Proceedings of ULAB XII

University of Edinburgh
9 - 11 April 2022
edited by **Caitlin Wilson**

the Undergraduate Linguistics Association of Britain

ISSN 2752-7417

Undergraduate Linguistics Association of Britain

ULAB XII Proceedings
University of Edinburgh
9-11 April 2022

Editor: Caitlin Wilson
Copyeditors: Beatrix Livesey-Stephens, Andrew Tobin, Hui Zhu
Cover design: Lydia Wiernik

ISSN 2752-7417

Contents

SECTION B: Squibs and Write-ups

ULAB National Committee 2021-22

Cliodhna Hughes, *University of Edinburgh*	Chair
Hafren Vaughan, *University of Aberdeen*	Vice Chair
Louis Van Steene, *University of Cambridge*	Secretary
Beatrix Livesey-Stephens, *University of Aberdeen*	Treasurer
T.R. Williamson, *University of Cambridge*	Journal Editor
Stephanie Jat, *University of Cambridge*	Magazine Editor-in-Chief
Roma Dhasmana, *University of Aberdeen* **Beatrix Livesey-Stephens**, *University of Aberdeen*	Social Media Coordinators
Lydia Wiernik, *University of Edinburgh*	Archivist
Louis Van Steene, *University of Cambridge*	Webmaster
Eloise Parr, *University of Birmingham*	Events and Opportunities Coordinator
Caitlin Wilson, *University of Edinburgh*	Institutional Representative Coordinator
Michael Goessler, *University of Edinburgh* **James Morley**, *University of Cambridge* **Roma Dhasmana**, *University of Aberdeen* **Grace Cotton**, *University of Ulster* **Eloise Parr**, *University of Birmingham* **Charlie Cross**, *Bangor University* **Jaidan McLean**, *University of Oregon* **Maggie Mi**, *University of Lancaster* **Cecilia Tang**, *University of Manchester* **Caren Chua**, *King's College London* **Emily Shepherdson**, *University of York* **Simran Hussain**, *University College London*	Institutional Representatives

ULAB Local Committee 2022

Caitlin Wilson	Chair
Nicole Chan	Vice Chair
Darja Prudcenko	Secretary
Elif Yildiz	Treasurer
Oliwia Szczerbakiewicz **Keyu Dong**	Workshop and Panel Coordinators
Clara Braun **Cliodhna Hughes**	Raffle and Merch Coordinators
Joana Počopkaitė	Social Secretary
Trudy Kalvanaite **Kelly Jin**	Publicists
Michael Goesseler **Nicole Chan**	Tech Coordinators

Foreword from the National Chair, 2021-2022

Cliodhna Hughes, University of Edinburgh

The 2021-2022 academic year has seen a continuation of momentum for ULAB, in finding and creating further opportunities to fulfil our aims of bringing undergraduate linguists together, recognising and showcasing their research and helping them achieve their linguistics-related career goals. This academic year we ran a buddy scheme for the first time, which saw an incredible 49 sign-ups and which we are planning on running again in the 2022-2023 academic year, in new and improved form! Another ULAB first was our 2021-2022 Undergraduate Essay Competition, which saw some incredible submissions from talented undergraduate authors, studying at a variety of institutions around the UK, in the areas of syntax, discourse analysis and language evolution. We also secured funding to print some of our U-Lingua issues, which we then distributed to universities around the country. We ran a successful series of online workshops, aimed at helping undergraduate linguists improve their CVs and application-writing skills, and informing them of linguistics-related career opportunities beyond university.

In the 2021-2022 academic year we also made a successful return to in-person events following the COVID-19 pandemic. We did, however, retain a hybrid aspect to our 2022 conference, which allowed it to remain accessible to students who could not attend in-person. Thanks to the hard work of Caitlin, our newly-elected archivist, you will be able to read about lots of the work presented at the conference in these proceedings, which we hope will inspire current and future undergraduates.

I want to thank everyone on the ULAB 2021-2022 Committees and Subcommittees for their hard work and dedication throughout the year. I wish the best of luck to the newly-elected Committees and I look forward to seeing what ULAB does next!

Foreword from the Local Chair, 2021-2022

Caitlin Wilson, University of Edinburgh

The ULAB 2022 conference brought us many firsts as well as bringing back lots of classic ULAB experiences that we missed over the course of the pandemic. This year was ULAB's first ever hybrid conference. Carrying-on from the success of the virtual ULAB conference held in 2021 at the University of Aberdeen, we decided to embrace the benefits of virtual talks – receiving talks from speakers as far as the United States and Australia – as well as welcoming both new and old faces in person at the University of Edinburgh.

Another first included inviting secondary school students to present posters at the conference which turned out to be great fun and an excellent way to introduce younger minds to the exciting possibilities held in a Linguistics degree. Our in-person attendees also got to experience the first ULAB ceilidh. This was a wonderful evening of fun and laughter in which we thoroughly enjoyed introducing everyone to a bit of Scottish culture.

We were very grateful to receive talks from distinguished guests such as Graeme Trousdale who led a workshop on puzzle building, as well as presentations from up-and-coming researchers who presented plenary talks and workshops alike. Undergraduate speakers joined us from over 20 universities in the UK and abroad, giving talks on everything from Syntax and Sign Language to Pragmatics and Psycholinguistics.

Overall, the conference was a great success and I am forever indebted to my amazing Local Committee who worked tirelessly to make these three days run as smoothly as possible. I can't wait to see what's next for ULAB and am excited to see you all at ULAB 2023 in Manchester.

Introduction to the Proceedings

Caitlin Wilson, University of Edinburgh

It is such an honour to be able to present to you the Proceedings of the 2022 ULAB conference. As the outgoing Local Chair, I am particularly attached to the papers in these Proceedings. The conference held at the University of Edinburgh was everything ULAB stands for: we had speakers from all corners of the United Kingdom as well as further afield presenting the finest that undergraduate research in Linguistics has to offer.

Guests and speakers alike were able to attend the conference both in-person and online enabling us to spread our passion for Linguistics far and wide. The virtual aspects of the conference were held on Teams, allowing speakers to phone in and present their talks to an online audience as well as being projected on to screens in Edinburgh for our in-person attendees. Similarly, in-person talks were live-streamed via Teams to all our virtual attendees. This hybrid model, while proving to be quite the technical challenge, was something the Local Committee was committed to seeing through to make this conference accessible to people outside the country. Overall, the model worked well with few glitches across the three days conference and allowed us to welcome speakers from four contients who would have otherwise been unable to attend.

I am incredibly honoured to have been elected as this year's Archivist for ULAB. I hope to carry on the great legacy left by Lydia Wiernik, to whom I am incredibly grateful for all their support and guidance during these past few months. My first big project as Archivist was publishing the Proceedings from ULAB II, held in 2012. This was an incredible project spearheaded by Lydia and Richard Littauer, co-founder of ULAB. It was fascinating to read papers presented a decade ago and see how far ULAB has come as an organisation. I highly recommend you give it a read once you've finished these Proceedings.

I can now add this publication to ULAB's archives to sit alongside our other amazing projects such as the U-Lingua Magazine and JoULAB, as well as the many other ULAB archives consisting of conference programmes, presentation slides, constitutions, and many, many photos of happy times spent together.

These Proceedings are composed of 15 articles written by undergraduate students who presented their research at ULAB XII. They are organized as follows: Section A comprises full length papers, and Section B includes short squibs and write-ups. Full length papers have all been written as part of the work needed to achieve completion of a university course and have been awarded a Class I or II.1. Papers in Section B are a combination of shorter assessed work as well as research carried out independently of a university programme which therefore may not have been assessed. This does not diminish the quality of these papers as they were all reviewed and received approval during the submission process for the conference.

These papers span a wide range of sub-fields, and their authors are all at different stages of their undergraduate career. This, I hope, will make for a truly interesting read as you work your way through these Proceedings.

The journey to these completed Proceedings has been long and helped by many people to whom I am very grateful. Special thanks must be given to the incredible copyediting team who helped me compile, copyedit, and format these articles: Beatrix Livesey-Stephens, Andrew Tobin, and Hui Zhu. Thank you once again to Lydia Wiernik for their encouragement and for the beautiful cover design. I also give thanks to my Vice-Chair in the Local Committee, Nicole Chan, who supported me during the conference and stepped up to every challenge sent her way. Lastly, a huge thank you to our outgoing National Chair, Cliodhna Hughes, who led ULAB for the past two years with fervent dedication.

Please enjoy the 2022 Proceedings of ULAB.

SECTION A

ASSESSED CONTRIBUTIONS

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

What's So Hard About *That*? Demonstratives as a Locus for Cross-Linguistic Interaction in Bilinguals

Ariane Branigan
University of Edinburgh

Abstract. Three criteria have been established for determining whether a faculty or aspect of language is a locus for cross-linguistic interaction (CLI) in bilinguals: crosslinguistic performance congruity, intragroup homogeneity, and intergroup heterogeneity (Jarvis and Pavlenko, 2008, p. 41). This study aims to address all three in order to determine whether demonstratives - which are both linguistic universals and also subject to great cross-linguistic variation – are a possible locus for CLI (Diessel, 1999). Accordingly, we collected data from L2 speakers of six languages - three with 2-term systems (English, Italian, and Chinese) and three with 3-term systems (Thai, Japanese, and Spanish) - and compared them to L1 speakers of these same languages. Our results meet all three of Jarvis and Pavlenko's (2008) criteria suggest that demonstratives are indeed a locus for CLI.

Keywords: second language acquisition; demonstratives; bilingualism; deixis

1 Introduction

Today, it is well-accepted in the field of bilingual studies that a speaker's first language (L1) can affect the acquisition and usage of their second (L2). This phenomenon has been given a variety of names; it will be referred to here as *cross-linguistic influence* (CLI) (Jarvis, 2016, p. 608). Regardless of nomenclature, the phenomenon is striking due to its pervasiveness: not only has it been attested in every aspect of language, from syntax to pragmatics to phonology, but it also affects both comprehension and production (Jarvis & Pavlenko, 2008, p. 111).

Despite the extensive and ever-growing body of research surrounding CLI, one area which has still been understudied thus far is that of demonstratives (i.e., *this* and *that*). These items are particularly interesting from a perspective of CLI because they are considered to be linguistic universals which also exhibit great typological variation. They therefore provide ripe ground for cross-linguistic comparison (Diessel, 1999, p. 472). Furthermore, because controversy still persists regarding the extent of CLI in various linguistic subsystems, an open question is whether these ubiquitous words are affected by CLI at all (Odlin, 2003, p. 439). Accordingly, this dissertation aims to advance our understanding of whether demonstratives are a locus for CLI in bilinguals. To do so, we devised a study which collected demonstrative judgements from L1 and L2 speakers of six typologically distinct languages – English, Chinese, Italian, Japanese, Thai, and Spanish. We used the three criteria set out by Jarvis and Pavlenko (2008) (cross-linguistic performance congruity, intragroup homogeneity, and intragroup heterogeneity) to compare the L2 groups to their L1 counterparts. The sections are organised as follows: Chapter 2 provides an overview of CLI across multiple domains. In Chapter 3, we discuss demonstratives from a broad crosslinguistic perspective, before focusing on the six languages examined in our survey. Chapter 4 details the survey itself, including the procedure and results. Finally, we discuss our results in Chapter 5.

2 Cross-Linguistic Influence

Adult L2 learners, unlike children learning an L1, already have full working knowledge of their L1's sounds, structures, and words – we may see CLI when this knowledge diverges from, or aligns with, a speaker's L2 (Appel, 2010, p. 385; Odlin 2003, p. 436). We now provide a summary of such influences across multiple linguistic domains.

Impressionistically, the obvious locus of CLI is phonology, as the influence of a speaker's L1 is clearly present in their L2 accent (James, 2010, p.293). Due to age-related constraints, late L2 speakers routinely struggle to correctly produce phonemes which are not in their L1 (Major, 2008, p. 66; Ruben, 1997, p.203). This is particularly pronounced when two phonemes are allophones in one language, but contrastive in another (Eckman, Elreyes, and Iverson, 2003, p. 171). Tones are also affected by CLI; native speakers of non-tonal languages perceive tones as melodic (i.e., non-linguistic) and are less sensitive to tonal differences as a result (Halle, Chang, and Best, 2003, p. 416).

We also see CLI in syntax and morphology. A speaker's L1 may hinder the usage of their L2 when both languages have similar, competing systems. Take grammatical gender: L1 German speakers produce many more agreement errors between Dutch nouns and determiners when there is gender incongruence across the two languages than when there is gender congruency (Lemhöfer, Spalek, and Schriefers, 2008). However, CLI in overlapping systems may also enable L2 speakers to behave *more* like native speakers. L2 speakers perform more like native speakers if there is a clear mapping between the morphemes in their L1 and L2 (Hawkins & Liszka, 2003; Jiang et al., 2011).

Although relatively understudied compared to phonological and morphosyntactic transfer, evidence also suggests that L2 speakers' lexicons are subject to L1 influence (Appel, 2010). In particular, speakers struggle to acquire semantic boundaries which do not exist in their L1. For example, L1 Spanish speakers have difficulties acquiring the nuances between stools, benches, and chairs in English, as Spanish only makes two distinctions (Graham & Belnap, 1986). The same behaviour has been observed in speakers of various unrelated L1s when naming everyday household items in English. Although high-proficiency speakers behave more similarly to native speakers than low-proficiency ones, the difference between L1 and L2 groups persists (Malt & Sloman, 2003).

All of the examples discussed have provided empirical evidence to support the intuitive belief that a speaker's L1 plays an influential role in the production and comprehension of their L2 in terms of syntax, morphology, phonology, and the lexicon (Jarvis, 2011, p. 1). Outside these realms, however, there is also a growing body of research surrounding "conceptual transfer", which explores the impact of a speaker's L1 on their L2 beyond purely "structural and semantic well-formedness" (Jarvis, 2011, p. 2). Conceptual transfer should not be confused with the Sapir-Whorfian hypothesis of linguistic relativity: while the latter focuses on *extralinguistic* behaviour such as colour, spatial, and temporal perception (cf. Winawer et al., 2010; Majid et al., 2004; Boroditsky, 2001, respectively), conceptual transfer refers to purely linguistic behaviour. Results from studies focusing on conceptual transfer convincingly point to language-specific categorization strategies, which take time to restructure when acquiring an L2 (Jarvis, 2016, p. 5). One such example is colour naming: L1 Greek speakers with long-term exposure to English slowly lose their sensitivity to the light/dark blue distinction which Greek encodes lexically (Athanasopoulos et al., 2010).

We have established that there is a rich empirical grounding for CLI, spanning multiple linguistic domains and competencies. We now ask: how do demonstratives fit into this framework? In particular, are they subject to semantic or conceptual transfer? To answer this question, context regarding demonstratives – both how they are used, and how they are acquired – is necessary.

3 Demonstratives

Demonstratives play an important role in natural language. Through deictic reference, they provide a joint focus of attention for interlocutors, enabling us to talk about the world around us (Diessel, 2014, p. 117; Dixon, 2003, p. 61). In addition to this so-called "exophoric" usage (Diessel, 2006, p. 470), they can also be used non-deictically for empathetic ("I hate **that** idiot"), recognitional ("I love **that**"), and anaphoric ("**this** is what I mean") functions (Levinson, 2004, p. 103; Chen, 1990, p. 140). It is widely accepted that the exophoric function of demonstratives is "unmarked" (Diessel, 1999; Levinson, 2004). Due to scope limitations, any mention of demonstratives going forward should be understood only in this capacity.

3.1 L1 Acquisition of Demonstratives

Demonstratives pose a particular challenge for children learning an L1, because they involve constant changes in reference and boundaries (Clark, 1978, p. 89). They also occasionally require hearers to take the perspective of their speaker; this requires executive function and inhibition control, both of which pose problems for young children (Charney, 1977, p. 79; Chu & Minai, 2018, p. 1361).

Despite these difficulties, demonstratives are still learned by children very early in the acquisition process (Clark & Sengul, 1977, p. 459; Diessel, 2006, p. 481). However, children do not necessarily understand that these words refer to a distance contrast; while demonstrative *forms* are acquired early, the distance features they encode are acquired over multiple years (Webb & Abrahamson, 1975, p. 350; Clark & Sengul, 1977, p. 459; Chu & Minai, 2018).

3.2 L2 Acquisition of Demonstratives

The challenges of acquiring demonstratives and other spatial terms in an L2 — as well as language as a whole – are obviously different to those faced by children acquiring an L1. As described in Section 2, the biggest problem for L2 learners in this dimension is what Klein (1986, p. 62) terms *the matching problem*: Because learners approach an L2 already equipped with knowledge of their L1, and all languages have demonstratives, the speakers will encounter competing representations.

Very little has been written regarding the L2 acquisition of demonstratives. Cursory evidence from Niimura and Hayashi's (1996) bidirectional investigation into L2 Japanese and English non-exophoric demonstrative usage suggests that demonstratives pose an issue even for advanced learners. The authors posit that this results from the different mechanisms involved when choosing the correct form in each language, as outlined in the preceding sections.

Despite this lack of literature, however, there is a large body of literature on a similar phenomenon: prepositions. Although prepositions differ from demonstratives in that they are not deictic, they still encode spatial reference and show cross-linguistic variation (De Villiers, 1974, p. 439). While infants are sensitive to prepositional contrasts which are not encoded in their L1, this starts to wane throughout childhood; accordingly, adults have reduced sensitivity to non-L1 spatial structures (Munnich & Landau, 2010; McDonough et al., 2003; Choi, 2006). Evidence from studies on Finnish (Jarvis & Odlin, 2000), Chinese (Li & Cai, 2015), Korean (Park & Ziegler, 2014), and Spanish and Danish (Alonso et al., 2016) suggest that prepositions are subject to CLI when the prepositions of a speaker's L1 do not match those of their L2. A mismatch in prepositions leads to less "nativelike" behaviour.

Like prepositions, demonstratives also bear spatial meaning, are learned early in the L1 acquisition process, and have considerable cross-linguistic variation. As such, it is reasonable to assume they would also be susceptible to CLI. Returning to the question we originally posed in Section 2, however – will demonstratives be subject to semantic or conceptual transfer? – we have seen that they "… straddle the boundaries between abstract semantic organization, and context-specific… interaction" (Kemmerer, 1996, p. 56). Therefore, it is still difficult to determine how exactly they will be subject to CLI. Accordingly, the scope of this dissertation is restricted only to determining whether they are a locus for CLI at all.

Having established demonstratives' function and acquisition, we will now provide a brief overview of demonstratives cross-linguistically, before focusing on the languages featured in our survey.

3.3 A Typology of Demonstratives

Although demonstrative systems share some features cross-linguistically (e.g., almost all encode a distance contrast[1]), they also vary greatly (Diessel, 1999, p. 38). For example, the number of forms in a language's demonstrative inventory can range from one to five (e.g., Koasati) (Kimball, 1991, p. 486). The majority of the world's languages are *bipartite* or *tripartite*, meaning that they have two or three demonstratives, respectively (Diesel, 2013).

In addition to the number of demonstrative forms a language can have, the type of distance contrasts encoded by those forms also varies. While bipartite languages mainly divide space depending on whether an object is close (*proximal*) or far (*distal*) from the speaker, tripartite languages differ in whether their third, intermediate (*medial*) term is determined by distance to the speaker (*distance-oriented*) or the hearer (*person-oriented*) (Diessel, 1999, p. 39; Levinson, 2018, p. 19).

3.3.1 English

English is bipartite; *this* and *that* refer to objects which are proximal and distal to the speaker, respectively (Wu, 2004, p. 31; Diessel, 2013). We will now discuss the languages addressed in our study.

3.3.2 Chinese

Chinese is bipartite, with *zhe* and *na* encoding proximal and distal relationships to the speaker. Spontaneous production data suggests that these forms are used similarly to English demonstratives (Wu, 2004, p. 70).

3.3.3 Italian

Italian is bipartite; *questo* "corresponds by and large to English *this*", while *quello* maps onto *that* (Maiden & Robustelli, 2013, p. 82).

3.3.4 Japanese

Japanese is tripartite. Its three forms – proximal, medial, and distal – are referred to as the *ko-, so-,* and *a-* series, respectively (Hoji et al., 2003). Traditionally, Japanese has been construed as a person-oriented demonstrative system (cf. Yoshimoto, 1986, p. 55; Oshima & McCready, 2017, p. 803;

[1] A notable example of a language without such a distance contrast is German, which only has one demonstrative: *dies-*. However, distance contrasts can still be expressed via the adverbial demonstratives *hier* and *da* (Diessel, 2013).

Wilkins, 2018, p. 60; Kuno, 1973, p. 27). However, this interpretation is not unanimous: Ono (1997) posits that the speaker's location in relation to the object is the only determining factor in demonstrative usage, making Japanese a distance-oriented system. Matsuoka (2000), on the other hand, argues for a hybrid system. They state that Japanese is person-oriented when speakers are opposite each other (*tairitsu*), while it is distance-oriented when speakers are next to each other (*yugo*).

3.3.5 Thai

Thai is tripartite. The forms *nii, naan,* and *noon* can act as demonstrative pronouns or adjectives, depending on the tone (Smyth, 2002, p. 224). There is little consensus in the English literature regarding the anchoring system of Thai. It has been described as both person- (cf. Nakagawa, 2002, p. 29; Dixon, 2003, p. 86) and distance-oriented (cf. Iwasaki & Dechapratumwan, 2021, p. 4).

3.3.6 Spanish

Spanish is tripartite and distance-oriented. *Este* refers to proximal objects, *ese* refers to medial objects, and *aquel* refers to distal objects (Levinson, 2004, p. 24).

4 The Present Study

The previous sections have shown that, despite an influx of recent studies, much is still unknown regarding the areas of language CLI affects – relating in particular to both the lexicon and conceptual organization (Appel, 2010, p. 382). We have observed that, while there is extremely limited information about the L2 acquisition of demonstratives, there is ample evidence that prepositions, which are similar to demonstratives in many ways, prove difficult for L2 speakers to master. Finally, we have seen that demonstratives are subject to great cross-linguistic variation.

Accordingly, the present study aims to provide an overview of how L1 speakers of six typologically distinct languages – English, Chinese, Thai, Japanese, Spanish, and Italian – use demonstratives in their L2s, with a view to determining whether demonstratives are a potential locus for CLI. Our methodology was shaped by Jarvis and Pavlenko's (2008, p. 41) assertion that there are three types of evidence which are crucial in determining whether a specific faculty or area of language is a locus for CLI:

Table 1: *Jarvis and Pavlenko's (2008) three criteria for CLI*

Type of evidence	Expected pattern if CLI exists	Impact on study design
Crosslinguistic performance congruity (Is there a link between how speakers behave in their L1 and their L2?)	Speakers' behaviour in their L2 is traceable to their L1; there is a clear mapping between their behaviour in their L1 and L2.	Gather L1 judgements from speakers of the same language – some of whom do the experiment in their L1, and some of whom do it in their L2.
Intragroup homogeneity (How do speakers of the same L1 behave in their shared L2?)	Speakers who share an L1 behave similarly to each other in their L2.	Gather L2 judgements from speakers who share an L1.

Intergroup heterogeneity (How do L1 speakers of different L1s behave in their shared L2?)	Speakers who do not share an L1 behave differently to each other in their L2, and these differences are traceable back to their respective L1s.	Gather L2 judgements from speakers of different, unrelated L1s.

4.1 Participants

Anyone over the age of 18 was eligible to take part in the experiment if they were a native speaker of any of the languages specified in Section 3, *or* a non-native speaker of English. The survey received a total of 2,189 complete submissions. We removed 20 participants who either completed the form in less than 1 minute or did not consent, and 514 participants whose first language was not one of the languages specified in Section 3.3. The remaining 1,654 participants were comprised of the following demographics:

- L1 English (N = 1,130)
- L1 Chinese (N = 51)
- L1 Thai (N = 283)
- L1 Japanese (N = 24)
- L1 Spanish (N = 105)
- L1 Italian (N = 61)

4.2 Procedures

The experiment took place in the form of an internet-based Qualtrics survey. There were 16 survey "flows": six control and ten experimental. The control surveys involved L1 speakers of the six languages performing the task in their L1, while the experimental surveys involved bilingual speakers performing the task in their L2. For L1 Chinese, Thai, Japanese, Spanish or Italian speakers, this was the English condition, while for L1 English speakers, this was whichever listed language they indicated knowledge of (see appendix B for more information). The language combinations, and the number of participants who took each survey, are shown below:

Table 2: *Possible speaker/survey language combinations*
(Red and blue cells indicate experimental and control groups, respectively)

Native Language	**Language of Survey**					
	English	*Chinese*	*Thai*	*Japanese*	*Spanish*	*Italian*
English	N = 452	N = 56	N = 34	N = 396	N = 135	N = 57
Chinese	N = 23	N = 28				
Thai	N = 129		N = 154			
Japanese	N = 10			N = 14		
Spanish	N = 49				N = 56	
Italian	N = 32					N = 29

All bilingual participants were asked to rate their reading proficiency in their selected second language(s), as previous evidence shows that proficiency levels may affect CLI (cf. Malt & Sloman, 2003). The decision to measure proficiency via reading level was informed by evidence that L2 reading ability is a good indicator of wider linguistic ability (Carrell, 1991), and follows other studies which match proficiency measures to modality of the study (cf. Montrul, 2018). The study elicited self-reported proficiency ratings using the following 6 levels adapted from the Common European Framework of Reference (CEFR) (Council of Europe, 2020):

(1) Very short, simple texts (e.g., simple postcards) (CEFR A1)
(2) Short, simple texts (e.g., adverts, menus, timetables) (CEFR A2)
(3) Letters, brochures, short official documents (CEFR B1)
(4) Articles, news items, reports and contemporary literature (CEFR B2)
(5) Complex literary texts, specialised articles, very technical instructions (CEFR C1)
(6) All forms of written language, including complex and abstract pieces (CEFR C2)

To ensure they understood the task, participants were given a survey in their L2 only if they rated themselves at level A2 or above in that language; otherwise, they were given the condition in their L1.

More information about how the survey was weighted to ensure participants were evenly distributed across conditions is in Appendix B.

4.3 Stimuli

4.3.1 Targets

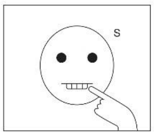
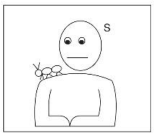
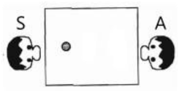
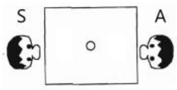
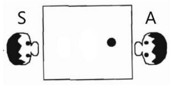
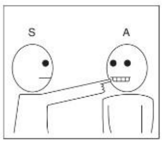
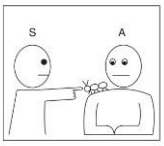
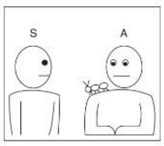
22 picture-and-sentence pairs were adapted from two existing questionnaires, Pederson and Wilkins (1996) and Wilkins (2018). Both of these questionnaires had been designed to elicit demonstratives from native speaker informants in the field, and depict the elicitation materials as line drawings which the researcher is encouraged to re-enact in person. For this online experiment, the materials were adapted into photographs.

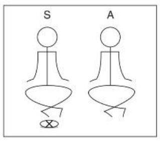
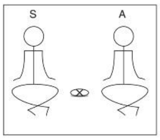
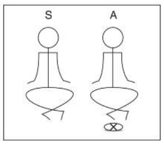
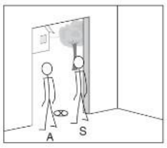
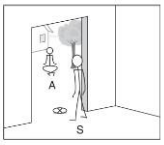
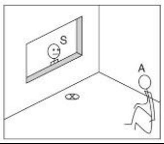
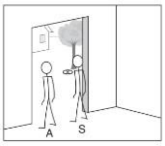
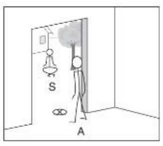
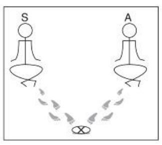
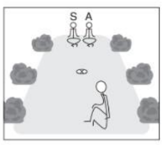
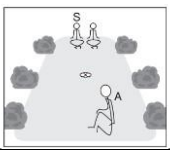
Each photo depicted an object whose position varied in relation to the speaker and addressee. The photo was accompanied by a sentence in the target language describing the scene, with a dropdown in place of the demonstrative. The participant was tasked with selecting the demonstrative from this dropdown which, in their opinion, best fit the scene. Both target and filler stimuli were randomised. Objects were chosen to feature in the target photos on the basis of on two criteria. Firstly, they had to be inanimate and harmless. This was due to constraints which were both semantic (Rocca, Tylén; Wallentin, 2019) and pragmatic (Nakagawa, 2002, p. 24). Additionally, they had to be objects whose names would be recognized by speakers of A2-level English (as determined by the 2012 Cambridge KET Vocabulary List). The objects and their various positions were identical across all surveys. However, the gesture associated with the utterance differed between the English, Spanish, and Italian conditions and the Chinese, Japanese, and Thai conditions due to cultural sensitivities around pointing (Axtell, 1998).

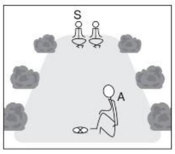
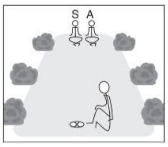
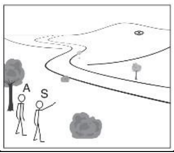
Examples of a photo-sentence pair for the English and Japanese conditions are shown in Figure 1 and 2, respectively. Table 3 provides the full list of items, with the original scene drawings from Wilkins' (2018) demonstrative questionnaire. A table containing all photo-sentence pairs featured in the survey is provided in Appendix A.

Figure 1: *This/that juice was expensive.* **Figure 2**: *この/その/あのジュースは高かった.。*

Table 3: *Target Sentence-Picture Pairs*

Item number	Scene	Prompt – this/that…
1		finger hurts
2		bracelet is lovely
3		juice was expensive
4		cake is sweet
5		coffee is tasty
6		arm is strong
7		necklace is pretty
8		drink is good

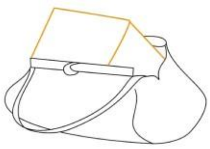
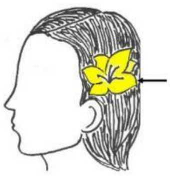
9		book is interesting
10		sandwich was cheap
11		magazine is boring
12		wallet is big
13		phone is small
14		laptop is broken
15		scarf is beautiful
16		notebook is useful
17		hat is yellow
18		bag is red
19		bottle is full

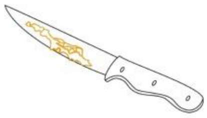
20		box is big
21		umbrella is black
22		church is old

4.3.2 Filler

As prepositions can pose particular difficulty for non-native speakers, our filler material consisted of eight additional picture-and-sentence pairs. The materials were adapted from both Bowerman and Pederson (1992) and Zhang (2013).

Table 4: *Filler sentence-picture pairs*

Item	Scene	Title
1		The box is (in/on) the bag
2		The flower is (in/on) your hair
3		The sweets are (in/on) the plate
4		The ring is (in/on) my finger
5		The toy is (in/on) the bowl

6		The jam is (in/on) the knife
7		The crack is (in/on) the cup
8		The cup is (in/on) the stool

4.4 Results

We will go through each type of evidence identified by Jarvis and Pavlenko (2008) in turn. Starting with *cross-linguistic performance congruity*, we will first establish how L1 bipartite speakers behave when doing the task in their L1s. We will use these data to make predictions about how L1 speakers of their languages may behave in their L2 if demonstratives are indeed a locus for CLI. We then test these predictions with the data for L1 bipartite speakers doing the task in their L2s and repeat this process for L1 tripartite speakers. Regarding *intragroup homogeneity*, we will compare the rates of agreement for L1 and L2 English speakers, to determine how similarly L2 speakers are behaving within their speaker group. Finally, to explore *intergroup heterogeneity*, we will compare patterns of L2 demonstrative usage across L1 bipartite and tripartite groups, to determine whether they are different or not.

4.4.1 Cross-Linguistic Performance Congruity (Chinese, Italian, English)

First and foremost, L1 speakers of bipartite languages agree fairly consistently with other members of their respective speaker groups when we average out the rate of agreement for each language across items; the average percentage of participants who chose the top demonstrative choice in their L1 across items was 79% for English, 80% for Chinese, and 84% for Italian speakers.

Figure 3 shows the behaviour of the three speaker groups for each individual item; dots with a bluish tint indicate that speakers preferred their L1's proximal form for the item in question, while dots with a reddish tint indicate that they preferred their L1's distal form. A more monochrome dot indicates that speakers had a high agreement rate for that item.

