and the perception experiment with context totally removed. Lowered pitch level and a flattened pitch contour (narrower span) are significant for longer utterances, and increased local duration of accented syllables (but not increased global duration) are significant for shorter ones. Further research on multi-sentence empathetic utterances could reveal yet another type of empathetic production in speech. Other elements that factor into empathetic production, such as awareness of producing the emotion. This would help illustrate the behaviour of cognitive versus affective empathy, which was not studied in this dissertation. The context of the utterance (e.g., everyday interactions versus clinical settings), has not been explored in detail in this dissertation either, but it is clear that this is another avenue of potential research.

By shedding light on the behaviour of empathy in speech, further questions about the behaviour of secondary emotions in speech as a whole have been raised. More work is needed on how subjective secondary emotions are generally, and how they behave across contexts, as well as how secondary emotions relate to one another. There are indications in this study that there is some cross-cultural consistency. This apparent consistency could be the result of a bias towards studying Western languages, and investigation of non-Western languages and cultures (such as Japanese or Korean) could change this idea of consistency. In terms of speech synthesis, even though the parametric approach is now outdated, this approach works well as a means of testing hypotheses about secondary emotions, but the naturalness of samples needs to be considered carefully for perception studies. Parametric analysis-by-synthesis techniques can also help to explain any potential perceptual issues with secondary emotions in end-to-end text-to-speech (or indeed speech-to-text), as the utterance length sensitivity found in this study may not be picked up on depending on input data. Clearly, therefore, this research is far-reaching in its potential.

6 References

Alam, F., Danieli, M., & Riccardi, G. (2018). Annotating and modeling empathy in spoken conversations. *Computer Speech & Language*, 50, 40–61.

Amazon. (2021). *Speechcon Reference (Interjections): English (US) | Alexa Skills Kit*. Retrieved from *Amazon*: <https://developer.amazon.com/en-US/docs/alexa/custom-skills/speechcon-reference-interjections-english-us.html>.

Asada, M. (2015). Towards artificial empathy. *International Journal of Social Robotics*, 7(1), 19–33.

Bangayan, P., Long, C., Alwan, A. A., Kreiman, J., & Gerratt, B. R. (1997). Analysis by synthesis of pathological voices using the Klatt synthesizer. *Speech Communication*, *22*(4), 343–368.

Becker, C., & Wachsmuth, I. (2006). Modeling Primary and Secondary Emotions for a Believable Communication Agent. In D. Reichardt, P. Levi, ns J-JC. Meyer (Eds.), *Proceedings of the First Workshop on Emotion and Computing in conjunction with the 29th Annual German Conference on Artificial Intelligence* (pp. 31–34). Bremen: Universität Bielefeld.

Becker-Asano, C., & Wachsmuth, I. (2008). Affect Simulation with Primary and Secondary Emotions. In H. Prendinger, J. Lester, & M. Ishizuka (Eds.), *Intelligent Virtual Agents* (pp. 15–28). Berlin: Springer.

Bezooijen, R. van. (2011). *Characteristics and Recognizability of Vocal Expressions of Emotion*. Berlin: Walter de Gruyter.

Boukricha, H., Wachsmuth, I., Carminati, M. N., & Knoeferle, P. (2013). A Computational Model of Empathy: Empirical Evaluation. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, 1–6.

Breazeal, C. (2001). Emotive qualities in robot speech. *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*, 3, 1388–1394.

Brenner, C. (1980). A Psychoanalytic Theory of Affects. In R. Plutchik & H. Kellerman (Eds.), *Theories of Emotion* (pp. 341–348). New York: Academic Press.

Burkhardt, F., & Sendlmeier, W. F. (2000). *Verification of Acoustical Correlates of Emotional Speech using Formant-Synthesis*. ISCA Tutorial and Research Workshop (ITRW) on speech and emotion.

Busso, C., Lee, S., & Narayanan, S. (2009). Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(4), 582–596.

Cahn, J. E. (1990). The Generation of Affect in Synthesized Speech. *Journal of the American Voice I/O Society*, 8(1).

Cohn, M., & Zellou, G. (2019). Expressiveness Influences Human Vocal Alignment Toward voice-AI. *INTERSPEECH*, 2019, 41–45.

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A Review of the Concept. *Emotion Review*, 8(2), 144–153.

d'Heureuse, C. (2021). *Chdh/klatt-syn* [TypeScript]. Retrieved from: <https://github.com/chdh/klatt-syn>. (Original work published 2019).

Ekman, P. (1973). *Darwin and facial expression: A century of research in review*. New York: Academic Press.

Emde, R. N. (1980). Levels of Meaning for Infant Emotions: A Biosocial View. *Development of Cognition, Affect, and Social Relations: The Minnesota Symposium of Child Psychology*, 13, 1–37.

Erickson, D. (2005). Expressive speech: Production, perception and application to speech synthesis. *Acoustical Science and Technology*, 26(4), 317–325.

Fant, G., Liljencrants, J., & Lin, Q. (1985). A four-parameter model of glottal flow. *STL-QSPR*, *4*, 1–13.

Fletcher, J. (2010). The Prosody of Speech: Timing and Rhythm. In W. J. Hardcastle, J. Laver, and F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 521–602). Chichester: John Wiley & Sons.

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?. *Computer Speech & Language*, 28(2), 580–597.

Gobl, C., Bennett, E., & Ni Chasaide, A. (2002). Expressive synthesis: How crucial is voice quality? *Proceedings of 2002 IEEE Workshop on Speech Synthesis*, 91–94.

Graham, C. (2014). Fundamental Frequency Range in Japanese and English: The Case of Simultaneous Bilinguals. *Phonetica*, 71(4), 271–295.

Hatfield, E., Bensman, L., Thornton, P. D., & Rapson, R. L. (2014). New perspectives on emotional contagion: A review of classic and recent research on facial mimicry and contagion. *Interpersona* 8(2), 159–179.

Huggins, A. W. F. (1972). Just Noticeable Differences for Segment Duration in Natural Speech. *The Journal of the Acoustical Society of America*, 51(4B), 1270–1278.

James, J., Watson, C. I., & MacDonald, B. (2018). Artificial Empathy in Social Robots: An analysis of Emotions in Speech. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2018, 632–637.

Johnstone, T., & Scherer, K. R. (1999). The effects of emotions on voice quality. *XIVth International Congress of Phonetic Sciences*, 4, 2029–2032.

Keating, P., Garellek, M., & Kreiman, J. (2015). *Acoustic properties of different kinds of creaky voice*. International Congress of Phonetic Sciences.

Kemper, T. D. (1987). How Many Emotions Are There? Wedding the Social and the Autonomic Components. *American Journal of Sociology*, 93(2), 263–289.

Kitahara, Y., & Tokhura, Y. (1992). Prosodic Control to Express Emotions for Man-Machine Speech Interaction. *IEICE TRANSACTIONS on Fundamentals of Electronics, Communication and Computer Sciences*, *E75-A*(2), 155–163.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America*, 67(3), 971–995.

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Li, H., Kang, Y., & Wang, Z. (2018). EMPHASIS: An Emotional Phoneme-based Acoustic Model for Speech Synthesis System. *INTERSPEECH*, 2018, 3077–3081.

McHenry, M., Parker, P. A., Baile, W. F., & Lenzi, R. (2012). Voice analysis during bad news discussion in oncology: Reduced pitch, decreased speaking rate, and nonverbal communication of empathy. *Supportive Care in Cancer*, *20*(5), 1073–1078.

Mozziconacci, S. J. L., & Hermes, D. J. (2000). *Expression of emotion and attitude through temporal speech variations*. Sixth International Conference on Spoken Language Processing.

Murray, I. R., & Arnott, J. L. (1995). Implementation and testing of a system for producing emotion-by-rule in synthetic speech. *Speech Communication*, 16(4), 369–390.

Park, J. W., Kim, W. H., Lee, W. H., Kim, J. C., & Chung, M. J. (2011). How to completely use the PAD space for socially interactive robots. *2011 IEEE International Conference on Robotics and Biomimetics*, 2011, 3005–3010.

Patterson, D. (2000). *A Linguistic Approach to Pitch Modeling* [Unpublished PhD Dissertation]. Edinburgh University.

Podesva, R. J. (2011). Gender and the social meaning of non-modal phonation types. *Annual Meeting of the Berkeley Linguistics Society*, 37(1), 427–448.

Runyon, R. P., Coleman, K. A., & Pittenger, D. J. (2000). *Fundamentals of behavioral statistics, 9th ed.* New York: McGraw-Hill.

Schröder, M. (2001). *Emotional Speech Synthesis: A Review*. 4.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156(1), 81–96.

Wang, Y., Skerry-Ryan, R. J., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Z., & Bengio, S. (2017). Tacotron: A fully end-to-end text-to-speech synthesis model. *ArXiv Preprint ArXiv:1703.10135*, *164*.

Weiste, E., & Peräkylä, A. (2014). Prosody and empathic communication in psychotherapy interaction. *Psychotherapy Research*, 24(6), 687–701.

Xiao, B., Imel, Z. E., Atkins, D. C., Georgiou, P. G., & Narayanan, S. S. (2015). Analyzing Speech Rate Entrainment and Its Relation to Therapist Empathy in Drug Addiction Counseling. *INTERSPEECH, 2015*, 5.

Yalçın, Ö. N. (2019). Evaluating empathy in artificial agents. *ArXiv Preprint ArXiv:1908.05341*.

Yiu, E., Murdoch, B., Hird, K., & Lau, P. (2002). Perception of synthesized voice quality in connected speech by Cantonese speakers. *The Journal of the Acoustical Society of America*, 112 (3), 1091–1101.

7 Appendices

7.1 Appendix One: Dialogues for Examination of Natural Correlates

Participants alternated between being Speaker A and Speaker B, according to the stimulus; the participant producing Stimuli 1 and 2 is highlighted in yellow, and the participant producing Stimuli 3 and 4 is highlighted in green.

Dialogue 1
(Speaker A is meeting Speaker B for coffee)
A: Hey, sorry I'm late. I sort of got caught up trying to finish my work.
B: No worries — are you feeling OK about your work?
A: Not really — I can't keep on top of it and I have too many deadlines to be able to do anything well enough.
B: It's tricky — I know where you're coming from. Hopefully your deadlines will clear up soon! The coffee's on me this week.

Dialogue 2
(Speaker A and Speaker B are discussing a piece of work)
A: What about trying to solve it with this formula instead?
B: That's hard. I think that's probably more difficult than the way I'm trying to solve it now.
A: Are you sure? I thought it was easier this way.
B: I mean I can try it your way, but I'm not sure it'll work.

Dialogue 3
(Speaker A is explaining a task to Speaker B)
A: It's fairly simple — what you have to do is sort all the emails by the date stamp in the top right corner.
B: I understand.
A: Do you have any questions about it?
B: No, it all makes sense.

Dialogue 4
(Speaker A and Speaker B are talking about a friendship)
A: I've been finding it really difficult to keep her in my life — it's too much of a burden on me.
B: Entirely fair — that's hard. It's tricky when you want someone as your friend but it isn't working.
A: I know, and I feel guilty for it.
B: Honestly, you shouldn't. Sometimes things just play out this way.

Dialogue 5

(Speaker A has called Speaker B on the phone)

A: I think I've taken a wrong turn — I'm on market street heading down towards the church, but I'm not sure how to get to where you are.

B: The church? I know where you're coming from. Turn right at the end of the road and you should see where I am straight away.

A: Thank you! Is it the red door?

B: Yes — just knock when you arrive.

Dialogue 6

(Speaker A is trying to do a video call with Speaker B)

A: Can you hear me? I'm not sure if my connection is working.

B: I hear what you're saying. The issue is your camera doesn't seem to be on. Can you try that?

A: Is this better?

B: That's fantastic — I can see you now.

Dialogue 7

(Speaker A is talking to Speaker B about pulling out of a commitment)

A: I'm really sorry, it's just that I have too much on and I don't think I'd do a good enough job of it at this point.

B: I understand. Is there anything we can do to keep you involved? Like a small side role?

A: I think that could be possible.

Dialogue 8

(Speaker A and Speaker B are discussing difficulties in a relationship)

A: I don't know, it feels like every time I try to bring up a problem, it's like he can't hear me.

B: It's difficult when things play out like that. I hear what you're saying. Is there anything I can do to help?

A: No, it's ok — just hearing me out is enough of a help.

7.2 Appendix Two: Tables of Production Study Results

Table A1: *Recordings of Pitch (Hz) to Calculate Level and Span.*

Speaker	Gender	Condition	Stimulus	Onset	iH*	L	H*	FL
1M	0	1	4	186.1836	314.6444	185.0633	188.2561	111.1049
2M	0	1	4	171.6812	243.7277	138.5375	170.7096	113.8491
3M	0	1	4	147.4196	223.7096		174.2714	99.3598
4F	1	1	4	243.054	280.0496	178.1281	246.6126	216.3599
5F	1	1	4	254.0662	317.7307	242.7403	304.5188	192.6017
1M	0	2	4	132.8034	177.2732	124.402	132.7516	106.3324
2M	0	2	4	128.6286	146.813	113.9586	133.9459	116.3762
3M	0	2	4	131.8342	159.1047			100.0057
4F	1	2	4	218.7229	213.1621	198.7427	183.1586	179.4615

5F	1	2	4	212.1362	252.3341	177.9944	214.1353	176.0216
1F	1	1	1	217.8663	239.151	196.0871	202.0269	168.3874
2F	1	1	1	209.0436	237.3267	215.7522	236.2947	155.474
3F	1	1	1	215.8901	252.5421	198.1099	215.138	182.8348
4M	0	1	1	182.4433	216.0916	160.7704	163.0261	126.3148
5M	0	1	1	154.7202	169.2912	133.2186	143.5788	113.3869
1F	1	2	1	219.2405	249.1521	198.3934	214.5653	193.2744
2F	1	2	1	210.2398	210.3221	185.1298	196.2252	166.1931
3F	1	2	1	197.1213	205.5711	188.0919	202.9164	175.7269
4M	0	2	1	132.0265	157.8302	133.2666	155.5577	127.1892
5M	0	2	1	128.8578	153.6241	133.3716	139.7998	106.7242

Speaker	Gender	Condition	Stimulus	Onset	H	L*
1M	0	1	3	134.9883	141.5012	101.2258
2M	0	1	3	134.632	142.3951	113.2533
3M	0	1	3	135.2472	141.0034	
4F	1	1	3	227.529	211.8973	
5F	1	1	3	242.4128	269.8936	219.5254
1M	0	2	3	169.9053	177.3776	116.2968
2M	0	2	3	154.2155	160.4436	109.4331
3M	0	2	3	135.5188	137.1657	96.61737
4F	1	2	3	246.2417	242.552	198.1112
5F	1	2	3	226.1611	230.4029	188.0907
1F	1	1	2	245.718	313.6697	263.0393
2F	1	1	2	213.0834	203.8372	179.3743
3F	1	1	2	196.4035	189.6419	164.1068
4M	0	1	2	179.4981	185.7792	119.0717
5M	0	1	2	177.4619	177.9571	112.9983
1F	1	2	2	243.702	228.1987	202.5876
2F	1	2	2			175.8134
3F	1	2	2	203.4197	188.9772	151.5105
4M	0	2	2	144.978	144.5198	
5M	0	2	2	162.8449	151.1602	121.0556

Table A2: *Difference in Global Duration between Non-Empathetic and Empathetic Conditions*

Stimulus	Participant	Deviation from N (%)	Difference (ms)
1	1F	+2.29	+2.2
	2F	+8.88	+9.2
	3F	+10.15	+8.9
	4M	+16.72	+17

	5M	-0.35	-0.3
2	1F	-9.76	-5.7
	2F	+8.05	+4.7
	3F	+2.66	+1.5
	4M	-24.95	-13.7
	5M	+4.88	+2.2
3	1M	+14.07	+8.2
	2M	-13.47	-9.5
	4F	-12.86	-7.6
	3M	+11.61	+72
	5F	+20.17	+11.6
4	1M	-14.57	-13.1
	2M	+7.09	+6.6
	4F	+7.85	+7.1
	3M	-33.16	-31
	5F	+1.33	+1.1

Table A3: *Difference in H1-H2 between Non-Empathetic and Empathetic Conditions.*

Long Stimuli			
Speaker	Stim.	Non-Emp. H1-H2 (dB)	Empathetic H1-H2 (dB)
1F	1	-15.5	4.8
2F	1	2.9	-15.4
3F	1	-9.8	8.6
5M	1	-16.3	-7.7
1M	4	-7.8	-11.2
2M	4	-9.7	-14.7
4F	4	-1	-16
5F	4	10.9	11.7

Short Stimuli			
Speaker	Stim.	Non-Emp. H1-H2 (dB)	Empathetic H1-H2 (dB)
1F	2	0.4	3.9
2F	2	7.2	-10.9
3F	2	3.8	10
5M	2	-8.5	-9.2
1M	3	-4.7	-9.7
2M	3	-13.6	-15.9
4F	3	-9.7	-7.7
5F	3	6.5	4.9

7.3 Appendix Three: Survey Designs

Below are listed the introductions to the pilot and main experiments, detailing the instructions that participants were given:

Pilot Experiment

This survey is designed to collect information on perceptions of whether an extract of speech sounds 'normal' or 'abnormal' (unnatural). By participating in this experiment you consent to the recording of your gender, if you are a native speaker of English, and your responses to whether or not the audio sounds normal/abnormal. No other information will be recorded.