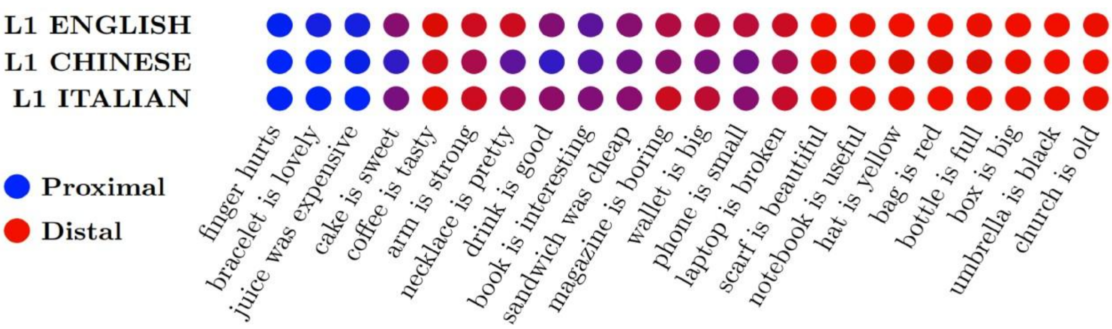


Figure 3: *Demonstrative Choices of L1 Bipartite Control Groups in their respective L1s, per Item.*

Here, we can see that L1 speakers of English, Italian and Chinese all decisively use the proximal form for the items which Wilkins (2018) identifies as either part of the speaker's body or in contact with the body (e.g., *this finger hurts*). They also use the distal form consistently for items where the objects are meters or tens of meters away from the speaker (e.g., *this/that umbrella is black*). However, their preferences diverge in the middle of the spectrum – those where the objects are "within arm's reach of speaker" or within a few steps of the speaker's personal space (Wilkins, 2018, p. 53). The items where the three speaker groups disagreed on their top demonstrative choice are as follows:

Table 5: *Items where L1 English and L1 Chinese Speakers Disagree on Their Top Demonstrative Choice*

<table>
<tr><th>Item</th><th>L1 English speakers</th><th>L1 Chinese speakers</th></tr>
<tr><td>4: Cake is sweet</td><td rowspan="5">Distal</td><td rowspan="5">Proximal</td></tr>
<tr><td>6: Arm is strong</td></tr>
<tr><td>7: Necklace is pretty</td></tr>
<tr><td>8: Drink is good</td></tr>
<tr><td>10: Sandwich is cheap</td></tr>
<tr><td>12: Wallet is big</td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td>13: Phone is small</td></tr>
</table>

Table 6: *Items where L1 English and L1 Italian Speakers Disagree on Their Top Demonstrative Choice*

<table>
<tr><th>Item</th><th>L1 English speakers</th><th>L1 Italian speakers</th></tr>
<tr><td>4: Cake is sweet</td><td rowspan="4">Distal</td><td rowspan="4">Proximal</td></tr>
<tr><td>7: Necklace is pretty</td></tr>
<tr><td>8: Drink is good</td></tr>
<tr><td>10: Sandwich is cheap</td></tr>
</table>

The general trend is that the English proximal form is used in much more restricted contexts than either the Chinese or Italian proximal form. To visualise this pattern more clearly, we plotted the percentage of speakers who chose the proximal form for each question; as seen in Figure 4, the trend of Chinese speakers using the proximal form more than English and Italian speakers is apparent even when the three groups preferred the distal form overall for that item; Chinese speakers' usage of the proximal form matches, or surpasses, that of the English and Italian speakers' for every item.

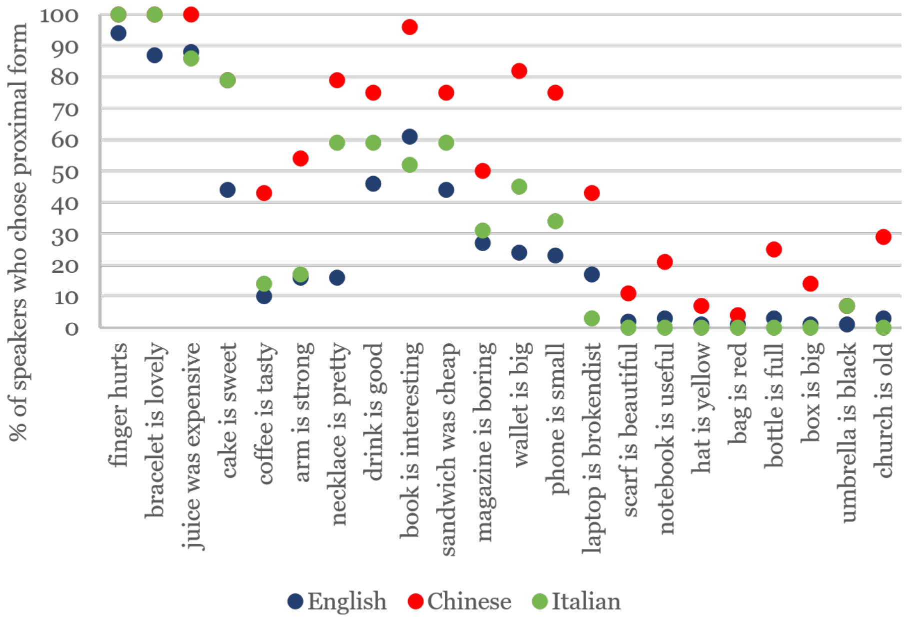


Figure 4: *% of L1 Bipartite Control Group Speakers who Chose the Proximal Form in Their L1, per Item.*

By comparing where these L1 speakers diverge, we can make some generalisations about their respective demonstratives and the spaces they encode in a manner which will be relevant when examining L2 usage. Specifically:

- The Italian proximal form *questo*, unlike the English proximal form *this*, is preferred when referring to objects either in contact with the addressee's body (although *not* when referring to part of the addressee's body), or situated between the speaker and the addressee.
- The Chinese proximal form, *zhe*, can be used for a much wider range of distances than the English proximal form *this*; it is preferred when referring to objects which are in contact with, or part of, the addressee's body, and objects within reaching distance of the speaker. It is also used at higher rates for all other items than the English proximal forms is.

We can now make predictions about how L1 speakers of these bipartite languages may behave when doing the task in their L2. We posit that, if demonstratives are subject to CLI, we would expect to see:

- Prediction 1a: L1 English speakers doing the task in Chinese will consistently use the proximal form less than L1 Chinese speakers, but especially for items where the object is contact with, or part of, the addressee's body, and objects within reaching distance of the speaker (items 4, 6, 7, 8, 10, 11, 12, and 13); o Prediction 1b: L1 Chinese speakers doing the task in English will do the inverse (i.e. overuse the proximal in general, especially for the aforementioned items).
- Prediction 2a: L1 English speakers doing the task in Italian will use the proximal form less than L1 Italian speakers for items where the object is in contact with the addressee's body or between the speaker and addressee (items 4, 7, 8 and 10); o Prediction 2b: L1 Italian speakers doing the task in English will do the inverse (i.e., overuse the proximal for the aforementioned items).

Figure 5 shows how L1 English speakers doing the task in Chinese (L2 Chinese speakers) behave compared to the L1 Chinese control group:

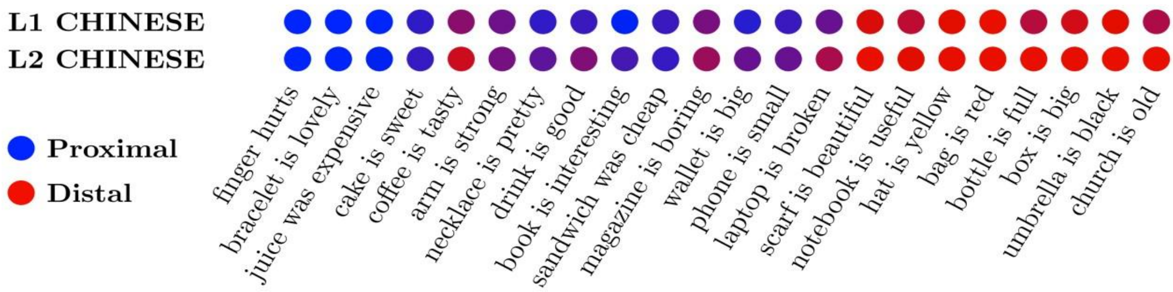


Figure 5: *Chinese Demonstrative Choices of L1 and L2 Chinese Speakers, per Item.*

We can see that L2 Chinese speakers have the same top choice of demonstrative as L1 Chinese speakers for the most proximal and most distal conditions. This was expected, as figure 3 has already shown that the control groups of L1 English and Chinese speakers doing the task in their respective first languages also have similar choices for these items. Unsurprisingly, we also see L1 and L2 Chinese speakers' choices diverging at the same point that L1 Chinese and L1 English speakers' choices diverge (i.e., items where Wilkins (2018) identifies the object as immediately reachable by the speaker, or reachable within a few paces).

Interestingly, however, L1 and L2 Chinese speakers only actually disagree on the top demonstrative choice for one item:

Table 7: *Items where L1 and L2 Chinese Speakers Disagree on Their Top Demonstrative Choice.*

Item	L1 Chinese speakers	L2 Chinese speakers
8: Drink is good	Proximal	Distal

We identified this item as one where L2 Chinese speakers may differ from L1 Chinese speakers, if CLI is in effect. However, we also predicted that L2 Chinese speakers would prefer the distal for another six items (4, 6, 8, 10, 12, and 13), yet instead they behave like L1 Chinese speakers in preferring the proximal for all of these. When we break down the items by percentage of speakers who preferred the proximal for each one, however, the pattern becomes a little clearer:

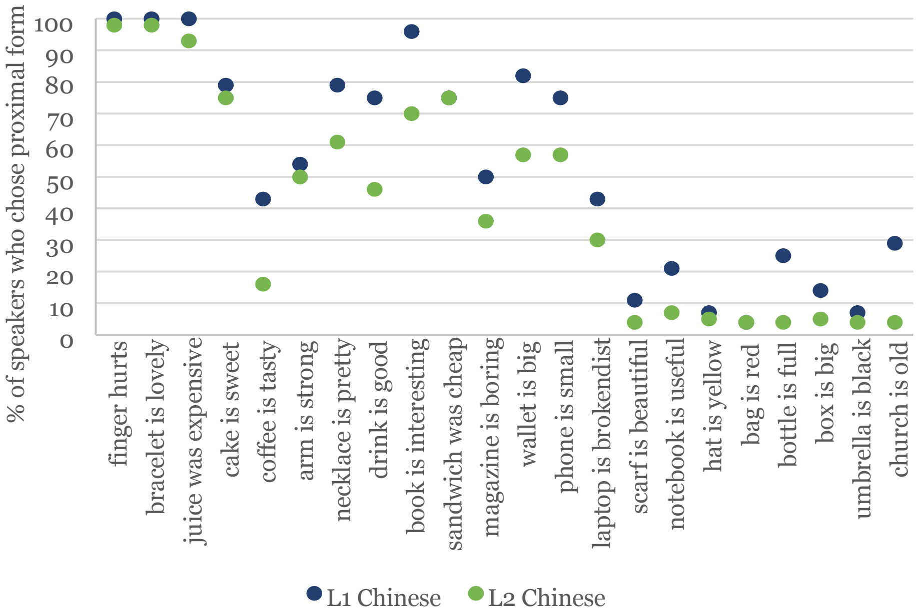


Figure 6: *% of L1 and L2 Chinese Speakers who Chose the Proximal Form in Chinese, per Item.*

Similar to what we saw in Figure 5, L1 Chinese speakers use the proximal form as much, if not more, than the L2 Chinese (i.e., L1 English) speakers for every item. This combination of results suggests that demonstratives may be subject to CLI in L2 Chinese speakers.

Turning next to L1 English speakers doing the task in Italian (L2 Italian speakers), we can see that the same pattern observed in the Chinese condition also holds. Broad agreement for very proximal and very distal items, but divergence for middle-distance items:

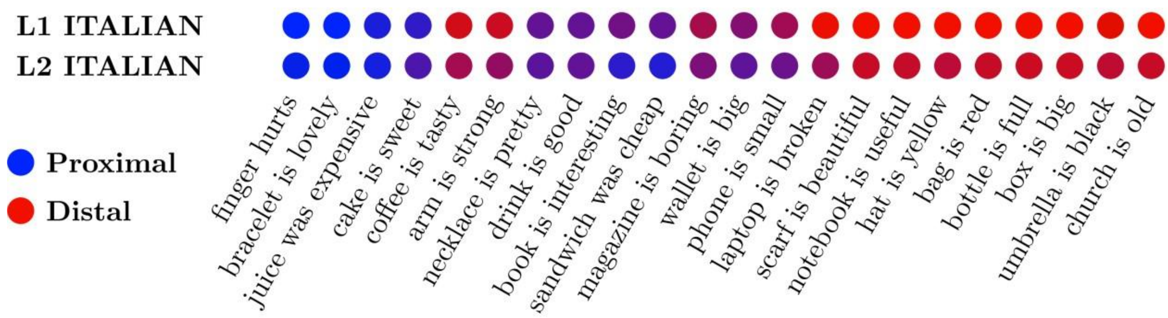


Figure 7: *Italian Demonstrative Choices of L1 and L2 Italian Speakers, per Item.*

In Section 4.4.1, we predicted that, if CLI affects demonstratives, we would expect to see L2 Italian speakers (whose L1 is English) preferring the distal form for items 4, 7, 8 and 10; these are the items where L1 Italian and L1 English speakers doing the tests in their respective L1s have different top choices. However, L1 and L2 Italian speakers only disagree on two items, and neither of those are items which we predicted would pose issues:

Table 8: *Items where L1 and L2 Italian Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 Italian speakers	L2 Italian speakers
Item 12: Wallet is big	Distal	Proximal
Item 13: Phone is small		

As these are items where L1 Italian and L1 English speakers both prefer the distal form in their respective L1s, it seems that this divergence does not necessarily stem from CLI.

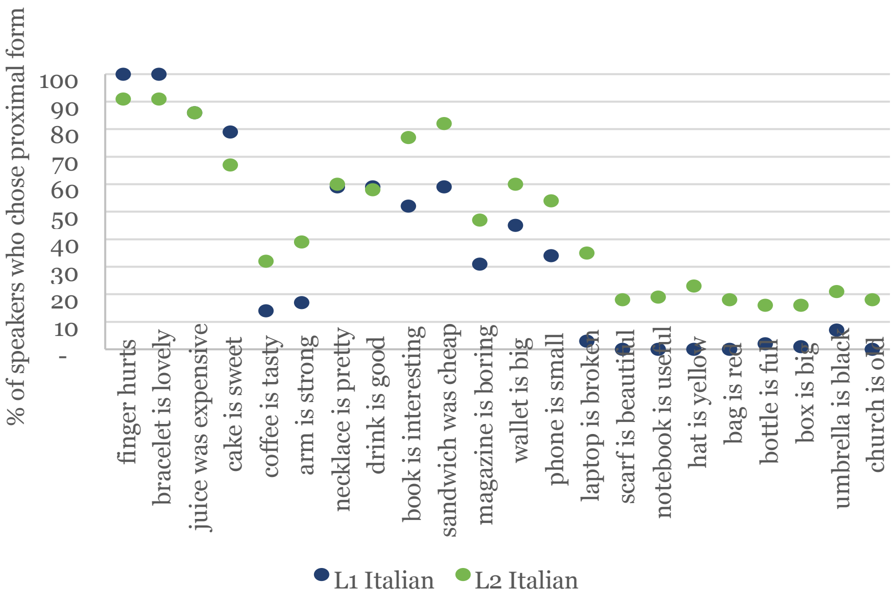


Figure 8: *% of L1 and L2 Italian Speakers who Chose the Proximal Form in Italian, per Item.*

Figure 8 shows that L2 Italian speakers are simultaneously underusing the proximal form when L1 speakers strongly prefer it and overusing the proximal form when L1 speakers strongly disprefer it. Some of this difference could be explained by the fact that L2 Italian speakers are less categorical in their first choices than L1 Italian speakers. When the percentage of speakers who chose the top choice of demonstrative is averaged out across all items, L1 Italian speakers' average agreement is 84%, while L2 Italian speakers' is only 73%. However, it is still unclear why L2 Italian speakers are overusing the proximal form, even in situations where L1 English speakers use it at a similar rate to L1 Italian speakers.

We will now examine L1 Chinese speakers doing the task in English (henceforth L1 Chinese-L2 English speakers). Figure 9 shows that agreement is high for the most proximal and most distal forms. We can also see immediately that L1 Chinese-L2 English speakers skew more towards the proximal form where L1 English speakers prefer the distal:

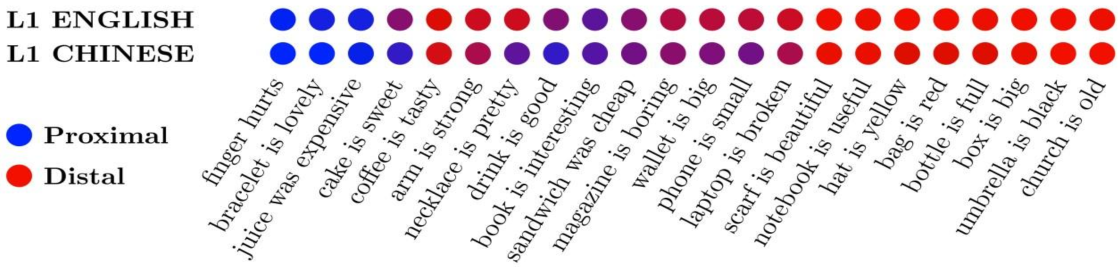


Figure 9: *English Demonstrative Choices of L1 English and L1 Chinese-L2 English Speakers, per Item*.

The five items where L1 English speakers and L1 Chinese-L2 English speakers disagree on the top choice of demonstrative are as follows:

Table 9: *Items where L1 English and L1 Chinese-L2 English Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 English speakers	L1 Chinese-L2 English speakers
4: Cake is sweet	Distal	Proximal
7: Necklace is pretty		
8: Drink is good		
10: Sandwich is cheap		
13: Phone is small		

These are all items which we previously identified in prediction 1b as those where L1 English speakers and L1 Chinese, L2 English speakers may diverge, due to different preferences in demonstrative form (proximal for Chinese, distal for English). While L1 English and L1 Chinese-L2 English speakers agreed on the distal form for the other items which we suggested may reveal differences in behaviour (items 6, 11, and 12), Figure 10 below shows that L1 Chinese-L2 English speakers still preferred the proximal form more than the L1 English speakers did for these items – and, indeed, all others:

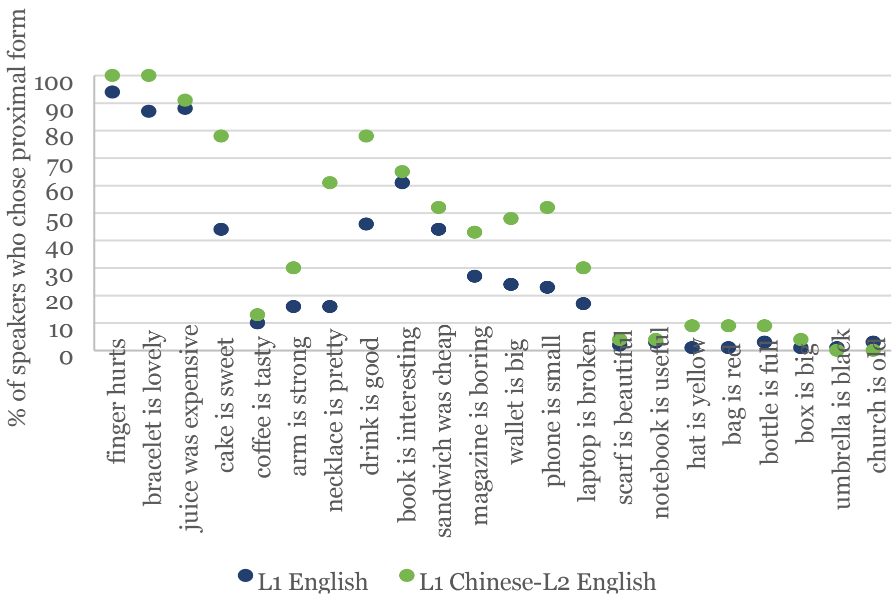


Figure 10: *% of L1 English and L1 Chinese-L2 English Speakers who Chose the Proximal Form in English, per Item.*

The combination of this evidence suggests that, although the majority of L1 Chinese-L2 English speakers have successfully learned that *this* has a narrower set of uses than *zhe* (hence choosing *that* as the top demonstrative choice for many of the same items that L1 English speakers do), they still consistently use *this* more than L1 English speakers (in line with prediction 1a).

Finally, we will examine the results of Italian speakers doing the test in English (henceforth L1 Italian-L2 English). Figure 11 shows the results broken down by item:

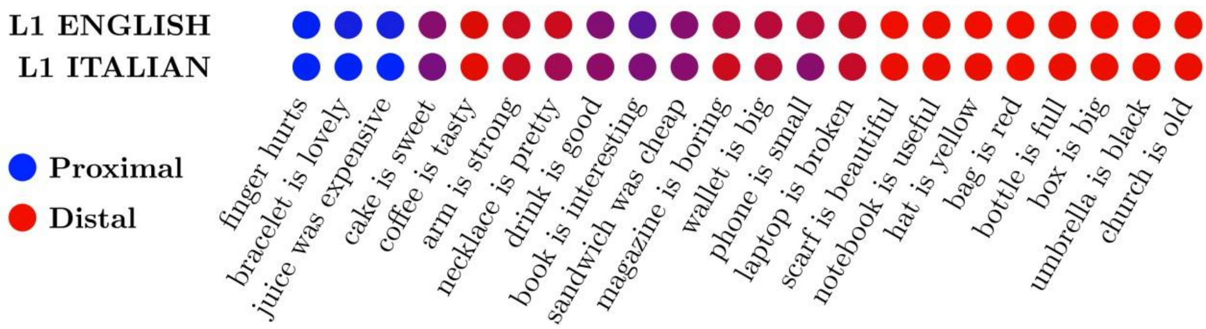


Figure 11: *English Demonstrative Choices of L1 English and L1 Italian-L2 English speakers, per Item.*

As suggested in prediction 2b, we can see that L1 Italian-L2 English speakers behave very similarly to L1 English speakers across the first three and last eight items. However, they only *disagree* with L1 English speakers for one item – and it is not one of the items where we predicted a disagreement:

Table 10: *Items where L1 English and L1 Italian-L2 English Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 English speakers	L1 Italian-L2 English speakers
9: Book is interesting	Proximal	Distal

Additionally, recall that we predicted L1 Italian-L2 English speakers would use the proximal form more than L1 English speakers for items 4, 7, 8 and 10; however, figure 12 shows that this only holds for items 4 and 7:

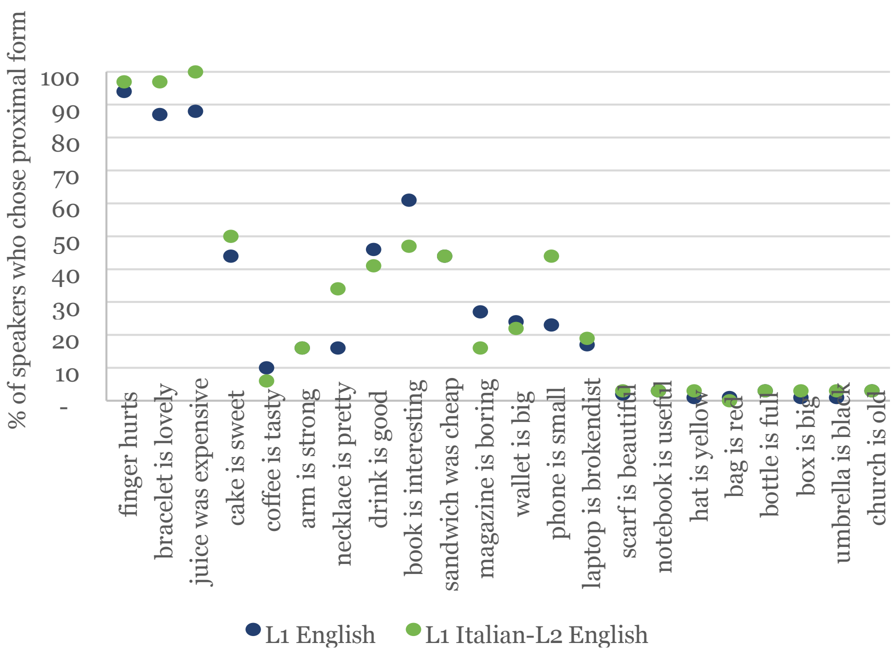


Figure 12: *% of L1 English and L1 Italian-L2 English Speakers who Chose the Proximal Form in English, per Item.*

The evidence is therefore unclear regarding demonstratives as a locus for CLI in L1 Italian L2 English speakers.

4.4.2 Cross-Linguistic Performance Congruity (Japanese, Spanish, and Thai)

Having established how L1 speakers of bipartite languages behave – both in their L1s and in their L2s – we now turn to our L1 tripartite control groups.

When we average out the rate of agreement for each language across items, L1 speakers of tripartite languages have less agreement within their respective speaker groups than L1 speakers of bipartite languages; even the highest rate of tripartite intragroup coherence (Japanese, 75%) is lower than the lowest bipartite rate (English, 79%). The average proportion of top demonstrative choice for Thai speakers was 71%, while Spanish speakers had the lowest agreement rate out of all six L1 groups with 68%. These agreement rates are not entirely surprising, given that speakers have three possible demonstrative forms to choose from, instead of two.

Turning now to demonstrative choice per item, the results are as follows:

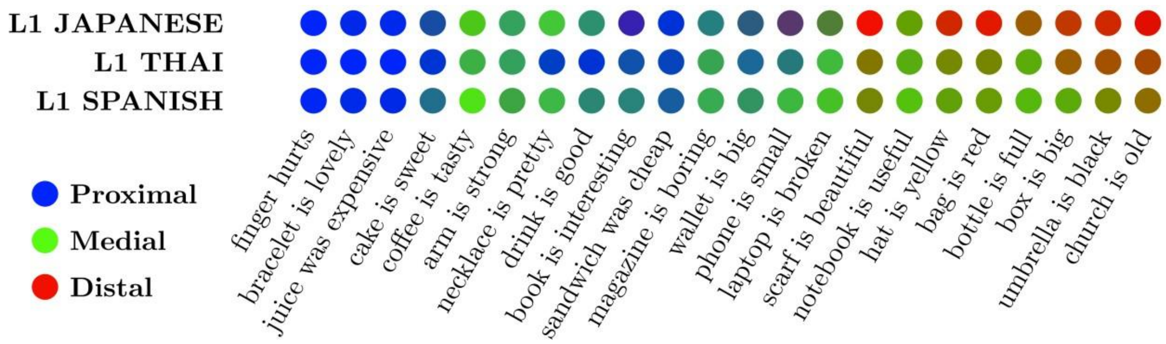


Figure 13: *Demonstrative Choices of L1 Tripartite Control Groups, Broken Down by Item.*

We can see that, in line with the agreement rates discussed above, L1 speakers of tripartite languages clearly use demonstratives less similarly to each other than L1 speakers of bipartite languages do. While there is still agreement regarding the most proximal conditions, judgements diverge from the middle of the spectrum all the way to the end.

The differing demonstrative preferences for the speakers in each group become clearer when broken down by demonstrative:

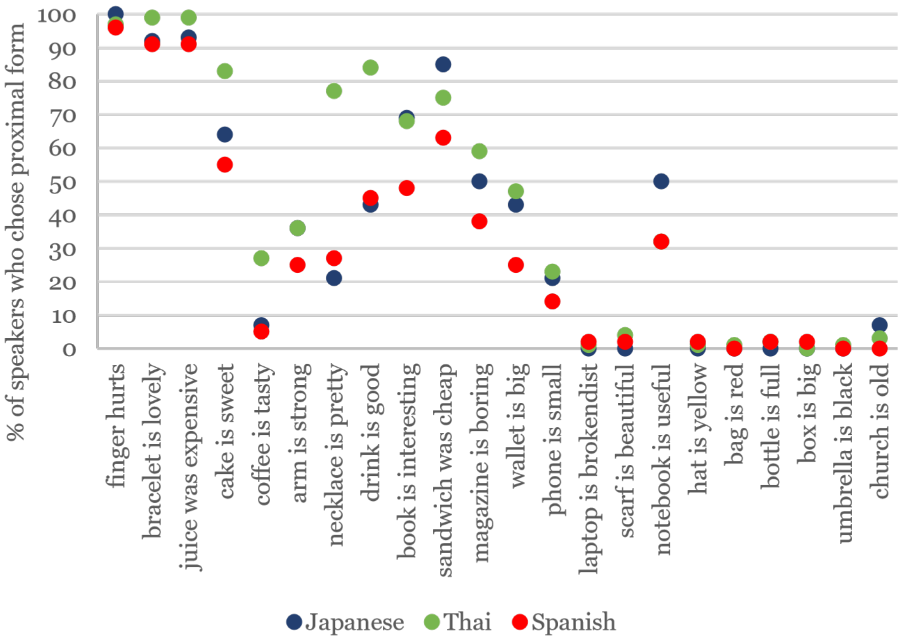


Figure 14: *% of L1 Tripartite Control Group Speakers who Chose the Proximal Form in Their L1, per Item.*

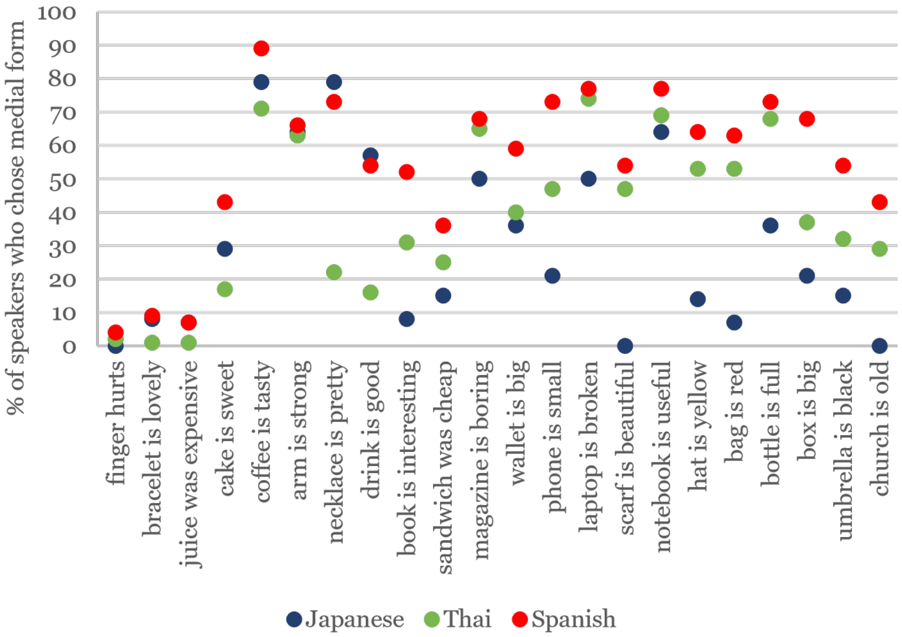


Figure 15: *% of L1 Tripartite Control Group Speakers who Chose the Medial Form in Their L1, per Item.*

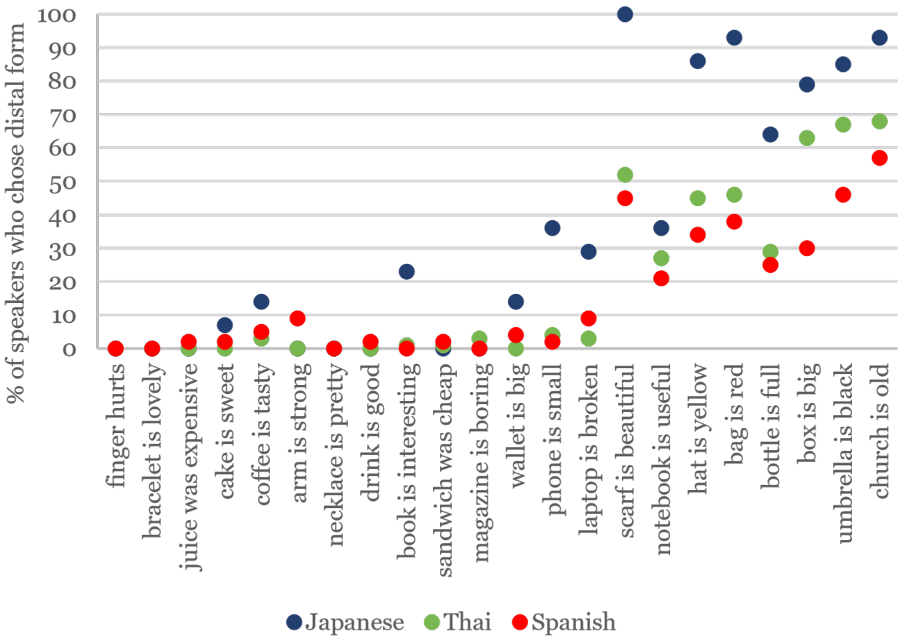


Figure 16: *% of L1 Tripartite Control Group Speakers who Chose the Distal Form in Their L1, per Item.*

Figure 14 shows that Thai speakers use the proximal form more than the other two speaker groups for three classes of items: those where the object is in contact with the speaker or addressee's body, within reaching distance of the speaker but closer to the addressee, and within a few steps of the speaker. Unsurprisingly, proximal usage drops off steeply for all three groups across items where the object was situated multiple metres away from the speaker.

Looking next at medial usage, Figure 15 reveals that Spanish speakers consistently use the medial form more than either of the other two groups, across all items. Japanese speakers use the medial more than the Thai speakers for the items where the object is either in contact with the addressee's body or within reaching distance of the speaker but closer to the addressee; this trend then is inverted for the rest of the items.

Finally, Figure 16 shows that Japanese speakers consistently use the distal form either as much as the other two groups or more so, for all items. In particular, for items 17-22, we see speakers forming a continuum in their choices; Japanese speakers use it the most, followed by Thai and then Spanish speakers.

Recall that, regarding the three bipartite speaker groups, our predictions for the L1 Chinese and Italian speakers were the inverse of our predictions for the L1 English speakers. For example, we suggested that L1 Chinese speakers would *over*use the English proximal form for the exact same items where L1 English speakers would *under*use the Chinese proximal form. However, this is not applicable when making predictions around how L1 speakers of tripartite languages will behave in English, and vice-versa, because there is a mismatch between the languages' demonstrative inventories.

Instead, we will make completely separate predictions for the two groups, starting with L1 Japanese, Thai and Spanish (i.e., L2 English) speakers. Firstly, contrary to previous studies, our data suggests that both Japanese and Thai have distance-oriented demonstrative systems, similarly to Spanish – see appendix C for a more detailed breakdown of how we reached this conclusion. Accordingly, we tentatively posit that, when speakers with a tripartite, distance-oriented L1 use a bipartite L2, they maintain their proximal judgements while collapsing their medial and distal judgements into one. This is motivated by various studies which suggest that humans make a fundamental, two-way distinction between near (i.e., proximal) and far (i.e., medial/distal) space (cf. Coventry et al., 2008, p. 895; Kemmerer, 1999, p. 56).[2]

We would then expect these speakers to behave similarly to L1 English speakers for the items where both groups would use either the medial or distal forms in their respective L1s. On the other hand, they would differ for items where L1 tripartite speakers prefer the proximal while English speakers prefer the distal (or vice-versa):

Table 11: *Items where L1 English and L1 Japanese Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 English speakers	L1 Japanese speakers
4: Cake is sweet	Distal	Proximal
10: Sandwich was cheap		
11: Magazine is boring		
12: Wallet is big		
13: Phone is small		Proximal/Medial

[2] This may also explain why the vast majority of demonstrative systems cross-linguistically are bipartite (Kemmerer, 1999, p. 56).