Main Experiment

This test is designed to measure perceptions of empathy in speech. Empathy is when a person can identify with and share the feelings of another person. For example, B is showing empathy in the following dialogue:

(Speaker A and Speaker B are talking about a friendship)
A: I've been finding it really difficult to keep her in my life — it's too much of a burden on me.
B: Entirely fair — that's hard. It's tricky when you want someone as your friend but it isn't working
A: I know, and I feel guilty for it.
B: Honestly, you shouldn't. Sometimes things just play out this way.

In this test, you will hear a number of different samples of voices, some male and some female. Your task is to give the voices a rating from 0 to 4 depending on how empathetic they sound. More simply put, give these voices a rating based on how much you believe they are being empathetic. If a voice does not sound empathetic, then please rate it 0. If it sounds slightly but not totally empathetic, select one of the middle values, and if the voice sounds fully empathetic, then select 4 (the highest value).

Acknowledgements

I would like to thank Calbert Graham for his invaluable advice and guidance through the whole experimenting and writing process, particularly for his tireless efforts during editing, particularly given the pandemic circumstances. Any remaining errors are my own.

The Unsolved Problem of Language Identification: A GMM-based Approach

Maggie Mi
Lancaster University

Abstract. Language identification (LID) systems attempt to identify a language from a series of randomly spoken utterances (Das & Roy, 2019), and this provides the foundation of many natural language processing (NLP) applications, such as multimedia mining, spoken-document retrieval, as well as multilingual spoken dialogue systems (Navratil, 2006). However, presently, the LID task is still an unsolved problem, often with increasing equal error rate (EER) as the duration and quality of the dataset decreases (Ambikairajah et al., 2011). The HMM-GMM (Hidden Markov Model-Gassian Mixture Model) approach taken in this paper involves building an acoustic model that uses probabilistic representations of speech datasets across 10 languages (Dutch, Russian, Italian, Portuguese, German, English, French, Turkish, and Greek). Through the exploration of the cross-linguistic features present in language families and the effect of the experimental parameters on the performance of the system, i.e., the length of the data recording, areas of weaknesses and corresponding means of improvement are therefore revealed.

Keywords: Spoken language identification; speech processing; HMM-GMM; computational linguistics; natural language processing (NLP)

---

1 Introduction

1.1 Background

Automatic Language Identification (LID) is defined as, 'the identification a language from random[ly] spoken utterances' (Das & Roy, 2019, p.81). Although this technology is also available for written text, for example, Google Translate's 'Detect language' function, the focus of this project is from a speech technology standpoint. The implementation of LID systems can be seen throughout many different areas. For instance, technological conglomerates require such a system to aid the categorisation of user data. Completing this manually will not only be labour intensive but also costly in terms of human resources and time efficiency (Baldwin & Lui, 2010). Telephonically, LID systems have been often used for emergency services and call centres, including parts of the tourism industry (Muthusamy, 1993). The shared purpose of these systems is the ability to route the caller to the right interpreter, or agents speaking the same language, to help the caller with their needs. Thus, the emphasis on accuracy is particularly important, especially, in life-and-death situations dealt with by emergency services.

What makes this field particular intriguing, is the attempt at capturing the characteristic and representative features of natural language, through speech data, which is highly idiosyncratic and inconsistent, with the quality of the data directly affecting the performance of such systems (Xu, Ding, &

Watanabe, 2019). Speech, being less tangible than textual data, is temporal. It is also liable to an individual's articulatory differences. For example, younger speakers tend to have high F1 and F1 values, similarly, in general, the vocal tract length of males is greater than females, thus, giving rise to inevitable variety (Kumar et al., 2011; Rastatter et al., 1997).

Taking the International Phonetics Alphabet as a framework, it is proposed that, there are in total 107 phones deemed producible by the human anatomy. Combined with 4 prosodic markers of intonation and 52 diacritics marks, this means, there are a finite number of possible combinations of sounds, and these combinations often overlap in many languages (Association & Staff, 1999). An example would be 'bùkě' (不可), meaning 'must not; do not', as opposed to 'book' in English. Being typologically different languages, Chinese, a Sino-Tibetan language, has a system of intricate tones, which are absent from the (Proto-)Indo-European languages, such as English. However, despite intonation differences, there also exists phonemically similar utterances across languages, regardless of the typological language family distinction. Even in English, phrases such as, 'Let's recognize speech' and 'Let's wreck a nice beach' shares a similar combination of phonemes and thus, at times, even humans can perceptively misinterpret the two. Therefore, the project aims to explore the following areas: the pattern of recognition performance between typologically similar languages, weaknesses of such a system and means to improve these shortcomings, and finally, whether experimental methodology and parameters, such as duration of the data would affect classification results. The approach taken to explore these research questions involves building a GMM-HMM LID system. Compared to rule-based systems, which aims to break down a language into different parameters and corresponding rules and assumptions, GMM-HMMs are conceivably more advantageous in terms of capturing an overall representation of a language (Ives, 1986; Zissman, 1993).

In the subsequent sections, the system design will be provided, setting out the inner workings of the system (Section 4). This is followed by descriptions of the systems' performance in Section 5. Sections 6 then aims to analyse the results and discusses ways to improve the shortcomings of the system. The goal of this project is to obtain an in-depth understanding of this statistical approach towards LID, and to overcome the issue of misclassification, as much as possible.

2 Literature Review

Initial studies of LID systems date back to the 1970s. Since then, a various range of techniques has been developed. In this section, a brief history of previous methodologies used in building LID systems will be analysed with reasons for choosing GMM-HMM modelling as the methodology of this project.
In order to narrow down the scope of the focus placed on the literature; certain factors were kept in mind when completing this review. These include:

(A) Number and similarity of languages (i.e., language families): The fewer the languages, the easier the recognition task; similar languages are harder for identification.
(B) Data used for the study: whether there is an overlap in speakers present in test and training data; gender-inclusiveness in the data; quality of data (i.e., recording conditions) …etc all these features will directly affect the meaningful conclusions, that can be drawn from a study.
(C) Methodology: HMMs-based approaches, clustering, neural networks…etc

2.1 Static Classification Methods

Leonard and Doddington (1974) were believed to be the earliest scholars of this field. Their approach to the problem involved extracting filter bank feature vectors and spotting regions of change and stability. Filter bank feature vectors is a type of quantified representation of an acoustic signal. A filter bank typically contains sets of bandpass filters that permit certain frequencies of a specified range but rejects frequencies outside that range. Each bandpass filter can be thought of as being able to 'select' or 'distinguish' a particular band of frequencies from the input signal. The feature vector produced by the filter bank is a vector, or conceptually, a string of numbers, representing 'the amount of energy in each frequency band' (King, 2017). Thus, by inspecting the filter bank feature vectors, such regions of change and stabilisation were taken to be indicative of a specific language. Accordingly, these regions were then used as templatic patterns for comparison with the test data.

Due to the classified nature of this research the languages used were not disclosed, nor was background information of the 100 speakers. Thus, it is difficult to come to evaluate the 70% correct classification across the five languages without additional information such as typological, or phonemic similarities in these languages, or metadata of the participants.

2.2 Importance/Presence of Linguistic Units

Later studies completed by Leonard and Doddington cemented the idea that linguistics units are most likely to differentiate languages. Due to the varied occurrences of linguistic units in different languages, it was hypothesised that using the distributions of frequency of such units, different languages can be distinguished. As sounds can be represented in terms of their acoustic representation, thus, extracting features from the speech signal retains the unique features of the language (Leonard & Doddington, 1974; 1978)[42].

Cimarusti and Ives (1982) found that discriminating features are not necessarily linguistic features. Their approach disregarded the idea of linguistic units, such as syllables or phones, and instead, they implemented pattern analysis techniques to acoustic features extracted from the speech signal. The overall result of the system was 84% performance accuracy, with scores ranging from 76.8% (American English) to 93.4% (Korean). From these results, it is plausible to conclude that, acoustic features alone can be used in LID. Although, this would need to be further tested as the dataset has a small sample size of 5 speakers per language. This also leads to the speculation as to whether the system is "speaker-independent", meaning the system could be picking out characteristics of the speaker as opposed to the language.

In a subsequent study conducted by Ives (1986), both prosodic feature vectors and formants values used as a basis for a LID system were explored. It was found that formant frequencies are more useful in distinguishing languages. Thus, classification was based on thresholds and densities of quantified representation of each language. LID was then based on sets of 'production rules', defined by variance, the value of F0 and variance of F2, amongst other parameters. Classification results of this system had an overall accuracy of 92%, and individual classification results ranging from 84% (Russian) to 99% (Vietnamese). Despite the promising results, the study lacked information on the amount of data and test/training data assignments. Moreover, the possibility of an overlap between training and test sets were not eliminated. Given the approach taken in building this system is based on formants, it is critical for the system to be exposed to speech produced by female speakers. This is because, acoustically, female speakers

[42] This foundational assumption has fuelled many other works in the field, such as Cimarusti and Ives (1982) and Foil (1986).

tend to have higher F0 values than males, thus, it would be interesting to see how the system would perform when exposed to female speakers (Muthusamy, 1993).

2.3 (GMM-)HHM-based Studies

A prominent study following Leonard and Doddington (1974) is completed by House, Neuberg, and Wohlford (1975). In this leading study, House and colleagues proved the possibility of using sequences of phonetic categories (stops, fricatives, vowels, silences) to tackle LID. Working under the assumption, that these categories can be demonstrated as a Markov process, and that the model's parameters could be projected from necessary training data, phonetic transcriptions were used for the following eight languages: Urdu, American English, Greek, Chinese, Japanese, Swahili, Korean and Russian. Phonetic category labels from these phonetic transcripts were then used to train statistical models. The underlying concept is given a sequence of category symbols U, and language L, P(U|L) is calculated for each language. 'U is said to represent the language L, for which this probability is a maximum' (House et al., 1975; Muthusamy, 1993, p.14).

The study of House et al. (1975) is an important underpinning foundational work for many future designs of the LID and speech recognition systems. However, in their actual experiment, speech recordings were not used, only phonetic transcriptions of texts from each of the eight languages. This is a major flaw in methodology design, as 'absolute', or 'perfect' acoustic segmentation of real speech data is very much unfeasible (Muthusamy, 1993). The continuous sinusoidal nature of speech signals' frequencies further proves this. Thus, this experiment is not representative of the performance and effectiveness of the system on actual speech data.

Li and Edwards further advanced the grounding HMM techniques by House et al. (1975) and used them in real speech data (Li & Edwards, 1980). They used a segmentation framework that contains six categories: syllabic nuclei, non-vowel sonorants, vocal murmur, voiced frication, voiceless frication and silence, or low energy segments. The segmental models characterised the likely, segmental sequences in the language, whereas the syllable model was split into two types: inter-syllable-nuclei sequences and intra-syllable-nucleus segment sequences. The former can be thought of as depicting the likelihoods of possible consonant clusters, and the latter, as the likelihood of a certain internal arrangement of a syllable.

The training data consisted of 200 minutes from 20 speakers of five languages, three Indo-European and two Asian, reading aloud. Due to the nature of the Asian languages being tonal and monosyllabic, the utterances follow straightforward consonant-vowel (CV) or consonant-vowel-consonant (CVC) formations. In contrast, the rest of the dataset is far more complex, with compounded consonant clusters and longer word lengths. A maximum of approximately 80% correct identification was gained using the inter-syllable model. Analysing the mistakes made by the system revealed that, this approach excellently differentiated the Asian languages from Indo-European languages (Li & Edwards, 1980).

As an extension on the work of House et al. (1975), this study supports the hypothesis that, 'broad phonetic category sequences do possess language discriminatory information' (Muthusamy, 1993, p.16). However, this study contains several caveats, such as the limited information on the languages used, quality of recording conditions…etc are all unknown factors. Moreover, the study lacks consideration of variability as expected in LID systems. Namely, the nature of reading speech and the absence of female speakers greatly limits the reproducibility of the results in real-world applications.

A final notable study that needs to be addressed is one completed by Nakagawa, Ueda, and

Seino (1992). In this study, four different approaches to LID are compared: vector quantisation (VQ), discrete HMM, continuous density HMM and a mixture Gaussian distribution model.

> 'A mixtured Gaussian distribution model is regarded as a special case of a continuous HMM with mixtured distributions, that is, the number of states corresponds to only one and the distribution is Gaussian.' (Nakagawa et al., 1992, p.1012).

Conceptually, this can be thought of as modelling the probability of a language occurring using normal distributions[43].

The dataset consisted of 750 recordings from four languages: English, Japanese, Mandarin Chinese, and Indonesian. Each language had 15 native speakers, producing 50 sentences e ach, with an average duration of the utterance at approximately, 3 seconds. Results from the continuous HMMs and GMM-HMM approaches yielded the best results, both at 81.1%, and were far more superior than the VQ (77.4%) and discrete HMMs approaches (47.6%). Despite this accomplished result, the generality of this study is once again limited by the lack of female speakers.

The studies reviewed above in 2.1-2.2, all primarily performed 'static classification'. HMMs, on the other hand, are more dynamic. They can 'model the sequential characteristics of speech production and have been used widely in speech recognition systems' (Zissman, 1993, p.399). As noted previously, languages differ from each other due to differences in phonemic inventories, as well as the realisation of similar phonemes in particular languages. For example, the fricative in German 'ich', has no Italian counterpart. Another example would be the /r/ in Spanish differing from its English counterpart (the former is realised as a trill, whereas the latter is a flap). Therefore, due to the complexity and intricacy of these small differences, it is more logical to use a comprehensive statistical modelling approach, to capture the characteristics of a language, rather than specified, preselected features.

Furthermore, in HMM-based LID systems, a language is identified by estimating the likelihood of each language occurring at 'contiguous, frames of the speech signal' (Radha, 2012, p.1101). Unlike static models, HMMs can be trained to represent different items, thus, this flexibility in defining the scope of a 'unit', means that HMMs are advantageous in building speaker-independent and text-independent systems. This, coupled with the temporal nature of speech data, and the ability to encode a language as a string of quantified representations, in the form of a spectral vector, makes GMM-HMM the preferred approach in building a LID system (Gales & Young, 2008).