Table 12: *Items where L1 English and L1 Thai Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 English speakers	L1 Thai speakers
4: Cake is sweet	Distal	Proximal
7: Necklace is pretty		
8: Drink is good		
10: Sandwich was cheap		
12: Wallet is big		
13: Phone is small		

Table 13: *Items where L1 English and L1 Spanish Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 English speakers	L1 Spanish speakers
4: Cake is sweet	Distal	Proximal
10: Sandwich was cheap		

We can formalise these suggestions as follows:

- Prediction 3: L1 Japanese speakers doing the task in English will use the proximal form more than L1 English speakers for items 4, 10, 11, and 12. They may also use the proximal more for item 13.
- Prediction 4: L1 Thai speakers doing the task in English will use the proximal form more than L1 English speakers for the same items as mentioned in prediction 3. They will also use the proximal more than L1 English speakers for items 7 and 8.
- Prediction 5: L1 Spanish speakers doing the task in English will use the proximal form more than L1 English speakers for items 4 and 10.

We now turn to L1 bipartite speakers doing the task in a tripartite L2. For our previous sections, we have either been able to draw clear parallels between a speaker's L1 and L2 due to a match between the inventories (i.e., L1 bipartite-L2 bipartite), or due to previous research on spatial perception (i.e., L1 tripartite-L2 bipartite). For L1 bipartite-L2 tripartite speakers, however, it is more difficult to make predictions around specific items; there is almost no published literature on how speakers restructure when presented with an additional demonstrative option, so this work is necessarily exploratory. Instead, we will describe some behavioural *trends* which may reflect how L1 English speakers behave in their tripartite L2, if demonstratives are a locus for CLI.

Firstly, as we saw in the bipartite-bipartite groups, we would expect speakers to maintain their most proximal (items 1 – 4) and distal (items 15 – 22) judgements at each end of the spectrum, as determined by Wilkins (2018). In terms of the "intermediate" space (items 5 - 14), the pattern most clearly indicating CLI would be English speakers exclusively using the distal form (i.e., simply mapping their English *this/that* judgements straight onto the L2 items). However, we doubt that this will be the case; a cursory examination of various Japanese, Thai, and Spanish reference grammars suggest that L2 speakers should have at least a basic awareness of the medial form (cf. Bowring and Laurie, 1992, p.35; Smyth, 2002, p. 35; Bradley & Mackenzie, 2004, p. 44, respectively). Instead, we tentatively posit that speakers may use the medial for *all* of these items – that is, not applying the nuance that we have seen L1 tripartite speakers show in their responses for intermediate items.

We will make this suggestion more specific, using the distances laid out by Wilkins' (2018) in his demonstrative questionnaire:

- Prediction 6: L1 English speakers doing the task in a tripartite L2 will maintain their L1 proximal judgements for items 1 – 4 and L1 distal judgements for items 15 – 22.[3] For items 5 – 14, where the object is in the intermediate distance (i.e., those which are within "easy access of the speaker)[4], they will prefer the medial form.

As suggested in prediction 3, we see that the most proximal and distal ends of the spectrum are uniform across both L1 English and L1 Japanese-L2 English, while the latter group skews proximal for many of the intermediate items.

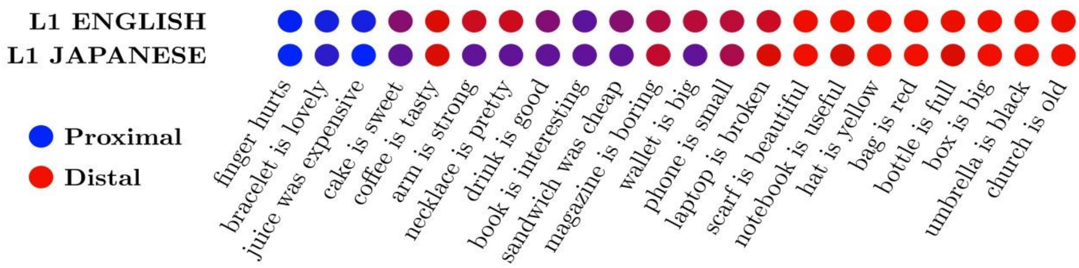


Figure 17: *English Demonstrative Choices of L1 English and L1 Japanese-L2 English Speakers, per Item*.

Table 14 lists the items where the two groups had a different top demonstrative choice:

[3] Wilkins (2018) categorised item 16 (*notebook is useful*) as "tens of metres away", but this was changed to be "within a few steps".
[4] Wilkins (2018) categorised item 17 (*hat is yellow*) as "within a few steps", but this was changed to be "tens of metres away".

Table 14: *Items where L1 English and L1 Japanese-L2 English Speakers Disagree on Their Top Demonstrative Choice.*

Item	L1 English speakers	L1 Japanese-L2 English speakers
4: Cake is sweet	Distal	Proximal
6: Arm is strong		
7: Necklace is pretty		
8: Drink is good		
10: Sandwich was cheap		
12: Wallet is big		
13: Phone is small		

In prediction 3, we identified four of these items (4, 10, 12, and 13) as ones where the two speaker groups might diverge in their top demonstrative choice if L1 Japanese speakers collapse their medial and distal judgements in English. Interestingly, however, there are an additional three times (6, 7, and 8) where the two groups diverge. We did not predict these items would be problematic, based on L1 behaviour:

Table 15: *Items where L1 English and L1 Japanese Speakers Disagree on Their Top Demonstrative Choice.*

Item	L1 English speakers	L1 Japanese speakers
6. Arm is strong	Distal	Medial
7. Necklace is pretty		
8. Drink is good		

This data suggests that, while we were correct in our assertion that Japanese speakers do combine their medial and distal judgements in English for almost all of the items where the two L1 groups diverge along these lines, Japanese speakers *also* combine proximal and medial judgements for items where the object is part of the addressee's body or in contact with it. Figure 18 shows the consistency with which they do so:

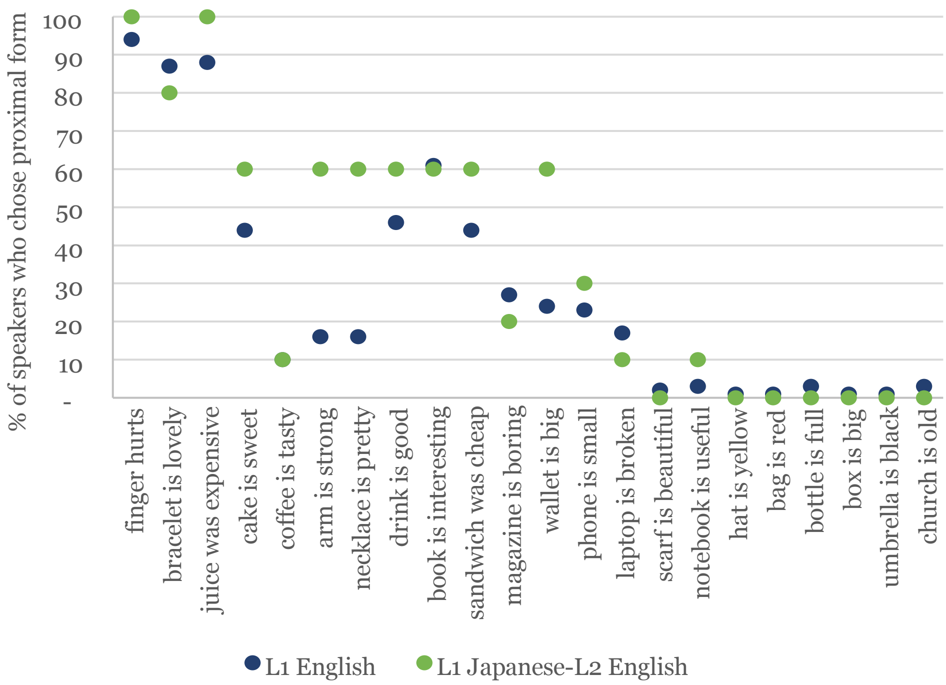


Figure 18: *% of L1 English and L1 Japanese-L2 English Speakers who Chose the Proximal Form in English, per Item.*

It seems, therefore, that there is evidence of CLI in L1 Japanese speakers' demonstrative usage in their L2. However, the manner in which their L1 affects their L2 is more complicated than we first thought.

Figure 19 shows that L1 Thai speakers doing the task in English (L1 Thai-L2 English speakers) use the proximal more than L1 English speakers for almost all of the items we identified in prediction 4. The only exception is item 5 (*coffee is tasty*).

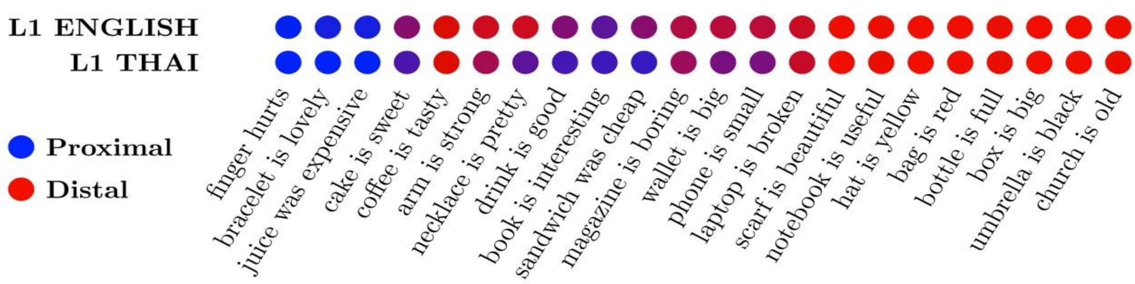


Figure 19: *English Demonstrative Choices of L1 English and L1 Thai-L2 English Speakers, per Item.*

Table 16 lists the items where the two groups diverge on their top demonstrative choice:

Table 16: *Items where L1 English and L1 Thai-L2 English Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 English speakers	L1 Thai-L2 English speakers
4: Cake is sweet	Distal	Proximal
7: Necklace is pretty		
8: Drink is good		
10: Sandwich was cheap		
12: Wallet is big		

These are five of the six items which we identified in prediction 4 as a sign of CLI, if Thai speakers are indeed collapsing their medial and distal judgements into one category, while preserving their proximal judgements.

Figure 20 also reveals that, for the only one of these six items where Thai and English speakers did *not* diverge on their top choice (13), the former group still had a much stronger preference for the proximal (49%) than the latter (23%). A chi-square test revealed this difference was significant ($p < 0.05$).

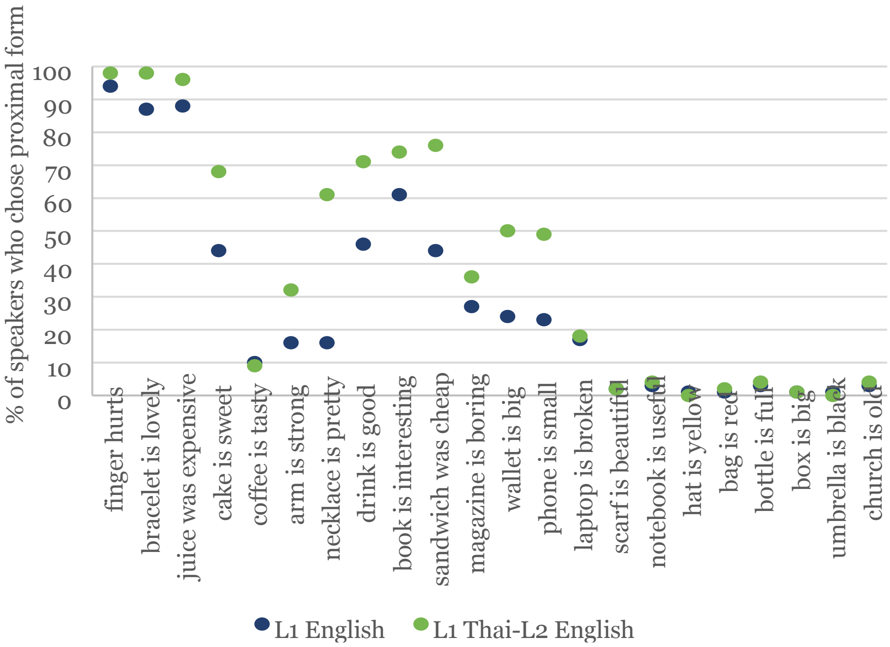


Figure 20: *% of L1 English and L1 Thai-L2 English Speakers who Chose the Proximal Form in English, per Item.*

These results support prediction 4 and – by extension – suggest that there is evidence of CLI in L1 Thai speakers' usage of English demonstratives.

As previously shown in Table 13, usage of the proximal form was very similar across both L1 English and L1 Spanish control groups. Therefore, if L1 Spanish-L2 English speakers maintain their

L1 proximal judgements in their L2 and collapse their medial and distal judgements into one (as suggested in prediction 5), we would expect them to use both English proximal and distal forms very similarly to the L1 English speaker group. Figure 21 supports this prediction:

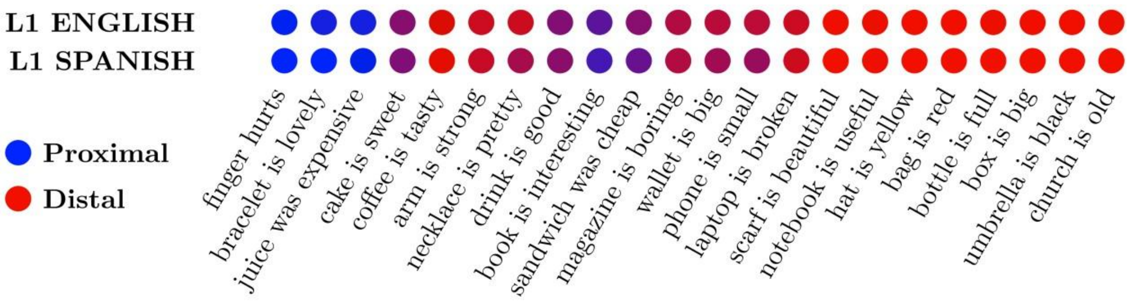


Figure 21: *English Demonstrative Choices of L1 English and L1 Spanish-L2 English Speakers, per Item.*

However, the two groups only disagree on their top choice for one of the two items we identified in prediction 5:

Table 17: *Items where L1 English and L1 Spanish-L2 English Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 English speakers	L1 Spanish-L2 English speakers
10: Sandwich was cheap	Distal	Proximal

Although the L1 English and L1 Spanish-L2 English groups had the same top choice of demonstrative for the other item (*cake is sweet*), a chi-square test revealed that there was not a significant difference between the L1 Spanish and L1 Spanish-L2 English speakers' demonstrative usage (p = 0.3) here. It seems, therefore, that prediction 5 is borne out. There is evidence of CLI in L1 Spanish speakers' usage of English demonstratives, as their use of the proximal in their L2 can be mapped onto their L1 behaviour.

Recall that, in prediction 6, we suggested that English speakers doing the task in Japanese (L2 Japanese speakers) may maintain their L1 proximal and most distal choices and use the medial indiscriminately for all intermediate items. However, Figure 22 shows this clearly is not the case.

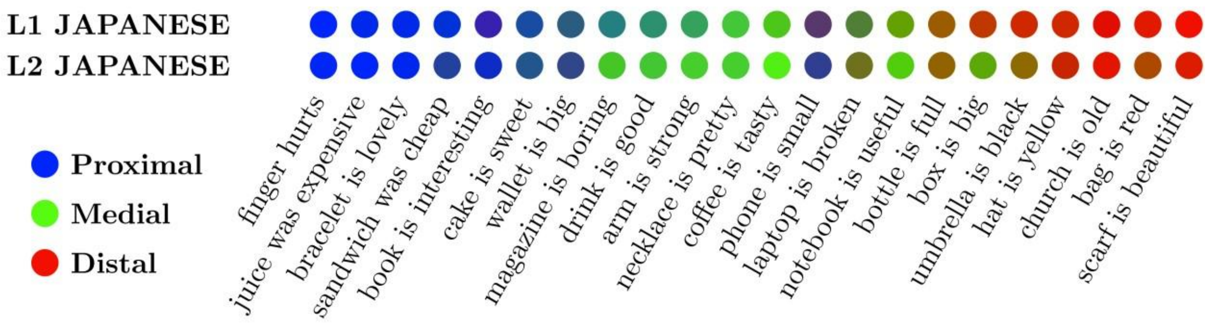


Figure 22: *Japanese Demonstrative Choices of L1 and L2 Japanese Speakers, per Item.*

Instead of only using the proximal for items 1-4 (like L1 English speakers doing the task in English), L2 Japanese speakers are also extending it to items 5-7. This matches L1 Japanese speakers' choices. The L2 Japanese speakers' demonstrative usage for the most distal items (16-22) also does not match that of L1 English speakers, as the former group uses the medial for 20 (*box is big*) where the latter used the distal. Finally, the L2 Japanese speakers clearly do not use the medial form indiscriminately for "intermediate distance" items – this is evidenced by 13 *(phone is small*), where both groups prefer the proximal form.

The two groups disagreed on the top demonstrative choice for only one item:

Table 18: *Items where L1 and L2 Japanese Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 Japanese speakers	L2 Japanese speakers
20: Box is big	Distal	Medial

As previously mentioned, we did not expect the L1 and L2 Japanese speakers to disagree on this item, if L2 Japanese speakers are carrying over distal judgements form their L1. However, when we break responses down by individual item and percentage of speakers who chose the proximal form, there is marginally more support for CLI.

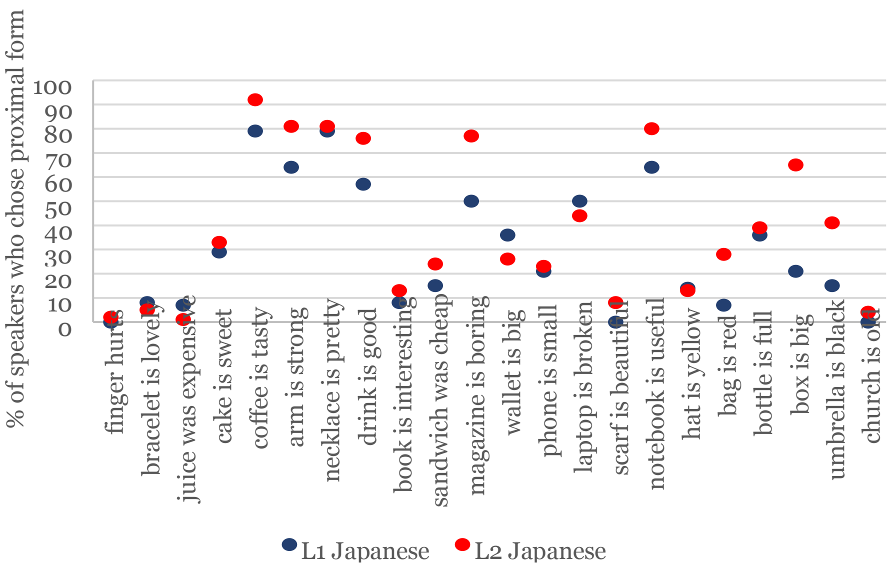


Figure 23: *% of L1 and L2 Japanese Speakers who Chose the Proximal Form in Japanese, per Item.*

Focusing on items 4 – 12, where L1 English speakers (doing the task in English) preferred the distal form, we see that fewer L2 Japanese speakers choose the proximal form than the L1 Japanese speakers for all but one item (*book is interesting*).

Breaking down responses by the percentage of speakers who chose the medial form also supports prediction 6: for the seven items which we identified as featuring objects in the intermediate distance, L2 Japanese speakers preferred the medial more strongly than L1 Japanese speakers for five of them. This suggests that at least some speakers are "defaulting" to the medial form when using a tripartite L2.

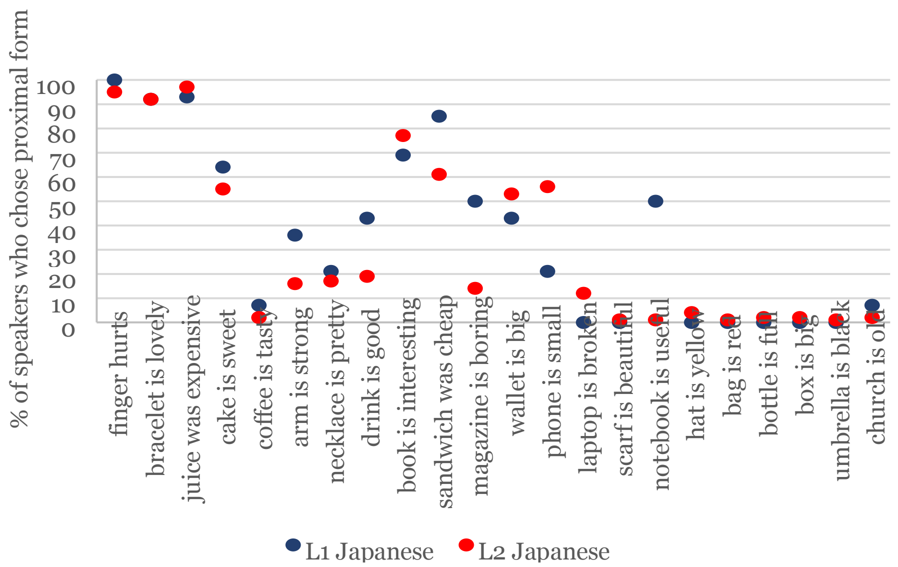


Figure 24: *% of L1 and L2 Japanese Speakers who Chose the Medial Form in Japanese, per Item.*

The data for L2 Japanese speakers, therefore, is not conclusive regarding demonstratives as a locus for CLI.

Looking at figure 25, it is striking how similarly L1 English speakers doing the task in Thai (L2 Thai speakers) behave in comparison to L1 Thai speakers. In particular, we can see that the gradients of each item match almost exactly for every item except one (*arm is strong*).

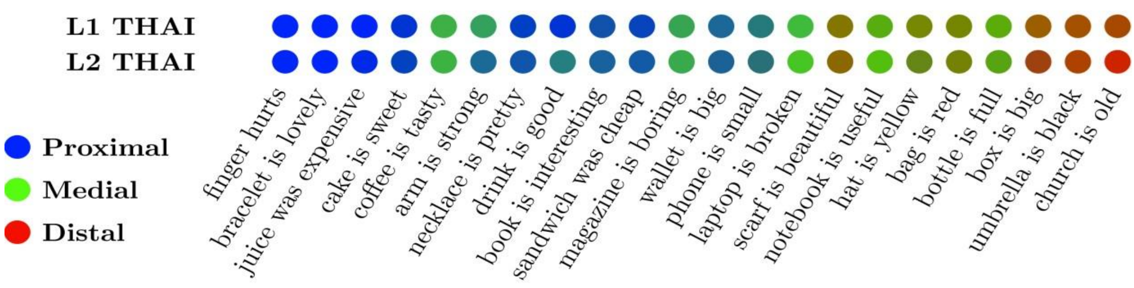


Figure 25: *Thai Demonstrative Choices of L1 and L2 Thai Speakers, per Item.*

Unsurprisingly, this is the only item where speakers disagreed with each other on their top choice:

Table 19: *Items where L1 and L2 Thai Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 Thai speakers	L2 Thai speakers
6: Arm is strong	Medial	Proximal

We expected that, if demonstratives are affected by CLI, L2 Thai speakers would actually agree with L1 Thai speakers in choosing the medial for item 6, as the object is in the "intermediate distance".

Additionally, we suggested that L2 Thai speakers would maintain the same distal judgements as in their L1 (i.e., by choosing the distal more than medial for items 15-22), but this is not the case for items 17-19. Instead, like the L1 Thai speakers, L2 Thai speakers prefer the medial. Finally, it is clear that L2 Thai speakers also know to use the proximal instead of medial for the same "middle-distance" items as L1 Thai speakers (9, 10, 12 and 13). This suggests that prediction 6 was incorrect and, by extension, CLI is not necessarily a factor here.

Figure 26 depicts how L1 English speakers doing the task in Spanish (L2 Spanish speakers) choose demonstratives, in comparison to L1 Spanish speakers. The fact that L2 Spanish speakers (i) do not have a clear medial preference for items 5 – 14 and (ii) do not have a clear distal preference for items 15-22 undermines prediction 6.

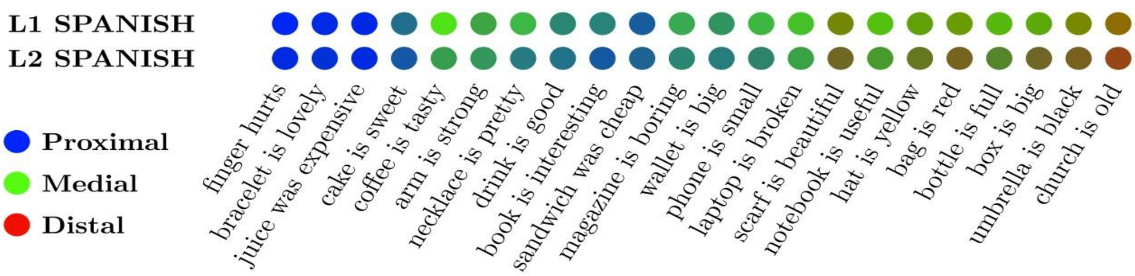


Figure 26: *Spanish Demonstrative Choices of L1 and L2 Spanish Speakers, per Item.*

There are seven items where L1 and L2 Spanish speakers disagree on their top choice of demonstrative. They are as follows:

Table 20: *Items where L1 and L2 Spanish Speakers Disagree on Their Top Demonstrative Choice*

Item	L1 Spanish speakers	L2 Spanish speakers
7: Necklace is pretty	Medial	Proximal
8: Drink is good		
9: Book is interesting		
15: Scarf is beautiful		Distal
18: Bag is red		
20: Box is big		
21: Umbrella is black		

This partly confirms prediction 6, which stated that L2 Spanish speakers would be expected to diverge on their top choice for items 15-22, because L2 Spanish speakers are transferring their L1 distal preference. We would expect the two groups to agree on items 7-9, however, if L2 Spanish speakers are indiscriminately using the medial form for intermediate items. Instead, we see L2 Spanish speakers preferring the proximal (which is not traceable to their L1).

Figure 27 emphasises the fact that our predictions around the medial form do not apply. L1 Spanish speakers are consistently using the medial form more than L2 Spanish speakers for the six of the seven items which we identified as featuring objects in the intermediate distance.

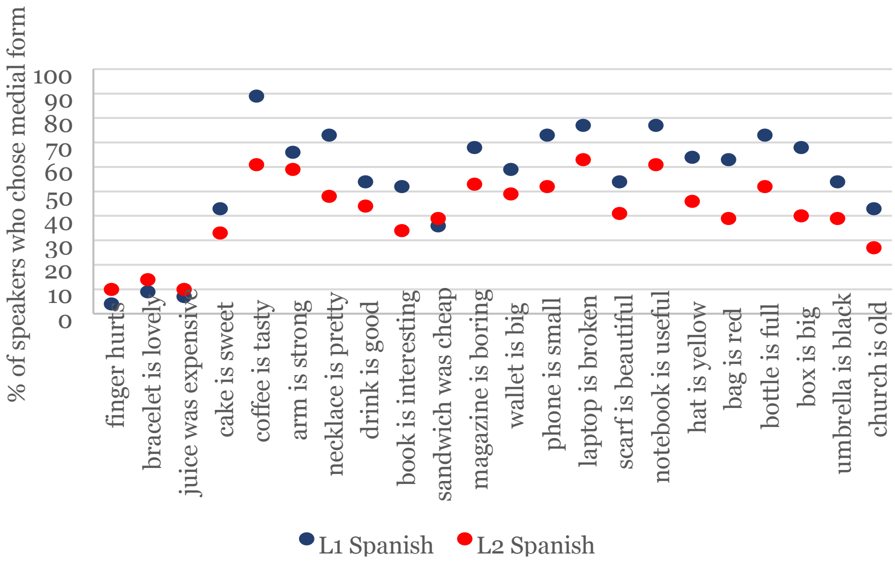


Figure 27: *% of L1 and L2 Spanish Speakers who Chose the Medial Form in Spanish, per Item.*

Much like both L2 Japanese and L2 Thai speakers, this evidence for CLI is therefore inconclusive.

4.4.3 Intragroup Homogeneity

When examining how similarly speakers of the same L1 behave in their L2, Jarvis and Pavlenko (2008, p. 48) note that participants should show “a level of uniformity in their use of the recipient language that is comparable to the level of uniformity they display in their use of the source language.” Accordingly, we compared the agreement rate for each L2 speaker group, averaged across items, to the agreement rate of speakers doing the task in their L1[5]. The results are as follows:

Table 21: *% of Speakers who Chose the “Top” Demonstrative Choice, Averaged across All Items*

Speaker’s L1	Task in L1	Task in L2 (English)
Chinese	80	81
Italian	84	83
Japanese	75	82
Thai	71	83
Spanish	69	84

Table 21 shows that, with the exception of the L1 Italian group, every group of speakers who did the task in their L2 (English) were *more* uniform in their demonstrative choices than those who did the task in their respective L1s. Although this is less meaningful for L1 speakers of tripartite languages, as agreement rates will be necessarily lower when split across three options, these results are still striking

[5] We were only able to compare intragroup homogeneity and intergroup heterogeneity between L2 English speakers, as there were no other groups to compare L1 English speakers against in their respective L2s (i.e., L1 English speakers were the only group who performed the task in Japanese as their L2).

when we take into account the fact that L1 English speakers' average rate of agreement when doing the task in English was only 79% – lower than any of the L2 speaker groups'. This suggests that the high level of homogeneity between L1 speakers of tripartite languages when doing the task in their L2 is not necessarily just because they are forced to be more categorical.

4.4.4 Intergroup Heterogeneity

In addition to showing that L2 speaker groups have *high* levels of intragroup homogeneity, Table 21 also shows that they have very *similar* levels of intragroup homogeneity to each other. Although this could raise the question of whether these speakers are behaving similarly to each other – thereby failing to show the intergroup heterogeneity required under Jarvis and Pavlenko's (2008) account of CLI – the evidence presented in preceding sections, repeated below for convenience and broken down by L1 inventory, refutes this concern.

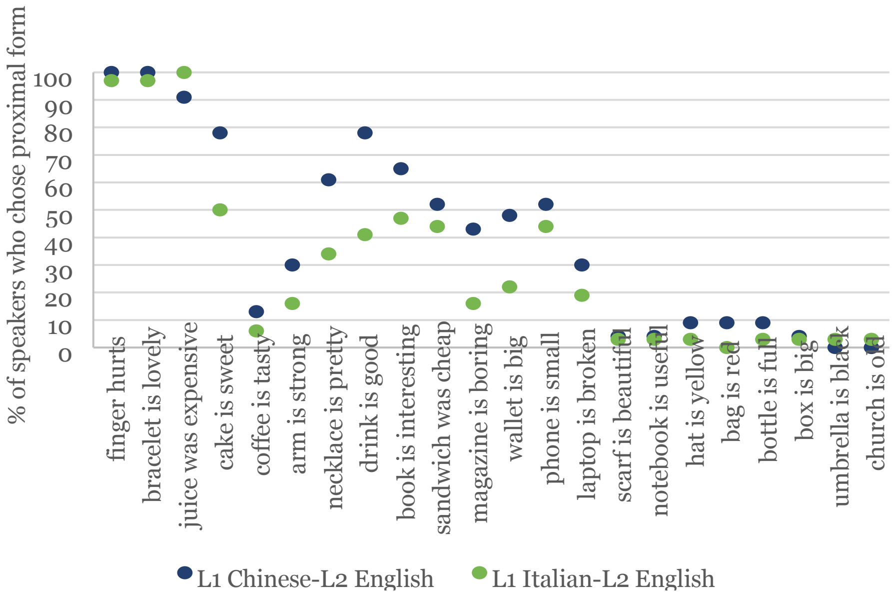


Figure 28: *% of L1 Chinese-L2 English and L1 Italian-L2 English Speakers who Chose the Proximal Form in English, per Item.*

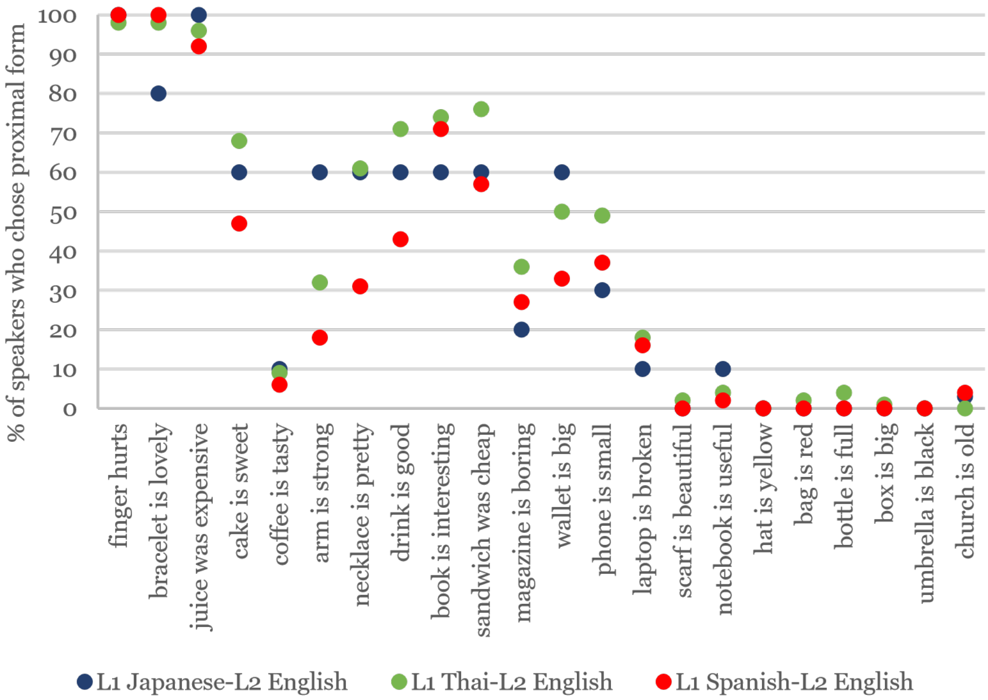


Figure 29: *% of L1 Japanese-L2 English, L1 Thai-L2 English and L1 Spanish-L2 English Speakers who Chose the Proximal Form in English, per Item.*

It is clear that L1 speaker groups are behaving differently from each other in their shared L2. Although we see a great deal of overlap in the most proximal and most distal forms, this was expected based on the L1 control groups' demonstrative usage. It is the middle-distance items where we can observe the unique patterns for each language. Looking at Figure 28, we can see L1 Chinese-L2 English speakers using the proximal much more than the L1 Italian-L2 English speakers for these items. Turning to Figure 29, L1 Thai-L2 English speakers use the proximal more than the other two groups for all but item 12 (*wallet is big*).