3 Data

A common trend that is visible across the literature, is that the number of languages present in these systems usually does not exceed eight. There is a visible balance between building a system that deals with a high degree of complexity (i.e., multi-lingual compatibility) and one that has a high accuracy of classification. In the GMM-HMM system of Zissman (1993), one of the datasets used to train and test the system was the 20 languages CCITT database. Notably, the classification result of this system was the lowest, at 54%. On average, however, previous studies completed tend to have 2-8 languages. Therefore, an important aspect is to explore the shortcomings of these systems that aims to classify numerous languages. With this goal in mind, this project aims to sample as many languages as possible, eventually arriving at 10 languages.

[43] Please see Section 4 for a detailed account of how GMM-HMMs work.

3.1 Data Collection

Voxforge is a free open-source speech database, where people from all around the world voluntarily contribute speech data for the development of open-source speech recognition systems[44]. The repository offers speech datasets in the following languages: Albanian, Bulgarian, Catalan, Croatian, Dutch, English, French, German, Greek, Hebrew, Italian, Mandarin Chinese, Persian (Farsi), Portuguese, Russian, Spanish, Turkish, and Ukrainian. However, there is great variability in the datasets in terms of availability. For example, at the time of completion of this project, the Mandarin Chinese dataset is not available for download. On a similar note, within the datasets that are available for download, there exists a clear discrepancy between the magnitude of the datasets- Croatian, for instance, only contains two speakers, whereas English has 1234.

Thus, certain parameters are needed to control the variability of the dataset, to ensure unbiased results. These parameters are also necessary for data selection and refinement. See Table 1.

[44] http://www.voxforge.org/

Table 1: *Data Selection Criteria and Remarks.*

Parameter	Requirement	Method Deployed	Remarks
Magnitude of Dataset (hours)	At least 2.5 hours of speech for each language. Less data would prove difficult to train up statistical models.	The metrics provided on Voxforge was used as a guide, to discard languages with small datasets	Persian (Farsi), Croatian, Hebrew, Albanian, Catalan, Ukrainian, Bulgarian, were dropped from the final dataset.
Number of Speakers	At least 40 speakers. To minimise the probability of the system being speaker-dependent, it is necessary to expose it with as many speakers as possible.	The metrics provided on Voxforge was used as a guide, to discard languages with a small number of speakers	Persian (Farsi), Croatian, Hebrew, Albanian, Catalan, Ukrainian, Bulgarian, were dropped from the final dataset.
Quality of Audio	Discernible speech, with minimal background noise	Random sampling technique – sampled 10% of the audio files of each language. Files with noisy data, singing, intangible utterances were discarded.	The total number of audio files added up to more than 339.9 hours of audio. Thus, it was more time-efficient to use random sampling to check through the datasets.
Sample Rate (kHz)	16 kHz was used, in line with industry standard.	Downloaded 16kHz .wav files directly from Voxforge, and/or passed through ffmpeg for conversion.	See Section 3.3.2

Table 2: *Summary of Dataset*

Language	Format	Overall Length (Hours)	Number of Speakers
Turkish	WAV	2.8	57
Greek	WAV	3.8	44
Dutch	WAV	10.6	103
Italian	WAV	20	203
Russian	WAV	24.8	207
French	WAV + FLAC	37.5	320
Spanish	WAV	52.4	477
German	WAV + FLAC	57.1	111
English	WAV	130.9	1234
Portugues	WAV	N/A	N/A
Total		339.9	2756

'N/A' labels were used to signify missing metric information from Voxforge. However, after downloading and unpacking the datasets, that did not have comprehensive metadata, such as Portuguese, it was found that, with manual inspection, these datasets were sufficient and fitted the selection criteria. Therefore, using the 'total' tally as a minimum count, the dataset included more than 339.9 hours of speech, provided by more than 2756 speakers. This gives rise to a total of 207,463 .wav files.

3.2 Data Assignment

Building such a LID system requires two stages: training and testing. The training stage involves exposing the 'blank' system to a set of data, so the audio files can be analysed, and statistical models are created. The testing stage, then, involves exposing to the system a set of unseen data, which offers a sort of confirmation to verify whether outputs from the models are the expected results. Thus, the data were split accordingly: 80.7% (167,496 .wav files) were assigned as training data; 19.3% (39,976 .wav files) were assigned as unseen, test data. In total, the dataset involves 207, 436 .wav files.

3.3 Data Preparation

3.3.1 Task Grammar

The goal of the system is to identify the language being spoken in an audio file. The system must be able to handle ten languages, without being confined by the content and vocabulary of the utterance. Therefore, a 'blueprint' is needed to inform the system as to what are the possible units that it should recognise and the possible context in which it can appear. It should be noted that by 'units', it means 'languages' in this instance, as the system is trying to represent each language. This 'blueprint' is known as the task grammar.

The Hidden-Markov Tool Kit (HTK) 'provides a grammar definition language for specifying simple task grammars' (Young et al., 2002, p.25). Importantly, the task grammar is directly tied to the acoustic models, thus, the languages encompassed by the WORD variable must be the languages that are used to train the acoustic model.

3.3.2 16kHz Sample Rate and Format Conversion

All data files must comply with HTK's configuration. This involves the conversion of all audio files to .wav format. Moreover, for consistency, the dataset must have a sample rate of 16kHz. 16kHz is an industry -as it is the threshold for which higher levels have no substantial effect on the system performance (Ashihara, 2007).

It is worth noting that some of the data downloaded from Voxforge are corrupted. Some are also saved as older formats, such as FLAC and AIFF. Some of these files encountered problems during the conversion process and thus, they were discarded.

4 System Description

The heart of all speech recognition systems is composed of a set of statistical models, that aims to represent the acoustic and probabilistic information of the dataset. Unlike the static classification methods mentioned above, GMM-HMMs can capture a more comprehensive representation of the audio, as it takes into consideration the entirety of the audio, as opposed to certain preselected components (Muthusamy, Barnard, & Cole, 1994). The following subsections aim to expand on this process in more detail and offer deeper insights into the previously mentioned concepts that are lacking explanation. Importantly, the architecture of this system is an actualisation of the methodologies described by Ambikairajah et al. (2011) and this paper was used as a fundamental guide in the system's blueprint.

4.1 System Design: An Overview

The basic rubrics of a LID system's front-end is illustrated in Figure 1.

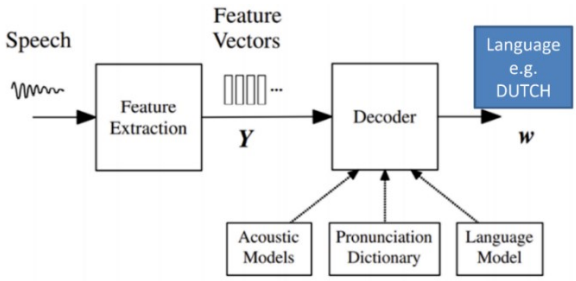


Figure 1: *System Design*

Adapted from (Gales & Young, 2008, p.201) The approach mainly consists of two components: feature extraction and decoding. Fundamentally, these two stages can be conceptualised as 'signal modelling' and 'pattern matching', respectively (Kesarkar & Rao, 2003, p.1).

Broadly speaking, the feature extraction stage involves converting the speech signal to a sequence of acoustic vectors, thus, parameterises the speech waveform. Parameterisation is required for extraction of the most relevant information from speech waveform and disregarding unnecessary noise. In practice, this achieved by converting the speech waveform, frame-by-frame, to a single N-dimensional vector, where $N=26$ for this study. Therefore, each utterance is converted into a sequence of vectors, $X = [x_1, x_2, \ldots x_{26}]$,

where x_{26} is a 26-dimensional vector. The decoder then finds the sequences of words matching to the acoustic vectors by consulting the acoustic models, lexicon, and language model. The latter of which are stored as $\{\lambda_l \mid l = 1,2, \dots L\}$, where L is the total number of languages, or in this case, 10. The set of features vectors, X, is then used to carry out model training, where a separate model, λ_l is generated for each possible language, l.

The feature extraction stage is then repeated for a set of unseen data, in the identification stage. This newly extract feature set is then compared to the model set, $\{\lambda_l \mid l = 1,2, \dots L\}$, to identify which l has the highest probability of producing the feature vector X. Mathematically, this involves pinpointing the language model λ_l that 'maximises a posteriori probability across the set of language models' (Ambikairajah et al., 2011, p. 84). Thus, giving rise to the model below:

$$\hat{l} = arg \max_{1 \le l \le L} P(\lambda_l | \boldsymbol{X}) \quad (1)$$

Applying Bayes' Rule to (1):

$$\hat{l} = arg \max_{1 \le l \le L} \frac{P(\boldsymbol{X}|\lambda_l)P(\lambda_l)}{P(\boldsymbol{X})} \quad (2)$$

Notably, the key assumption made by Ambikairajah et al. (2011) is that each language model has the same likelihood and that, irrespective of the language model, P(X) is the same. The task at hand can thus be thought as identifying the language model that corresponds to the highest probability of X occurring (Ambikairajah et al., 2011, p. 84):

$$\hat{l} = arg \max_{1 \le l \le L} P(\boldsymbol{X}|\lambda_l) \quad (3)$$

The following subsections will look at these components and stages in greater detail.

4.2 Feature Extraction

The feature extraction technique used by HTK is Mel-Frequency Cepstral Coefficients (MFCCs). In total, there four stages to feature extraction and they are: (1) pre-emphasis, framing and windowing, (2) Fast Fourier Transform, (3) Mel Filter Bank, and (4) Log() Compression and Discrete Cosine Transform.

4.2.1 Pre-emphasis, Framing and Windowing

The first step in MFCC extraction involves boosting the speech signal by passing it through a filter that emphasises higher frequencies. The goal of this step is to 'compensate the high frequencies that are suppressed during humans' sound production', and in turn, amplifying 'the importance of high-frequency formants' (Jang, 2011). As one would expect, there tends to be less energy at the higher frequencies, thus, this boosting is needed to overcome spectral tilt. In this specific scenario, the PREEMCOEF is set to 0.97, to achieve the necessary boost.

Windowing refers to segmenting the speech signal into small blocks of duration. These segments, or windows, are then used to determine each parameter vector (Singh & Rani, 2014). In this case, the window size is 25ms[45]. Importantly, to ensure continuity of the signal, each window overlaps with each other by 10ms. MFCCs are consequently extracted for each window.

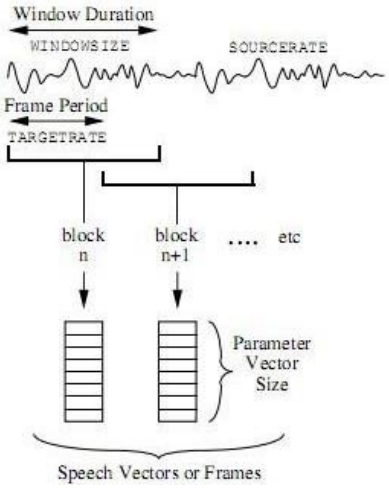


Figure 2: *Visualisation of Windowing* (Young et al., 2002, p.92)

4.2.2 Fast Fourier Transform (FFT)

Fast Fourier Transform is then applied to each frame. The purpose of FFT is for the conversion of the time domain into a frequency domain. Thus, the magnitude frequency response of each frame can be obtained (Singh & Rani, 2014).

4.2.3 Mel-Spaced Filter Bank

The human ear is, arguably, the best speech recognition system currently available. This is accredited to its ability to disentangle arbitrary frequencies across the audio spectrum (Shrawankar & Thakare, 2013). Thus, drawing inspiration from the human ear, the design of MFCC extraction aims to 'operate in a similar non-linear manner, to improve recognition performance' (Young et al., 2002, p.94).

Using HTK's filter bank, that is based on Fourier transform, to produce an equal resolution on a Mel-scale, triangular filters are applied, and they are positioned equally throughout the Melscale (Young et al., 2002, p.95). In this case, a Mel filter bank of 26 channels is applied to the signal, spaced equally throughout. The Mel-scale is used to imitate the non-linear human ear perception of sound, so, it is more discriminative at lower frequencies than higher frequencies (Fayek, 2016). This means that a Mel-spaced filter bank tends to identify fewer extraction points at the top of the spectrum. Taking the windows mentioned previously, each window is transformed using a Fourier transform, and the magnitude values are obtained. Conceptually, this can be thought of as an attempt to capture the energy values of each window, to capture characteristic information of the sound. These energies are known as filter bank energies (FBEs).

[45] HTK specifies all durations in 100 nano second (ns) standard units.

4.2.4 Log() Compression and Discrete Cosine Transform (DCT)

Finally, a Discrete Cosine Transform (DCT) compresses and ranks the values representing the 'power spectrum' to those which are most useful for speech recognition. The log-compressed FBEs' correlations are reduced during this process. Moreover, redundant noise from the speech signal is also refined. For this system, 12 cepstra are returned.

4.3 Decoder

4.3.1 Acoustic Models

In a LID system, the acoustic model represents the acoustic features of a language. Taking the MFCC vectors of the training dataset, acoustic models are constructed for each language, using a 3-state-HMM, with emission distributions in the form of Gaussian Mixture Models (GMMs).

The Hidden Markov Model is used to model sequential data. As implied by the name, much of the underlying data is hidden or remains unknown. This ties in with the main assumption of the approach – that, only the probabilities of the current state are needed to predict the next one. The Hidden-Markov chain used in this project consists of three states and transitions, as well as the probabilities associated with them. More specifically, the HMM is defined by the following set of parameters:

> **Initial state probability π:** a probability distribution that represents the probability of the HMM starting in each state.
>
> **Transition probability A:** a matrix that indicates the likelihood of transition to other states in the HMM, given the current state.
>
> **Emission probability B:** a group of multivariate probability distributions (one for each state), that indicates the probability of producing the current state occurring, given the observed sequence of acoustic vectors. Noticeably, the need of a multivariate emission distribution is dependent on the observations' multivariate nature.

It should be noted that two additional states are added, namely, an initial state and a final state are added. This is because they are the entry and exit states, which act as dummy-like, non-emitting states. When one of these states is encountered, it will directly skip to the next one. Fundamentally, 'these non-emitting states serve as the connection terminal for two HMMs, enabling continuous speech recognition' (Kasuriya et al., 2003).

In this system, the emission probability distribution is 'based on a weighted sum of multivariate Gaussian distributions' (Ambikairajah et al., 2011, p.90). To obtain a more characteristic representation of the language, a single Gaussian HMMs is converted to multiple mixture component HMMs. This process of increasing the number of components in a mixture is called 'mixture splitting' (Young et al., 2002, p.200). Conceptually, this can be thought of as using two Gaussians to model the same information as a single one did before, and this act of 'splitting' and break-down into smaller parts, means that there are more components to depict a better picture of the data.

Fundamentally, 'a Gaussian Mixture is a function that is comprised of several Gaussians, each identified by $k \in [1, \ldots, K]$, where K is the number of clusters in the dataset' (Carrasco, 2019).
Applying it to the LID task at hand, a Gaussian Mixture is composed of sets of normal distributions, or Gaussians, denoted by k. Each Gaussian k is defined the following parameters:

Mean (μ) defines its centre or the peak.

Variance (σ^2) defines the spread or width of the distribution curve.

A Gaussian bell curve, therefore, is a probabilistic distribution that expresses the probability of a value occurring.

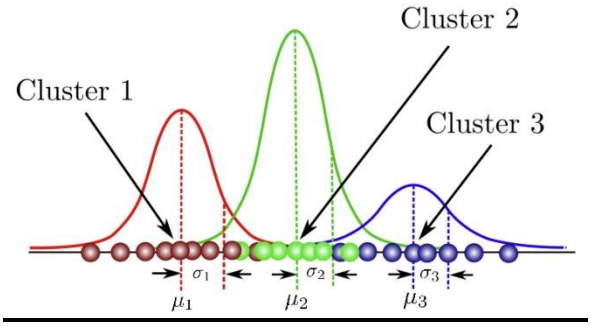


Figure 3: *Gaussians (Carrasco, 2019).*

For the purposes of this system, a more suitable approach is representing the emission distributions as a mixture of multiple multivariate Gaussian densities.

(4) $$g(x) = (\sum_{k=1}^{N} \lambda_k) N(x; \mu_k, \Sigma_k)$$

Or a more robust representation:

(5) $$b_m(x^{(t)}) = (\sum_{k=1}^{N} \lambda_k^{(m)}) \, N\left(x^{(t)}; \mu_k^{(m)}, \Sigma_k^{(m)}\right)$$

where $x^{(t)}$ is an observable vector at time t, whereas $\mu_k{}^{(m)}$ is the mean vector and $\Sigma_k{}^{(m)}$ is the covariance matrix of the k^{th} mixture component of the m^{th} state, and $\lambda(_k{}^{m)}$ is the mixture weight, constrained in such a way that:

(6) $$\sum_{k=1}^{N} \lambda_k^{(m)} = 1$$

4.3.2 Textual Representation Output

During recognition, clusters of features, that belongs to each model are then formed. Should the features extracted from the unseen, test samples, fall within the decision boundary of the language, the system would output these vectors as, for instance, 'German'. During this process, the language model, which quantifies all the possible sequences of acoustic features and it indicates which sequences are likely to occur, and which are less, are also consulted. (See Section 4.1).

Overall, a GMM system is very easy to train, as it does not require phonetic labelling nor orthographic transcriptions of the training speech.

5 Results

For ease of comparison of the results and further analysis, a way to group the data is needed and this categorisation is significant in decomposing the LID problem into manageable parts. Moreover, many of the recognition patterns also arise from the typological similarities that underpin these languages. Therefore, the 10 languages in the dataset are grouped according to their language family.

Table 3: *Language Family Categorisation.*

Germanic language(s)	Romance Language(s)	Turkic Language(s)	Slavic Language(s)	Hellenic Language(s)
German	Italian	Turkish	Russian	Greek
Dutch	Portuguese			
English	French			
	Spanish			

The results of the system performance are illustrated below in a confusion matrix. For visualisation purposes, a stacked box plot is also provided.

Table 4: *Confusion Matrix of Recognition Result.*

	Dutch	English	French	German	Greek	Italian	Portuguese	Russian	Spanish	Turkish	Performance (3.s.f)
Dutch	1030	250	56	22	81	99	34	23	83	33	60.2%
English	2066	7898	1611	363	1531	1227	586	454	1079	527	45.5%
French	293	839	1304	235	283	560	120	233	426	218	28.9%
German	313	658	303	4230	159	253	105	80	274	36	66.0%
Greek	10	61	23	1	47	19	7	12	80	10	17.4%
Italian	102	506	201	55	38	774	59	10	271	101	36.6%
Portuguese	24	145	89	25	43	57	253	29	53	17	34.4%
Russian	26	597	157	20	115	391	96	183	159	96	9.95%
Spanish	467	714	439	103	368	821	113	223	1135	318	24.1%
Turkish	0	23	23	9	25	44	15	49	49	92	28.0%

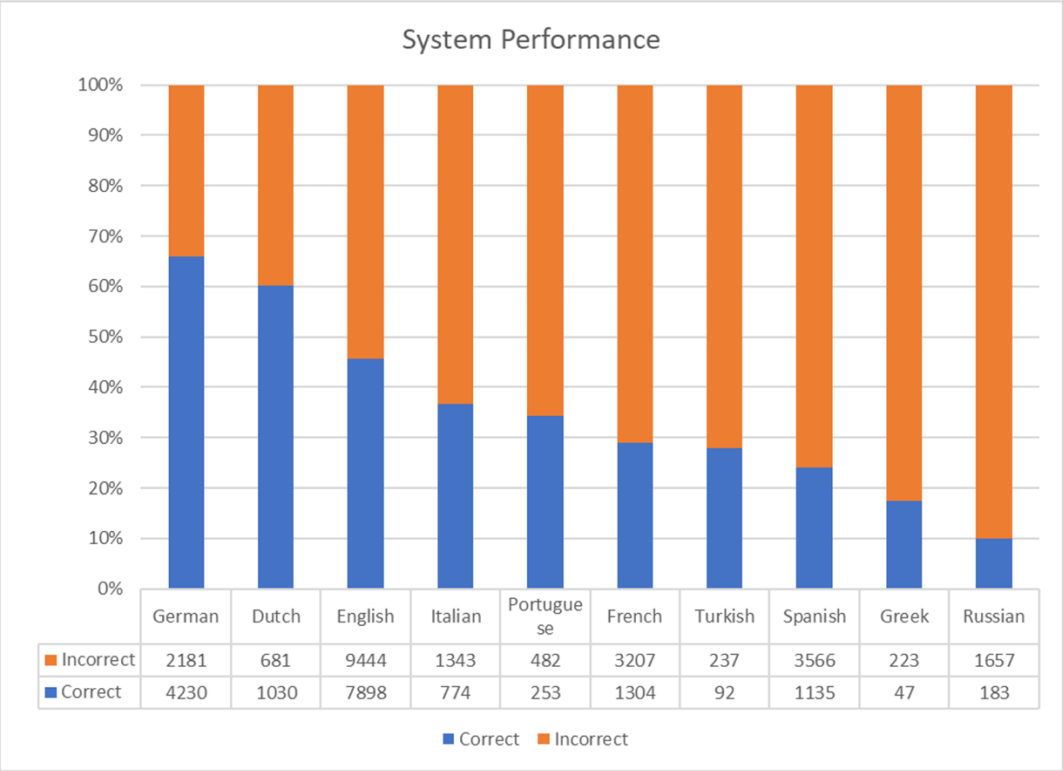


Figure 5: *Stacked Box Plot of Recognition Results.*

The confusion matrix shows the performance of the system. The overall accuracy of the system is 42.8%. whereas the average performance for a single language is 35.1%.

From the results, it is visible that, the system performed best in identifying German utterances. At a 66.0% correct recognition accuracy, this places German at the top of the ranking for performance, out of the ten languages. Notably, the two languages that follow German in classification performance are Dutch and English. Thus, there seems to be a pattern of the system excelling in the identification of Germanic languages. Dutch, similar to German, also has a high probability of being recognised. 60.2% of the Dutch test files were correctly identified, and 250 (14.6%) recordings were wrongly classified as English. For samples in English, the system correctly predicted 45.5% of the samples to be English. This 14.7% drop in

performance, mainly was due to the system misclassifying 2066 (11.9%) files of the English test dataset, like Dutch. On average, the German languages had a classification accuracy of 57.2%.

Romance languages have an overall lower classification accuracy of 31.0%. Out of the four languages, Italian had the lowest error rate, whereas Spanish had the highest. Portuguese and French followed Italian, respectively, but Spanish was preceded by Turkish. Compared to the Germanic languages, the spread of the performance accuracy is smaller, with a standard deviation of 5.63 compared to 10.6.

Interestingly, out of the three languages (Turkish, Greek and Russian), that are 'in isolation', that is, being the only languages representing a language family in the dataset, Turkish had the highest classification rate. With a classification accuracy of 28%, Turkish is ranked 7th preceding Greek (17.5%) and Russian (9.95%). Noticeably, from the Turkish results, equal amounts of data files (2 x 14.9%) were classified as Russian and Spanish, and similarly, for English and French (2 x 6.99%). These equal amounts of misclassification files could indicate the models have picked up similarities across these languages with Turkish, and thus, it was biased towards not classifying these samples as Turkish.

The system performed slightly below chance level for Russian. 'Chance level' is defined as the probability of an event occurring, in the context of an unbiased, random choice (Batanero, Henry, & Parzysz, 2005). Thus, the chance level threshold is 10.0%, as the system has a 1 in 10 chance of achieving correct classification. Visibly, out of the misclassifications, the system misclassified most of the Russian samples (32.4%) as English.

5.1 T-test on Duration

One of the experimental parameters that could impact the system's performance is the duration of the data samples. By taking a random sample of the durations of 40 audio files from each language, Student's t-test was carried out. The suspected correlation is shorter samples are more likely to yield incorrect results, due to the limited acoustic information they carry.

Therefore, a two-sample unequal variance one-tailed test was computed. A result of 0.0858 was obtained and testing at 5% significance level, this suggests that there isn't a significant effect on duration and recognition accuracy. However, this does not eliminate the probability that, there exists no correlation at all between the duration of the data and the accuracy of identification.

6 Discussion

6.1 Phenomenon of Grouping

The approach taken involves scrutinising the patterns and trends visible in different language families. As can be seen from Figure 5, the accuracy of the system forms clusters, centred around languages from the same language family. To explain this phenomenon of 'grouping', i.e., languages in the same family having similar results, the unique acoustic and vocalic features, possessed by these languages are explored. Similarly, there are also features that these languages share, which could be the reason behind misclassification.

6.1.1 West Germanic Languages

The system seems to be able to differentiate German, Dutch and English more accurately than the rest of the dataset. Coincidentally, these three languages are all Germanic languages, and they belong to the same language family. What was particularly interesting to see, apart from the high accuracy rate, was, in fact, the misclassifications the system made. When exposed to German samples, the most common mistake the system made was classifying 10.3% of the data as English. Similarly, when exposed to the English samples, the system classified 11.9% as Dutch. There seems to be a pattern in the system either choosing the correct language or one of its counterparts from the same family. To explain this phenomenon for the Germanic languages in question, it is important to consider the overlap in the phonetic inventory of these languages.

Historically, the West Germanic Languages developed in the regions of Elbe, Rhine-Weser and the North Sea. English is a direct descendent of the North Sea Germanic, with the most striking evolutions are the loss of nasals before voiceless fricatives, palatalisation of /k/ before fronter vowels and /j/, and palatalisation of /g/ before front vowels. Though not as 'deviant' as the changes in English, Dutch too, emerged as a distinct branch of West Germanic language, with its unique development of i-umlaut (Buccini, Moulton & Herzog, 2010). From a phonological perspective, despite these historical changes, the phonemic inventory of these languages is nonetheless similar due to descending from the same ancestor. This similarity could be captured by the MFCCs, during the parameterisation process. Speech waves are longitudinal in nature and differentiation of speech events can be achieved via acoustics according to frequency and amplitude components (Ambikairajah et al., 2011).

The misclassification of 10.3% of the German samples as English can, thus, be offered with an explanation. Some of the phonemes of German, English and Dutch could be potentially identified by the models, and similarity between German and English at an acoustic level is picked up. However, more data and a narrower in-depth study would be required to further prove this.

6.1.2 Romance Languages

Setting aside the phenomenon of overlaps in phonemes across the languages, another potential cause of error in recognition is the potential similarity in the acoustic information that the MFCCs aim to represent. Such information, that could be both advantageous and disadvantageous to task of recognition, manifests as pitch and prosodic patterns.

By comparing the frequency of boundary cues of Catalan, Spanish, South European Portuguese and North European Portuguese, Sónia, d'Imperio, Elordieta, Prieto, and Vigário (2007) found that prosodic breaks are usually marked by a High boundary tone in all five languages. The term 'High boundary tone' (HBT) can be understood to mean the surge in pitch that occurs at the start or end of an intonational phrase (Pierrehumbert, 1980). Despite having this common feature, nuclear pitch accent choice, that is, the accent of the head of the prosodic phrase divides these languages into two sets. More specifically, Portuguese and Italian were found to sustain the pitch after L+H* (low tone + accented high tone) configurations and a continuation rise in pitch is observable for L*+H (accented low tone + high tone) configurations (Sónia et al., 2007). However, the same patterns were not found for Spanish and Catalan.

This could be an implausible explanation as to exactly which features the statistical models have picked out from the speech datasets, but it is an explanation, nonetheless, that sheds light on the common acoustic features present in Italian and Portuguese. It is also plausible in explaining why Spanish had a noticeably lower accuracy rate, with 17.5% of its samples classified as Italian. Again, the idea of the system recognising samples as another 'related' language within the same family, hints at its ability in discerning prosody and intonation.

6.2 Individual Languages

Greek, Turkish and Russian are the three remaining languages in the dataset. They are ‘isolates’, in the sense that, they are the only members in the dataset of a particular language family. Thus, these datasets are more independent, as they are not affected by similar languages.

6.2.1 Turkish

Turkish is the only Turkic language in the dataset, and the system was able to classify 28.0% of the Turkish samples correctly. Phonologically, Turkish has sounds that overlap with Russian and Greek such as the velarized alveolar lateral approximant /ɫ̪/ and voiced palatal plosive /ɟ/ (Yavuz & Balcı, 2011). Yet, the recognition rate of Turkish was still quite high, relative to the other ‘language isolates’. This perhaps is due to prosodic features that mark Turkish apart from the other languages. It would be interesting to see how the classification accuracy would change if another Turkic language were to be introduced. Such a set of results would be able to confirm whether languages from the same family aids identification, or whether it is a hindrance.

According to Torres-Carrasquillo et al. (2002), a GMM is thought to approximate ‘the acoustic phonetic distribution of a language’. Manchala, Prasad, and Janaki (2014) further agrees with this idea, and so, it is believed that ‘each Gaussian density captures some broad phonetic class’ (p. 100).

Strikingly, the system did not identify any of the Turkish samples as Dutch and a small proportion of the samples were identified as German. Thus, from a phonological standpoint, this shows that, the system is perhaps capable of identifying small differences that sets Turkish apart from the Germanic languages. This phonological feature could be final devoicing in Turkish (Hulst & Typology of Languages in Europe (Project), 1999) and the phenomenon of aspirated stops and pre-voiced stops, which are absent from German, Dutch and Russian (Petrova et al., 2006).

6.2.2 Greek

Greek was the smallest dataset out of then 10 languages. Although the size of the dataset does not introduce bias, since the statistical models are trained for each language, having a small dataset means that the models are exposed to limited acoustic characteristics of the language. Thus, the presence of more data would be vital in building a more characteristic model.

6.2.3 Russian

The performance of Russian at chance level could be explained by the methodological practices taken. For this specific dataset, the parameter for mixture splitting was defined as ‘2’. From the results of the preliminary study, it is apparent that this parameter directly affects the classification result for some languages (Mi, 2021). From the initial study, German also had a chance level performance rate, at 3.2%, however, this was changed to 66.0% once the mixture splitting parameter was applied. Theoretically, this means that splitting the mixture into two was beneficial in capture the likely acoustic shape of the signal, for some languages, but not others. Therefore, to obtain the optimal system performance, a set of suitable mixture splitting parameters, such as [16, 32, 64, 128], should be tested to find the best parameter.

6.3 Data Imbalance

An important factor to mention is the existing data imbalance in the study. This is a prominent factor in this system as each individual language is not represented by equal amounts of data. Thus, the performance of one language, such as the likes of the West Germanic languages might seem impressive, but this is also because they have been trained on relatively large amounts of data, whereas the languages with lower recognition accuracy have not. Thus, perhaps direct comparison of the system's accuracy should be completed under more context and scope.

Moreover, the 'metadata' of the dataset remains unclear; many factors, such as age and gender of speakers who contributed the data. This means that the possibility of the system being gender/age- biased cannot be completely ruled out. To eliminate this uncertainty, again, emphasises the importance of a balanced dataset and its necessity for such projects.