4.4.5 Proficiency

In Section 2, we mentioned that previous studies such as Athanasopoulos et al. (2007) and Malt and Sloman (2003) found an interaction between CLI and proficiency, with high proficiency L2 speakers showing less CLI than low-proficiency L2 speakers. Accordingly, to test for a main effect of proficiency (CEFR level A2 vs B1/2 vs C1/2), an ANOVA was conducted to predict the rate of matching with a single 3-level factor for proficiency. The data were aggregated by participant. The results show a main effect of proficiency ($F(2,917) = 19.01$, $p<0.001$). Pair-wise comparisons confirm significant differences between the A2 and B1/2 groups ($F(1,707) = 20.72$, $p<0.001$) and between the B1/2 and C1/2 groups ($F(1,685) = 7.155$, $p<0.008$).

5 Discussion

We set out to determine whether demonstratives are a locus for cross-linguistic influence (CLI), based on the three criteria outlined by Jarvis and Pavlenko (2008). Our results are mixed, but taken together provide convincing evidence that demonstratives are indeed a locus for CLI. Cross-linguistic performance congruity is the area with most variability in the results: while the majority of speaker groups showed a clear link between demonstrative usage in their L1 and L2 (i.e., L2 Chinese, L1 Chinese-L2 English, L1 Japanese-L2 English, L1 Thai-L2 English), others were inconclusive (i.e., L1 Italian-L2 English, L2 Japanese) and some did not meet our predictions at all (i.e., L2 Thai, L2 Spanish). Our tests for intragroup homogeneity and intergroup heterogeneity were more consistent, however, and suggested that participants with the same L1 behave similarly to each other in their shared L2, while also behaving differently to those with different L1s.

While this study is important in showing that demonstratives are a locus of CLI, many unanswered questions remain. Firstly, despite examining six languages with different demonstrative inventories, this study did not investigate all known demonstrative systems. For example, some languages distinguish between demonstratives based on the visibility of the object, which our study did not address (Wilkins, 2018, p. 43). The same applies to one-term systems such as German (Diessel, 2013). Additionally, the study was limited to exophoric demonstrative usage. As noted in section 3, this is just one of many possible ways demonstratives are used (Levinson 2018, p. 11). Our survey also did not control for factors such as type of L2 input and educational background, both of which Jarvis (2000, p. 260) identifies as having a significant effect on CLI. Future work should also aim to determine whether the CLI involved is semantic or conceptual. Jarvis (2016) has outlined a variety of study designs to delineate the two.

6 Conclusion

This dissertation aimed to determine whether demonstratives are a locus for CLI. Accordingly, we first provided an overview of previous work on both CLI and demonstratives, motivating our research by highlighting the latter's similarity to prepositions – another part of speech which is subject to CLI – and emphasising the lack of existing literature on the subject. After providing an overview of the demonstrative systems for six languages relevant to our study, we summarised the methodology we devised to address this gap in the research. In particular, we focused on how it enabled us to gather crucial evidence pertaining to CLI as outlined by Jarvis and Pavlenko (2008). We broke down the results of our study by language and showed that CLI was present for multiple speaker groups. Although not definitive, the data suggest that demonstratives are indeed a locus for CLI, and provide a solid foundation upon which future research can build.

7 References

Alonso, R. A., Cadierno, T., and Jarvis, S. (2016). *Crosslinguistic Influence in the Acquisition of Spatial Prepositions in English as a Foreign Language*. In R. Alonso Alonso (Ed.), *Crosslinguistic Influence in Second Language Acquisition* (pp. 93-120). Bristol: Blue Ridge Summit: Multilingual Matters.

Appel, R. (2010). The Lexicon in Second Language Acquisition. In P. Jordens and J. Lalleman (Eds.), *Investigating Second Language Acquisition.* Berlin, New York: De Gruyter Mouton.

Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J., and Thierry, G. (2010). Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition*, 116, 437-443.

Axtell, R. E. (1998). *Gestures*. New York: John Wiley & Sons, Inc.

Boroditsky, L. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology*, 43, 1-22.

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

Bowring, R. J., and Laurie, H. U. (1992). *An Introduction to Modern Japanese, Book 1: Grammar Lessons.* Cambridge: Cambridge University Press.

Bradley, P. T. and Mackenzie, I. E. (2004). *Spanish: An Essential Grammar*. London: Routledge.

Charney, R. (1977). The comprehension of "here" and "there". *Journal of Child Language*, 6, 69-80.

Chen, R. (1990). English Demonstratives: A Case of Semantic Expansion, *Language Sciences*, 12(2/3), 139-153.

Choi, S. (2006). Influence of language-specific input on spatial cognition: Categories of containment. *First Language*, 26(2), 207-232.

Chu, C.-Y., and Minai, U. (2018). Children's demonstrative comprehension and the role of non-linguistic cognitive abilities: A cross-linguistic study. *Journal of Psycholinguistic Research,* 47(6), 1343–1368

Clark, E. (1978). From gesture to word: On the natural history of deixis in language acquisition. *Human Growth and Development*, 85-120.

Clark, E. V., and Sengul, C. J. (1977). Strategies in the acquisition of deixis. *Journal of Child Language*, 5, 457-475.

Conventry, K. R., Valdés, B., Castillo, A., and Guijarro-Fuentes, P. (2008). Language within your reach: Near–far perceptual space and spatial demonstratives. *Cognition*, 108, 889-895.

Cook, V., Bassetti, B., Kasai, C., Sasaki, M., and Takahashi, J. A. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism*, 10(2), 137-152.

Council of Europe (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment – Companion Volume*. Strasbourg: Council of Europe Publishing.

De Villiers, P. and De Villiers, J. G. (1974). On This, That and the Other: Nonegocentrism in Very Young Children. *Journal of Experimental Child Psychology*, 18, 438-447.

Diessel, H. (1999). *Demonstratives: Form, Function and Cognition.* Amsterdam: John Benjamins Publishing Company.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, 17(4), 463-489.

Diessel, H. (2014). Demonstratives, Frames of Reference, and Semantic Universals. *Language and Linguistic Compass,* 8(3), 116-132.

Directorate-General for Communication (2005), Europeans and their Languages. Retrieved from *European Data*: <https://data.europa.eu/data/datasets/s518_64_3_ebs243?locale=en>.

Dixon, R. M. W. (2003). Demonstratives. *Studies in Language*, 27(1), 61-112.

Eckman, F. R., Elreyes, A., and Iverson, G. K. (2003). Some principles of second language phonology. *Second Language Research*, 19(3), 169-208.

Graham, C. R. and Belnap, R. K. (1986). The Acquisition of Lexical Boundaries in English by Native Speakers of Spanish. *International Review of Applied Linguistics in Language Teaching*, 24(4), 275-286.

Grimes, B. F. (Ed.). (2000). Ethnologue: Languages of the World (14th Ed.). Dallas, Texas: SIL International.

Halle, P. A., Chang, Y., and Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32, 395-421.

Hawkins, R. and Liszka, S. A. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In R.van Hout, A. Hulk, F. Kuiken, and R. J. Towell. (Eds.), *The Lexicon–Syntax Interface in Second Language Acquisition* (pp. 21-44). Amsterdam: John Benjamins.

Hoji, H., Kinsui, S., Takubo, Y., and Ueyama, A. (2003). *The demonstratives in modern Japanese*. In A. Simpson, A. L. Yen-hui (Eds.), *Functional Structure(s), Form and Interpretation* (pp. 97-128). London: Routledge.

Iwasaki, S. and Dechapratumwan, P. (2021). Creating versatility in Thai demonstratives. *Studies in Language*. Amsterdam: John Benjamins.

James, A. (2010). Second language phonology. In P. Jordens and J. Lalleman (Eds.), *Investigating Second Language Acquisition* (pp. 293-320). Berlin, New York: De Gruyter Mouton.

Jarvis, S. (2000). Methodological Rigor in the Study of Transfer: Identifying L1 Influence in the Interlanguage Lexicon. *Language Learning*, 50(2), 245-309.

Jarvis, S. (2011). Conceptual transfer: Crosslinguistic effects in categorization and construal. *Bilingualism: Language and Cognition*, 14(1), 1-8.

Jarvis, S. (2016). Clarifying the Scope of Conceptual Transfer. *Language Learning*, 66(3), 608-635.

Jarvis, S. and Odlin, T. (2000). Morphological Type, Spatial Reference, and Language Transfer. *Studies in Second Language Acquisition*, 22(4), 535-556.

Jarvis, S. and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition.* London: Routledge.

Jiang, N., Novokshanova, E., Masuda, K., and Wang, X. (2011). Morphological Congruency and the Acquisition of L2 Morphemes. *Language Learning*, 61(3), 940-967.

Kemmerer, D. (1999). “Near” and “Far” in Language and Perception. *Cognition*, 73, 35-63.

Kimball, G. (1991). *Koasati Grammar*. Lincoln, Nebraska: University of Nebraska Press.

Kuno, Susumo. (1973). *The Structure of the Japanese Language.* Cambridge, Massachusetts: Massachusetts Institute of Technology Press.

Lemhöfer, K., Spalek, K., and Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, 59, 312-330.

Levinson, S. (2018). Introduction: Demonstratives: Patterns in Diversity. In S. Levinson, S. Cutfield, M. Dunn, N. Enfield, & S. Meira (Eds.), *Demonstratives in Cross-Linguistic Perspective, Language Culture and Cognition* (pp.1-42). Cambridge University Press.

Levinson, S. (2004). Deixis. In Horn, L. R. and Ward, G. (Eds.) *Handbook of Pragmatics* (pp. 97-121). Blackwell Publishing.

Levinson, S. C., & Enfield, N. J. (Eds.). (1996). *Manual for the field season 1996*. Nijmegen: Max Planck Institute for Psycholinguistics.

Li, J. and Cai, J. (2015). L1 transfer in Chinese learners’ use of spatial prepositions in EFL. In L. Yu and T. Odlin (Eds.), *New Perspectives on Transfer in Second Language Learning* (pp. 63-75). Bristol, Blue Ridge Summit: Multilingual Matters.

Maiden, M. and Robustelli, C. (2013). *A Reference Grammar of Modern Italian* (2nd ed.). London: Routledge.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., and Levinson, S. C. (2004). Can language restructure cognition? The case for space. *TRENDS in Cognitive Sciences*, 8(3), 108-114.

Major, R. C. (2008). Transfer in Second Language Phonology. In J. G. H. Edwards and M. L. Zampini (Eds.) *Phonology and Second Language Acquisition* (pp. 63-945). Amsterdam: John Benjamins Publishing Company.

Malt, B. C. and Sloman, S. A. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition*, 6(1), 47-67.

Matsuoka, H. (2000). *Shokyuu o oshieru hito no tame no nihonngo bunpo Handobukku*. Tokyo: Suruee Nettowaaku Co. Ltd.

McDonough, L., Choi, S., and Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, 46, 229-259.

Montrul, S. (2018). Heritage language development: Connecting the dots. *International Journal of Bilingualism*, 22(5), 530-546.

Munnich, E., and Landau, B. (2010). Developmental Decline in the Acquisition of Spatial Language. *Language Learning and Development*, 6(1), 32-59.

Nakagawa, S. W. (2002). Cross-Cultural Practices: A Comparison of Demonstrative Pronouns in Japanese and Thai. *NUCB JLCC*, 4(1), 21-31.

Niimura, T., and Hayashi, B. (1996). Contrastive Analysis of English and Japanese Demonstratives from the perspective of L1 and L2 Acquisition. *Language Sciences*, 18(3/4), 811-834.

Odlin, T. (2003). Cross-linguistic Influence. In Doughty, C. J. and Long, M. H. (Eds.), *The Handbook of Second Language Acquisition* (pp. 436-486). Blackwell's Publishing.

Ono, K. (1997). Territories of Information and Japanese Demonstratives. *The Journal of the Association of Teachers of Japanese*, 28(2), 131-155.

Oshima, D. Y., and McCready, E. (2017). Anaphoric demonstratives and mutual knowledge: The cases of Japanese and English. *Natural Language & Linguistic Theory*, 35(3), 801-837.

Park, H. and Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean–English bilinguals. *Bilingualism: Language and Cognition*, 17(2), 410-430.

Pederson, E., & Wilkins, D. (1996). *A cross-linguistic questionnaire on 'demonstratives'*. In S. C. Levinson (Ed.), *Manual for the 1996 Field Season* (pp. 1-11). Nijmegen: Max Planck Institute for Psycholinguistics.

Rocca, R. Tylén, K. and Wallentin, M. (2019). This shoe, that tiger: Semantic properties reflecting manual affordances of the referent modulate demonstrative use. *PLOS ONE*, 14(1), e0210333.

Ruben, R. J. (1997). A Time Frame of Critical/Sensitive Periods of Language Development. *Acta Oto-Laryngologica*, 117(2), 202-205.

Stowe, L. A. and de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 22(1), 1-29.

Sakuma, K. (1936). Daimeishi no mondai. *Gendai Nihongo no hyoogen to gohoo (pp. 33-36)*. Tokyo: Kooseikaku.

Smyth, D. (2002). *Thai: An Essential Grammar*. London: Routledge.

University of Cambridge (2012). *Key English Test*.

Webb, P. A. and Abrahamson, A. A. (1975). Stages of egocentrism in children's use of 'this' and 'that': a different point of view. *Journal of Child Language*, 3, 349-367.

Wilkins, D. (2018). The Demonstrative Questionnaire: "THIS" and "THAT" in Comparative Perspective. In S. Levinson, S. Cutfield, M. Dunn, N. Enfield, & S. Meira (Eds.), *Demonstratives in Cross-Linguistic Perspective, Language Culture and Cognition* (pp. 43-71). Cambridge: Cambridge University Press.

Winawer, J., Witthoft, N., Wulund, L., and Boroditsky, L. (2010). Effects of language on color discriminability. *Journal of Vision*, 3. 711-711.

Wu, H. (2004). *Spatial Demonstratives in English and Chinese*. Amsterdam: John Benjamins Publishing Company.

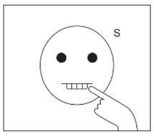
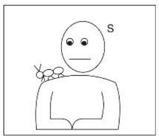
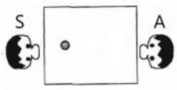
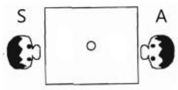
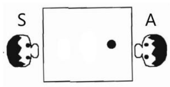
Yoshimoto, K. (1986). On Demonstratives KO/SO/A in Japanese. *Genko Kenkyu*, 90, 48-72.

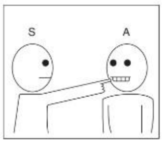
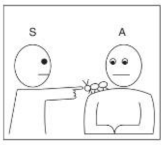
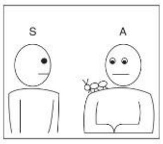
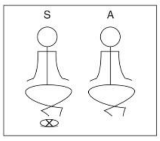
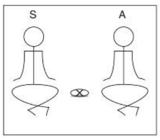
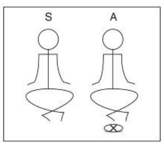
Zhang, Y. (2013). *Spatial Representation of Topological Concepts IN and ON: A Comparative Study of English and Mandarin Chinese* [Unpublished doctoral dissertation]. Concordia University.

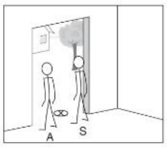
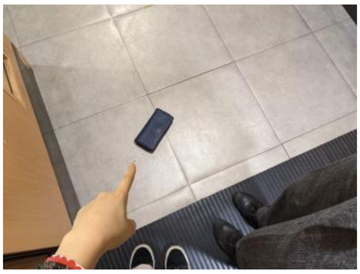
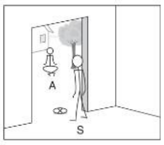
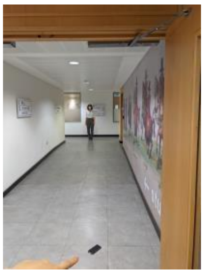
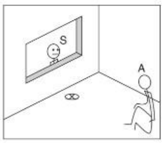
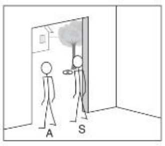
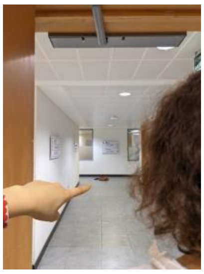
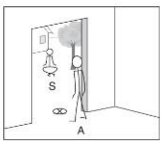
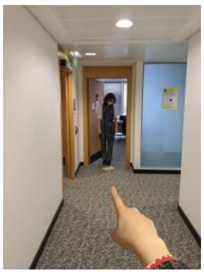
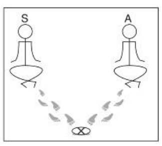
8 Appendices

8.1 Appendix A

Table 22: *Target Photo-Sentence Pairs*

Item	Scene	Photo	Prompt – this/that…
1			finger hurts
2			bracelet is lovely
3			juice was expensive
4			cake is sweet
5			coffee is tasty

6			arm is strong
7			necklace is pretty
8			drink is good
9			book is interesting
10			sandwich was cheap
11			magazine is boring

12			wallet is big
13			phone is small
14			laptop is broken
15			scarf is beautiful
16			notebook is useful
17			hat is yellow

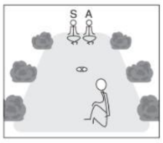
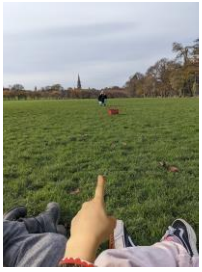
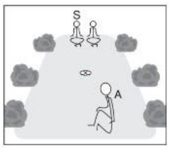
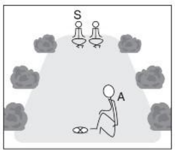
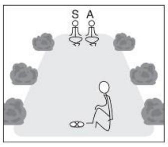
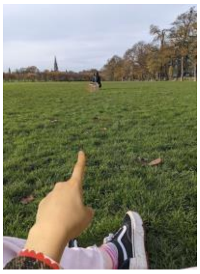
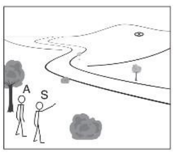
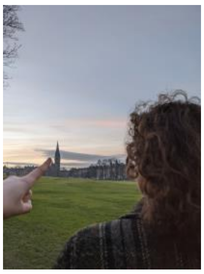
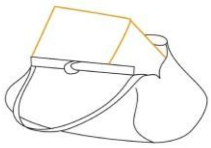
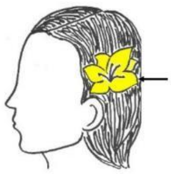
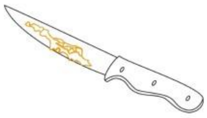
18			bag is red
19			bottle is full
20			box is big
21			umbrella is black
22			church is old

Table 23: *Filler Photo-Sentence Pairs*

Item	Scene	Photo	Title
1			The box is (in/on) the bag
2			The flower is (in/on) your hair
3			The sweets are (in/on) the plate
4			The ring is (in/on) my finger
5			The toy is (in/on) the bowl

6			The jam is (in/on) the knife
7			The crack is (in/on) the cup
8			The cup is (in/on) the stool

8.2 Appendix B

L1 English speakers

Due to discrepancies between the number of L1 English speakers with a working knowledge of the different target L2s (Europeans and their Languages, 2005), the form was weighted to prioritize some L2s over others if a participant indicated knowledge of both. This ensured that bilingual English participants were equally distributed among all L2 conditions.

The order of languages – from least to most widely spoken by L1 English participants – was set as follows:

Thai → Italian → Japanese → Chinese → Spanish

For example: if a participant indicated that they knew both B2 Italian and C1 Spanish, they would be assigned the Italian condition.

L2 English speakers

Because English is spoken by over three billion people worldwide (Ethnologue, 2022), the experiment was weighted so that L1 speakers of Chinese, Thai, Japanese, Spanish, and Italian who met the minimum requirements for the English condition (i.e., rated themselves at CEFR level A2 and above) were randomly assigned a condition either in English or their L1. This ensured that a sufficient number of non-L1 English participants ($N = 281$) performed the experiment in their L1 (the control task).

Other L1

If the participant indicated that their L1 was not listed, they were asked to indicate their L1 and their proficiency in English. Individuals whose English proficiency was reported as level A2 or above were then assigned the English survey to complete. If not, they were directed to end the task.

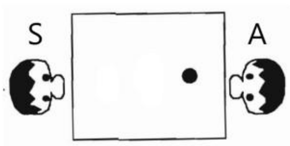
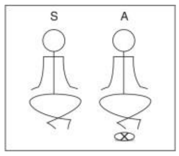
8.3 Appendix C

Japanese as a Distance- or Person-Oriented System

In order to determine whether the Japanese demonstrative system is person- or distance-oriented – or a combination of the two – we identified two pairs of items in the study where the location of the object is stable in relation to the speaker, but the location of the addressee varies.

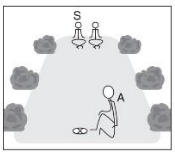
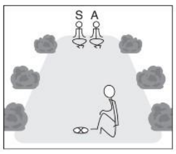
For the first pair, the object is closer to the addressee in both items, but still within easy reaching distance of the speaker.

Table 24: *Items where the Location of the Object Differs in Relation to the Addressee, But Not the Speaker*

Tairitsu (speaker and addressee are opposite each other)	**Yugo** (speaker and addressee are next to each other)
 4: This/that coffee is tasty.	 10: This/that magazine is boring.

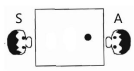
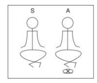
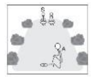
For the second pair, the object is closer to the addressee in one item, and equally distant from both interlocutors in the other – but very far away from the speaker in both.

Table 25: *Questions where the Location of the Object Differs in Relation to the Addressee, But Not the Speaker*

Tairitsu (speaker and addressee are opposite each other)	**Yugo** (speaker and addressee are next to each other)
 Item 20: This/that box is big.	 Item 21: This/that umbrella is black.

By manipulating the position of the speaker and addressee, we can determine whether the addressee's location has an effect on L1 Japanese speakers' demonstrative choice – all three accounts have differing predictions across the question pairs:

Table 26: *Predicted Demonstrative Preferences, per Anchoring System.*

	Item 4 	**Item 10** 	**Item 20** 	**Item 21** 
Person-oriented	Medial	Medial	Medial	Distal
Distance-oriented	Proximal/Medial	Proximal/Medial	Distal	Distal
Hybrid	Medial	Proximal/Medial	Medial	Distal

If Japanese has a person-oriented system, we would expect to see L1 speakers using the medial form for both 10 and 4, as the object is closer to the addressee than speaker. On the other hand, participants should use the distal form more frequently for 21 (where the speaker and addressee are sitting next to each other, equally far away from the object), and the medial form for 20 (where the addressee is closer to the object, facing the speaker).

On the other hand, if Japanese is distance-oriented, we would speakers to prefer using either the same proximal or medial form across 10 and 4; a priori, it is difficult to predict which form speakers would prefer. We would also not expect there to be a difference in demonstrative usage across 21 and 20, because the position and relative distance of the object from the speaker is identical.

Finally, if Japanese uses a hybrid distance- and person-oriented system (as suggested by Matsuoka, 2000), we would expect to see a combination of results across the four questions. For 10 and 21, where the speaker and addressee are next to each other, speakers should follow the distance-oriented pattern of proximal/medial and distal, respectively. For 4 and 20, however, the speaker should follow the person-oriented pattern and prefer the medial form across both questions, because they are positioned opposite the addressee.

Results from the four items are given below:

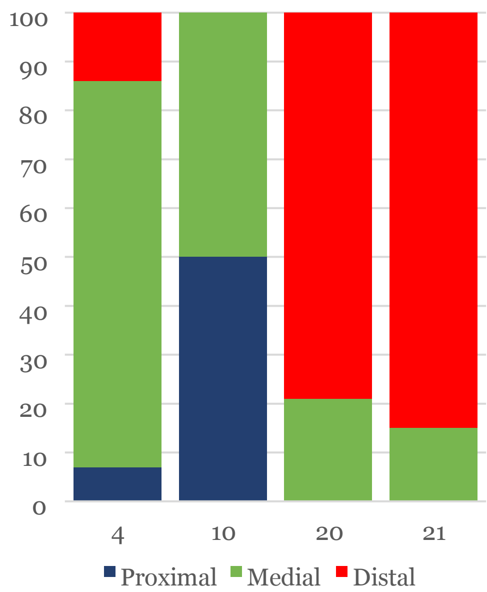


Figure 30: *Demonstrative Choices (%) of L1 Japanese Speakers, per Item.*

Table 27: *Demonstrative Choices (%) of L1 Japanese Speakers.*

Item	Proximal	Medial	Distal
4	7	79	14
10	50	50	0
20	0	21	79
21	0	15	85

While L1 Japanese speakers are split evenly between proximal and medial demonstratives in item 10, they have a strong preference for the medial form in item 4; a chi-squared test revealed this difference in chosen form between 4 and 10 was significant ($p = 0.02$). The fact that 50% of speakers chose the proximal form for item 10 – despite the object clearly being closer to the addressee – cannot be explained by the person-oriented and hybrid models. However, the significant difference in behaviour between the two items is also not explicitly predicted by the distance-oriented account, as the objects are the same distance away from the speaker. Assuming the distance-oriented account is correct, the discrepancies between the two questions could be explained by the location of the scene – while 10 takes place outdoors, 4 is situated indoors, meaning speakers' perceptions of "near" and "less near" may have been affected accordingly.

Turning to items 20 and 21, we see that L1 Japanese speakers overwhelmingly prefer to use the distal form (79% and 85%, respectively) over the medial form in both questions.

Although the medial form is used slightly more in 20 than the 21, a chi-squared test revealed that these differences in chosen form for items 20 and 21 were not significant ($p = 0.68$). As the person-oriented and hybrid accounts predict a strong preference for the medial form in 20, the results suggest that, for these two questions, Japanese speakers select demonstratives regardless of addressee location and orientation.

Thai as a Distance- or Person-Oriented System.

To determine whether Thai is a distance- or person-oriented system (or hybrid, although there is no suggestion of this in the published literature), we selected the same questions as those used in the previous section with Japanese.

Results from the four items are given below:

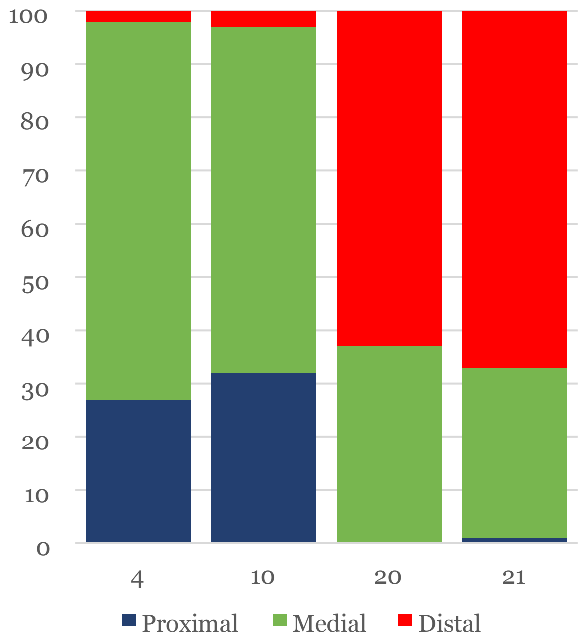


Figure 31: *Demonstrative Choices (%) of L1 Thai Speakers.*

Table 28: *Demonstrative Choices (%) of L1 Thai Speakers.*

Item	Proximal	Medial	Distal
4	27	71	3
10	32	65	3
20	0	37	63
21	1	32	67

Unlike Japanese speakers, Thai speakers are consistent in their preferences for the medial form in both 10 and 4; although their preference for the proximal form is slightly higher in 10 than 4, a chi-squared test revealed this difference was not significant ($p = 0.2$). As discussed in the previous section, while a strong proximal preference would suggest that Thai is a distance-oriented systems all three accounts can account for a medial preference (either because the objects in 10 and 4 are neither near nor far from the speaker, or because they are closer to the addressee).

On the other hand, the results for questions 21 and 20 support the distance-oriented account much more definitively. L1 Thai speakers prefer the distal over the medial form in both questions – although not as strongly as Japanese speakers do – and their behaviour is not significantly different across the two items, despite the addressee's change in position ($p = 0.4$). This suggests that, similar to Japanese speakers, Thai speakers' demonstrative usage is determined not by the position of the object in relation to the addressee, but instead by its position in relation to the speaker.

Acknowledgements

This dissertation would have been impossible without the help of many, many people.

First and foremost, to my incredible supervisor Dr Hannah Rhode, whose feedback and support was invaluable from the very start – thank you for helping me to refine a loose collection of tangents into a (hopefully!) cohesive dissertation. I am also indebted to Dr Rob Truswell, for encouraging me to follow this topic all the way back in March 2021, and Dr Linda van Bergen, for being such a wonderful Personal Tutor throughout my four years at the University of Edinburgh.

To Freyja Steindórsdottir, Orestis Michaelides, Rachel Noach, Pramay Rai, and the countless others who kept me company on the library mezzanine while I wrote all 10,000 words, you made a seemingly Sisyphean task possible – thank you for making my university experience what it was.

To my family, thank you for everything (obviously). Rorie, your midnight facetimes never fail to cheer me up. Dad, I appreciate your unerring faith in me, in all of my endeavours. And Mum, you're my inspiration – not only for academia, but everything else.

To Taiga Taguchi, Jose Neira, Yuning Liu, and Ponrawee Prasertsom – thank you for patiently answering my (many) questions and providing endless grammaticality judgements.

And finally, to all 2,189 friends, family members, and kind strangers who took the time to answer my survey – thank you, 谢谢, grazie, ありがとうございます, ขอบคุณ, gracias.

This dissertation is dedicated to my wonderful grandparents – Nong and Keith Branigan, and Charlotte and Douglas Boutell.

Passives in Extended Dynamic Antisymmetry

Evelyn R. Burrows
University of Cambridge

Abstract. This dissertation proposes a new derivation for English passives in accordance with Extended Dynamic Antisymmetry (EDA) (Moro & Roberts, 2021) and Problems of Projection (Chomsky, 2013; 2015). In these theories, movement is triggered by the labelling requirement in symmetric (XP YP or X Y) configurations. Such approaches to movement are arguably theoretically superior to traditional feature-driven approaches, but their empirical coverage is currently minimal. I apply EDA to passives, arguing that the passive derives from the same underlying structure as the active, but that they differ as to which constituent moves to resolve the symmetric structure. This translates den Dikken's (2020) approach to actives and passives as canonical and reverse predication structures into an EDA framework, as predication structures are represented as symmetric structures in EDA. To analyse passives in this way, it is necessary to consider the *by*-phrase an argument as in Collins (2005) as opposed to an adjunct. I argue that binding and control data support this hypothesis. I suggest two derivations for English passives. These make use of atomisation (Fowlie, 2013) – a process transforming an XP into a syntactic atom at Spellout – and smuggling (Collins, 2005) – an operation enabling a constituent to move over an intervener – respectively. I extend these analyses to passives of ditransitives, i.e., the passive counterparts to IO and *to* constructions. The smuggling approach appears superior as the atomisation approach violates RM, however, I believe that atomisation has promising applications to EDA, as it effectively turns an XP into an X.

Keywords: syntax; movement; passive; Dynamic Antisymmetry; Problems of Projection

1 Introduction

The passive construction[6] has various definitions and cross-linguistic instantiations. I focus on the passive construction in English, the operation whereby the patient is promoted to surface subject, the agent is absent from the surface structure or realised as an optional *by*-phrase, and the verb takes passive morphology. I analyse the passive construction as deriving from the same underlying structure as the active. I argue that movement arises from symmetry-breaking according to the Extended Dynamic Antisymmetry (EDA) framework (Moro & Roberts, 2021), and that actives and passives differ in which constituent moves to resolve the symmetric structure. My aim is to analyse passive constructions more elegantly, according to the minimalist goals of theoretical elegance and explanatory value, without neglecting empirical coverage and accuracy. I do not investigate passive constructions in languages other than English, however, I consider evidence concerning the structure of passives, and particularly of the *by*-phrase.

First, I discuss the evidence in favour of approaching movement as a symmetry-breaking phenomenon, specifically to resolve labelling failures as in Problems of Projection (PoP) (Chomsky, 2013; 2015) and EDA (Moro & Roberts, 2021). Second, I summarise previous analyses of the passive, arguing that there are strong theoretical advantages to analysing *by*-phrases as arguments. Third, I discuss empirical evidence for the structure of *by*-phrases, arguing that such evidence largely aligns with the theory. Fourth, I combine these insights on passives with the theory of EDA, suggesting two

[6] Special thanks go to Ian Roberts for his invaluable comments while supervising this dissertation. Any remaining errors are my own.

possible derivations for English passives. These make use of atomisation (Fowlie, 2013) in a new way, as well as smuggling. I extend these approaches to passives of ditransitives, i.e., the passive counterparts to IO and *to* constructions. I conclude by discussing theoretical and empirical consequences of the approaches, including possible concerns and extensions.

2 Extended Dynamic Antisymmetry

2.1 Does movement need to be triggered?

Many minimalist theories still struggle with movement. The Minimalist Program replaced the operation Move with Internal Merge (IM)[7]. As IM and External Merge (EM) are the same operation, they should operate on the same principles; IM should only have a trigger if EM does, and their triggers should be similar. However, it is still standard for features to trigger IM.

Whether EM is triggered is debated, as summarised by Fukui (2011). The minimalist guideline that every operation must be justified motivates approaches arguing that Merge must be triggered. In such approaches, many ungrammatical structures are simply not generated as the triggers restrict what is merged. In approaches where Merge is not triggered, the narrow syntax is simplified, instead, the interfaces do most of the work of ruling out impossible structures (although some restrictions remain, for instance, those involving locality). The preferred theoretical solution is still debated, and beyond the scope of this dissertation. I assume that Merge, and consequently IM, are free to apply, but must apply when there is a trigger (e.g., Fukui (2011); Collins & Seely (2020)).

Thus, IM and EM must function analogously. In many current approaches, EM is free[8] and IM must be triggered, violating this requirement. The idea that movement is more complex than EM or needs more justification is an outdated relic from earlier frameworks[9]. Furthermore, if both IM and EM can involve triggers, these triggers must be similar. I argue that symmetry-breaking as a trigger for IM is a desirable candidate to fulfil this requirement.

2.2 Unifying movement triggers

Movement triggers form a heterogenous class, with different triggers proposed in different approaches. Standardly, movement occurs to pick up or check features and affixes, e.g., Case features. Movement might also be triggered by symmetry-breaking, and/or linearisation and labelling. A related classification is that of base- versus target-driven movement (Bošković, 2020). In base-driven movement, the trigger is the element which undergoes movement, or something about the configuration in which it has been merged, such as cases of symmetry breaking. Target-driven movement, however, is movement triggered by an element in or associated with the position where the moved element will merge. This includes feature-driven movement. Although both types may be compatible, base-driven movement seems preferable in minimalism, which builds the syntax bottom-up, in terms of issues like cyclicity and the extent to which Merge can access information about merged structure. However, target-driven phenomena are fairly standard, including Agree, and the extent to which Merge can view previously merged structure is debated (e.g., Fukui (2011).

Symmetry-breaking as a movement trigger has much greater similarity to possible triggers for EM than feature-based triggers do. Though it is unclear how to measure such similarity, feature-based

[7] I use the terms 'IM' and 'movement' interchangeably.
[8] That EM applies freely is not necessarily stated explicitly, however, there does seem to be more focus given to justifying instances of IM than those of EM.
[9] Indeed, Chomsky (2021) argues that IM is actually simpler than EM as it involves least search.

triggers for EM (e.g., edge or c-selectional features) are either outdated or disfavoured. If features do not trigger EM, they cannot trigger IM (Chomsky, 2013). Symmetry-breaking accounts, on the other hand, refer to configurational factors. This is closer to EM since it concerns structure building. Moro (2000) also argues against these 'morphological' triggers for movement in favour of configurational ones, as we shall see when discussing Dynamic Antisymmetry.

As symmetry-breaking accounts are new, they cannot yet account for all the data; numerous triggers may have been proposed because it is difficult to give a simple account for movement. Nevertheless, symmetry-breaking is a desirable solution if having some types of movement triggered and others untriggered is considered unproblematic when unifying analyses of movement.

2.3 Symmetry-breaking accounts

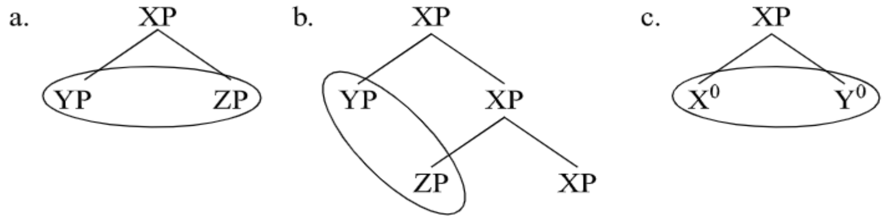
In symmetry-breaking accounts, movement is triggered by the need to break symmetry. In newer versions of minimalism, Merge can produce structures of the form {XP YP} or {X Y}, where one of the elements must move — either because syntax dislikes symmetry, or because symmetry is problematic for linearisation or labelling and thus for the interpretation of the structure at the interfaces. Movement creates the configuration {<XP> YP} or {<X> Y}, resolving the symmetric structure. Crucially, it must be assumed that copies are irrelevant for linearisation and labelling (meaning that the unmoved element can label the structure) (Chomsky, 2013).[10]

In addition to this, labelling-based accounts allow for instances of feature sharing (see Section 2.3.3). Since there is no a priori reason why syntax should dislike symmetry (after all, Merge is symmetrical and I assume that it can apply freely), and feature sharing is useful and relevant, labelling can sensibly be considered in addition to symmetry. Approaches differ on whether symmetry, linearisation or labelling is most important. I focus on labelling, but not necessarily to the exclusion of linearisation.

2.3.1 Dynamic Antisymmetry

In Moro's (1997; 2000) Dynamic Antisymmetry (DA), movement is driven by the search for antisymmetry, required for the structure to be linearised according to Kayne's (1994) LCA. Three structures are considered points of symmetry (PoS) or triggers for movement: small clauses, multiple spec constructions and clitic complementation (1a-c). Small clauses, the most relevant for this dissertation, have been analysed in various ways (e.g., Citko (2011)). For Moro, small clauses have the simple symmetric XP YP structure of (1a).

(1)



(Moro, 2000, p. 32)

[10] The implications of this assumption for e.g., the distinction between copies and repetitions are beyond the scope of this dissertation.

DA aims to account for movement without reference to morphological triggers, although morphological restrictions remain relevant to the grammar. Moro argues that DA is both theoretically and empirically superior to theories involving morphological triggers for movement. A key empirical argument comes from canonical versus inverse copular sentences in Italian. Moro argues that the canonical copular sentence in (2) (where the raised DP is the subject of predication) and the inverse copular sentence in (3) (where the raised DP is the predicate) are derived from the same underlying structure in (4).

(2) *Questa foto era la causa della rivolta*
this picture was the cause of.the riot
'This picture was the cause of the riot'

(3) *La causa della rivolta era questa foto*
the cause of.the riot was this picture
'The cause of the riot was this picture'

(4) BE [α [DP *questa foto*] [DP *la causa della rivolta*]]
Moro and Roberts (2021, pp. 3-4)

The constituent α in (4) is a small clause and thus a PoS. The inverse and canonical copular sentences are derived through movement of one of the two DPs. Since either DP can move, this cannot be neatly accounted for by a Case-driven approach.

2.3.2 Labelling versus linearisation

Recent symmetry-breaking approaches refer to labelling instead of, or as well as linearisation since labelling appears to be a superior trigger. Labelling is important to both interfaces (e.g., Fukui & Narita, 2014), and arguably also to the narrow syntax, e.g., for locality restrictions and for creating points of spell-out (Bošković (2015; 2016)). Linearisation, however, is solely a property of PF in minimalism (Chomsky, 1995). Having a PF requirement as the trigger for IM, a narrow syntactic operation, is undesirable. In terms of timing and motivations, it seems strange for interface requirements to drive IM; they would require some kind of look-ahead (Moro & Roberts, 2021). Adapting DA so that the trigger is labelling rather than linearisation may resolve these issues. Empirically, the predictions are slightly different; multiple spec constructions are no longer points of symmetry. Nevertheless, adapting the approach is fairly unproblematic, hence the development of both Chomsky's (2013; 2015) *Problems of Projection* (PoP/PoP+) and Moro & Roberts' (2021) Extended Dynamic Antisymmetry (EDA).

2.3.3 Problems of Projection

Thus, PoP is a labelling interpretation of DA, in the context of new versions of minimalism in which structures need not be endocentric. Merge does not involve projection, instead, labelling is done by the Labelling Algorithm (LA), which is reduced to minimal search, a third-factor condition sensitive to hierarchy, which takes the most prominent lexical item (LI) as the label (Chomsky, 2013). When Merge generates symmetric structures, the LA cannot label the constituent as both LIs are equally prominent. Instead, one of the constituents must move, while the other labels the structure. The moved constituent keeps moving until it reaches a configuration where there is no longer a labelling failure, i.e., a configuration of Spec-head feature sharing. This agreement provides a label and has a morphological realisation. Furthermore, the moved constituent is frozen in place: this is criterial freezing (Rizzi, 2015). This approach has been applied to VISH and successive-cyclic wh-movement (Chomsky, 2013).

In the original formulation of PoP, IM happens only at the phase level (Chomsky, 2013); in PoP+ (Chomsky, 2015), however, IM applies freely, which is preferable given that IM and EM are the same in nature.

2.3.4 Extended Dynamic Antisymmetry

Both EDA (Moro & Roberts, 2021) and PoP involve interpreting DA in terms of labelling. EDA considers both labelling and linearisation relevant for movement. As in DA, IM is dissociated from morphological factors. Hence there is no need for elements such as feature strength and uninterpretability (Moro & Roberts, 2021), although arguably feature strength remains, since only some features are strong enough to label on their own without feature sharing (as discussed in PoP; Chomsky (2015)).

EDA can be applied to the same phenomena as other symmetry-breaking accounts, furthermore, Moro and Roberts argue that the different repair strategies for symmetric X X structures may form the basis for an analysis of head-initial and head-final typologies.

2.4 Summary

Given that IM and EM are the same operation, I have argued that configurational triggers, in particular symmetry, are preferable to feature-driven or morphological triggers. Various symmetry-breaking theories exist, particularly (E)DA and PoP, which refer to labelling and/or linearisation. As these approaches are new, much data remains to be accounted for. To analyse passives in the EDA framework, I now discuss existing theories of passivisation.

3 Theories of the passive

In an EDA analysis, actives and passives differ in which constituent moves to subject position to resolve the PoS. Thus, *by*-phrases and transitive active subjects must be generated in the same underlying position. Here, I review analyses of the passive, arguing that approaches in which *by*-phrases are arguments generated in the same position as transitive subjects are theoretically superior for independent reasons.

3.1 Early approaches

In Chomsky's (1957) early approach to the passive, transitive subjects and the *by*-phrase originate in the same position. As syntactic theory developed, however, this idea was abandoned. Instead, in P&P approaches, the passive *by*-phrase was analysed as an adjunct, and a process was posited to link the external argument and the *by*-phrase. One reason for considering the *by*-phrase as an adjunct comes from its apparent optionality, as shown by short passives (e.g., *the food was eaten*).

Jaeggli (1986) proposed one of the earliest[11] P&P theories of the passive, in which the defining aspect is theta-role absorption; the passive -EN morpheme absorbs the external theta-role of the verb. The verb is unable to assign objective Case as -EN absorbs this too. The underlying direct object,

[11] The earliest P&P theory of the passive was Burzio (1981/1986), as discussed in Chomsky (1981).

therefore, moves to Spec TP to receive Nominative Case. The *by*-phrase is an adjunct, linked to the external argument through the rather complicated operation of theta-role transfer, whereby the theta-role absorbed by -EN is transferred to the NP in the *by*-phrase. This approach to passives was new because it reduced the construction to other phenomena. Nevertheless, the process of theta-role absorption, then theta-role transmission, is undesirably complex. Theta-role absorption also appears redundant, as it duplicates what could be done transformationally (Bowers, 2010). Furthermore, this approach violates UTAH by assigning the external theta-role differently in the active and the passive, although this approach predates UTAH (Baker, 1988). Empirically, Jaeggli's approach does not handle impersonal passives or passives of ditransitives well, amongst other things (e.g., Bowers (2010)).

Baker et al (1989)'s P&P approach modifies Jaeggli's account but is otherwise similar. Instead of theta-role absorption, the defining characteristic of passives is that -EN is the external argument. -EN is generated in I(nfl) and lowers to V in order to get Case, which replaces Case absorption. The *by*-phrase can double the theta-role in a process analogous to clitic doubling. This approach is theoretically superior to Jaeggli's account, as there is no need for theta-role and Case absorption. Nevertheless, many of Jaeggli's issues remain; and the clitic doubling analysis of *by*-phrases is perhaps not ideal, as it would be a unique instance of clitic doubling in English (Bowers, 2010). Furthermore, downward movement is not allowed in modern syntactic theory.

3.2 Adjunct approaches

Analysing the *by*-phrase as an adjunct is common in current theorising. I focus on the theories of Bruening (2013) and Legate (2014), although many similar approaches exist, such as Alexiadou et al (2015) and Legate et al (2020). In these approaches, unlike in P&P, the agent is syntactically unprojected, however, as in P&P, there are various ways of linking the *by*-phrase with the external argument.

Bruening (2013) proposes a uniform analysis of *by*-phrases in all contexts, including nominals. His analysis captures similarities between *by*-phrases, instrumentals and comitatives.[12] According to Bruening, the head Pass selects a projection of Voice which has not yet projected its external argument. Instrumentals, comitatives and *by*-phrases all adjoin to this projection of Voice; the *by*-phrase is thus an alternative realisation of the EA. Bruening's approach accounts for the empirical similarities between certain constructions, nevertheless, it is complicated when compared to argument approaches, and is heavily reliant on s-selection, which may be problematic if Merge is to apply freely. Furthermore, like P&P approaches, adjunct approaches such as Bruening's violate UTAH.

Legate's (2014) approach to passives resembles Bruening's. Legate argues that the Voice head has phi-features which semantically restrict the EA (the 'initiator'), but do not saturate it. The *by*-phrase and EA are linked as *by* assigns an initiator theta-role to the DP constituent, which is semantically linked to the theta-role introduced by Voice. Legate's analysis is based on the Malayo-Polynesian language Acehnese and comparisons between the passive and object voice in this and similar languages.

3.3 Argument approaches

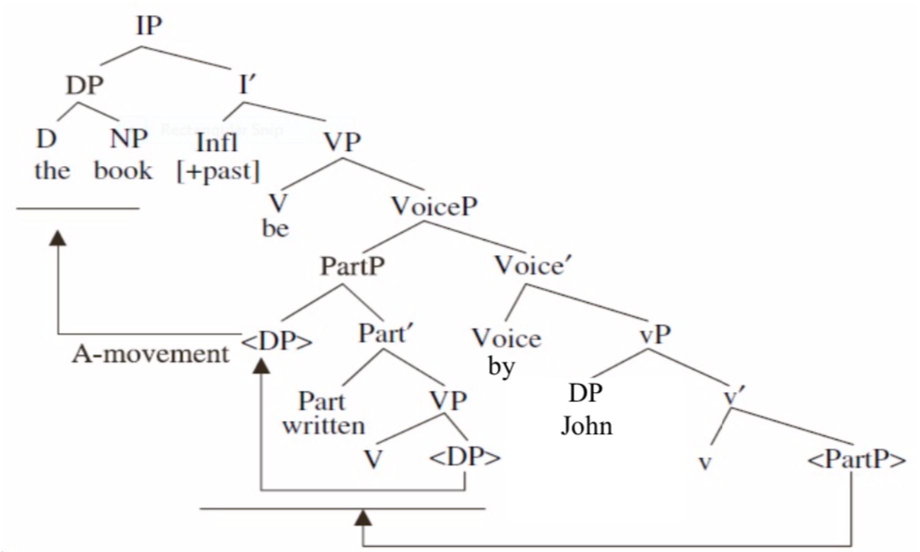
It remains, therefore, common to analyse the *by*-phrase as an adjunct. The following section discusses approaches in which it is analysed as an argument. These approaches comply with UTAH, as the EA and *by*-phrase are generated in the same underlying position. As in P&P approaches, there is an argument present, although the argument is the *by*-phrase instead of the passive morpheme. Since

[12] Such data cannot be explained as well in argument approaches and would benefit from further study.

argument approaches are important for my analysis, I discuss some of these in greater detail; firstly, those involving smuggling, and secondly, those which generate the EA lower than the IA.

Collins (2005) proposes a smuggling analysis of English passives. For Collins, the EA is merged in Spec vP in both the passive and the active, partly to obey a strong form of UTAH. To derive passives, the internal argument (IA) must raise to Spec TP, however, if it were to simply raise, there would be a Relativised Minimality (RM) violation given the presence of the EA. Instead, Collins suggests that the Participle Phrase (PartP) containing the IA moves over the *by*-phrase, after which the IA can raise to Spec TP. This is referred to as smuggling as the IA is 'smuggled' over the EA. The derivation is summarised in (5), for *the book was written by John*.

(5)



Collins (2005, p. 90)

Collins suggests that the *by*-phrase is not a constituent. Indeed, according to a strong form of UTAH, the EA must be a DP not a PP. Collins (2018a) later adjusts his theory so that the *by*-phrase is a constituent, since this was an important criticism of his approach. Collins' analysis is simpler than adjunct approaches as no operation linking the *by*-phrase and the EA is required. It is the earliest argument approach to passives. Nevertheless, the theory makes incorrect empirical predictions, for instance predicting incorrect word orders[13]. Collins acknowledges many of these issues and provides solutions for them, though they appear *ad hoc*. Bowers (2010) argues that this theory is thus unfalsifiable. Additionally, smuggling approaches such as this may be problematic in terms of freezing. Indeed, it is generally assumed that constituents are frozen when in a derived position where they are fully licensed. In particular, den Dikken (2020) argues that the IA should not be able to move out of the VP, as VP in a derived specifier position is an island. This issue is relevant to all smuggling approaches; I return to this in Section 5.1.

Roberts (2019) offers another smuggling-based approach to passives. This derivation includes a Voice head, which has formal features probing the closest argument in its c-command domain. In actives, v inherits these features, and can license the internal argument. The defining characteristic of passives is that Voice withholds its formal features. Thus, the IA cannot be licensed, instead, Voice

[13] For instance, it predicts the grammaticality of the following:
1. *Written the book was by John
2. *There will a book written be by John (Bowers, 2010:57-58)

probes its closest argument, the EA. The IA is probed by T and raised to Spec TP in a smuggling derivation. Roberts extends this approach to account for typological variation in passives.

Other less common approaches generate the EA lower than the IA, at least in passives. Bowers (2010) argues that the Universal Order of Merge is different to that standardly assumed, with Agents being generated the lowest. The agent of transitive active sentences and the passive *by*-phrase are both merged as the lowest argument in Spec Ag(ent)P. Bowers argues that this approach is simpler, since there is no need for smuggling. Nevertheless, the empirical data are questionable.

Den Dikken (2020) bases his analysis on Collins (2005) but derives passives without smuggling. This is done partly to deal with the freezing problem. Den Dikken has developed a theory of predication, which he adapts to passive sentences. Unlike Moro, for whom small clauses are symmetric constituents, den Dikken considers small clauses a projection of a functional category, RELATOR. He argues that actives derive from a canonical predication structure whereas passives derive from a reverse predication structure, as in (6-9) (den Dikken, 2020, p. 8). The patient DP can move out of the VP to subject position without a freezing problem, since the VP was base-generated in the relevant position instead of being moved there.

(6) [RP [Specifier SUBJECT] [R' RELATOR [Complement PREDICATE]]]
(7) [RP [Specifier PREDICATE] [R' RELATOR [Complement SUBJECT]]]
(8) [TP Spec [T' [RP [DP John] [R' RELATOR [VP kissed Mary]]]]]
(9) [TP Spec [T' [RP [VP kissed Mary] [R' RELATOR [DP John]]]]]

As in Collins (2005), the *by*-phrase is not a constituent, instead *by* is the RELATOR head. Den Dikken acknowledges that his approach violates a strict formulation of UTAH (as in Collins), but argues that UTAH is about relations rather than absolute positions. Although den Dikken considers it an advantage that the *by*-phrase is not a constituent in his approach, the analysis inherits Collins' *by*-phrase problems. These cannot be solved as easily as in Collins, as the predication structure is central to the analysis. Other issues, including potentially problematic assumptions about LF-raising, are discussed in Sections 4.3.2–4.3.3. Nevertheless, den Dikken's predication approach to passives is interesting, given the relevance of predication structures or small clauses in EDA. Indeed, I make use of this idea when analysing passives in EDA.

3.4 Summary

Theories of passivisation have become increasingly elegant over time. Indeed, P&P analyses involve complex operations relating the external theta-role with the *by*-phrase. Modern adjunct approaches improve upon this; nevertheless, I consider argument approaches theoretically superior, as they are simpler and obey UTAH. These analyses must deal with the RM issue, either by smuggling or generating the EA below the IA. The extent to which smuggling is a problem depends on the validity of the criticisms concerning freezing, which I return to in Section 5.1. Overall, argument approaches seem theoretically superior, and are the kind of structure required for an EDA approach to passives. In the following section, I discuss the empirical validity of such approaches; this is important since it remains standard to treat the *by*-phrase as an adjunct.

4 Is the by-phrase an argument?

The nature of the *by*-phrase is determined by considering theta-roles and UTAH, and c-command phenomena, especially binding and control. Before discussing the data, some preliminary comments:

grammaticality judgements are unclear or vary hugely for some sentences in this section; authors disagree on the precise definitions of certain terms, e.g., UTAH; and in P&P analyses of the passive, the EA is an argument but the *by*-phrase is an adjunct. In later adjunct approaches, there is no syntactically present EA. This distinction is relevant for certain c-command phenomena.

4.1 Theta-roles and UTAH

By-phrases bear the same range of theta-roles as the transitive active EA. As can be seen in (10-13) (Jaeggli, 1986, p. 599), the DP in the *by*-phrase is interpreted differently depending on the sentence. The *by*-phrase can even be an idiomatic subject, as discussed in Collins (2005, p. 83) ((14-15); originally from Postal (2004, p. 256)).

(10) Bill was killed by Mary (Agent)
(11) The package was sent by John (Source)
(12) The letter was received by Bill (Goal)
(13) That professor is feared by all students (Experiencer)
(14) Ted was bitten by the lovebug
(15) I was told that by a little bird

For Jaeggli, this is evidence for theta-role transmission. For Collins, *by* is a dummy preposition which does not assign its own theta-role; the *by*-phrase has the same theta-role as the EA precisely because it is the EA. This seems simplest, evidence from compositional theta-roles strengthens this claim.

4.1.1 Compositional theta-roles

Compositional theta-role, as discussed in Marantz (1984), is the phenomenon whereby the entire predicate marks the EA. Roberts (1987, p. 27) demonstrates that this is present in the passive, as shown in (16-20), where the *by*-phrase has the same exact compositional theta-role as is assigned to the EA by the predicate:

(16) A baseball was thrown by Fernando
(17) Support was thrown behind the candidate by the CIA
(18) The match was thrown by the prizefighter
(19) The party was thrown by the department
(20) A fit was thrown by the Countess

The data follow naturally from argument approaches, as the EA and the *by*-phrase are underlyingly the same. Adjunct approaches such as Bruening (2013) argue that the compositional theta-role is the contextual interpretation of the EA theta-role assigned by Voice. Roberts (2019) argues that it would be a huge complication for *by* to have to track the compositional theta-role assigned to the EA. Thus, theta-role data demonstrate that there is a link between the EA and the *by*-phrase, which is best explained by argument approaches.

4.1.2 UTAH

A theoretical claim for argument approaches comes from UTAH. According to a strong form of UTAH, the EA and *by*-phrase must be underlyingly the same. In minimalism, there is no independent UTAH, instead, its effects result from a strict theory of theta-role assignment. This is the configurational approach to theta-role assignment, where if a given position is not projected, the theta-role cannot be assigned. If this view is taken, adjunct approaches are ruled out, for instance, part of Bruening's (2013) definition of the passive is that the EA is unprojected. This strict interpretation of UTAH is a key argument for Collins (2005).

4.2 C-command phenomena

Binding phenomena provide the most relevant data concerning the status of the *by*-phrase. This is because c-command relations are possible only from A-positions. Sentences where the *by*-phrase c-commands something else, e.g., a reflexive or PRO, constitute much of the data in favour of argument approaches. However, in some argument approaches (notably Bowers (2010) and den Dikken (2020)), the *by*-phrase is generated lowest; movement (e.g., smuggling) is required to account for such data. In this section I focus on binding and control.

4.2.1 Binding

The key argument from binding is that the *by*-phrase seems able to bind a non-logophoric[14] reflexive or reciprocal, whereas adjunct PPs are not. This can be seen in (21-23), where there is a Principle A violation in (22) but not in the sentence with a *by*-phrase.

(21) The packages were sent by the children$_i$ to themselves$_i$ /each other$_i$
(22) *The packages were sent for the children$_i$ to themselves$_i$ / each other$_i$
(23) The packages were sent by / for the children$_i$ to their$_i$ mothers
(Collins, 2018a, pp. 2-3)

To rule out the possibility of sentences like (21) involving logophoricity, Angelopoulos et al (2020, p. 11) use inanimate DPs, as in (24) ((25) is mine).

(24) The planets were drawn by the black hole$_i$ into itself$_i$[15]
(25) *The planets were drawn on behalf of the black hole$_i$ into itself$_i$[16]

To account for these phenomena, the DP of the *by*-phrase must be an argument c-commanding the anaphor. For Collins (2005), this is clearly the case, as the *by*-phrase is not a constituent. In approaches such as Collins (2018a) or Roberts (2019), the *by*-phrase is a constituent. This complicates c-command as the DP does not c-command the reflexive. Instead, it is suggested that the entire PP can bind, either for semantic reasons or because it is a KP (Case Phrase). Roberts (2019) argues that *by* can be considered a case marker, the head of KP and a manifestation of probing. It is an argument bearing the relevant features for c-command relations (e.g., phi-features for binding). In adjunct analyses, however, the *by*-phrase should not be able to bind the anaphor, not only because it is an adjunct but also as it does not c-command the anaphor. Semantic and agreement issues also exist. Except for Legate (2014), these

[14] Anaphors are often considered exempt from Condition A when behaving as logophors, i.e., reflexives with discourse antecedents external to the clause. Logophors must have perspective centres (e.g., deictic centres) as antecedents (Angelopoulos et al, 2020).
[15] (24) appears slightly degraded to me. Collins (2018b:14) discusses why the short passive equivalent of (24) is ungrammatical.
[16] Semantic factors may also be relevant here.

analyses don't tend to mention binding phenomena; adjunct analyses do not satisfactorily account for these facts.

The c-command data so far suggest that the *by*-phrase is an argument, generated (we assume for simplicity) in the same place as the transitive subject. Analyses such as Collins (2005) do not hold up for certain other binding phenomena; I do not discuss these for space reasons.

4.2.2 Binding: Alternative position of the by-phrase

Other argument approaches generate the *by*-phrase elsewhere. In Bowers (2010), it is generated as the lowest argument; in den Dikken (2020) it is the lowest argument in passives, but not in actives, due to the predication structure used. I now discuss how such approaches deal with the previous section's c-command phenomena and additional data. I argue that the advantages of these approaches do not outweigh the arguments for generating the *by*-phrase as in Collins (2005).

According to Bowers (2010, pp. 33-34), NPI data like (26-27), where the *by*-phrase is c-commanded by the other arguments, suggest that the *by*-phrase is an argument, and that it is merged the lowest (the wrong order would be predicted if it were an adjunct).

(26) Money was given to no student by any professor
(27) *Money was given to any student by no professor

However, NPI can also show the opposite effect, as in (28-29) (my examples):

(28) Money was given by no student to any professor
(29) *Money was given by any student to no professor[17]

Bowers shows that sentences with an adjunct c-commanding the *by*-phrase are grammatical as well, e.g., (30). As NPI-licensing can take place from A'-positions (e.g., (32), where the licensor is in Spec CP), this data is not particularly useful when it comes to discussing the argument/adjunct status of *by*-phrases (33-34).

(30) The performers were hit with no rotten eggs by any spectators
(31) *The performers were hit with any rotten eggs by no spectators
(Bowers, 2010, p. 34)
(32) At no time did anyone speak
(33) The performers were hit by no spectators with any rotten eggs
(34) *The performers were hit by any spectators with no rotten eggs

Bowers argues that the *by*-phrase / EA must be generated in the lowest argument position in all cases; in the sentences where it c-commands other elements, it must have moved (schematically represented in (35); my example). However, the same facts can be captured by the derivation in (36) (in both cases, movement is optional).

(35) The books were given by Mary to John (by Mary)
(36) The books were given to John by Mary (to John)

[17] Both (26) and (28) sound somewhat contrived to me, but they are certainly better than (27) and (29).

Due to issues with NPI and for consistency, I now return to binding. I discuss (35-36) further as well as the Principle A data in (37-41).

(37) The packages were sent by the children$_i$ to themselves$_i$ / each other$_i$
(38) *The packages were sent for the children$_i$ to themselves$_i$ / each other$_i$
(Collins, 2018a, p. 2)
(39) The books were given to mei by myself$_i$
(40) ??The books were given to myself$_i$ by me$_i$
(41) *Mary was shown himself$_i$ by John$_i$
(Bowers, 2010, p. 35)

Collins (2005) argues that (39) is slightly degraded, and that (40) is suboptimal but allowed as there is reconstruction. Bowers (2010), however, argues that (39) is fine in the right context. (40), though, should not be allowed, as the *by*-phrase is never in a position above the indirect object (IO). It is marginally acceptable because of logophoricity, as opposed to (41). This fits with Bowers' general approach, nevertheless, it does not account for the data in (37-38).

Returning to (35-36), Bowers argues that the order in (39/35) is unmarked relative to (37/36) based on evidence from movement. However, I consider (37) better (and thus less marked) than (39). Indeed, grammaticality judgements about many of the sentences are debatable, and much of the data can be analysed in multiple ways. Bowers' approach does not account for the binding data at the beginning of this section and disregards the standard analysis of clause structure. Unless there is a very good reason not to, it seems best to stick with standard analyses.

4.2.3 Binding: Passive and object voice

Legate's (2014) study of passives focusses on Malayo-Polynesian languages, which have both an object and a passive voice. She provides data distinguishing the two; in the object voice, the initiator (external theta-role) is an argument, whereas it is considered an adjunct in the passive. Such distinctions are not visible in languages without the object voice. Legate's data are interesting, assuming that the passive construction in these languages can be equated with the English construction and analysed in a sufficiently similar way, and that it has been analysed correctly.

Concerning binding, in the object voice, the initiator binds the subject and everything in the verb phrase under reconstruction. If the passive initiator is in Spec VoiceP (the position of the EA), we would expect the same binding relations as in the object voice, however, this is not the case. Indeed, the Indonesian passive (42) is ungrammatical, but the object voice (43) is grammatical, as the theme is bound in the object voice but not in the passive.

(42) ?**Diri-nya* *di-serah-kan* *ke* *polisi* *oleh* *Amir*
self-3 PASS-surrender-APPL to police by Amir
'Himself was surrendered to the police by Amir'

(43) *Diri-mu* *mesti* *kau-serah-kan* *ke* *polisi*
self-2 must you-surrender-APPL to police
'You must surrender yourself to the police'
(Legate, 2014, p. 71)

Legate's data is evidence for *by*-phrases being adjuncts; or arguments generated lower than the initiator in object voice constructions (as in Bowers or den Dikken). This data remains a problem for analyses such as Collins or my EDA approach to passives[18].

4.2.4 Control

Object voice comments aside, binding phenomena seem compatible with the *by*-phrase being an argument generated in the same position as the active transitive EA. I now discuss control phenomena.

Den Dikken (2020) and Bowers (2010) predict no binding or control effects. Den Dikken suggests that this explains Visser's Generalisation, the principle that subject control verbs resist personal passivisation. He argues that the *by*-phrase never c-commands PRO in (45), hence its ungrammaticality.

(44) John$_i$ promised the politicians$_j$ PRO$_i$ to reward them$_j$
(45) *The politicians$_j$ were promised PRO$_i$ to reward them$_j$ (by John$_i$)

However, this does not explain the ungrammaticality of (46). Unless it can be shown that the *by*-phrase cannot move to a c-commanding position, Visser's Generalisation does not help den Dikken's claim.

(46) *The politicians$_j$ were promised by John$_i$ PRO$_i$ to reward them$_j$

I return to control in Sections 4.3.2–4.3.3 since most of the data is for short passives.

4.3 Implicit arguments

The perceived optionality of *by*-phrases is used as support for the adjunct approach. Short passives, where the *by*-phrase is (at least overtly) absent, are perfectly grammatical in English. If the *by*-phrase is an argument, it must be syntactically present in some form, thus implicit (external) arguments (IEAs)[19] must be syntactically projected as implicit and overt arguments should not be too qualitatively different. Bhatt and Pancheva (2006) discuss approaches to IEAs, showing that the IEA is generally considered semantically present; whether it is syntactically active and projected is debated. Bhatt and Pancheva argue that IEAs are syntactically active based on c-command data, but that it is less clear whether they are syntactically projected. It is, however, undesirable (or actually impossible in bare phrase structure) to posit a syntactically active but not projected element. I therefore assume that c-command data is relevant to whether the IEA is syntactically projected. I argue that the IEA is indeed projected as an argument and discuss how different approaches account for various interpretive phenomena.

4.3.1 More binding data

It has been argued that the IEA can bind reflexives (e.g., Collins, 2005). For instance, (47) from Baker et al (1989, p. 228)[20] is grammatical as the IEA ('IMP') c-commands the anaphor.

---

[18] However, the position where the anaphor is base-generated is unclear, which weakens this argument somewhat.
[19] I use this abbreviation to distinguish implicit arguments from internal arguments (IAs).
[20] The argument is '-EN' in their approach, so it is not really an 'implicit argument', however, the data can be easily transferred to a theory like Collins (2005), hence my sentence annotation.

(47) Such privileges should be kept IMP$_i$ to oneself$_i$

Alexiadou et al (2015) argue that such examples may be logophoric; Collins (2018b; in press) shows that they cannot all be accounted for in this way. While logophors can be replaced with pronouns, (49) is ungrammatical.

(48) John$_i$ kept those privileges to himself$_i$
(49) *John$_i$ kept those privileges to him$_i$
(Collins, 2018b, p. 2)

Nevertheless, the data are controversial and not uniform across languages. For instance, Schäfer (2012) argues that the IEA cannot bind reflexives, based on Dutch examples such as (50). However, (50) is an impersonal passive, the nature of which and their relation to transitive passives is a matter of debate. Indeed, it has been known since at least Roberts (1987) that *by*-phrases in impersonal passives differ from those in ordinary passives. Using impersonal passives to discuss ordinary passives is problematic.