6.4 Means of Improvement

This section explores ways in which the overall classification error rate of 57.2% can be reduced. In practice, there are various ways to achieve this, the points explored below are by no means exhaustive.

6.4.1 Variability of Dataset

Speech data is very intricate as it is not consistent like other, more tangible, datasets. This means that speech data samples are varied in nature and this variability directly affects the systems' performance and training result.

The core values of the Voxforge dataset relies on it serving as an open-source speech repository. Practically, users are encouraged to contribute to the project, via submission of their speech. This kind of 'data collection' method is extremely inconsistent, as there is no feasible way in assuring that, all the data is recorded under the same conditions, using the same technical setup. Therefore, the dataset is incredibly varied in quality, both in terms of acoustics and ambient background noise. For some languages, some users have contributed a lot of data to the repository. If the conditions in which they completed the recording is noisy, then this would directly impact the trained acoustic model, and introduce an element of bias.

Furthermore, the absence of users' metadata regarding age and gender means that meaningful comparisons cannot be made. For example, it is not possible to eliminate the probability that, the misclassification of the data could be due to the models being more in conformity with younger, female speakers, who prototypically have higher formant values.
Therefore, to eliminate the possibility of 'gender-dependence', metadata is needed for this dataset.

6.4.2 Use of MFCCs

Phonetically, speech consists of sequences of sound units (phonemes) produced by the excitations of the vocal tract. Each of these phonemes is characterised by sets of formant frequencies, which corresponds to the resonances of the vocal tract (Ogden, 2017). Therefore, formants are one of the most prominent acoustic

cues for identification purpose. However, despite this revelation, formant frequencies have seldomly been used for LID.

Consequently, Manchala et al. (2014) proposed a new feature vector, that extracts both the MFCCs of the speech signal and also formant frequencies, through linear prediction analysis. They then proceeded to use this approach in a GMM-based system, and it was found that MFCC features vectors coupled with formant values gave a more comprehensive representation of the acoustic features of the speech signal, which in turn, improved the LID performance (p.104). Therefore, drawing inspiration from Manchala et al. (2014), a similar approach could be taken to improve the performance of this system.

6.4.3 Noise in Dataset

With the goal of increasing the robustness of LID systems in mind, Deshwal, Sangwan and Kumar (2020) investigated the performance of a GMM-based system under conditions of background noise. It was found that the efficiency of such a LID system was strongly affected by noise, as the accuracy of the system with noisy samples decreased by 40%, compared to that of clean data. Furthermore, spectral processing techniques such as Spectral Subtraction (SS) and Minimum Mean Square Error (MMSE) were found to be particularly useful, as they improved the system by 17-20%.

The process of SS involves observing the noise spectrum from periods where the signal, or speech, is absent and only noise is present. Once this estimation of the average noise spectrum is obtained, this is then subtracted from the overall signal, compensating the magnitude of the speech signal (Vaseghi, 1996). MMSE method, on the other hand, is a probabilistic estimator with minimum mean squared errors (meaning it is optimal) (Zhou & Chen, 2015, p.2), and it plays a direct role in the calculation of the vectors during the feature extraction stage. By taking these spectral enhancements and noise suppression methods into consideration, the system can be drastically improved.

6.4.4 Other Speech Information for LID

Hockett (1963) made an empirical generalisation that intonation is universal to all languages, regardless of origins. The term ‘intonation’ can often be founded to appear interchangeably with the term ‘prosody’. For clarity, ‘intonation’ is defined in a broad sense, composing of factors such as word-stress, tone, duration, pitch and intensity.

Often the duration characteristics of phonemes shared across different languages are determined by the different phonetic constraints of the language. All languages utilise pitch to convey surprise, irony or to pose questions. For some languages, tonal pitch variations are crucial in the identification of the language, such is the case for Mandarin Chinese, Thai and Vietnamese. For others, patterns of stress can be used to identify languages that has a word-final stress, such as French, and languages with word-initial pattern, such as Hungarian (Schultz & Kirchhoff, 2006).

From previous research completed on pitch movements in Dutch (Hart & Collier, 1975) and in English (Willems, 1983), it was found that rising and falling pitch movements, i.e., frequency values, are crucial in modelling intonations. This generalisation is also believed to be true for German (Adriaens, 1984). Adriaens (1984) stated that ‘the mean value of the fundamental frequency slowly decreases at a rate which depends on the utterance length’ (p. 37). Adriaens then proceeds to map this as ‘declination lines’. He found that German, like British English, distinguishes between three levels of pitch, whereas Dutch only

possesses two. Moreover, it was discovered that German has a significantly steeper declination of pitch than Dutch (Adriaens, 1984).

Therefore, prosody parameters can be taken into consideration for improvements of a LID system. Although the system captures limited prosodic patterns by means of the MFCCs, it does not explicitly try to capture prosodic information, nor does it make the best use of them. To utilise prosodic information, tone can be translated to acoustic parameters, such as pitch or F0, rhyme can be translated into duration sequence and intensity can be used to parameterise stress (Ambikairajah et al., 2011).

7 Conclusion

To conclude, this Independent Study project explored automatic language identification systems. A 10-languages GMM-HMM system was built using HTK. The overall system accuracy was found to be 42.8%. This was attributed to the system misclassifying samples from the same typological language family and experimental shortcomings in the methodology. Ways of improvement have been provided as well as enhancements strategies in dealing with the varied nature of speech data. As a further study, different numbers of mixtures should be tested and exploring the performance of such a system on varying parameters such as distance from microphone, and application of other feature extraction techniques (Mi, 2021).

8 References

Ambikairajah, E., Li, H., Wang, L., Yin, B., & Sethu, V. (2011). Language identification: A tutorial. *IEEE Circuits and Systems Magazine, 11*(2), 82-108.

Ashihara, K. (2007). Hearing thresholds for pure tones above 16 kHz. *The Journal of the Acoustical Society of America, 122*(3), EL52-EL57.

Association, I. P., & Staff, I. P. A. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*: Cambridge University Press.

Baldwin, T., & Lui, M. (2010). *Language identification: The long and the short of the matter.* Paper presented at the Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In *Exploring probability in school* (pp. 15-37): Springer.

Buccini, A. F., Moulton, W. G., & Herzog, M. I. (2010). West Germanic Languages. In *Britannica*.

Carrasco, C. O. (2019). Gaussian Mixture Models Explained. Retrieved from <https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95>

Cimarusti, D., & Ives, R. (1982). *Development of an automatic identification system of spoken languages: Phase I.* Paper presented at the ICASSP'82. IEEE International Conference on Acoustics, Speech, and Signal Processing.

Das, H. S., & Roy, P. (2019). Chapter 5 - A Deep Dive Into Deep Learning Techniques for Solving Spoken Language Identification Problems. In N. Dey (Ed.), *Intelligent Speech Signal Processing* (pp. 81-100): Academic Press.

Deshwal, D., Sangwan, P., & Kumar, D. (2020). Language Identification Performance Evaluation Using Spectral Processing. *Available at SSRN 3734808*.

Fayek, H. (2016). Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between. Retrieved from <https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>

Foil, J. (1986). *Language identification using noisy speech.* Paper presented at the ICASSP'86. IEEE International Conference on Acoustics, Speech, and Signal Processing.

Gales, M., & Young, S. (2008). The application of hidden Markov models in speech recognition: Now Publishers Inc.

Hockett, C. F. (1963). The problem of universals in language. *Universals of language, 2*, 1-29.

House, A., Neuberg, E., & Wohlford, R. (1975). Preliminaries to the automatic recognition of speech: language identification. *The Journal of the Acoustical Society of America, 57*(S1), S34-S34.

Hulst, H. V. D., & Typology of Languages in Europe (Project). (1999). *Word prosodic systems in the languages of Europe*. Berlin ; New York: Mouton de Gruyter.

Ives, R. (1986). A minimal rule AI expert system for real-time classification of natural spoken languages. *Proc. of 2 nd Artificial Intelligence Advanced Computer Technology*, 337-340.

Jang, J.-S. R. (2011). Audio signal processing and recognition. *Roger Jang's Homepage*.

Kasuriya, S., Sornlertlamvanich, V., Cotsomrong, P., Kanokphara, S., & Thatphithakkul, N. (2003). *Thai speech corpus for Thai speech recognition.* Paper presented at the Proceedings of Oriental COCOSDA.

Kesarkar, M. P., & Rao, P. (2003). Feature extraction for speech recognition. *Electronic Systems, EE. Dept., IIT Bombay*.

King, S. (2017, 07/11/2017). Re: Filter bank vs. filter coefficients. Retrieved from <https://speech.zone/forums/topic/filter-bank-vs-filter-coefficients/>

Kumar, P., Jakhanwal, N., Bhowmick, A., & Chandra, M. (2011). *Gender classification using pitch and formants.* Paper presented at the Proceedings of the 2011 International Conference on Communication, Computing & Security.

Leonard, R. G., & Doddington, G. R. (1974). *Automatic Language Identification*. Retrieved from

Leonard, R. G., & Doddington, G. R. (1978). *Automatic Language Discrimination*. Retrieved from

Li, K., & Edwards, T. (1980). *Statistical models for automatic language identification.* Paper presented at the ICASSP'80. IEEE International Conference on Acoustics, Speech, and Signal Processing.

Manchala, S., Prasad, V. K., & Janaki, V. (2014). GMM based language identification system using robust features. *International journal of speech technology, 17*(2), 99-105.

Mi, M. (2021). The Unsolved Problem of Language Identification: A GMM-based Approach. Paper presented at the T.W.I.S.T. Conference 2021, Leiden University. Retrieved from <https://conference.studieverenigingtwist.nl/2021/student-speakers>

Muthusamy, Y. K. (1993). A segmental approach to automatic language identification. Citeseer,

Muthusamy, Y. K., Barnard, E., & Cole, R. A. (1994). Reviewing automatic language identification. *IEEE Signal Processing Magazine, 11*(4), 33-41.

Nakagawa, S., Ueda, Y., & Seino, T. (1992). *Speaker-independent, text-independent language identification by HMM.* Paper presented at the Second International Conference on Spoken Language Processing.

Ogden, R. (2017). *Introduction to English Phonetics*: Edinburgh University Press.

Petrova, O., Plapp, R., Ringen, C., & Szentgyörgyi, S. (2006). Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* Massachusetts Institute of Technology,

Radha, V. (2012). Speaker independent isolated speech recognition system for Tamil language using HMM. *Procedia Engineering, 30*, 1097-1102.

Rastatter, M. P., McGuire, R. A., Kalinowski, J., & Stuart, A. (1997). Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatrica et Logopaedica, 49*(1), 1-8.

Schultz, T., & Kirchhoff, K. (2006). *Multilingual speech processing*: Elsevier.

Shrawankar, U., & Thakare, V. M. (2013). Techniques for feature extraction in speech recognition system: A comparative study. *arXiv preprint arXiv:1305.1145*.

Singh, P. P., & Rani, P. (2014). An approach to extract feature using mfcc. *IOSR Journal of Engineering, 4*(8), 21-25.

Sónia, F., d'Imperio, M., Elordieta, G., Prieto, P., & Vigário, M. (2007). The phonetics and phonology of intonational phrasing in Romance. *Prosodic and Segmental Issues in (Romance) Phonology. John Benjamins (Current Issues in Linguistic Theory)*, 131-153.

Torres-Carrasquillo, P. A., Singer, E., Kohler, M. A., Greene, R. J., Reynolds, D. A., & Deller Jr, J. R. (2002). *Approaches to language identification using Gaussian mixture models and shifted delta cepstral features.* Paper presented at the Seventh international conference on spoken language processing.

Vaseghi, S. V. (1996). Spectral subtraction. In *Advanced Signal Processing and Digital Noise Reduction* (pp. 242-260): Springer.

Xu, H., Ding, S., & Watanabe, S. (2019). *Improving end-to-end speech recognition with pronunciation-assisted sub-word modeling.* Paper presented at the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Yavuz, H., & Balcı, A. (2011). Turkish phonology and morphology. *Turkish Phonology and Morphology. Eskisehir: Anadolu Universitesi*.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., . . . Povey, D. (2002). The HTK book. *Cambridge university engineering department, 3*(175), 12.

Zhou, B., & Chen, Q. (2015). A tutorial on Minimum Mean Square Error Estimation.

Zissman, M. A. (1993). *Automatic language identification using Gaussian mixture and hidden Markov models.* Paper presented at the 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing.

SECTION B

Squibs and Write-ups

The following contributions have been assessed by academic institutions and awarded a Class I, Class II.1, or international equivalent.

Bringing Your Homeland with You: First Generation Migrants from the Balkans' Decision to Pass on Their Language to Their Descendants (Austria)

Dina Stankovic
University of Vienna
Study hosted on https://kinguistics.wordpress.com/.

Abstract. Migrations have always been a part of history, especially for people from the Balkan Peninsula. Due to economical, political, or military events, people from the Ex-Yugoslavian republics reached out to many western European countries, hoping they will help them find a peaceful home in the future. One of these countries is certainly Austria – only in 1991, over 198.000 people from former Yugoslavian countries fled to Austria (Bonifazi, Mamolo, 2004) , mainly to its capital – Vienna. These people, most of them without ever encountering with German, made various decisions when it came to passing the essential part of their heritage to their descendants – their mother tongue. Some did not encourage their children in keeping the language alive, resulting in them being semilingual. What the main goal of this study is, is to find out what attitude towards passing on their mother tongue to their children is, and what results did it bring. Through a questionnaire, fifty people between the ages of 28 – 68 decided to share their point of view. The study shows that the majority of people find it really important for their children to understand and speak Bosnian/Serbian/Croatian – almost 90%, although only a bit more than 40% speak only Bosnian/Serbian/Croatian to their children. This also results in the fact that only 35% of children speak only their mother tongue with their parents, and over 20% of children speak only German to one another.

Keywords: Balkans; Austria; migration; language; integration

---

1 Introduction

People from former Republics of Yugoslavia have migrated a lot – you can find their minorities in Germany, France, Russia, USA, Argentina, Australia… One particular city holds a large community of Yugoslavian people – Vienna. In 2020, 5.3% of Vienna´s population was from Serbia; 2.1% from Bosnia; and 1.5% from Croatia. Of course, you have to take into the consideration people with Austrian citizenship whose country of origin is still one of the former Yugoslavian Republics. So let's back up a bit — there are several reasons why these people moved — the first wave in 60s and 70s was labour induced. Later, in the 90s, people sought a home while fleeing from a war. In the 2000s, the bad economic situation made people reach for Vienna once again. A lot of them started their families in Vienna and raised their children differently.

My study wanted to know — what are their reasons to pass or not to pass their mother tongue (in future will be called — BCS (stands for Bosnian, Croatian, Serbian) to their children. 55 people were questioned, and their average age is — men 44, women 45. Their mother tongue is BCS, and they all immigrated as adults.

2 Study

This first graph (Figure 1) shows you their approach to German before moving to Austria. As you can see, women paid more attention and their answers are more detailed, while men answered with sufficient or no knowledge at all. Women show that their interest in language and culture before moving was higher, indicating that they were more interested in fitting in and finding their place in a new environment than men. When taking a look at their current knowledge, we can see that while men are mostly good and really good, 20% of women are still not fully comfortable with German. This may be due to the fact that after the migration wave in the 90s, during the Yugoslavian war, men were the ones working, while most women stayed at home with children.

Figure 1: *Participant´s knowledge of German before moving to Austria (men & women).*

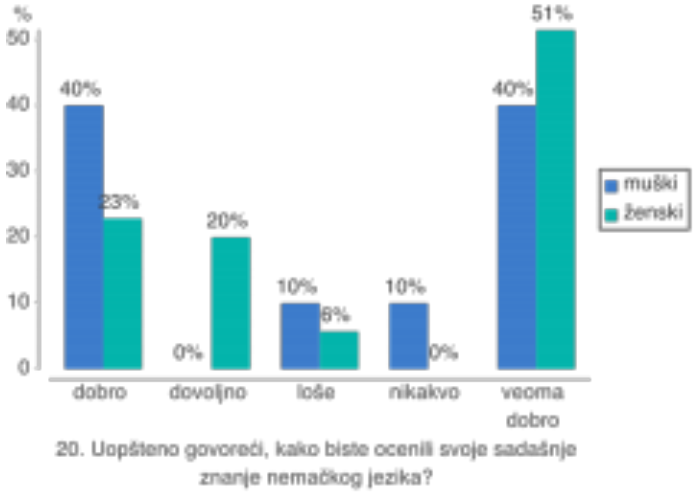


Figure 2: *Participant´s knowledge of German now (men & women).*

When taking a look at the language of a partner and its connection to the parent´s decision to pass on BCS to their children, people whose partners grew up with German really wanted their children to learn BCS,

while the ones who´s partners grew up speaking both German and another Slavic languages found it relatively important. 94% of people whose partners grew up speaking BCS as well succeeded in passing on BCS to their children as well, but some of them didn´t find it as important as one might assume.

Column %	Engleski	Makedonski/Njemacki	nemački	rumunski	srpski/hrvatski/bosanski	NET
relativno važno	0%	100%	0%	0%	3%	4%
važno	0%	0%	0%	50% ▲	3%	4%
veoma važno	100%	0%	100%	50% ▼	94%	92%
NET	100%	100%	100%	100%	100%	100%

Vertical: relatively important, important, very important
Horizontal: English, Macedonian/German, German, Romanian, BCS (language of the partner)

Figure 3: *Language importance.*

Now let´s take a look at the relationship between the language a parent spoke to its child and in what language does the child speak to the parents now. Great majority speaks only BCS to their kids — 44%, but 62% of children speak more German that BCS to their parents, although BCS was mostly the only language spoken by their parents at home. 1% of parents spoke only German, and so did their kids, but according to data, no kids favoured when it wasn't spoken by parents at all.

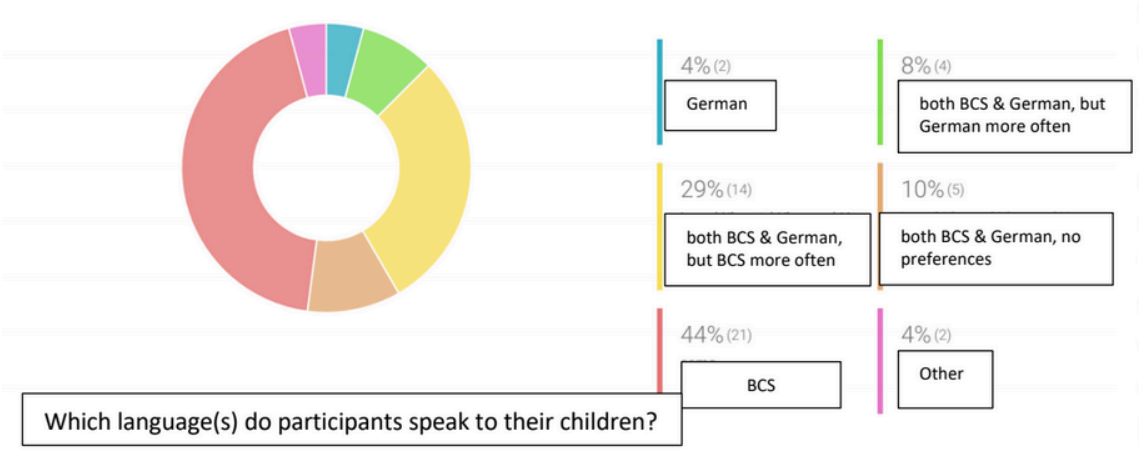


Figure 4: *Languages spoken with children.*

A bit more than 4% of parents were favouring German when talking to their children, and this got me thinking — why? Well, when looking at the data, the main reasons these people left were war induced. We can then look at this from a perspective of a post war trauma — by deciding to favour German and speak less or no BCS with their children, people are actively trying to cut all their ties to the past.

I also took a look at how much are the parents bothered when their children make mistakes in BCS. Of course, the ones who speak German with each other (not with parents) tend to make more mistakes and their parents tend to correct them more often – 40% of children whose parents correct their mistakes very

often speak only German with each other, while 50% of children whose parents corrected them rarely speak mainly BCS with each other. Only 2% of people said that passing on BCS to their children is not important. Again, those are the people who were forced to leave due to the war, so yes.

Column %	No, never	sometimes	regularely	Very often	Very rare	NET
English	0%	8%	0%	20%	0%	5%
BCS & German, but German more often	0%	38%	20%	40%	25%	27%
BCS & German, but BCS more often	33%	0%	0%	0%	50% ▲	11%
Both BCS & German, no preferences	0%	23%	27%	0%	0%	16%
German	33%	23%	13%	40%	13%	20%
BCS	33%	8%	33%	0%	13%	18%
NET	100%	100%	100%	100%	100%	100%

45. Which language(s) do your kids speak with eachother? by 48. Do you ever correct your kids while speaking BCS?
sample size = 44; total sample size = 56; 12 missing; 95% confidence level

Figure 5: *Languages spoken between children.*

When comparing the level of education and the existence of nostalgia, I found that the more people are educated, the less nostalgic they feel – probably because they moved because of a good job opportunity and are aware of the current situation in their home country. People with only high school tend to be more nostalgic, but they also favour German in their everyday life, indicating that they want to fit into the community. When looking at Figure 2, the level of education compared to the language participants spoke to their children, we confirm the theory that high school educated people have a desire to fit in – 100% of them who only spoke German to their kids are only high school educated, and 50% of the ones who prefer German over BCS. When taking a look at college/university educated participants, none of them exclude BCS from the communication, although we can see that a large amount of them choose German over BCS in everyday conversation. A lot of educated people hold a great amount of grieve and anger that, although they spent years getting educated in their home country, they still had to leave and find a better place to live. Some of them also found partners whose mother tongue isn´t BCS, so that could be the reasons for preferring German, although a high amount also speaks only BCS to their children as well.

This next fact was really surprising! Although TV channels in BCS were still watched at home after moving to Austria, the majority of children from those households speak German with each other. This may be due to the fact that no German knowledge was shown by the pa rents, therefore, children used it as a secret language in order to talk between each other. Again, if we take a look at the number of people who would love to, but don´t watch TV in BCS, their children also prefer German, which indicates that the lack of media support can be sensed in children´s language preference.

Column %	yes	no	NET
College/University	38% ▼	73% ▲	46%
Elementary school	10%	0%	8%
High school	38%	18%	34%
Higher education	13%	9%	12%
NET	100%	100%	100%

What level of education do you have? By Do you feel nostalgic about your homecountry?

Column %	BCS & German, but German more	BCS & German, but BCS more	Both BCS& German, no preference	German	BCS	NET
College/University	50%	43%	20%	0%	50%	43%
Elementary school	0%	7%	0%	0%	15%	9%
High school	50%	43%	40%	100%	20% ▼	36%
Higher education	0%	7%	40%	0%	15%	13%
NET	100%	100%	100%	100%	100%	100%

What level of eeducation do you have? By Which language(s) do you speak to your children?

Figure 6: *Levels of education compared to language(s) spoken.*

Column %	yes	no	I would love to, but I don´t	NET
English	3%	20%	0%	5%
BCS & German, but German more	30%	0%	50%	27%
BCS & German, but BCS more	14%	0%	0%	11%
Both BCS & Germna, no preference	16%	20%	0%	16%
German	22%	0%	50%	20%
BCS	16%	40%	0%	18%
NET	100%	100%	100%	100%

45. Which language(s) do your children speak with eachother? by
58. Do you ever watch TV in BCS sample size = 44; total sample size = 56; 12 missing; 95% confidence level

Figure 7: *Language preference and TV programmes.*

This last table shows how religion is connected to the children's interest in BCS language and culture. It affects it for certain, but it is not a key. You can see that no children whose parents are going to church or mosque are not entirely interested in the culture, while the ones that don't do rarely, have children who are not very fond of their heritage. It is also to be seen that 44% of children whose parents never go to church speak only German with each other. Then again, when looking at the column no. 8, which stands for really interested, 67% of these children´s parents don´t go to church. So, as I said, it can make a change but it doesn´t mean it will entirely affect the child´s interest.

Column %	0	1	3	4	5	6	7	8	9	10		NET
yes	0%	0%	0%	0%	20%	67%	75% ▲	0%	25%	42%		30%
never	50%	0%	0%	50%	40%	0%	25%	67%	50%	25%		35%
sometimes	50%	100%	100%	50%	40%	33%	0%	33%	25%	33%		35%
NET	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		100%

18. Do. you ever go to church/mosque in Austria? by 73. How high is your children´s interest in BCS culture and language?
sample size = 40; total sample size = 56; 16 missing; 95% confidence level

Figure 8: *Percentage of children who attend church and their interest in it.*

3 Conclusion

My conclusion is that, although people moved due to so many different reasons and most of them weren´t pretty, the majority of them still brought a little piece of their home with them and are willing to pass it on. The trauma induced by the war can be a factor that makes people want to forget their language, but these people are in a minority. When looking at data about their children's BCS knowledge, I am not surprised to find that, even they are interested in the culture and language, some of them still identify themselves with German more than with BCS, especially when it comes to talking to their siblings.

4 References

Brizić, K. (2007). Das geheime Leben der Sprachen. Vol. 465. Waxmann Verlag.

Bonifazi, C. & Mamolo, M. (2004). Past and current trends of Balkan migrations. *Espace, Populations, Societies,* 3, 519-531.

Fassmann, H., & Reeger, U. (2008). Austria: From guest worker migration to a country of immigration. *IDEA Working Papers*, 1, 1-39.

Memić, N. (2012). Serbische und ex-jugoslawische Migranten in Österreich und ihr politisches Potenzial. *Der Donauraum* 52(1), 105-112.

Wojnesitz, A. (2009). Sprachbewusstsein und Einstellungen zur Mehrsprachigkeit an Wiener AHS im Kontext von Migration.

The *heheo*: is That Even a Thing? A Sociolinguistic Approach from Loja (Andalusia)

Eva Aguilera Parejo
University of Grenada

Abstract. This sociolinguistic study focuses on an underanalysed phonetic feature of the Andalusian dialect (Spanish) which can be found in the town of Loja (Andalusia): the 'heheo'. The 'heheo' is the sporadic and lexically conditioned substitution of the /s/ and/or /θ/ sounds by /h/ at the beginning of a word or syllable. In spite of being widespread all-over Latin America and Southern Spain, this linguistic feature is characterised by its lack of linguistic prestige. This study can be of an interest to a diverse population: ranging from academics due to the terminological chaos around this phenomenon, to students of linguistics as this phenomenon actually constitutes a research gap. And of course, actual users of heheo, because of their low level of linguistic awareness concerning this stigmatized phenomenon. During the development of this research, a detailed analysis conducted on the sociolinguistic perception about this phonetic phenomenon. For this purpose, an online survey was designed and launched to a wide range of Spanish speakers from the town of Loja. Thanks to it, the three main research questions were answered: (1) What are the language attitudes towards the heheo in the town of Loja? (2) What is the level of linguistic awareness of the user of heheo? (3) Does a linguistic identity exists among the users of heheo? The only goal of this study is to provide new information about this research gap in order to contribute to a deeper understanding of the heterogeneity of the Andalusian dialect.

Keywords: 'heheo'; sociolinguistics; dialectology; Andalusian; Spanish

---

1 Introduction

This sociolinguistic study focuses on an underanalysed phonetic feature of the Andalusian dialect (of Spanish) which can be found in the town of Loja: the 'heheo'. And more specifically, it is focused on the linguistic perception and language attitudes towards the heheo in this town. This study follows the synchronic sociolinguistics' methodology by launching an online survey.

In relation to the 'heheo's' stay of play, it exists a lack of bibliographical references, of quantitative data and even of terminological consistency. This fact only demonstrates what a huge research gap the 'heheo' constitutes in itself, resulting in the necessity of collecting more real data about this phenomenon. Therefore, an online survey was launched to a wide range of local Spanish speakers from the town of Loja (Andalusia, Spain).

After all, the present study's only goal is to provide new information about this research gap in order to contribute to a deeper understanding of the heterogeneity of the Andalusian dialect.

So, at the beginning of this research it was hypothesised that: In spite of the 'heheo' being a stigmatized feature, the language attitude of the user of 'heheo' is not so negative.
In addition, three research questions were posed:

1. What are the language attitudes towards the ‘heheo’ in the town of Loja?

In other words, what people think about this linguistic feature and how these relate to attitudes about this use of the Spanish language.

2. What is the level of linguistic awareness of the user of ‘heheo’?

This linguistic or language awareness can be defined as explicit knowledge about language, and a conscious perception and sensitivity in language use.

3. Does a linguistic identity exist among the users of ‘heheo’*?*

Linguistic identity refers to a person’s identification as a member of one or more groups of speakers sharing common linguistic features. The linguistic identity is often an important part of one's personal identity.

2 Definition

The ‘heheo’ is the sporadic and lexically conditioned substitution of the /s/ sound and/or /θ/ sound by /h/ at the beginning of a word or syllable. For instance, a user of ‘heheo’ would pronounce these three words the same way: ‘casa’, ‘caza’*,* and ‘caja’ as /káha/.

In spite of being widespread all-over Latin America and Andalusia, this linguistic feature is characterised by its lack of linguistic prestige. In fact, the ‘heheo’ suffers from such a strong stigmatization that it is hardly recognised among a wide range of the population.

It is a phonetic feature of the Andalusian dialect that usually can be found in familiar communicative situations among users of ‘ceceo’*,* which is another phonetic feature of the Andalusian dialect, meaning the pronunciation of the /s/ sound as /θ/. It is usually heard in informal contexts since in formal registers users of ‘heheo’ tend not to use it due to the negative connotations linked to it.

3 Location

The ‘heheo’ can be found both in the Southern Spanish region of Andalusia and also in Latin America. The ‘heheo’ in America has been a bit more analysed than in Andalusia. A few studies properly describe the ‘heheo’ in Mexico, El Salvador, Honduras and Colombia. While in Andalusia, studies suggest that it can be heard in the provinces of Granada, Malaga, Cordoba, Seville, Huelva and Cadiz.

It makes perfect sense that the heheo, being a feature of the Andalusian dialect, is also found in American Spanish. It may be rooted in the great number of migrants from Andalusia to America that settled in the so-called New World and taught their dialect among the new populations.

4 Methods

As it has been already mentioned at the beginning of this paper, this study was conducted by following the synchronic sociolinguistics' methodology. The lack of information about the ‘heheo’ made crucial to collect more data so an online survey was launched in order to do so. 150 participants from the town of Loja

completed the survey. The variables designed were both quantitative and qualitative, so the type of analysis conducted was a mixed one.

As at the beginning of every sociolinguist study, the social variables were established. Being the age, sex and academic background the most important ones.