(50) **Er* *werd* *zich* *gewassen*
there was REFL washed
'People washed (themselves)'
(can't be treated as *Er werd IMP$_i$ zich$_i$ gewassen)*
(Reinhart & Siloni, 2004, p. 169)

4.3.2 Depictive control

Depictive control, i.e., whether the IEA can license depictive secondary predicates, is particularly contested, with grammaticality judgements varying hugely. Collins (2005) argues that the IEA can license depictives, hence the grammaticality of (51-52).

(51) The book was written drunk (i.e., The book was written IMP$_i$ PRO$_i$ drunk)
(52) At the commune, breakfast is usually eaten nude
(Baker, 1988, p. 318, in Collins, 2005, p. 101)

Bhatt and Pancheva (2006) argue that (51-52) are exceptional, and that most such sentences are ungrammatical. Landau (2010, in Alexiadou et al, 2015, p. 131) gives the following examples, suggesting that there are both strong and weak implicit arguments; only strong ones, which don't include the passive agent, can license depictives. I consider (53-54) degraded but not entirely ungrammatical (they might be improved by context), however, (55) is ungrammatical.

(53) *The room was left angry
(54) *The game was played shoeless
(55) *It's impossible for me to be visited together

Den Dikken argues that depictives are grammatical when the IEA moves to Spec TP at LF. This seems to fit the data, assuming grammaticality judgements about depictives are remotely reliable, however, it is unfalsifiable and problematic as it involves movement to a filled position.

Lastly, Roberts (1987, p. 94) shows that equivalent sentences with overt *by*-phrases are ungrammatical, as (56-59) show:

(56) *John was seen by Mary$_i$ drunk$_i$
(57) *Fred was kissed by Sue$_i$ happy$_i$
(58) *Tom was met by Bill$_i$ angry$_i$
(59) *Sue was arrested by the police$_i$ drunk$_i$

Adjunct approaches predict this c-command data, however, given the unreliability of evidence from depictives, this is not sufficient to abandon the argument analysis. Instead, these data show that depictives cannot provide information regarding the status of IEAs, since the corresponding sentences containing overt *by*-phrases are also ungrammatical.

4.3.3 Other types of control

Collins (2018b) argues that the IEA can participate in obligatory control (hence (60)). Den Dikken (2020), however, argues that such examples are grammatical because the IEA moves to Spec TP at LF, as with depictives.

(60) Once it had been decided IMP$_i$ PRO$_i$ to build a walled enclosure, new possibilities opened up.
(Collins, 2018b, p. 7; silent categories added)

Control into adjunct clauses has similarly been discussed. This refers to sentences such as (61), where the IEA is assumed to control PRO.

(61) This bureaucrat was bribed IMPi PROi to avoid the draft
(Baker et al, 1989, p. 221; silent categories added)

For such examples, den Dikken (2020) argues that c-command data is not relevant since it is not a case of obligatory control; the validity of this is questionable.

Thus, though control data is debatable and therefore not particularly helpful for either position, some data appears to be best explained by an analysis like Collins (2005).

4.3.4 Existential interpretation of the implicit argument

Various interpretive phenomena are relevant to the presence or absence of IEAs in the syntax. Bruening (2013) shows that the IEA cannot be bound or controlled, though the overt *by*-phrase can be. Indeed, (62/64) cannot be interpreted as (63/65), yet (63/65) are grammatical ((62/64) are from Bruening (2013, p. 19; (63/65) are mine).

(62) Every journalist wants the president to be interviewed
(63) Every journalist$_i$ wants the president to be interviewed by him$_i$
(64) John wants Mary to be seen
(65) John$_i$ wants Mary to be seen by him$_i$

Instead, the IEA means approximately ‘by someone’. In adjunct analyses e.g., Bruening (2013), this is evidence that passives do not syntactically project the IEA, instead, the EA is existentially quantified over. In argument analyses, the IEA tends to be considered a quasi-existential arbitrary PRO (PRO$_{Arb}$), or similar. (62-65) are explained using features: PRO_{Arb} and the binder/controller cannot be coindexed, as the binder/controller is not arbitrary. The phenomenon of disjoint reference, i.e., the fact that the passive IEA must be interpreted non-reflexively (e.g., Alexiadou et al, 2015) also helps explain (62-65).

5 An EDA approach to passives

Thus, it appears that data from long and short passives is best accounted for by an argument analysis of the passive construction. This type of construction, as discussed, also appears to be the most theoretically elegant. Furthermore, it is the starting point for an EDA approach to passivisation.

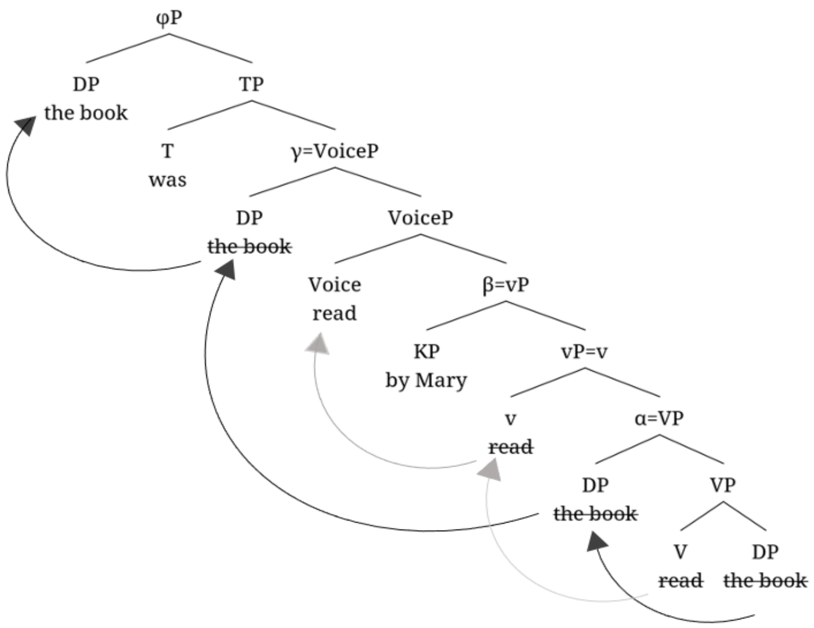
The basic idea of this approach is that actives and passives derive from the same underlying structure, with different constituents in each case moving to break symmetry. This approach is similar in some ways to that of den Dikken (2020), in which actives and passives are canonical and reverse predication structures respectively; in EDA, small clauses are a source of symmetry. The structure of small clauses in den Dikken and Moro is different; as mentioned, small clauses for Moro have the symmetric structure in (66), whereas den Dikken considers small clauses a projection of RELATOR (67-68).

(66) [$_{\alpha}$ XP [YP ZP]]
(67) [RP [$_{\text{Specifier}}$ SUBJECT] [$_{\text{R'}}$ RELATOR [$_{\text{Complement}}$ PREDICATE]]]
(68) [$_{\text{RP}}$ [$_{\text{Specifier}}$ PREDICATE] [$_{\text{R'}}$ RELATOR [$_{\text{Complement}}$ SUBJECT]]]

An advantage of my approach is that it keeps the standard clause structure for actives, instead of rethinking it drastically. First, I illustrate my structure for passives. Then I extend the approach to account for ditransitive constructions in the passive. My analysis does not discuss morphology (i.e., the passive suffix and the auxiliary) as my focus is elsewhere. CP is omitted as it is not relevant to the discussion.

5.1 Transitive passives

(69)



The derivation of an ordinary passive sentence is represented in (69). First, the IA merges with the verb; the constituent is straightforwardly labelled VP. The IA is then internally merged with VP (movement can apply freely, as discussed)[21]. This creates our first point of symmetry, α. This is followed by V-to-v movement, forming vP. This vP merges with the EA, a KP, creating our second point of symmetry, β. The IA also needs to move through this site to obey locality restrictions, the obvious problems with this are discussed later. Movement of the IA enables α to be labelled VP. The verb then moves to a higher verbal projection which I call Voice. The IA is then internally merged with VoiceP, which forms the PoS γ. *Was* is then externally merged to the structure. The constituent is straightforwardly labelled TP[22]. After this, the IA finally merges with TP; the symmetric structure thus created is resolved through phi-feature sharing. Indeed, the constituent can be labelled ϕP, as in POP (Chomsky, 2013). The movement of the IA means that γ can be labelled VoiceP.

At this point, the derivation is complete except for one key point: β remains to be resolved. To do this, as the EA must stay in situ, the vP constituent must either move or be altered in some way. One solution is to move the vP and label β KP. I return to this option after finishing this derivation. Alternatively, the vP could become a v head in some sense. This can happen straightforwardly according to Fowlie's (2013) Multiple Multiple Spellout, where Spellout consists of two operations, LINEARISE and ATOMISE. ATOMISE, based on Nunes and Uriagereka's (2000) Spellout, sends features of a constituent to PF and LF, rendering it inaccessible to the syntax. What remains is a syntactic atom which behaves as an LI.[23] Spellout domains roughly correspond to phases, ATOMISE happens once the next phase is complete to allow for movement (e.g., wh-movement). Depending on when and whether LINEARISE, ATOMISE or both occur, constituents corresponding to phases may

[21] Alternatively, the IA can be base-generated in Spec VP, as in certain approaches to ditransitive constructions (see Section 5.2).
[22] I use TP instead of, for instance, PredP as morphological details are not my central concern.
[23] The argument structure is retained as Fowlie posits both PF and LF-Spellout.

or may not be accessible and linearised. This can be parametrised to account for variation within and across languages. For our purposes, I assume that vP is atomised.[24]

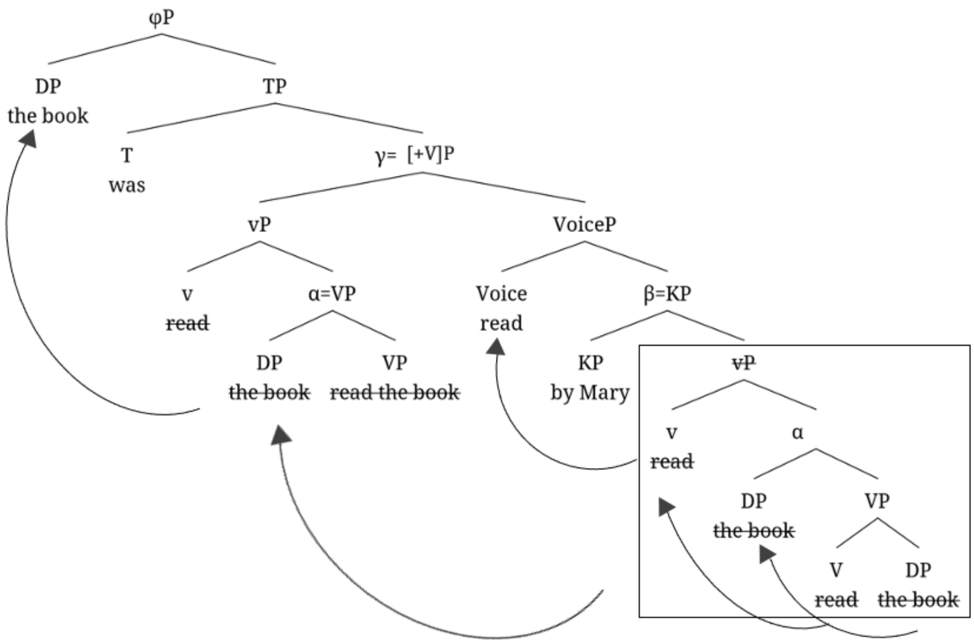
ATOMISE can be applied to EDA as it effectively turns an XP into an X; indeed, applying ATOMISE (or similar) to create asymmetries which act as a movement trigger has already been done by Hsieh and Sybesma (2011) for Taiwanese CPs. We thus assume that the vP phase is atomised; what remains in the syntax is a v head. This head can then label PoS β vP. This happens once the next phase, presumably CP, is completed. Crucially, the IA must have moved out before atomisation. It is, however, unclear how the vP which was previously PoS β is not also atomised. It might be possible to state that whatever gives v phasehood status has been deactivated or stripped away at Transfer; alternatively, it is possible that atomisation of all relevant constituents occurs at the same stage of the derivation, which would keep vP (PoS β) intact.

At this point, finally, c-command relations derive the correct linear order for the passive sentence. Nevertheless, there are various things worth pointing out about this approach. Firstly, the implicit argument must be a maximal projection, indeed, it is crucial for it to be present in short passives to create the PoS β. This is not an issue, as it can have a complex internal structure, e.g., Roberts' (2019) view of *pro*. In terms of actual problems with the derivation, however, this approach suffers from the same RM issue which I have already pointed out is a problem in argument approaches. Indeed, the IA cannot move over the EA intervener without causing a RM violation.[25] We have seen that this issue tends to be dealt with either with smuggling or by generating the EA lower. Another solution is to suggest, as does Belletti (2017), that intervention (though in the context of Agree, rather than movement) might apply representationally instead of derivationally. In other words, since the IA moves, only a copy and not the entire chain is in this position, so, representationally, there is no intervention. However, the intervener is the EA rather than the IA, so again the RM issue is not actually solved. Even if it were to work, it would appear ad hoc.

Resorting to smuggling, another way of dealing with the RM issue is to adopt a solution similar to that of Collins (2005), as represented in (70). This approach has the same derivation as the previous one up until the point where v moves to Voice, but crucially without movement of the IA to the position occupied by KP. This is then followed by movement of the entire vP to form a constituent with VoiceP. β is labelled KP; γ is labelled by feature sharing as a verbal category which I refer to as [+V]P[26]. After this, following the merger of *was*, the IA moves and forms a constituent with TP, which is then labelled through feature-sharing. This enables PoS α to be labelled VP.[27]

[24] I assume that atomisation does not happen to other phases such as DP/KP as that would cause problems with symmetry/labelling.
[25] Aside from the RM problem, there is an issue with deriving the tree at this point in terms of labelling. Indeed, if the IA is merged with PoS β (in a multiple-specifier construction), the resulting constituent will be labelled DP, which is highly undesirable.
[26] The fact that vP only contains copies of the verb at this point (given movement of the verb to Voice) is not a problem for labelling, as the vP label itself is still present, which is what is necessary for feature sharing.
[27] It is also possible that the IA never moves to Spec VP. It might be necessary for it to move through Spec VP in order to move to its final position, which is why I added in this movement. However, it would be simpler to omit this movement, as this would mean that PoS α does not appear. Indeed, this PoS is resolved rather belatedly and seems unnecessarily complex. This movement may not actually be necessary, e.g., Bošković (2020). For consistency with my previous derivation, however, I have kept it. In an approach without movement to Spec VP, movement of the IA to its final position is not triggered by symmetry but rather applies freely.

(70)



Despite the advantage in terms of RM, this approach has various additional issues. Firstly, movement out of a moved element is traditionally banned (e.g., Ross, 1967; den Dikken, 2020, as mentioned in Section 3.3). This so-called freezing ban is less problematic than a RM violation, so, in this way, this approach is still better than the previous one. In any case, the existence of this freezing ban is debated (e.g., Müller (2010)). For instance, Bošković (2020) argues that extraction out of moved elements is generally allowed, freezing data where YP seems unable to move out of XP (as in (71)) arises because XP is unlabelled (due to absence of agreement). Unlabelled elements are not allowed to move, so XP is not allowed to move in the first place. Rizzi's (e.g., 2014) criterial freezing approach also allows for movement out of moved elements, as even if the moved element itself is frozen in place (as it has satisfied a relevant criterion), elements within it can still move out to satisfy a different criterion.

(71) … YP … [XP <YP>] … <[XP YP]> …

Secondly, this approach makes use not only of smuggling, but also of remnant movement. Remnant movement is essentially smuggling with the order of operations reversed: in the configuration [XP [YP]], YP moves out of XP, after which XP moves to a higher position (e.g., Belletti & Collins (2020). In the case of (70), remnant movement occurs as the verb moves out of vP before vP moves, as illustrated in (72-73).

(72) *movement of v to Voice* [VoiceP [**Voice**] [β [KP] [vP **<[v]>** [α [DP] [VP]]]]]
(73) *movement of vP* [γ **[vP]** [VoiceP [Voice] [β [KP] **<[vP]>**]]]

Whether remnant movement causes any problems (or unnecessary complications) when it comes to c-command relations between the copies of YP (in our case, the verb) remains to be determined (e.g., Grewendorf, 2015). Since remnant movement has been used and applied to various phenomena, including VP fronting in German (den Besten & Webelhuth, 1987) and various aspects of interrogative syntax in Romance (Poletto & Pollock, 2015), I assume that it is not an issue worth discussing further in this dissertation.

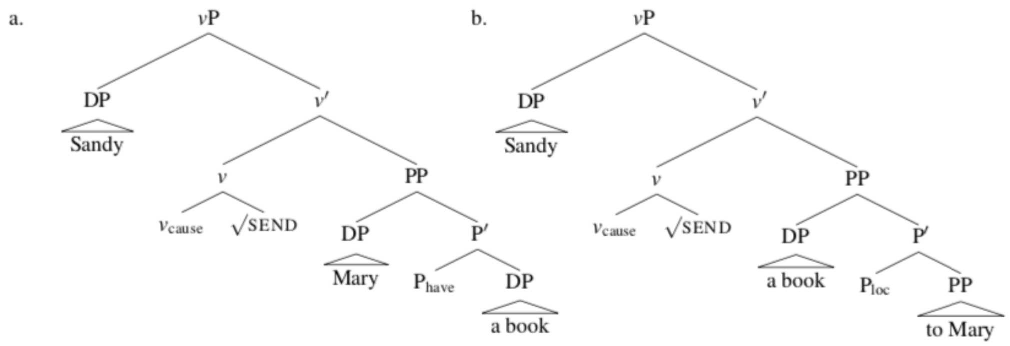
5.2 Extension to ditransitive constructions

Thus, I have begun to sketch an EDA approach to passivisation, suggesting one derivation involving atomisation and another involving smuggling. In this section, I extend these analyses to ditransitive constructions of both the *to* and IO variety in the passive, i.e., of sentences of the type (74-75), which I assume are the passive counterparts to (76-77) respectively. In other words, both the active and passive sentences can be derived from the same structure, depending on which constituent moves to resolve the PoS, as in the transitive.

(74) The book was given to John by Mary
(75) John was given the book by Mary
(76) Mary gave the book to John
(77) Mary gave John the book

For my approach to ditransitive passives, I am assuming an approach such as Harley (2002) or Pesetsky (1995) for the position of the direct object (DO) and the IO. In these approaches, the DO and the IO are generated in different configurations for the two constructions. For instance, (78) shows Harley's analysis of the two constructions.

(78)



The evidence for this is far beyond the scope of this dissertation and thus not discussed, however, it seems the most straightforward way to account for ditransitive passives.[28]

I argue that the passive counterparts to IO and *to* constructions can be generated from the same underlying structure as is used in the active, which I assume is similar to that of Harley's.

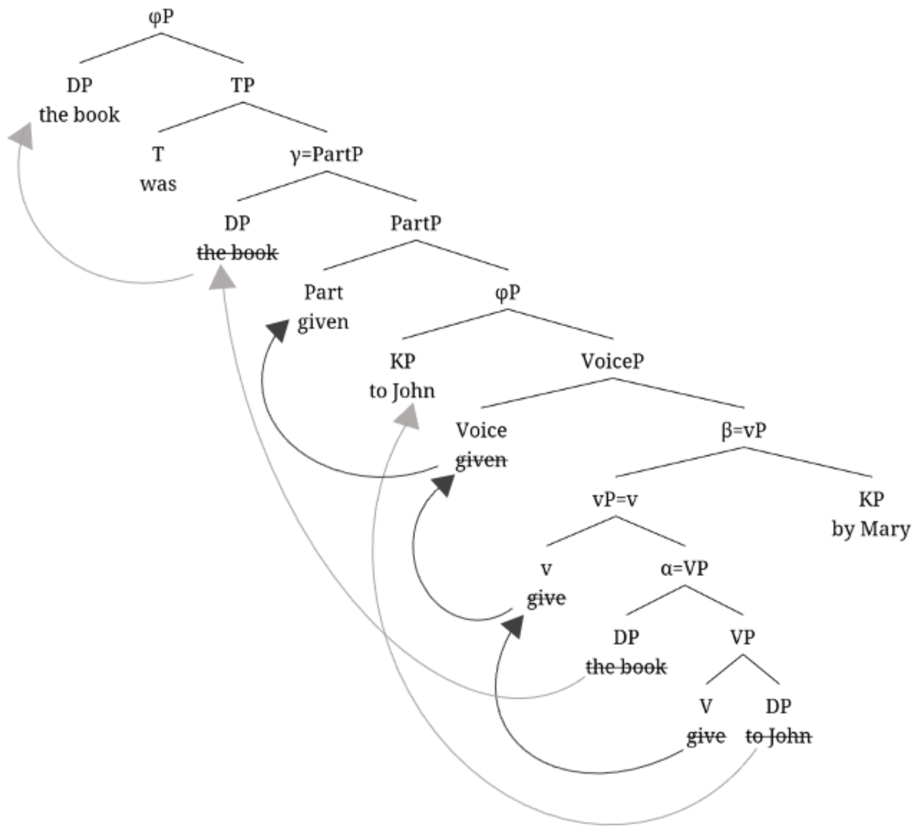
5.2.1 The to *construction in the passive*

Beginning with the *to* construction, I discuss sentences (74) and (76) (repeated in (79-80)), as represented in (81-82). I discuss the atomisation approach followed by the smuggling approach.

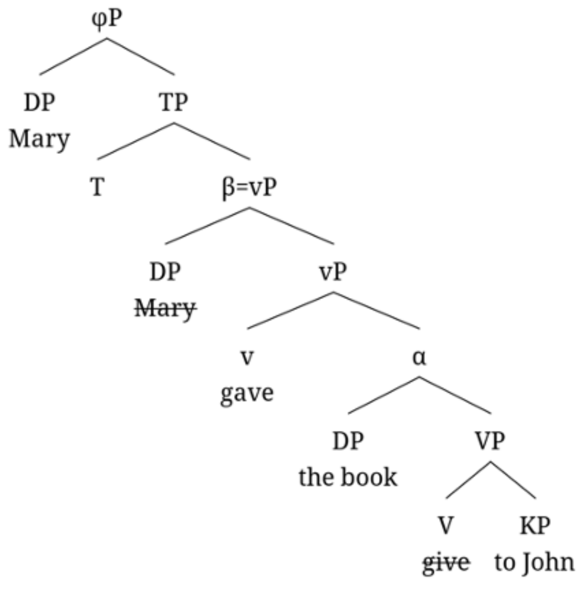
(79) The book was given to John by Mary
(80) Mary gave the book to John

[28] It is possible to derive IO and *to* passives from a structure in which the DO and the IO are base-generated in the same position in both cases, but this will lead to even more RM violations than the present analysis has.

(81)



(82)



[29]

The derivation precedes as follows. First, the IO merges with the verb; the constituent is straightforwardly labelled VP. Then the DO merges with VP, forming PoS α. This is followed by V-to-

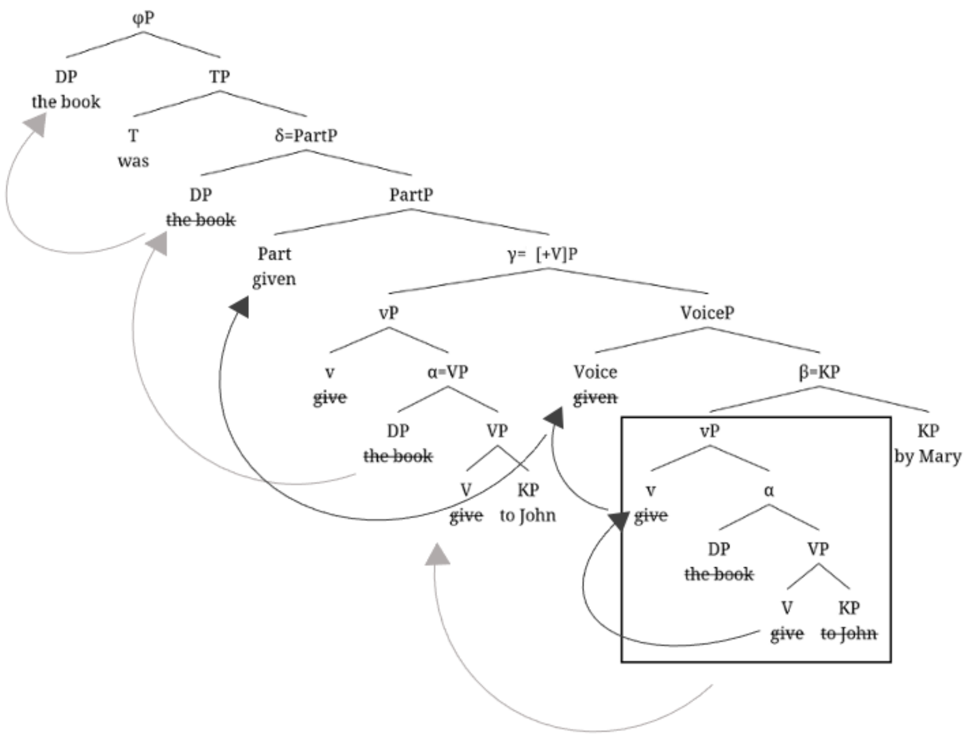
[29] I do not discuss the active sentence as it is not the focus of the dissertation. I do not have a solution for labelling α other than atomising DP, which seems extremely ad hoc, or moving DP to a higher position where the resulting PoS is solved through agreement.

v movement and labelling of the constituent as vP. Then, the vP merges with the EA[30], forming PoS β (which is resolved through atomisation of vP). Subsequently, the verb moves to Voice, forming VoiceP. The IO then merges with VoiceP. This creates a symmetric structure, however, the IO must remain in situ. This is somewhat problematic, the only solution seems to be to posit agreement of some kind here (e.g., phi-features, Dative case), though this is not particularly convincing[31]. I return to this issue later. After this, in order to derive the correct word order, the verb needs to move to a higher projection, which I have named PartP. PoS α is resolved as the DO merges with PartP, creating a third PoS, γ. This is followed by 'was' being externally merged, the labelling of TP, and, finally, the merger of the DO with TP, forming φP through feature-sharing. This enables γ to be labelled PartP. Once again, c-command relations give the correct linear order.

This derivation is even more problematic than that of ordinary passives. Indeed, the RM issues and other problems with the passive derivation remain. The RM issue is even worse in this sentence, however, given that the DO in Spec VP is an intervener for the IO. Moreover, the EA acts as an intervener for both the IO and the DO, meaning that these movements both violate RM.

For these reasons, I suggest an alternative analysis involving smuggling, represented in (83).

(83)



The derivation proceeds as in (81) up until the movement of v to Voice, forming VoiceP (with the difference that PoS β is not resolved through atomisation). This is followed by movement of the entire vP, which forms a constituent with VoiceP. This symmetric structure must be resolved by feature sharing as some kind of verbal category, as in the transitive. β is labelled KP. The verb then moves from Voice to Part, forming PartP. Subsequently, the DO moves, resolving α but forming PoS δ. Following

[30] The fact that KP is represented to the right of vP in the tree is purely for readability and is of no theoretical significance.
[31] An alternative solution would be to adopt the idea that KP (the IO would have to be considered a KP) functions as an anti-labelling device, as suggested in Saito (2016) for Japanese. In other words, KPs do not participate in labelling. This would also have interesting implications for the external argument, but I do not discuss these here for space reasons.

the merger of *was*, the DO moves and forms a constituent with TP, labelled through feature-sharing. PoS δ is labelled PartP.

The advantage of this analysis over (81) is that the RM problem is avoided, as the DO is smuggled over the EA. (83) is not without problems, however. Indeed, the complications of both smuggling and remnant movement which were seen in (70) remain. Furthermore, movement of the verb from Voice to Part requires it to cross earlier copies of itself, which seems complicated, if not problematic. In terms of empirical coverage (81) also fares better than (83). Indeed, the atomisation approach can be extended to sentences such as (84).

(84) The book was given by Mary to John

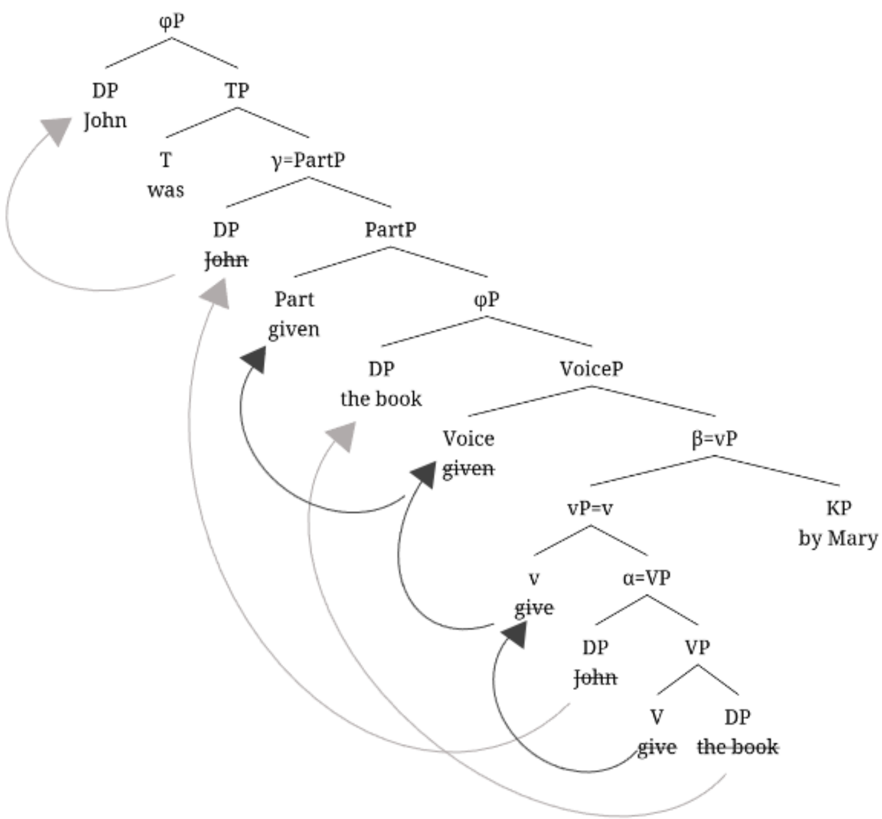
Such sentences can be derived if we say that movement of [$_{KP}$ to John] is optional. The smuggling approach cannot account for such sentences, since the KP necessarily moves above the EA. Nevertheless, RM violations are a significant concern, meaning that the smuggling analysis still seems theoretically superior.

5.2.2 The IO construction in the passive

(75), repeated in (85), shows the IO construction in the passive. Smuggling and atomisation approaches to this construction are practically identical to (83) and (81) respectively, the only difference being in where the arguments are merged. (86) represents the atomisation approach.

(85) John was given the book by Mary

(86)



Aside from the problems of (81), which obviously remain, a comment is worth making here. If the feature-sharing between DP and VoiceP involves Case, allowing Accusative and Dative feature sharing to occur in the same configuration seems problematic. The obvious solution would be to suggest that the relevant Voice category is slightly different in each case.

6 Conclusion

In this dissertation, I have argued for an EDA analysis of passivisation in English. First, I showed that EDA is a promising approach to movement with many theoretical advantages. Second, I surveyed existing theories of passivisation, demonstrating that argument approaches are the most theoretically elegant. Third, I discussed various empirical data, showing that argument approaches once again provide the best solution, although empirical issues remain. In the fourth section, I sketched two possible preliminary analyses of passives of transitives and ditransitives, taking insights both from previous approaches to passives and from EDA. The analyses translate den Dikken's (2020) idea about canonical and reverse predication into an EDA framework and make use of Fowlie's (2013) ATOMISE and smuggling as in Collins (2005). I concluded that the smuggling approach fares better than the atomisation one theoretically, as the atomisation analysis provides no satisfactory solution to RM problems. Nevertheless, the smuggling approach has some theoretical complications and may not provide as much empirical coverage. For these reasons, it would be worth refining this analysis to account for its issues. Alternatively, it could be interesting to consider approaches such as den Dikken (2020) and Bowers (2010), in which the EA is generated below the IA, in more depth, despite them not conforming to traditional assumptions about clause structure.

For space reasons, there is a lot which I have been unable to discuss. It would be interesting to examine what my approaches predict in terms of specific c-command relations, particularly when it comes to the issues faced by Collins' approach. Moreover, as my analysis is restricted to specific constructions in a single language, it would be worthwhile to attempt to extend this approach to other structures and languages to broaden its empirical coverage. In terms of English constructions, it would be interesting to examine pseudopassives; cross-linguistically, impersonal passives would be of interest. Lastly, I believe that Fowlie's (2013) ATOMISE has promising applications to EDA, as it effectively turns an XP into an X; it would be worth examining the implications of this in more depth.

7 References

Alexiadou, A., Anagnostopoulou, E., & Schäfer, F. (2015). *External Arguments in Transitivity Alternations: A Layering Approach*. Oxford: Oxford University Press.

Angelopoulos, N., Collins, C., & Terzi, A. (2020). Greek and English passives, and the role of by-phrases. *Glossa*, 5, 1–29.

Baker, M. (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Baker, M., Johnson, K., & Roberts, I. (1989). Passive Arguments Raised. *Linguistic Inquiry*, 20, 219–251.

Belletti, A. (2017). Labeling (Romance) Causatives. In E. Aboh, E. Haeberli, G. Puskás, & M. Schönenberger (Eds.), *Elements of Comparative Syntax: Theory and Description* (pp. 13–46). Berlin: De Gruyter.

Belletti, A., & Collins, C. (Eds.). (2020). *Smuggling in Syntax*. New York: Oxford University Press.

Bhatt, R., & Pancheva, R. (2006). Implicit Arguments. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (pp. 558–588). Chichester: Wiley-Blackwell.

Bošković, Ž. (2015). From the Complex NP Constraint to everything: On deep extractions across categories. *The Linguistic Review, 32*, 603–669.

Bošković, Ž. (2016). On the timing of labeling: Deducing Comp-trace effects, the Subject Condition, the Adjunct Condition, and tucking in from labeling. *The Linguistic Review, 33,* 17–66.