Then, the linguistic variables were presented in the shape of the online survey questions:

(1) Do you consider the way people talk in Loja to be peculiar? If so, in what sense?
(2) Do you think that other (Spanish) speakers consider the way people talk in Loja to be inferior? If so, how do you notice it?
(3) Do you think that you are speaking badly by talking the way people do in Loja? Could you explain your reasons?
(4) Do you find it difficult to be understood when talking to Spanish speakers who are not from the town of Loja? Yes/No
(5) Have you ever found yourself in a situation where you had to change your accent in order to be understood? If so, in which situations?
(6) Do you think that the speech found in Loja has negative connotations? If so, could you mention some of them? Do you feel ashamed of your local accent?
(7) Do you think that the speech found in Loja has negative connotations? If so, could you mention some of them? Do you feel proud of your local accent?
(8) Do you have any interesting comment about the way people talk in Loja that you may want to share?

5 Discussion

The most interesting aspect of this piece of research is how to dismantle the main prejudice concerning the ‘heheo’.

Quoted from one of few authors on whom I could base my sociolinguistic study, Juan Pablo Rodríguez Prieto: “[The ‘heheo’] is often used by speakers with a low educational level or from low socio-economic backgrounds, in informal or neglected speech”.

Is the ‘heheo’ only used by speakers from low academic and socio-economic backgrounds?

According to the survey I launched to the local population of the town of Loja, the average respondent's profile is a young woman with university studies. This fact demonstrates that it exists no direct link between the *heheo* and the user's academic background. Therefore, the main prejudice of the *heheo* is overturned.

The *heheo* is a transversal phenomenon that is not only used by "uneducated” speakers. To say otherwise only denotes a stale classism. And even if this were the case, there is still no reason to disregard the way people talk. This stigmatization is only provoked by social classism, and it is directly linked to the status of the Andalusian dialect, which is still stigmatized nowadays.

6 Conclusions

Thanks to the online survey, the three main research questions were answered:

(1) What are the language attitudes towards the heheo in the town of Loja?

They can be easily distinguished depending on whether the speaker is a user of ‘heheo’ or not, finding that this phenomenon is notably more stigmatized among speakers that do not use this phonetic feature.

(2) What is the level of linguistic awareness of the user of ‘heheo’?

It can be stated that it exists up to a certain level since the users of ‘heheo’ are actually aware of the use of this phonetic feature, although a majority of them do not know the specific term which describes it.

(3) Does a linguistic identity exist among the users of ‘heheo’?

To some extent, it does; they strongly identify their accent with their cultural values and roots. To sum up, it is the great variety found in a very same language what makes it worth learning about it. I have learnt during these years that your accent is also an important part of who you are and there is nothing more beautiful and moving than being proud of your local community and roots.

Social Manipulations in Advertising: Impact on Consumers' Perception of a Product

Valeria Penzina, Polina Reshetnikova, Dilnoza Umarova
National Research University Higher School of Economics

Abstract. Manipulations in advertising are rarely regarded as a sociolinguistics factor of influence. Having found the disparate classification of impact on people's desire to buy products, we aim to provide a full description list of manipulation features in the field of advertisement. It presents deeper comprehension of marketing discourse of psychological impact and protects from spontaneous decisions those consumers who pretend to obtain each thing that was promoted. After regarding the theoretical specification of manipulations in printed and video advertising and its general usage by involving the buyer in the trading process, we identify the most common classification of ways to influence the consumer including peculiarities of verbal and written utterance. These designations became the theoretical basis for analysis of advertisements while questioner was devoted to find out social awareness of Russians about marketing impact and its ways of resistance. Hence, we propose the description of the most used varieties of manipulations in advertising both in the domestic and foreign markets with an extensive analysis of examples of such promotions and with the assumption that it is necessary to develop ways to counter such manipulations.

Keywords: manipulation; advertisement; consumers; brand; market; impact

---

1 Introduction

There is a growing mainstream among different marketers and organizations to use a multitude of manipulations in their advertisements. A large number of studies reveal that various manipulations impact on consumer's perception of the product significantly. Complementing these investigations of manipulation's influence is research that examines the types of manipulations and their process. It is paramount to observe that the vast majority of manipulative advertisements are tough not only to recognise but also to prove their existence as their nature and content is controversial.

Many studies have been conducted on the topic of manipulation in advertising, but none of them gave a complete overview of the reasons for the use of manipulation and did not also give a complete qualification. Some of them have demonstrated that whenever these advertising efforts are focused on getting the consumers to do what the advertiser wants through subversive manners that lack the truth. It can be said that manipulation exists there. If the manipulator takes up the case, then a variety of falsification of facts about the product takes place in many cases (Danciu, 2014).

This research supposes a double objective, which includes examining how marketers and multiple organizations get consumers' heads to buy their production, despite their quality, price, effectiveness, and usefulness. In this case, the analysis of three types of advertisement will demonstrate manipulative tricks where its marketers are in vogue in recent time. The second intention is to examine Russian people's awareness of how they can be manipulated through viewing commercial of various types. This investigation

brings to a certain conclusion that could be interpreted as positive or negative depending on the obtained results.

The whole study can become an important implication for several stakeholder groups. First, our findings may help to recognise manipulation tricks in advertising and will teach not to yield to its influence. Hence, our results may be of use to individuals by enabling them to comprehend the ways and the processes of manipulation. Finally, such knowledge may help not only younger but also older generations — individuals who are mostly and easily susceptible to leverage — always to be on the alert.

2 Manipulation's Usage in Advertising

Nepriahin (2018) in his book identified manipulation as a people's desire to control others. This kind of obsession becomes a result of various psycho-diseases when a person manipulates for life, lives for manipulation. However, this definition cannot be used for the field of advertising as a whole, although it is crucial to notice here that the main reason for manipulation in advertising comes from the purpose to manage people's minds and involve them in a certain domain of the market. Low involvement is the main reason for the PR and marketing department to include psychological aspects for the purpose to engage the audience. Laczniak and Mueheling (2013) distinguished two stages of manipulation production, naming them as peripheral and central manipulation. Where the first term describes the quality of colour, background and primary characteristics provided on non-brand processing, whereas the second statement is devoted to the stage when a brand chooses more specific characteristics for the advertising and complement the peripheral manipulation. Nevertheless, it often happens that excessive use of manipulation leads to advertising that is not directly related to the product. It negates most types of manipulation, causing rejection of the product from customers. This phenomenon is explained by the occurrence of logical errors, which in turn are a clear example of the advertiser's incompetence.

Thus, the brand must carefully choose the advertising company. Moreover, it was noted that the same products of competing brands are not the subject of advertising production by the same company (Kent & Allen, 1994). Besides, different levels of brand fame are also one of the reasons for choosing an advertiser carefully. This is since low although the involvement in such situations, the less the brand and its product are known in the market, the more difficult it will be to choose the types of manipulation since the target audience will be more biased than if the brand has its influence in the market. In this way, the right usage of manipulation must include the awareness of brand manager that in many cases all that was shown in advertising remains on the periphery of consciousness and is not transmitted to the main memory of a person. Hence, four levels of ad perception were highlighted for a clear picture of how to attract the audience's attention (Greenwald & Leavitt, 1984). There is preattention when a person watches peripheral manipulation of colour, music, etc. Focal attention is considered that a person pays more attention to a product then comes comprehension of a brand and elaboration which becomes the main reason of desire to get a product. Thus, we can conclude that to successfully use manipulation in advertising it is necessary to know its classification and application in a particular case.

3 Classification of Manipulations in Advertising

Referring to a well-known fact, there are types of advertising that have different conceptual components. However, any kind of manipulation can be applied in one way or another to any of them. Thus, the issue

of classification becomes one of the most important and complex. This is since few scientists and analysts have tried to create a complete list of manipulations that advertisers use in their work.

Melnichuk and Klimova (2019) identified 6 types of manipulations that are more often used in brand newsletters. The first is 'Substitution of the beneficiary' which implies the omission of facts about the quality of a product or service and includes such phrases as 'benefit', 'advice', 'help', 'easy', 'two for the price of three', as well as semantic negotiation 'not only...but also'. 'Insistent invitation' is used to avoid contacting a specific person. Thus, the brand uses the imperative to appeal to everyone and no one at the same time, encouraging the audience to take action. 'Sociocultural significance' is used to emphasise the importance of the buyer, while the manipulation strategy 'global brand' applies to the product's position in the global market. It is usually implied such expressions as 'geographic position', 'worldwide', 'global market'.

The other type that is closely related to the previous two terms is called the 'growth indicator' and informs customers about what the company has achieved or is going to achieve by offering its products. Besides, this manipulation technique can be supplemented by 'witness testimonies' which are citations of famous people or workers of a company.

'Modelling a community of like-minded people' is the final term of given classification which uses generalizations of the audience and the certain brand throughout such words as 'we', 'ours', and foreign borrowings for showing people that they are a part of a big community.

Speaking about video advertising, Victor Danciu offered several effective types of manipulations (2014). It is impossible to face them separately but necessary to identify them. The first is deceitful advertising which is used to confuse the audience by shuffling the facts and providing a more profitable slogan formulation for the brand. This type of manipulation is often used together with a bad argument that has nothing to do with the brand or its products. Moreover, addressing the emotional background of customers is the third component of successful manipulation. Moreover, this can be either a well-staged speech of the actor or a peripheral manipulation, which, as it was said earlier, includes the background, music, and colour.

The manipulation of the size and the price of the product is another good technique (Li et al., 2019) that is always provided by advertisers who want to increase customers' involvement especially if the brand introduces a new product to the market. First, they advertise the product using all the previously listed classifications, and then the brand announces discounts, often increasing the original cost and leaving the old price as a favourable offer.

Therefore, there are many different manipulations, and we have reviewed the more widely used varieties for a clearer understanding of how the advertising manipulation system works.

4 Advertisement Analysis

Three types of advertising were taken for a complete analysis of manipulations. It is worth noting that these varieties are the most common in the market of increasing customers' engagement. Furthermore, it is possible to observe their different form, but the principle of their creation remains similar to those that will be analysed below. It is paramount to observe that the analysis of advertisements can give individuals an opportunity not only to comprehend the manipulations better but also to escape its impact on product perception.

The first type of advertising can be identified as an e-mail newsletter for registered clients of shops, magazines or banks. For the analysis for this study were taken written advertisement of Russian VTB bank.

In spite of the fact that the text of the advertisement was written in Russian, the sense in translated form will be transferred in a quite clear way.

This advertisement aimed to attract new investors in bonds of VTB bank. Importantly, the salutation was personalised. In many other cases it could be possible to regard the impersonal treatment as an introduction to something global or 'insistent invitation*'.* Nevertheless, the last type of manipulation is present in the selected text and will be reviewed later. Now, it is important to highlight another type of control consumers' engagement which is 'modelling a community of like-minded people'. The expressions '**we** will tell', '**our** analytics', '**our** subscribers' show how a potential novice investor can become a significant part of the VTB community.

Thus, coming back to 'insistent invitation' we can observe such imperatives as 'download our app', and 'keep abreast of developments' which manipulates the desire to be more informed and getting more opportunities than it was before. The VTB bank client is driven by the desire to become more secure begins to invest in the bank's bonds in which the following manipulation of 'price and size occurs'. In this case, it is mostly the size of provided percentage. Bank promoters attached structured data where it is easy to understand the benefits of the offer and memorised it:

- Coupon income – 4,7 % annually (which is a good percentage for Russian's investors)
- Frequency of payments – two times per year
- Period – a year
- Nominal price of one bond – 1000 robles.

The second type of advertisement that was chosen for analysis is a modern realm of phone games promoting. It supposes people's engagement in a process of video advertising itself. The advertiser does not leave any possibility to stop watching the video until the potential client completes a certain set of actions.

The game 'Project Makeover' can become a good example of such a subtype of video advertising. This game video promotion was chosen for a reason that people who do not understand the types of manipulation can reveal the influence on their desires. Developers manipulate them overly obvious that is a positive aspect of this advertising in the context of our research and provides an opportunity to show how the advertisers use manipulations.

Firstly, it is necessary to note the peripheral manipulation with animation of characters and music. A person can observe avatars of the main characters. The same can be said about the background. The advertiser provides the main storyline of the game where the plot revolves around a girl. Promotions allow choosing her future actions thereby manipulating the viewer and forcing him to play. It is worth noting that this type of advertising does not have the usual slogans. A person can only observe the names of the actions proposed by the developers and be sure that they are given a choice. Nevertheless, this is a prime example of a new subtype of manipulation, which we called 'imaginary free choice'. Since no matter what set of actions a person chooses, the result will always be the same. The user will be taken to the game page in the App Store or Google Play.

Finally, the third type of advertising is the most common and popular among a certain number of companies. It is video promotions on YouTube and Instagram platform or TV channels. It is worth noting that in such videos, it is necessary to pay a lot of attention not only to such types of manipulation as price and size, which may not be present in advertising but also to peripheral processing. Thus, we have chosen

famous parfum advertising by Lancôme. Namely, it is Lancôme Idol of 2019 with Hollywood actress Zendaya.

It is necessary to observe that the music that has been used is contemporary (Sia's 'Unstoppable'). In this connection, the singer and the song that have been chosen are of great importance and popularity, especially among the younger generation. The words of the song have a huge positive socio-cultural significance. Nevertheless, they become a good way for manipulating customers' minds against the backdrop of this promotion. The words 'unstoppable' and 'invincible' push people to some specific actions. It creates a realm of perception in which they believe that that they listen to themselves, and not someone's imposed opinion. A harmonious colour scheme selected for the background action strengthens the words presented in the ad. We should not exclude the fact that the main character, presented in a beautiful light dress on a white horse, embodies freedom, strength, and a sense of self sufficiency. This central element most influences the perception of the product. In this case, the potential customer will pay attention not to the real quality of the product (in this case, it is the durability and taste of the smell) but to the socio-cultural element embedded in the advertising itself. Hence, we observe two types of manipulation: 'sociocultural significance and deceitful advertising'.

4 Survey Results

The survey was made to see the level of awareness of manipulations used to promote and sell a product. 107 Russian citizens aged 18-24 were asked 10 questions about their experience with manipulation in the advertisement.

Some results were expected (respondents have a superficial knowledge of the topic and are not aware of ways to resist manipulation) and some were unexpected (respondents are aware of being manipulated and can point out some techniques if a list of them was provided).

It is worth noticing that the vast majority of respondents are women (Figure 1), which confirmed a greater interest in buying a particular product through watching advertisement. This happened because advertising captured the attention of the buyer, suppressing his critical thinking, and the purchase was made only through advertising (Figure 2).

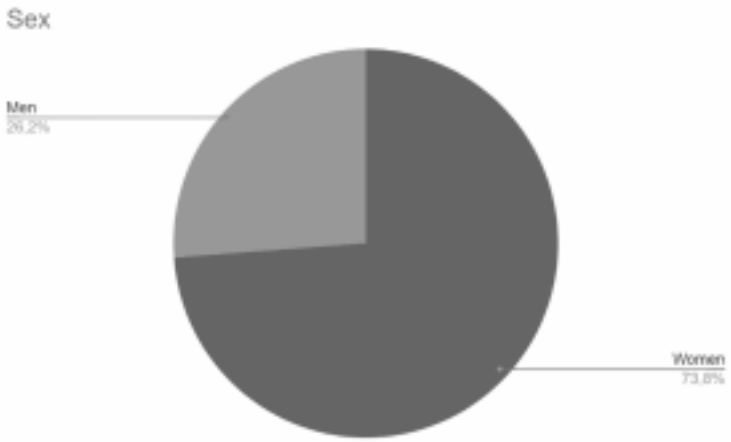


Figure 1: *The ratio of female and male respondents.*

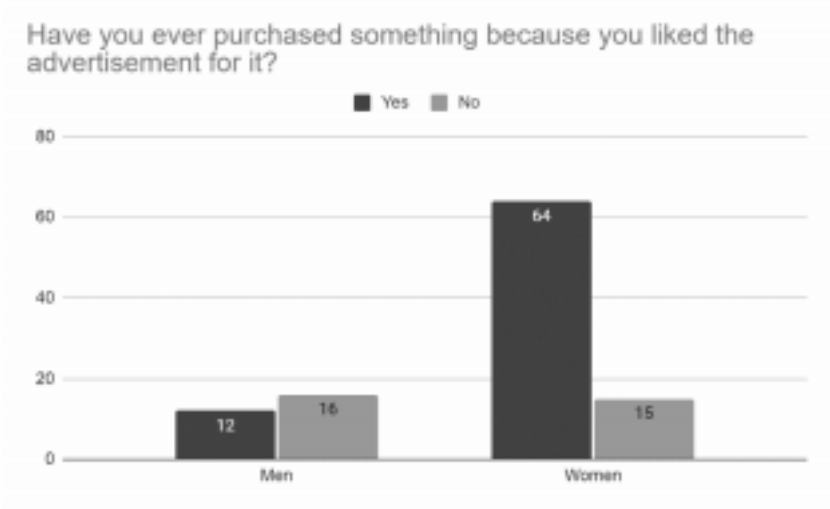


Figure 2: *The ratio of men and women realising attractiveness of an advertisement for future purchase.*

Thus, when asked whether the respondents were aware that they were being manipulated through advertising, the majority answered positively. They mentioned emotional and psychological pressure and sales as the reasons that made them buy the product. However, almost half of respondents did not know what manipulation techniques were used, although it was a manipulation.