Bošković, Ž. (2020). On smuggling, the freezing ban, labels, and tough-constructions. In A. Belletti & C. Collins (Eds.), *Smuggling in Syntax* (pp. 53–95). New York: Oxford University Press.

Bowers, J. (2010). *Arguments as Relations*. Cambridge, Massachusetts: MIT Press.

Bruening, B. (2013). By Phrases in Passives and Nominals. *Syntax*, 16, 1–41.

Burzio, L. (1986). *Italian syntax: A government-binding approach*. Dordrecht: Reidel.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1981). *Lectures of government and binding*. Dordrecht: Foris Publications.

Chomsky, N. (1995). *The minimalist program*. Cambridge, Massachusetts: MIT Press.

Chomsky, N. (2013). Problems of projection. *Lingua*, 130, 33–49.

Chomsky, N. (2015). Problems of Projection: Extensions. In E. Di Domenico, C. Hamann, & S. Matteini (Eds.), *Structures, Strategies and Beyond: Studies in honour of Adriana Belletti* (pp. 29–42). Amsterdam: John Benjamins Publishing Company.

Chomsky, N. (2021). Minimalism: Where Are We Now, and Where Can We Hope to Go. 言語研究 *(Gengo Kenkyu)*, 160, 1–41.

Citko, B. (2011). Small Clauses. *Language and Linguistics Compass*, 5, 748–763.

Collins, C. (2005). A Smuggling Approach to the Passive in English. *Syntax*, 8, 81–120.

Collins, C. (2018a). *Is the Passive by-Phrase an Adjunct?* [Unpublished manuscript]. New York University. Retrieved from: <lingbuzz/004032>.

Collins, C. (2018b). *On the Phi-Features of the Implicit Argument in the Passive*. [Unpublished manuscript]. New York University. Retrieved from: <lingbuzz/004331>.

Collins, C. (In press). *Principles of Argument Structure: A Merge-Based Approach*. Retrieved from: <lingbuzz/006409>.

Collins, C., & Seely, T. (2020). *Labeling without Labels*. [Unpublished manuscript]. New York University, Eastern Michigan University. Retrieved from: <lingbuzz/005486>.

den Besten, H., & Webelhuth, G. (1987). *Adjunction and Remnant Topicalization in the Germanic SOV-Languages*. Presented at the Generative Linguists of the Old World conference, March 30–April 2, Venice.

den Dikken, M. (2020). Canonical and reverse predication in the syntax of the active/passive diathesis alternation. In A. Belletti & C. Collins (Eds.), *Smuggling in syntax* (pp. 147–187). Oxford: Oxford University Press.

Fowlie, M. (2013). Multiple multiple spellout. In T. Biberauer & I. Roberts (Eds.), *Challenges to Linearization* (pp. 129–170). Boston: De Gruyter, Inc.

Fukui, N. (2011). Merge and Bare Phrase Structure. In C. Boeckx (Ed.), *The Oxford Handbook of Linguistic Minimalism* (pp.73-95). Oxford: Oxford University Press.

Fukui, N., & Narita, H. (2014). Merge, labeling, and projection. In A. Carnie, Y. Sato, & D. Siddiqi (Eds.), *The Routledge Handbook of Syntax* (pp. 3–23). Abingdon: Routledge.

Grewendorf, G. (2015). Problems of remnant movement. In G. Grewendorf (Ed.), *Remnant movement* (pp. 3–32). Berlin: De Gruyter.

Harley, H. (2002). Possession and the double object construction. *Linguistic Variation Yearbook*, 2, 31–70.

Hsieh, F., & Sybesma, R. (2011). On the Linearization of Chinese Sentence-final Particles: Max Spell Out and Why CP Moves. 중어중문학, 49, 53–90.

Jaeggli, O. A. (1986). Passive. *Linguistic Inquiry*, 17, 587–622.

Kayne, R. S. (1994). *The antisymmetry of syntax*. Cambridge, Massachusetts: MIT Press.

Legate, J. A. (2014). *Voice and v: Lessons from Acehnese*. Cambridge, Massachusetts: MIT Press.

Legate, J. A., Akkuş, F., Šereikaitė, M., & Ringe, D. (2020). On passives of passives. *Language*, 96, 771–818.

Marantz, A. (1984). *On the Nature of Grammatical Relations*. Cambridge, Massachusetts: MIT Press.

Moro, A. (1997). Dynamic Antisymmetry: Movement as a Symmetry-breaking Phenomenon. *Studia Linguistica*, 51, 50–76.

Moro, A. (2000). *Dynamic Antisymmetry*. Cambridge, Massachusetts: MIT Press.

Moro, A., & Roberts, I. (2021). *Unstable structures and Generalized Dynamic Antisymmetry*. [Unpublished manuscript]. IUSS Pavia, University of Cambridge. Retrieved from: <lingbuzz/005515>.

Müller, G. (2010). On Deriving CED Effects from the PIC. *Linguistic Inquiry*, 41, 35–82.

Nunes, J., & Uriagereka, J. (2000). Cyclicity and Extraction Domains. *Syntax*, 3, 20–43.

Pesetsky, D. (1995). *Zero syntax: Experiencers and cascades*. Cambridge, Massachusetts: MIT Press.

Poletto, C., & Pollock, J.-Y. (2015). Arguing for remnant movement in Romance. In G. Grewendorf (Ed.), *Remnant movement* (pp. 135–178). Berlin: De Gruyter.

Postal, P. M. (2004). *Skeptical Linguistic Essays*. New York: Oxford University Press.

Reinhart, T., & Siloni, T. (2004). Against the unaccusative analysis of reflexives. In A. Alexiadou, E. Anagnostopoulou, & M. Everaert (Eds.), *The Unaccusativity Puzzle* (pp. 159–180). Oxford: Oxford University Press.

Rizzi, L. (2014). Some Consequences of Criterial Freezing. In P. Svenionius (Ed.), *Functional Structure from Top to Toe: The Cartography of Syntactic Structures* (Vol. 9, pp. 19–54). New York: Oxford University Press.

Rizzi, L. (2015). Cartography, Criteria, and Labeling. In U. Shlonsky (Ed.), *Beyond Functional Sequence: The Cartography of Syntactic Structures* (Vol. 10, pp. 314–338). New York: Oxford University Press.

Roberts, I. (1987). *The Representation of Implicit and Dethematized Subjects*. Berlin: De Gruyter Mouton.

Roberts, I. (2019). *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.

Ross, J. R. (1967). *Constraints on Variables in Syntax* [Unpublished doctoral dissertation]. MIT.

Saito, M. (2016). (A) Case for labeling: Labeling in languages without ϕ-feature agreement. *The Linguistic Review*, 33, 129–175.

Schäfer, F. (2012). The passive of reflexive verbs and its implications for theories of binding and case. *The Journal of Comparative Germanic Linguistics*, 15, 213–268.

Codeswitching in the Pacific Northwest: A Sociophonetic Analysis of Mexican American Speakers in the Yakima Valley

Diana V. Davidson
University of Washington

Abstract. Numerous works (Poplack, 1980; Myers-Scotton, 1997; Muysken, 2000; Torres Cacoullos and Aaron, 2003; Aaron 2015) have shown that noun phrases are the most often switched lexical items — most commonly the insertion of bare nouns and bare noun phrases. Blom and Gumperz (1972) defined metaphorical codeswitching as constituting language switches that relate to types of topics of conversation. While studies have focused on English-Spanish codeswitching in places like New York (Poplack, 1980) and Southern California (Fought, 2003), little work has been done on codeswitching in the Pacific Northwest. The Yakima Valley in Eastern Washington has a long and rich history of Mexican Americans forming a sizable part of the community since the 1930s (Gamboa, 1981). By using conversational speech data from the larger Pacific Northwest English (PNWE) study (Wassink, 2015), I examined potential phonological incorporation of Spanish phonetics in intrasententially-codeswitched sentences, focusing on the variation of /ɛ/ and /eɪ/ on stressed syllables and potential monophthongization of /eɪ/ in English, along with the degree of phonological incorporation of Spanish allophones of /t/ (like apico-dental [t̪]). I also examined the types of words being codeswitched and what topics of conversation exhibit codeswitching. This study hypothesized English-origin words with /eɪ/ in the same sentence as a Spanish codeswitched token will undergo monophthongization; that English-origin words with /t/ in the same sentence as a Spanish codeswitched token will become apico-dental; nouns will be the most frequently codeswitched token; and topics of conversation highlighting the speakers' ethnic identity will pattern with codeswitching (Fought, 2003).

Keywords: sociophonetics; sociolinguistics; codeswitching; Spanish, Mexican American

---

1 Research Goals

The aims of the research are broken down into several parts. Firstly, one aim is to investigate a relatively understudied variety of American English – namely Pacific Northwest English, specifically English spoken by Mexican American speakers living in the Yakima Valley. This study hopes to add to the body of work done on codeswitching by looking at data from an underrepresented and understudied region. Looking at the integration of Spanish phonology into English speech, and codeswitching could shed light on the influence of Spanish in the Yakima Valley and what topics the codeswitches highlight.

2 Background

2.1 Literature Review

Poplack's (1980) seminal study of Spanish/English bilingual New York Puerto Ricans' speech was one of the first studies that characterized codeswitching in a structural way. Poplack denoted intersentential codeswitching (switching from one language to another across sentence boundaries) and intrasentential codeswitching (switching from one language to another within a sentence). Poplack found that speakers who were 'true bilinguals' tended to use more intrasentential codeswitching than those who were more dominant in one language versus another. Numerous works (Poplack, 1980; Myers-Scotton, 1997; Muysken, 2000; Torres Cacoullos and Aaron, 2003; Aaron, 2015) have shown that noun phrases are the most often switched lexical items — most commonly the insertion of bare nouns and bare noun phrases. An example of an NP switch from Aaron (2015) is *mucha gente juega muchos cards ahi en las casas, no?* ('lots of people play a lot of cards there in the houses, right?') (p. 460). Myers-Scotton (1997) characterizes codeswitching with a matrix language and embedded language dichotomy. The matrix language (ML) is the language that dictates morphosyntax, while the embedded language (EL) contains content morphemes from the EL to be inserted into ML frames. Muysken (2000) interprets code-switching, or what Muysken calls code-mixing, as a continuum of insertion, alternation, and congruent lexicalization through the lens of sociolinguistic and linguistic factors. Insertion, constituents from one language inserted into the frame provided by another language (much like Myers-Scotton's ML and EL framework), is frequent in colonial settings and recent migrant communities, where there is a considerable asymmetry in speakers' proficiency in the two languages. Blom and Gumperz (1972) pioneered the analysis of codeswitching in terms of interaction. Situational codeswitching is used when the definition of the social event changes in the same social setting, while metaphorical codeswitching 'allows for the enactment of two or more different relationships among the same set of individuals' (p. 127) and the social rules are related to the function of language.

Fought's (2003) book on Chicano English gives a comprehensive look at what Chicano English is and is not, specifically in a Los Angeles context, considering Mexican American immigration to the United States, the history of a dialect that grew from contact between English and Spanish, and social contexts that could interact with usage of Chicano English (p. 9). Chicano English is not the same as Spanglish; Chicano English speakers can go without uttering a single word of Spanish, but the occasional Spanish word can be seen to occur in Chicano English.

Chicano English, as defined by Fought (2003), is a 'non-standard variety of English, influenced by contact with Spanish, and spoken as a native dialect by both bilingual and monolingual speakers' (p. 1). Chicano English is only spoken by native English speakers, but those speakers could either be monolingual English speakers or bilingual in English and Spanish (Santa Ana, 1993, p. 15). A native English speaker in this definition means having acquired English during the critical period of language acquisition, whether it be at home or through schooling. Chicano English speakers can be described as those with Mexican ancestry; they share a 'historical and cultural provenance' (Santa Anna, 1993, p. 22), meaning that Chicano English speakers share a common history and share cultural practices. Chicano English is a sociolect of English that can have influence and/or incorporation from Spanish at many levels of the grammar, whether phonetic, syntactic, prosodic, lexical, or otherwise (Fought, 2003).

Although many elements of Chicano English can reflect the influence of Spanish, other elements of Chicano English may come from contact with other dialects (Fought, 2003, p. 63). In the case of Chicano English in the Yakima Valley, that could mean there is phonetic influence from the Anglo population, the Indigenous Yakama population, or other ethnic groups living in the region.

Fought (2003) didn't observe a lot of codeswitching in interviews, but she noticed emblematic switches intended to highlight ethnic identity in the speech of Chicano English speakers (p. 158). Emblematic switches denote words that peer into one's ethnic identity: Fought found this in kinship terms like when speakers referred to their *nina* ('godmother') (2003, p. 158). A 45-year-old man exhibited an emblematic switch in the sentence 'but I am the only one that came out *músico'* ('musician') (Fought, 2003, p. 159). The man referred to himself as a *músico* to highlight what it means for him to be a musician with specific ties to his Chicano American identity. Fought found that while codeswitching was frequent when Spanish was the matrix language, there were notably few instances of codeswitching when English was the matrix language. These speakers who switched from English to Spanish were a small group of older speakers from around the age of 45 (Fought, 2003, p. 159).

More generally, phonological features of Chicano English can include less frequent vowel reduction, frequent lack of glides, tense realization of /ɪ/, consonant cluster reduction, and glottalization of final voiceless stops (Fought, 2003, p. 64-69). Phonetic variation that is associated with Chicano English by Los Angeles speakers includes /æ/ backing, /æ/ raising, and /u/ fronting (Fought 2003, p. 126). Fought (2003) found that bilinguals didn't differ from monolinguals in their use of any of the /u/ fronting or /æ/ backing or raising phonetic variables and has yet to identify any phonological or phonetic element that distinguishes the two groups (p. 137). This finding was something to consider as I did further analysis with /t/, /ɛ/ and /eɪ/.

According to Fought (2003), apicodentals were shown to be a feature of Chicano English, but they were mostly used to substitute for interdental fricatives (p. 67). Examples of this phenomenon include 'something' (/səmθɪŋ/) realized as [səmt̪ʰɪn] and 'then'(/ðɛn/) realized as [d̪ɛn] Some speakers in Fought's (2003) study showed this variation almost categorically, along with being a feature that was prominent in Chicano English speakers who showed very few of Chicano English's syntactic features (p. 67).

Torres Cacoullos and Aaron's (2003) study examined the nature of lone English-origin nouns (LEON) in bilingual Spanish-English speech from New Mexico. Torres Cacoullos and Aaron (2003) concluded that since the English-origin words were behaving grammatically like Spanish even if there was English phonology in the word, that the LEONs were a strong proponent of Sankoff et al.'s (1990) nonce loan hypothesis. Sankoff et al. (1990) defined 'nonce' borrowing as a type of borrowing that does not require phonological incorporation or widespread use within an individual or community (p. 74). The nonce loan hypothesis stated that 'loanwords are distributed quantitatively among syntactic slots in the same way as native words' and that the 'variability of morphological marking is statistically parallel for borrowing and native words' (p. 76).

2.2 History of Mexican Americans in the Pacific Northwest

Mexican people migrated to the Pacific Northwest as early as the 1800s (Gamboa, 1981, p. 121). The migration, although fluctuating over the decades, was linked the Southwest and the Northwest through livestock drives and other economic activities (Gamboa, 1981, p. 121). During World War II, 'unprecedented numbers of Mexican people were systematically recruited' to Washington states to work in the agriculture sector as braceros from Mexico and the American Southwest (Gamboa, 1981, p. 121). The *braceros* took the position that white farmers left to pursue more lucrative jobs like manufacturing or transit, or pursue 'national defense jobs' (Gamboa, 1981, p. 121). The *braceros'* labour was part of a bilateral agreement between the United States and Mexico, allowing bracero labor as an emergency wartime measure — they were a short-term solution to a longstanding problem of agriculture worker shortage in hop crops (Gamboa, 1981, p. 124).

Communities in the Yakima Valley were 'ambivalent toward the *braceros'*; the *braceros* faced blatant discrimination in public spaces and yet were also celebrated for their economic contribution to the agricultural sphere (Gamboa, 1981, p. 126). After the war, agricultural production went down, and the prices of hops, asparagus, and sugar beets rose (Gamboa, 1981, p. 127). *Chicano* laborers (migrants mostly with families) had clear advantages over braceros because they offered farmers stable labor force and freed them of annual practice of contracting Mexican nationals (Gamboa, 1981, p. 127). Many of the Chicano migrants who came to the Yakima Valley during the 1940s came from Colorado or Wyoming and Texas, states that had a surplus of Chicano workers (Gamboa, 1981, p. 128). Tejanos generally settled in Granger and Sunnyside, while Chicanos from Colorado and Wyoming settled in Toppenish and Wapato (Gamboa, 1981, p. 128). The southern and rocky mountain Chicano groups formed two separate social communities of Chicanos, each with distinct cultural activities and kinship

(Gamboa, 1981, p. 129). Even though there are few public records demonstrating the presence of the Chicano communities in the Yakima Valley, an increased number of Spanish-speaking families in the Yakima diocese of the Catholic church and the labor camps in the 1940s serve as presence of Chicanos (Gamboa, 1981, p. 129). The 'social, political, and economic alienation from established Yakima Valley communities only served to increase the need' for Chicanos to have their own cultural practices and activities, and the role of 'familia' ("family") filled that gap, as well as being a 'transmitter of culture' (Gamboa, 1981, p. 130). The Chicanos who moved to the Pacific Northwest were of lower socioeconomic classes — with their migration 'their economic and social status did not change appreciably' (Gamboa, 1981, p. 130). 1950 brought the first Spanish-language program airing in Washington state, the first Spanish-speaking Catholic priest in the Yakima Valley, and the first tortilla factory in 1951 (Gamboa 1981, p. 130). Changes in agricultural practices and immigration laws in the 1980s encouraged many former migrant workers to settle permanently in the Yakima Valley (Kershner, 2009). By the 1980s, Yakima County's Hispanic/Latino population hit 14.8%, the largest percentage of any county in Washington state at the time (Kershner, 2009). By 2000, 33% of the population were Hispanic/Latino, compared to 7.5% statewide (Kershner, 2009). Mexicans and Mexican Americans living in the Yakima Valley have and continue to play a crucial role in the community.

3 Research Questions and Hypotheses

My research questions and hypotheses are three-fold, as they relate to my phonetic variables of interest.

a. What kind of variation of /ɛ/ and /eɪ/ on stressed syllables is exhibited between heritage and receptive speakers of Spanish – in their English speech? In what environments would there be possible /eɪ/-monophthongization (a possible sign of Spanish phonology in English) if any? My hypothesis is that English-origin words with /eɪ/ in the same sentence as a Spanish-origin word/codeswitched token will undergo monophthongization.

b. To what degree is there phonological incorporation of Spanish in the English speech of Mexican Americans in the Yakima Valley – in terms of /t/ word-medially and word-finally? In Spanish origin words like plato ("plate"), Centro de la Raza (Spanish name of an organization) and Latino, [t̪] is not uncommon as alveolars in Spanish are often apico-dental. My hypothesis is that English-origin words with /t/ in the same sentence as a Spanish-origin word/codeswitched token will become apico-dental.

c. What tokens are being codeswitched, if any? Which topics of conversation exhibit codeswitching? My hypothesis is that consistent with prior research, nouns will be the most frequently codeswitched item, and topics of conversation that highlight the speakers' ethnic identity will pattern with codeswitching, as found by Fought (2003).

4 Methodology

For each of the speakers' sound files, I used the Praat pre-processing script 'clox-preprocessing-trim-silences-1' revised by Rob Squizzero on April 24, 2020 and modified by Jake McManus in May 2020. I used CLOx 2.0 (Wassink et al., 2018), an automatic speech recognition system powered by Microsoft, to generate text output. I then took the generated .txt files, aligned them with the corresponding sound file, and made separate ELAN files with for each segment by using ELAN Linguistic Annotator version 5.9.

After going through error coding analysis of each conversational segment of the interviews, I generated the utterances that had Spanish words included in them into a separate Excel spreadsheet called Codeswitching Tokens, organized by speaker. As a native speaker of Spanish, I was able to judge when Spanish words were introduced into the otherwise English speech, along with when Spanish phonology was potentially influencing English words. However, as I am a native speaker of Argentine Spanish, there were some words introduced in the conversation portion like *mana* ‘sister’, (short for *hermana*), *brecas* ‘breaks’, and *maneas* ‘mannerisms’ specific to Mexican Spanish, where I referred my intuitions to a native speaker of Mexican Spanish.

I then went back through the sound files where each speaker said a Spanish word and located it in the Praat grid. From there I generated narrow transcriptions of each Spanish word uttered and the surrounding words in the sentence. I focused on the nature of /t/ in Spanish words like ‘Districto Federal’, *toros* ‘bulls’, *Puerto Rico, plato* ‘plate’, *traste* ‘cooking utensil’, *tina* ‘dishes’, *Latinos, Centro de la Raza* (name of an organization in Washington state), along with English words with word-medial and word-final /t/ that occurred in the sentences with Spanish codeswitches.

For the /ɛ/ and /eɪ/ portion of the study, I looked at Spanish-origin words introduced in the otherwise English speech of the Mexican American speakers, along with English words with /ɛ/ and /eɪ/ in the sentence where codeswitching took place. I took the 50% point and the 80% point for the two vowels to look for potential smoothing of /eɪ/ and plotted them into Norm. Along with plotting /ɛ/ and /eɪ/ for the speakers who codeswitched, I created vowel plots on Norm for the speakers who codeswitched. For the purpose of these proceedings, I recreated the vowel plots in R for better legibility. The vowel plots can be found in the Appendix (Section 9).

5 Speaker Demographics

Table 1: *List of Mexican American speakers with conversation analysis completed*

Speaker	Location	Ethnicity	Gender	Generati-on	Languages Spoken at Home	ESP words present
YY35HF3A	Yakima	Hispanic	F	3	ESP, ENG	No
YY36HF3B	Yakima	Hispanic	F	3	Primarily ESP, some ENG	No
YY37HF2C	Yakima	Hispanic	F	2	ENG	No
YY38HF2D	Yakima	Hispanic	F	2	ENG	Yes
YY39HF3D	Yakima	Hispanic	F	3	ENG, ESP	Yes
YH47HM3H	Harrah	Hispanic	M	3	ESP	No
YH48HF3H	Harrah	Hispanic	F	3	ENG (mother) ESP (father)	No
YY60HF2I	Yakima	Hispanic	F	2	ESP, ENG	Yes
ESP71HF2R	Spokane	Hispanic	F	2	ENG, ESP	No
YUG127HF2AB	Union Gap	Hispanic	F	2	ENG, ESP	No
YT129HM2AC	Toppenish	Hispanic	M	2	N/A (no demographic information collected)	Yes
YY131HF1AD	Yakima	Hispanic	F	1	ESP	Yes
YY132HF1AD	Yakima	Hispanic	F	1	ENG, ESP	Yes

Speakers who uttered Spanish words with [t] in otherwise English speech were YY60HF2I, YUG127HF2AB, YT129HM2AC and YY132HF1AD. Speakers who uttered Spanish words that did not have /t/ were YY38HF2D, YY39HF3D, and YY131HF1AD, as seen in Table 2. Speakers who uttered Spanish words with [ɛ] in otherwise English speech were YY38HF2D, YY39HF2D, YY60HF2I, YY129HM2AC, YY132HF1AD.

Table 2: *Speakers who used Spanish words in conversational speech*

Speaker	Spanish word w/ /t/ (ESP phon)	Spanish word w /ɛ/ (ESP phon)	Other Spanish words (ESP phon)	Spanish word w/ ENG phonology
YY35HF3A	N/A	N/A	N/A	Mexico [mɛksɪkoʊ]
YY36HF3B	N/A	N/A	N/A	Mexico [mɛʔsɪkoʊ]
YY38HF2D	*Toros* 'bulls', Puerto Rico	N/A	*Charro* [ʧaro], *mana* 'sister' [mana], *aqui* 'here' [aki], *dolor* 'pain' [dolor]	Rica [ɹikə] (second half of Costa Rica)
YY39HF3D	N/A	N/A	Los Oros [los oɾos]), Cinco de Mayo [sinko dɛ maɪjo]	Mexico [mɛksɪko]
YY60HF2D	*Que tal* 'how are you', 'Distrito Federal' [de Mexico] (CDMX)	*En* 'in', *alberca* 'swimming pool', *maneas*, **breicas* (*brecas*), Mexico	*La*, *piscina* [pisina] 'swimming pool', *las*	Antonio [æntʰoʊnjo]
ESP71HF2R	N/A	N/A	N/A	San Antonio [sæn æntʰoʊnjoʊ];
YUG127HF 2AB	*Latino, Latinos*	N/A	N/A	N/A
YY129HM2 AC	*Plato, tina, traste* (variations of 'plate')	N/A	N/A	N/A
YY131HF1 AD	*Latino*	N/A	Luis	Buena [bjuænə], Ramirez
YY132HF1 AD	*Latina, Centro de la Raza*	*De Raza* [dɛ raza], 'Centro de la Raza' [sɛnθro dɛ la rasa]	*Burros* 'bulls' [buros]	*Burros* [bʌɹoʊz] 'bulls', *Reyes* [ɹeɪæz]

Speakers who uttered Spanish-origin words with English phonetics were YY35HF3A, YY36HF3B, YY38HF2D, YY60FS2I, ESP71HF2R, YY131HF1AD, YY132HF1AD Some Spanish-origin words like 'Buena' (location), 'Antonio' (person's name), 'San Antonio' (location), and 'Mexico' (location), 'Reyes' (last name) were uttered in English phonology, along with the word *burro* ('bull'). What is interesting with *burro* is that YY132HF1AD said it first with Spanish phonology, and then later in the same sentence went on to say it with English phonology.

6 Results

6.1 /ɛ/ and /eɪ/

The speakers who codeswitched and had words with /ɛ/ or /eɪ/ on stressed syllables more than once were YY38HF2D, YY39HF3D, YY60HF2I, and YY132HF1AD. Although speakers YY129HM2AC and YY131HF1AD exhibited codeswitching, the Spanish word and the surrounding English words in the sentence did not contain /ɛ/ or /eɪ/, so the two speakers do not have Norm vowel plots specific to /ɛ/ or /eɪ/ production beyond the general vowel plots. Vowel plots can be found in the Appendix (Section 9).

In the general vowel plots of YY38HF2D, YY39HF3D, and YY60HF3D there is significant overlap between [i] and [eɪ] — in the three speakers [eɪ] is raised to the typical vowel space of [i] and exhibits clear examples of upgliding. YY38HF2D's plot that compares [eɪ] and [ɛ] tokens also shows that [eɪ] and [ɛ] inhabit the same general vowel space. Overall, the [eɪ] tokens are higher in the vowel space between 394 Hz and 476 Hz for F1 and between 2024 Hz and 2725 Hz for F2 at the 50% mark, and between 380 Hz and 470 Hz for F1 and between 1939 Hz and 2552 Hz for F2 at the 80% mark. For YY39HF3D, [eɪ] ranges from 430 and 473 Hz for F1 and between 2288 Hz and 2543 Hz for F2 at the 50% mark, and between 392 Hz and 464 Hz for F1 and between 2220 Hz and 2226 Hz for F2 at the 80% mark. For YY60HF2I, [eɪ] ranges from 411 and 447 Hz for F1 and between 2382 Hz and 2624 Hz for F2 at the 50% mark, and between 324 Hz and 424 Hz for F1 and between 2455 Hz and 2598 Hz for F2 at the 80% mark.

In the utterances where codeswitching was present, YY38HF2D had 11 relevant words with [eɪ] and [ɛ]; two Spanish-origin and 9 English words. The words that occupy the uppermost part of the vowel space for YY38HF2D are the Spanish-origin words 'en' ([ɛ]) with 476 and 479 Hz for F1 at the 50% and 80% mark and 'Mexico' ([ɛ]) with 603 and 630 Hz for F1 at the 50% and 80% mark; 'forget' ([ɛ]) at 422 and 577 Hz for F1 at the 50% and 80% mark; 'they' ([eɪ]) at 416 and 470 Hz for F1 at the 50% and 80% mark; 'they're' ([eɪ]) at 545 Hz and 542 Hz for F1 at the 50% and 80% mark; 'they' ([eɪ]) again at 394 Hz and 380 Hz for F1 at the 50% and 80% mark; 'they' for a third time at 452 Hz and 414 Hz for F1 at the 50% and 80%; and 'famous' ([eɪ]) at 476 Hz and 479 Hz for F1 at the 50% and 80% mark. In the lower end of the [eɪ] ~ [ɛ] vowel plot for YY38HF2D for the English words 'friends' ([ɛ]), 'dress' ([ɛ]), and 'else' ([ɛ]), the F1 values ranged from 746 to 828 Hz at 50% and 676 to 753 Hz at the 80% mark. These three words hover around the same vowel space as the same ([ɛ]) tokens from English words where there was no codeswitching in the sentence –- the two token words 'friends' and 'respect' had F1 at values at 877 Hz at the 50% and 755 Hz at the 80% mark, and 700 Hz at the 50% and 582 Hz at the 80% mark respectively. Therefore, for most of the sentences where codeswitching occurs, the F1 values of ([ɛ]) of the English words place the words significantly higher in the vowel space and shows phonological incorporation of Spanish phonology. Even with learning Spanish as a teenager without being exposed to it at home, YY38HF2D showed phonological incorporation of Spanish in English words.

While YY39HF3D did codeswitch in the sentence where she mentions 'there's the Cinco de Mayo stuff that happens downtown', the [ɛ] was shorter than 50 milliseconds, so that [ɛ] was excluded from analysis. In the utterance 'a lot of people don't get education from Mexico', the /eɪ/ in the stressed syllable of 'education' is monophthongised and produced as an [ɛ], as evident from the small difference in F1 and F2 at the 50% and 80% marks: 395 Hz and 405 Hz for F1 and 2654 Hz and 2561 Hz for F2. The 20 Hz difference in F1 and 97 Hz difference in F2 shows further evidence for monophthongisation of /eɪ/. Vowel plots for YY39HF3D can be found in the Appendix, Figures 8 and 9.

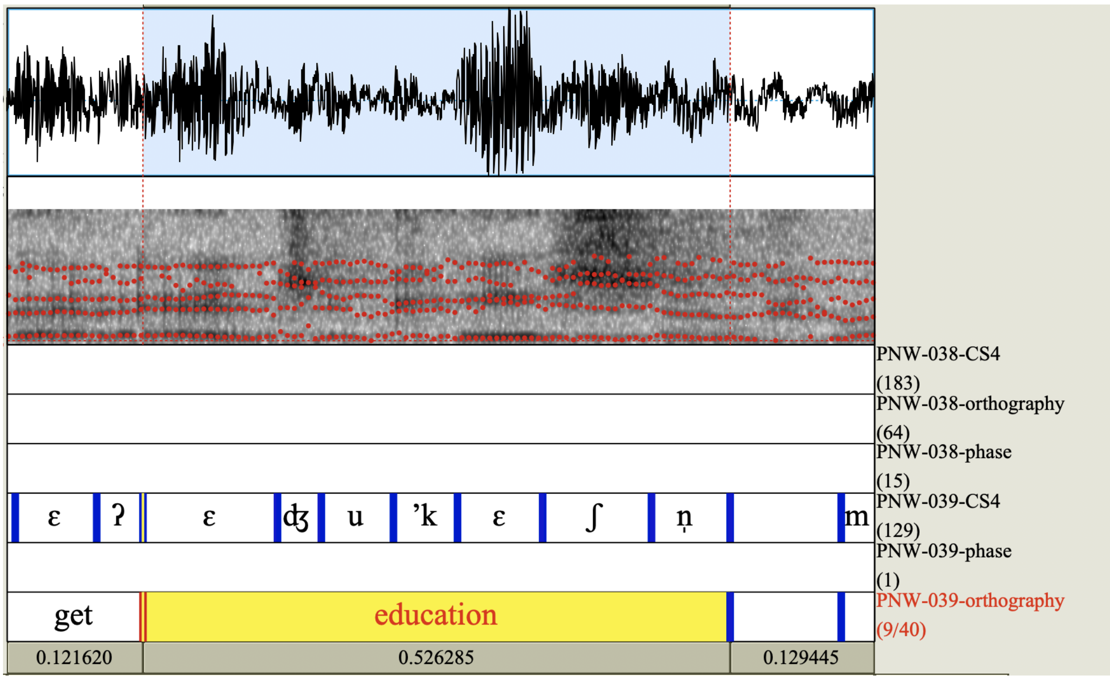


Figure 1: *YY39HF3D says the sentence 'A lot of people don't get their education from Mexico, so their Spanish is not like the correct type of Spanish.' This figure shows an example of /eɪ/ monophthongisation to [ɛ].*

Along with the /eɪ/ in 'education' monophthongising, the F1 for the 50% and 80% marks of the vowel locate the vowel particularly high in the vowel space: 395 Hz and 405 Hz respectively. Since Spanish /ɛ/ are more tense and raised compared to the English /ɛ/ — like an /e/ — it's safe to say that 'education' has phonological incorporation from Spanish. The word 'get' also has F1 values of [ɛ] at 534 Hz and 550 Hz for the 50% and 80% mark respectively. On the other hand, YY39HF3D's prototypical [ɛ], without any Spanish-origin words or codeswitching involved, hover lower and more central in the vowel space with F1 values for 'else' at 613 Hz and 563 Hz at the 50% and 80% mark and 'yet' at 678 Hz and 674 Hz at the 50% and 80% mark respectively. YY39HF3D said in the conversational portion of the interview that her first language was Spanish and both English and Spanish were spoken at home, but as soon as she went to school she started talking more in English.