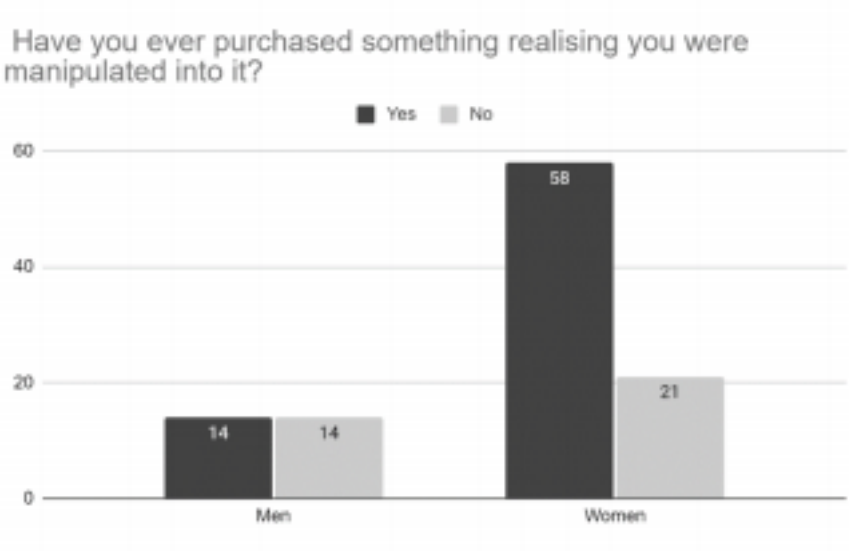


Figure 3: *The awareness of respondents being manipulated by advertisement.*

It is seen from the answers that respondents can spot manipulation if paying attention, but they lack structural knowledge of manipulation techniques and ways to resist it (Figure 4). Those few who answered positively could not give an exact answer, what can be opposed against manipulation. Simple ignoring was suggested most often.

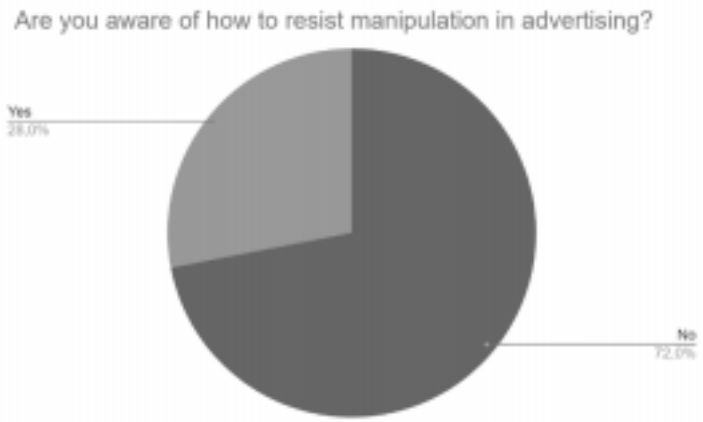


Figure 4: *The awareness of respondents of how to resist manipulations in advertisement.*

During the survey, respondents were asked to view the selected ads and tell us what types of manipulation were used in them. Figure 5 shows the variable responses to the advertisement from Lâncome. Global brand in advertising is a major point that people could indicate. Nevertheless, such a criterion is erroneous, due to ignorance about the details of the division into types of manipulations in advertising. This cannot be said of the following frequently chosen type. Sociocultural significance represents 20% of the total number of responses collected while the other right identified factor comprises less than 10%.

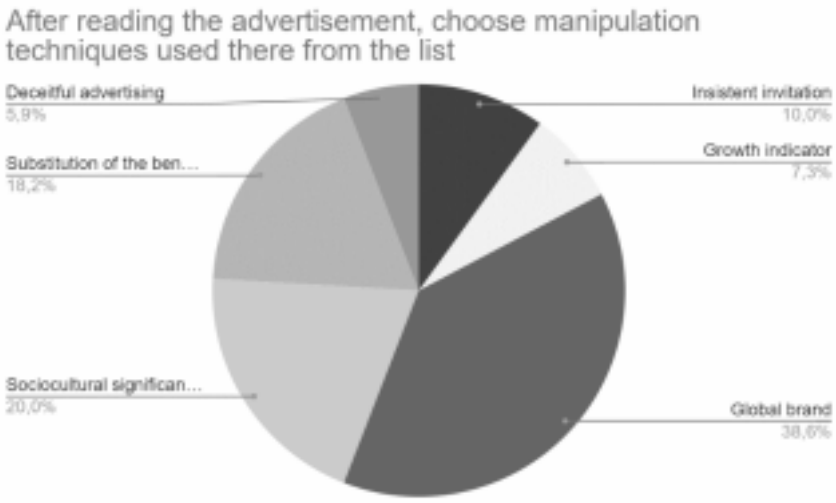


Figure 5: *Chosen types of manipulation in advertisement of Lâncome.*

It was necessary to confirm initial assumptions about respondents' awareness of manipulation techniques in any types of promotion. Therefore, we provide the extra question, where asked to make a brief analysis of the VTB bank e-mail advertisement (Figure 6). The results again confirmed the superficiality of the respondents ' knowledge. Percentage of 'growth indicator' and 'price and size' appears almost the same

although one of criterion shouldn't have been marked intensely. This tiny difference between the two types of manipulations forces participants to pay more attention to numbers and superficially read the other points that were demonstrated in the e-mail. Thus, the point of 'modelling a community of like-minded people' was dismissed in many cases. People have not seen any references for choosing this variant although they were apparent.

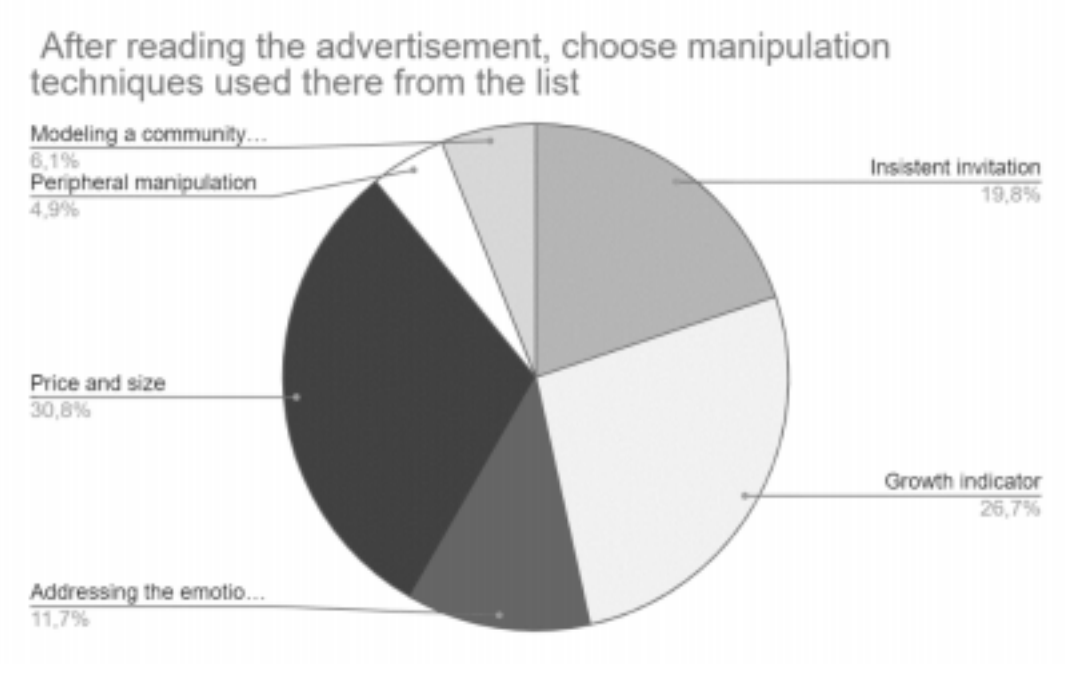


Figure 6: *Chosen types of manipulations by VTB in advertisement.*

Hence, despite the fact that the results were expected they have brought more structural information that consequently allowed to come to certain conclusions.

5 Conclusion

Social manipulation in advertising has an unambiguous intention to control people's desires and needs. However, it is impossible to name one approach that would be known to everyone. Manipulation is multi-faceted. This has a double meaning. Hidden or peripheral approaches can weaken mindfulness, while overlaid features become a call to action. Hiding the facts about the product or giving the opportunity to feel important distorts the real characteristics of the product. They allow seeing what the consumer need. This is the main reason why it was necessary to identify the main types of manipulations.

The analysis demonstrates that it is necessary to pay great attention to the wording, especially where the pronoun We is present. It makes clear to the buyer that they are not alone and is universal in the use of utterances in all types of advertising. In addition, where the pronoun We is present, it is possible to find the imperative mood, which encourages to be engaged. Nevertheless, when identifying manipulations in advertising, it is essential to pay attention not only to the linguistic component but also to the music and

background. In this case, socio-cultural significance can be laid down in the basis, for example, contained in the words of a song.

These and many other techniques for intentionally involving consumers in the shopping process led to the need for simultaneous viewing of advertising. The survey showed unambiguous results. Russian young people know about manipulations superficially, which forces them to realise that the perception when buying a product was violated through external influence. Moreover, it is not worth saying that the respondents would be able to identify specific phrases from the context of advertising if they were not provided with a list.

These facts lead to the need to study the types of social manipulation in advertising that were presented in this study by Russian citizens. Hence, it provides a perspective for further research, in particular, for the development of counteraction methods that could reduce the level of exposure while viewing ads.

6 References

Grenawald, A.G. & Leavitt, C. (1984). Audience involvement in advertising: Four levels. *Journal of Consumer Research*, 11(1), 581-592.

Melnichuk, M. V., & Klimova, I. I. (2019). Manipulative means of language in the advertising discourse of financial companies. Humanities. *Bulletin of the Financial University*, 9(6).

Mel'nichuk, M. V., & Klimova, I. I. (2019). Manipulyativnye sredstva yazyka v reklamnom diskurse finansovyh kompanij. Gumanitarnye nauki. *Vestnik finansovogo universiteta*, 9(6).

Li, K., Chen, Y., & Zhang, L. (2020). Exploring the influence of online reviews and motivating factors on sales: A meta-analytic study and the moderating role of product category. *Journal of Retailing and Consumer Services*, *55*, 102107.

Wang, A. (2006). Advertising engagement: A driver of message involvement on message effects. *Journal of Advertising Research*, *46*(4), 355-368.

Laczniak, R. N., Muehling, D. D., & Grossbart, S. (1989). Manipulating message involvement in advertising research. *Journal of advertising*, *18*(2), 28-38.

Danciu, V. (2014). Manipulative marketing: persuasion and manipulation of the consumer through advertising. *Theoretical and Applied Economics*, *21*(2), 591.

Nepryakhin, N. (2018). I manipulate you: Methods to counteract hidden influence. Alpina Publisher. // Nepryahin, N. (2018). YA manipuliruyu toboj: Metody protivodejstviya skrytomu vliyaniyu. Al'pina Pablisher.

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, *64*, 131-148.

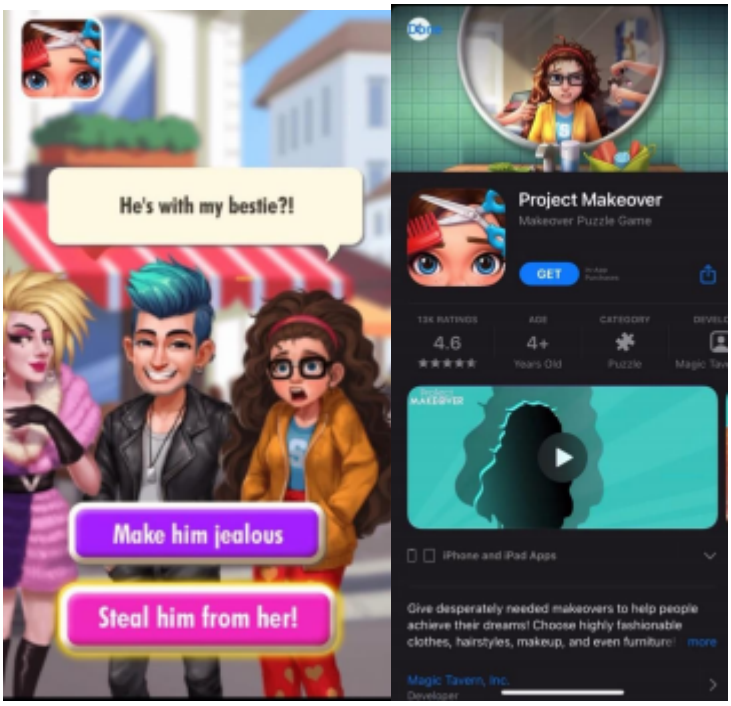
Kent, R. J., & Allen, C. T. (1994). Competitive interference effects in consumer memory for advertising: The role of brand familiarity. *Journal of marketing*, *58*(3), 97-105.

Crain, M., & Nadler, A. (2019). Political Manipulation and Internet Advertising Infrastructure. *Journal of Information Policy*, *9*, 370-410…

Lancôme Idol (2019) < https://www.youtube.com/watch?v=qHnb9A8pVVI >

7 Appendices

7.1 Appendix One: 'Project Makeover' – Screens of Video Advertising



7.2 Appendix Two: The VTB bank e-mail (translation)

Bonds are a great tool for a novice investor. In this letter, we will tell you how you can make money by buying VTB bonds.
Projected income and low risk
1. Twice during the term, you receive a coupon (interest) income.
2. At the end of the term, you will be refunded the entire amount invested.
3. You can return the money at any time without losing the coupon income.
The return occurs through the sale of the bond at the market price, which may differ from the purchase price.
What VTB bonds are available now
• Coupon income – 4,7 % annually
• Frequency of payments – two times per year
• Period – a year
• Nominal price of one bond – 1000 robles

7.3 Appendix Three: Buy Bonds: Install the App

"VTB My Investments" to:
• get a selection of the best securities – our analysts regularly manually select the best securities and give their recommendations.
• invest with the help of a robot adviser-it will select a personal portfolio for free and will send you tips on how to manage it.

• change the currency without leaving your home. Exchange at the exchange rate-from 1 USD; • study analytics from VTB Capital Investment Bank, use a virtual assistant and have access to 10,000 securities on 33 exchanges

CONTENTS