YY60HF2I had 22 relevant tokens with /eɪ/ and /ɛ/ in utterances where she codeswitched. She said *Mexico* 6 times, along with other words containing /ɛ/ like 'DF' (pronounced 'de efe'), *en* ('in'), *maneas* ('mannerisms') and *que tal* ('how are you'). *Breicas** is the only Spanish token that has /eɪ/, although it's a disfluency of the word *brecas* ('break'). The [eɪ] in *breicas** occupies the same general vowel space as the English [eɪ] in the words 'they', 'say', 'Spain' and 'breaks', the latter word occurring in the same sentence as *breicas**. The difference of F1 and F2 in *breicas** is 44 Hz and 56 Hz respectively, so even though acoustically it sounds like a diphthong, formant values show that it could be monophthongal [ɛ]. The difference of F1 for the English words with [eɪ] were between 21 Hz and 92 Hz. In the instances that YY60HF2I said 'Mexico', F1 for [ɛ] differed by a range of 5 and 154 Hz, and F2 differed by a range of 10 Hz and 129 Hz. Overall, the Spanish tokens with [ɛ] sit higher in the vowel space than the English tokens, except for one production of *Mexico* with F1 at 695 Hz at 50% and 541 Hz at 80%. The English tokens of 'then' and 'whatever' sit in the same vowel space as the Spanish tokens, with F1 values at 469 Hz and 508 Hz at the 50% mark and 499 Hz and 490 Hz at the 80% mark respectively. The other two English [ɛ] tokens — 'next' and 'Texas' had F1 values at 690 Hz and 692 Hz at the 50% mark and 537 Hz and 558 Hz at the 80% respectively. There is some overlap between English productions of [ɛ] and Spanish productions of [ɛ], but not as much as compared to YY38HF2D and YY39HF3D. YY60HF2I is a heritage speaker of Spanish as she grew up with it at home, but she doesn't show as much phonological incorporation of Spanish in her English speech as YY38HF2D who learned Spanish as a teenager. Vowel plots for YY60HF2I can be found in the Appendix, Figures 10 and 11.

Unlike YY38HF2D, YY39HF3D, and YY60HF2I, YY129HM2AC doesn't show complete overlap between [i] and [eɪ]. Interestingly, YY129HM2AC's [i] is retracted compared to YY38HF2D, YY39HF3D, and YY60HF2, but his [eɪ] is around the vowel space of the three speakers: hovering around 400 Hz for F1 and the range of 2000-2500 Hz for F2. Even though YY129HM2AC exhibited codeswitching when he said, 'or you call a plate, you know some people in Spanish call it a *plato,* some people call it a *tina'*, the word 'plate' is the only sample of /eɪ/ within a codeswitched sentence. That sentence is also the only utterance in which YY129HM2AC codeswitched, so sufficient conclusions about phonological incorporation cannot be made with only one data point.

YY131HF1AC showed limited overlap between [i] and [eɪ] in her general vowel plot. Even in the general vowel plot, YY131HF1AC shows significant movement of (ɛ) toward the back of the vowel space. YY131HF1AC produced sentences where there were Spanish-origin words, but as they were pronounced with English phonology, like *Buena* [bjuænə] and *Ramirez*, the surrounding English words with /eɪ/ or /ɛ/ were not included in the analysis. Although YY131HF1AC did say 'Luis' with Spanish phonology, there was only one word in the partial sentence 'Luis said that this company from…' so the token [ɛ] in the word 'said' was also excluded from analysis. Vowel plots for YY131HF1AC can be found in the Appendix, Figure 12.

YY132HF1AD's [ɛ] from her general vowel plot ranged from 610 Hz to 633 Hz for F1 and 1712 Hz and 1939 Hz for F2 at the 50% mark, and 478 Hz to 641 Hz for F1 and 1771 Hz and 1882 Hz for F2 at the 80% mark. Her [eɪ] tokens ranged from 494 Hz to 596 Hz for F1 and 2482 Hz and 2610 Hz for F2 at the 50% mark, and 348 Hz to 358 Hz for F1 and 2619 Hz and 2826 Hz for F2 at the 80% mark. All the tokens except for one are Spanish words: *de* ('of') three times and *Centro* ('Centre') twice, with the lone token of [eɪ] from the English word 'raising' in the codeswitched sentence 'we're raising our kids like *burros'* ('bulls'). In the Spanish words, the range of F1 is between 463 Hz and 715 Hz at the 50% mark and 479 Hz and 581 Hz at the 80% mark. Although the difference in F1 in the word 'raising' is only 25 Hz, the difference in F2 is significant at 284 Hz, well within the range of perceiving a diphthong. YY132HF1AD's Spanish tokens are generally lower in the vowel space than YY38HF2D, YY39HF3D, and YY60HF3D, and overlap with her English productions of [ɛ].

Contrary to my hypothesis, /eɪ/ in codeswitched sentences rarely underwent monophthongisation. More notably, however, is that there was variable integration of Spanish [ɛ] in English words where a codeswitch had taken place — [ɛ] were raised and tensed for YY38HF2D, YY39HF3D, and YY132HF1AD.

6.2 /t/

From the sentences that contained Spanish-origin words and /t/ word-medially or word-finally, there were 67 tokens. Tokens include Spanish-origin words like *plato* 'plate'*, traste* 'plate'*, Latino,* and *Centro*, along with English words that contained /t/ word-medially or word-finally in the same utterance as the Spanish-origin word. Spanish words with word-initial /t/ like *tina* 'plate', *toros* 'bulls', and the first /t/ in *traste* 'plate' were left out of the analysis.

Table 3: *Frequency table of Spanish allophones and English allophones of /t/ based on position in the word. Allophones highlights in orange are Spanish allophones and allophones highlighted in blue are English allophones*

Allophone	Word-medial (WM)	Word-final (WF)	Total
[t̪]	0	9	9
[θ]	2	0	2
[tˣ]	1	0	1
[ʧ]	1	0	1
[t̚]	2	4	6
[t]	2	1	3
[ʔ]	1	15	16
[tʰ]	7	5	12
[ɾ]	7	5	12
[Ø]	0	5	5
Total	23	44	67

ESP71HF2R for example, said 'San Antonio' [sæn æntʰoʊnjoʊ] with English phonology. It's possible that San Antonio was pronounced that way because the name of the city in Texas is well-incorporated into the English lexicon. San Antonio came up in conversations about travel and differentiating the speaker's Pacific Northwest identity from people who live in San Antonio. Speaker ESP71HF2R exhibited English allophones of /t/ in both San Antonio and English words like 'that' [θæʔ], 'hot' [hɑʔ], and 'just' [ʤʌs] — glottal stops, consonant cluster simplification, and an unreleased [t̚] are to be expected word-finally. 'Concrete' [kɑnkrit̚] and 'get' [gɛt̚] end with an unreleased [t̚] word-finally, which can be expected in English. ESP71HF2R reported being able to 'somehow understand [her] grandma when she speaks to [her] in Spanish' and can understand it but doesn't report being able to speak Spanish herself.

YUG127HF2AB had three relevant tokens of /t/. When he pronounced 'Latino' and 'Latinos', the word-medial /t/ were clearly produced as [t̪], which is to be expected in Spanish. In a conversation about leisure where he said, "do you know that there was Latinos…", while there is the word-medial [t̪] in 'Latinos', in the word 'that', the word-final /t/ is a [ʔ], which is to expected word-finally in English. YUG127HF2AB didn't report being able to speak Spanish, but she reported listening to the radio in Spanish in her free time.

YT129HM2AC pronounced the ward *plato* 'plate' and *traste* 'plate' with [t̪] word-medially, just like how speaker YUG127HF2AB pronounced 'Latino'. Like speaker YUG127HF2AB, YT129HM2AC's /t/ in English words pattern with English phonology. 'It' [ɪɾ] and 'what' [wʌɾ] were pronounced with a tap due to the following word being a vowel: "Some people in Spanish call it a *plato*, some people call it a 'tina'… others call it a *traste* and you know, you're thinking what you are referring to." These Spanish words occurred in a conversation about Spanish language features. YY129HM2AC grew up with both English and Spanish at home and reported learning them at the same time and kept Spanish and English allophones of /t/ separate.

YY38HF2D has quite the range of /t/ allophones. The most frequent is the [ɾ] in English words like 'totally' [toʊɾəli] and 'bullfighting' [bʌlfaɪɾiŋ], but also word-finally in 'bit' as the following word — 'of' — started with a vowel and was in a spontaneous string of speech. Spanish-origin words like the first part of the country's names 'Puerto Rico' and 'Costa Rica' were pronounced with [t̪] and [ʧ] respectively. Like YT129HM2AC, YY38HF2D exhibited consonant cluster simplification in words like 'different' [dɪfɹn̩] in two instances, but also pronounced 'different' with an aspirated voiceless stop [tʰ] word-finally — both to be expected in English. While she learned Spanish as a teenager, she kept her Spanish and English allophones of /t/ separate. This is different from findings in the previous section about /ɛ/ and /eɪ/.

While YY39HF3D uses an emblematic switch in conversations about leisure and language features: "There's the Cinco de Mayo stuff that happens downtown…"; "a lot of people don't get education in *Mexico*, so their Spanish is not like the correct type of Spanish"; and "there's a big hall right here called *Los Oros* that…", the surrounding English words with /t/ word-medially and word-finally still retain expected English allophones. 'That', 'get' and 'right' all end with glottal stops [ʔ], 'stuff' has [t] word-medially, and the last syllable in 'correct' [kəɹɛk] undergoes consonant cluster simplification. The word 'right' in 'there's a big hall right here called 'Los Oros' that…' ended with [t̚]. There is little incorporation of Spanish allophones of /t/ in YY39HF3D's speech, which is the same from findings in the previous section about /ɛ/ and /eɪ/.

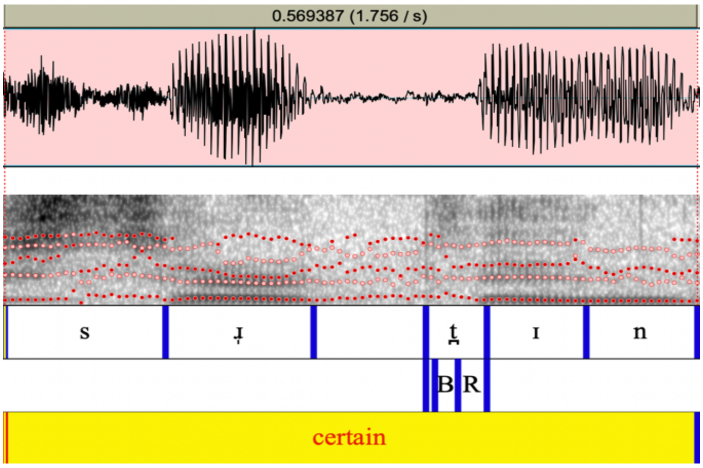


Figure 2: *Speaker YY60HF2I's first production of 'certain' (with apicodental [t̪]) shortly after 'Mexico', which retained Spanish phonology.*

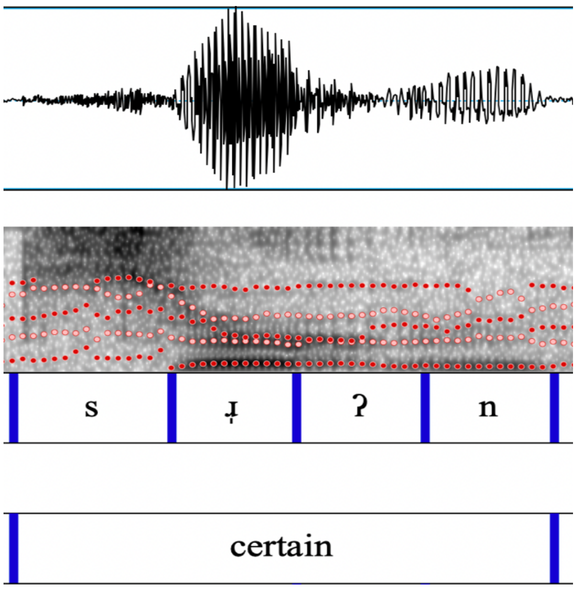


Figure 3: *Second production of 'certain' again, this time with an [ʔ].*

YY60HF2I shows similar patterns to other speakers, in terms of English words like ‘it’, ‘different’, ‘part’ and ‘went’ ending with [ʔ] in the word-final position. For Spanish-origin words like ‘Antonio’, the /t/ was produced as a [tʰ] rather than a [t̪], as would be expected for a Spanish-origin word. ‘Antonio’ [æntʰoʊnjo] is fully incorporated with English phonology. *Distrito* [ditˣito] in ‘Distrito Federal’ (Federal District, name for Mexico City), however, shows a mix of Spanish phonology and English phonology with [tˣ] and [t], respectively, word-medially. Speaker YY60HF2I shows alternations between Spanish and English iterations of /t/ when she repeats the word ‘certain’ in talking about language features related to Spanish: “You can tell by the way they say certain, certain things about how they’re explaining something.” The first time, she said ‘certain’ as [sɹ̩t̪ɪn], the second time as [sɹ̩ʔn].

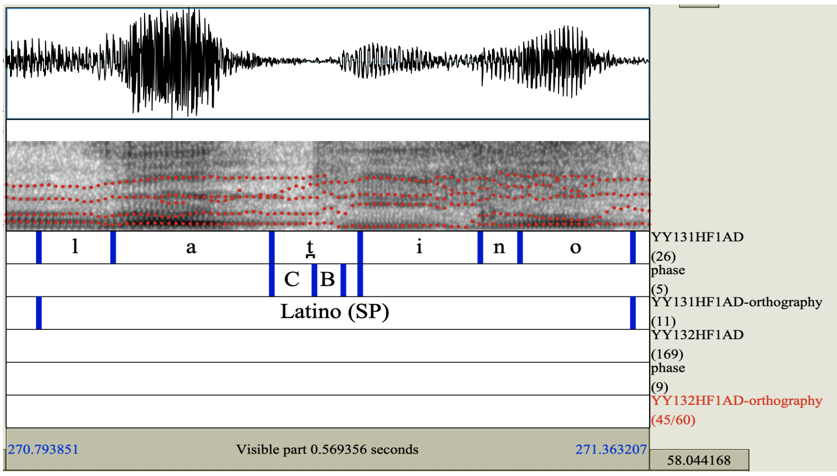


Figure 4: *YY131HF1AD’s production of ‘Latino’ in a dialogue with YY132HF1AD.*

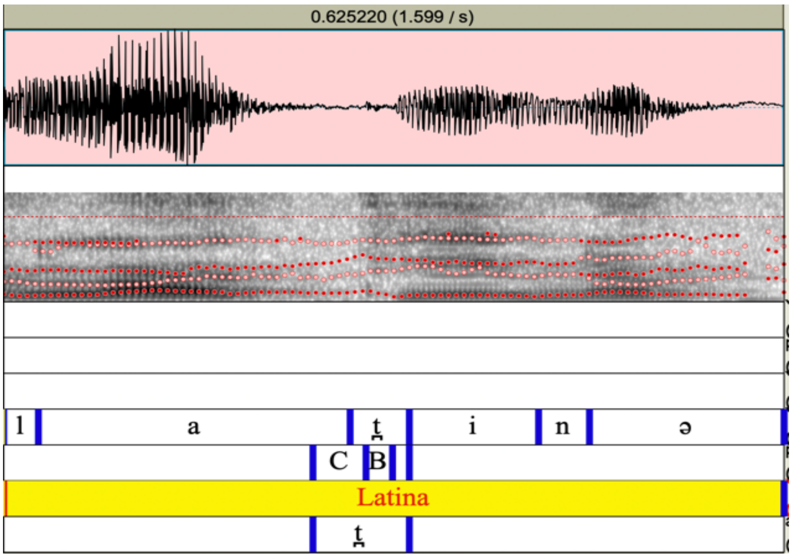


Figure 5: *YY132HF1AD’s production ‘Latina’ with a mix of English and Spanish phonology. The /t/ in both speakers are nonetheless [t̪] word-medially.*

With a sentence that has an emblematic switch such as “he helped organize, what was that, Hispanic, uh ‘Latino’ club?”, YY131HF1AD also shows English allophones of /t/ in words like ‘what’ and ‘that’, where word-finally the /t/ is a glottal stop [ʔ]. However, she also ends ‘helped’ [hɛlptʰ] with an aspirated voiceless stop [tʰ] like YY38HF2D. In this part of the interview, Speaker YY131HF1AD is talking about her family's history and community impact in the Seattle area. Her pronunciation of 'Latino' is clearly using Spanish phonology, as she says the word with a voiceless apicodental stop [t̪]. Spanish and English allophones of /t/ are kept separate for YY131HF1AD.

YY132HF1AD shows similar patterns to that of YY131HF1AD. In the same conversation about family history and community impact, YY132HF1AD says *Latina* with a voiceless apicodental voiceless stop [t̪] in a word-medial position, but also says *Centro* ('centre') in 'Centro de la Raza'(name of a community organization) with a voiceless interdental fricative [θ] in a word-medial position. Although *Latina* is produced with an apicodental voiceless stop [t̪], the final vowel is more like a schwa, showing partial incorporation of English in the Spanish-origin word, as can be seen in Figure 7. In sentences where YY132HF1AD pronounces words like *Buena* (location in Washington state) with English phonology, the English words surrounding the utterance "in Buena… And mom's kids went to school with the Wilcox kids! And my mom was just so impressed,' the /t/ in 'impressed' is aspirated word-finally. While YY60HF2I had an instance where Spanish phonology shows up in the first instance of an English word being uttered before she corrects herself, YY132HF1AD shows the opposite case where she says a Spanish word first and then produces it again with English phonology in the same sentence when talking about her family history: "We're raising our kids like burros, like *burros*. She says you can go, she says, but I want my kids in school." In this sentence where YY132HF1AD says *burros* 'bulls' first with Spanish phonology and then in English very soon after, the English word 'want' in the sentence retains English phonology in that word-finally the /t/ is a glottal stop [ʔ].

Table 4: *Frequency Table of the relationship between the topic of conversation and the types of /t/ allophones exhibited in the conversational portion of the interview*

	CS in utterance		Allophones		
Topic of conversation	**Yes**	**No**	**English**	**Spanish**	**Total**
Language features	24	12	28	8	36
Leisure	13	0	9	4	13
Family history	7	4	7	4	11
travel	0	2	1	1	2
Pacific NW Identification	0	4	3	1	4
Change in community	1	0	0	1	1
Total	45	22	48	19	67

As seen in Table 4, there were overall more English allophones of /t/ present, but that is due to most of the token words being English words. Consistent findings from Table 4 about the relationship between topic of conversation and codeswitched tokens, language features are the topic of conversation that brings about Spanish phonology in Spanish-origin words. However, YY60HF2I saying 'Antonio' [æntʰoʊnjo] and ESP71HF2R saying 'San Antonio' [sæn æntʰoʊnjoʊ] with aspirated voiceless alveolar stops [tʰ] shows full incorporation of English phonology in these Spanish-origin words. With the exception of 'certain' produced as [sɹ̩t̪ɪn], with a Spanish allophone of /t/, before being repeated as [sɹ̩ʔn] with an expected English of /t/, there was not any incorporation of Spanish allophones of /t/ in English words in a sentence that exhibited codeswitching. Consonant cluster simplification word-finally was commonly found across speakers; it is a widespread phenomenon in American English, including Chicano English.

Contrary to my hypothesis is that English-origin words with /t/ in the same sentence as a Spanish-origin word/codeswitched token rarely became apico-dental except for the instance of 'certain' for YY60HF2I. Spanish allophones and English allophones were largely kept separate.

6.3 Codeswitched Tokens

Consistent with previous research (Poplack, 1980; Myers-Scotton, 1997; Muysken, 2000; Torres Cacoullos and Aaron, 2003; Aaron 2015), an overwhelming majority of Spanish-origin words in the conversation portion of the interviews were nouns (55/62; 88.7%). Out of the 55 nouns, 39 were proper nouns and 16 were common nouns.

Table 5: Frequency table of Spanish-origin words and their corresponding topic of conversation in codeswitched sentences

Topic of conversation	Common Noun	Proper Noun	Article	Preposition	Discourse Marker	Adverb	Total
Language features	11	12	2	1	1	3	30
Family history	2	14	0	0	0	0	16
Leisure	3	7	0	0	0	0	10
Occupation	0	3	0	0	0	0	3
Change in community	0	1	0	0	0	0	1
PNW identification	0	1	0	0	0	0	1
Travel	0	1	0	0	0	0	1
Total	16	39	2	1	1	3	62

The topics of conversation that shows the most Spanish-origin words are language features, family history, and leisure. In the conversations about language features, even though the interviewers asked the Mexican American speakers specifically about whether they noticed differences in the way Pacific Northwesterners speak, the Mexican American speakers often responded by talking about specific features of Mexican Spanish being different from other Spanish-speaking countries, or even different regions of Mexico. This could point to the speakers more readily identifying as Mexican American than as a Pacific Northwesterner. In the conversations about language features, YY60HF2I talked about how one can tell by the way people 'say certain things' in 'Distrito Federal en Mexico' (Mexico City) and how some people say, "I went to *la alberca*, other people call it *piscina* ('swimming pool')." YY129HM2AC responded similarly to YY60HF2I when asked the same question; he noticed some people call a plate a *plato,* some people call it a *tina,* some call it a *traste* and that there can be some confusion as to what the different words refer to if it's not addressed between two people having a conversation. Speaker YY38HF2D mentioned how by just saying a few words like the pain, *aqui, dolor, aqui* 'here, pain, here' as a doctor talking to a patient, one is signaling that they know Spanish.

When asked about the speaker's family history (which ranged from being strictly related to one's family to how family members impacted the greater community), speakers mentioned Spanish-origin proper nouns like 'Mexico', 'Buena', 'Centro de la Raza', and San Antonio. In contrast with the Spanish-origin tokens that came up in conversation about language features, however, the proper nouns 'Mexico, 'Buena,' and 'San Antonio' were fully phonologically incorporated into English; only 'Centro de la Raza' retained its Spanish phonology. It's interesting to note that YY132HF1AD said 'Buena' [bjuænə] with English phonology, but later in the conversational portion said 'something *de Raza'* [dɛ raza] and 'Centro de la Raza' with its Spanish phonology ([sɛnθro dɛ la rasa]). Even in the same utterance, speaker YY132HF1AD said "we're raising our kids like *burros*… like *burros*" – the first time with Spanish phonology ([buros]) and the second with English phonology ([bʌɹoʊz]), possibly repeating the word with an English pronunciation so that the interviewer would understand.

YY38HF2AD was the speaker who used Spanish-origin words in the context of leisure-related conversations the most. She talked about how people in Yakima Valley wear *charro* ([ʧaro]) dresses and go horseback riding or bull fighting with *toros* ([t̪oros]). She also mentioned how a music group that appeared on TV are 'really famous *en Mexico*' ([ɛn mɛksiko]) and coming to Yakima. All these tokens were pronounced with Spanish phonology. YY39HF3D also talked about what people do for fun in the area, mentioning a big hall for rodeos called 'Los Oros' ([los oɾos]) and activities celebrating 'Cinco de Mayo' ([sinko de maɪjo]) downtown. YY36HF3B talked about the prevalence of soccer in the Yakima Valley comes from young Mexican Americans living in the area, but she said 'Mexico' ([mɛʔsɪkoʊ]) with English phonology.

I tallied up the topics of conversation in which Spanish-origins words occur in otherwise English speech. Relating to Blom and Gumperz' (1972) definition of metaphorical codeswitching, in which the language switch relates to particular kinds of topics or subject matter rather than a change in social situation, the use of Spanish-origin words in conversation highlights the cultural significance of Spanish — and the impact Mexican Americans have — in the Yakima Valley.

The pattern of code-switching happening with the Spanish-origin words closely match the insertion of constituents as proposed by Muysken's continuum of code-switching. The insertions are overwhelmingly nouns, and the matrix language in the code-switched utterances is clearly English; the Spanish nouns neatly fit within the English matrix frame. In the instances where more than a single Spanish-origin word is used, such as in utterances like 'you know, the breaks, *maneas'*, 'someone calling [it] *las *breicas*', 'really famous *en Mexico*', and 'DF which is 'Distrito Federal en Mexico', the preposition *en* ('in') and pronoun *las* ('the' pl. f.) are following their order according to Spanish syntax, and neatly fit within the English utterance. When trying to recall the name of a place called 'Centro de la Raza', YY132HF1AD at first says 'Yeah the 'Latina' club and the … something 'de Raza'?' before remembering and saying the whole name 'Centro de la Raza'. There is a clear omission of the article 'la' here and violates Spanish grammar.

Table 6: Frequency table of Spanish-origin words uttered and whether there was an emblematic switch present in an utterance[32]

Speaker	Spanish-origin word	Utterances with Spanish-origin tokens	Emblematic switch present	Non-emblematic switch present	No switch present
YY35HF3A	3	3	0	0	3
YY36HF3B	2	2	0	0	2
YY38HF2D	14	8	5	2	1
YY39HF3D	3	3	1	2	0
YY60HF2D	20	11	5	3	3
ESP71HF2R	4	4	0	0	4
YUG127HF2AB	2	2	2	0	0
YY129HM2AC	3	3	0	3	0
YY131HF1AD	3	3	0	1	2
YY132HF1AD	8	7	3	2	2
Total	62	46	17	14	15

The Spanish-origin words that most often did not induce a switch were proper nouns like 'San Antonio' (location), 'Mexico' (location), 'Buena' (location), 'Antonio' (first name), and 'Ramirez' (last name), but the common noun 'burro' ("bull") showed an instance of being pronounced with English phonology as well. Notably proper nouns were excluded from Torres Cacoullos and Aaron's (2003) LEON study as they cited Poplack et al. (1998) in stating that proper nouns may be treated differently than common nouns in terms of processes of integration. However, proper nouns like 'Latino', 'Latina', and 'Latinos' were always pronounced with Spanish phonology, namely apicodental [t̪] — by YY131HF1AD, YY132HF1AD, and YUG127HF2AB. As the word 'Latino' addresses the speakers' ethnic identity broadly, sentences with 'Latino' pronounced in Spanish count as emblematic switches. Other common nouns like *plato, traste, charro, maneas,* and *dolor* were always pronounced with Spanish phonology. When YY38HF2D talks about how a Mexican American friend of hers like to wear 'charro' dresses and that people in the community like to do the '*toro*, bull fighting', both intrasentential switches are emblematic as well. When YY38HF2D talks about how people can say the word *mana* 'sister' differently depending on where they're from, the use of this shortened version of *hermana* 'sister' is particular to Mexican Spanish and highlights the speaker's ethnic identity. Mentions of the 'Centro de la Raza' when talking about YY132HF1AD's family history and community impact highlights her ethnic identity as well, and she switched into Spanish when mentioning it. Utterances that I categorized as not having emblematic switches include YY132HF1AD's commentary about her mother saying she raised her kids like *burros*; YT129HM2AC's listing off different ways of saying plates or cooking utensils generally (*plato, traste, tina);* YY131HF1AD saying the name 'Luis' with Spanish phonology; YY38HF2D talking about a music group being really famous *en Mexico* 'in Mexico' and listing Puerto Rico as a place she's been to; and YY39HF3D's mention of a hall called 'Los Oros' and people not getting 'a good education in Mexico'.

The most frequent switches I saw from the data collected are from YY38HF2D and YY60HF2I with 14, YY132HF1AD with 5, YY129HM2AC and YY39HF3D with 3, YUG127HF2AB with 2, and YY131HF1AD with 1. The majority of the speakers who used emblematic codeswitching were in

[32] The table shows a count of the instances of Spanish-origin words, and some words (ie. Mexico, San Antonio) were used multiple times. The same utterance sometimes contained multiple Spanish-origin words, hence the utterance's column total not adding up with the Spanish-origin word column total.

generations 1 and 2 — meaning that speakers born between 1900-1950 and 1951-1976 respectively used emblematic codeswitching more frequently than those in the generation 3 (born in 1976 to age 18 at time of study). While YY39HF3D shows 3 instances of emblematic codeswitching, YY35HF3A and YY36HF3B – also in generation 3 — said Spanish-origin words with English phonology and didn't show any instances of emblematic switching in the conversational portion of their interviews. These results show similar findings to that of Fought (2003), who found the older generation surveyed (around 45 years old at the time of data collection) used Spanish switches as opposed to the teenagers who only very rarely used emblematic switches.

Volunteered attitudes of localness and connectedness to the community were collected through the conversational portion of the speakers' interviews. Attitude toward the Yakima Valley did not have an effect on codeswitching or phonological incorporation of Spanish into English words. YY38HF2D, who reported that the Yakima Valley was 'OK' and didn't really have anywhere else to compare to, while listing both positives and negatives of the area, showed the most switches and significant phonological incorporation of Spanish in English words containing [ɛ] in a codeswitched sentence. YY39HF3D reported generally positive attitudes toward the area but cited that gang violence had been around for a long time; she showed little phonological incorporation of Spanish into English words in a sentence where a codeswitch was present and only one utterance had an emblematic switch. YY60HF2I stated that she liked Yakima and that it's a good-sized town to live in, and she exhibits the most switches and minimal phonological incorporation of Spanish in English words: none for /ɛ/ and /eɪ/ and very little in /t/. YY131HF1AD reported the Yakima Valley being a nice place to live but didn't really elaborate on why that was the case. She is also the only speaker in the sample who is particularly active in the community — in her case the local church. YY131HF1AD didn't show signs of phonological incorporation of Spanish in English words and only had three utterances where a codeswitch was present. YY132HF1AD said that the Yakima Valley 'has its issues but it's a good place to live', that there isn't enough to do in the area, but there is a lot of diversity, so her attitude was mixed. YY132HF1AD shows overlap in her [ɛ] in Spanish and English words, more raised toward where a typical Spanish [ɛ] would be, but allophones of /t/ for the two systems are kept separate.

7 Discussion

This study investigated several variables. It investigated the relationship between proficiency in Spanish and amount of codeswitching and found that both heritage and receptive speakers of Spanish exhibited codeswitching, specifically insertional codemixing as expected in recent migrant communities (Muysken, 2000). Although there was one instance of /eɪ/ monophthongisation, there were overlaps between Spanish and English productions of [ɛ] in codeswitched sentences, showing potential evidence for phonological incorporation of Spanish in English speech. Spanish allophones and English allophones of /t/ were largely kept separate, but topics of conversation surrounding language features, leisure, and family history showed to be prevalent for both English and Spanish allophones of /t/.

A couple limitations to this study involved using data from a corpus such as the Pacific Northwest English study. The small sample of codeswitched tokens contributes to the limitations of the study and there's no way of knowing if perhaps these speakers codeswitch more in other settings like in a home environment or in a social setting that celebrates Mexican American identities. Collecting data in other types of social environments could be a path for future study to look at patterns of codeswitching. I also didn't have a quantifiable way of determining whether speakers were native or heritage speakers of Spanish – I worked off the descriptions the speakers used for themselves and proficiency wasn't always asked about explicitly in the interviews.

Exclusion of word-initial /t/ in words like *traste, tina, toros,* and *que tal,* could be incorporated in a future study to see if the trend of apicodental [t̪] in Spanish-origin words holds word-initially. A wider range of vowels could be studied as well to get a better picture of degrees of phonological incorporation of Spanish in otherwise English speech. Variables like th-stopping, /æ/ raising, /æ/ backing, /ɪ/ tensing, /ʊ/ backing, and consonant cluster reduction would be of interest as Fought (2003) found these occurring at varying degrees among Chicano English speakers in Southern California. Because Chicano is very specific to Southern Californians of Mexican ancestry (Fought, 2003), I hesitate to say that Mexican Americans in the Pacific Northwest also speak Chicano English, both due to the label and a need for more investigation of sociophonetic variables.

The potential evidence for phonological incorporation of Spanish in English speech is supported by other recent works investigating speech of Mexican Americans in the Yakima Valley. The knowledge of a heritage language could play a role in the degree non-white speakers participate in sound changes in the Yakima Valley and Washington state at large (Wassink & Squizzero, 2022). A follow-up I would want to do in the future involves looking at the relationship between codeswitching in conversation speech and the number and types of errors the automated speech recognition system powered by CLOx. Wassink (2020) found that Mexican American speakers of Pacific Northwest English show a higher error rate than Anglo speakers, and it would be worth investigating if possible phonological incorporation from Spanish could also play a role in the error rate.

8 References

Aaron, J. E. (2015). Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism,* 19(4), 459–480. <https://doi.org/10.11-77/1367006913516021>.

Blom, J. P., & Gumperz, J. (1972). Social meaning in linguistic structure: Code-switching in Norway. In L. Wei (2000), *The Bilingualism Reader* (pp. 111-136). London: Routledge.

ELAN (Version 5.9) [Computer software] (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>.

Fought, C. (2003). *Chicano English in Context*. New York: Palgrave MacMillan.

Gamboa, E. (1981). Mexican American Migration into Washington State: A History, 1940-1950. *The Pacific Northwest Quarterly*, 72(3), 121-131.

Kershner, J. (2009, October 16). Yakima. Retrieved from *Thumbnail History*: <https://www.historylink.org/File/9187>.

Muysken, P. (2000). *Bilingual Speech*. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1997). *Duelling Languages: Grammatical Structure in Codeswitching* (2nd ed.). Oxford University Press.

Poplack, S., Sankoff D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation*. Linguistics*, 26, 47-104. New-York: Mouton de Gruyter.

Sankoff, D., Poplack, S., & Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change,* 2(1), 71-101. <https://search.proquest.com/scholarly-journals/casenonce-loan-tamil/docview/85508092/se-2?accountid=14784>.

Santa Anna, O. (1993). Chicano English and the Nature of the Chicano Language Setting. *Hispanic Journal of Behavioral Sciences*, 15, 3–35. <https://hdl.handle.net/2027/uiug.30112006363086>.

Torres Cacoullos, R., & Aaron, J. E. (2003). Bare English-origin nouns in Spanish: Rates, constraints, and discourse functions. *Language Variation and Change,* 15(3), 289–328. <http://dx.doi.org/10.1017/S0954394503153021>.

Wassink, A.B., Squizzero, R., Fellin, C. & Nichols, D. (2018). *Client Libraries Oxford (CLOx): automated transcription for sociolinguistic interviews* [computer software]. <https://clox.ling.washington.edu>.
Wassink, A. B. (2020). *Uneven success: Automatic speech recognition and ethnicity-related dialects*. Talk presented at AAAS, Seattle, WA.
Wassink, A. B. & Squizzero, R. (2022, Jan. 9). *Back away from the rest of the West: ethnic minorities' participation in a Washington English vowel pattern.* [Presentation]. American Dialect Society 2022, Washington DC, United States.

9 Appendix

9.1 Vowel Plots of Speakers who Codeswitched

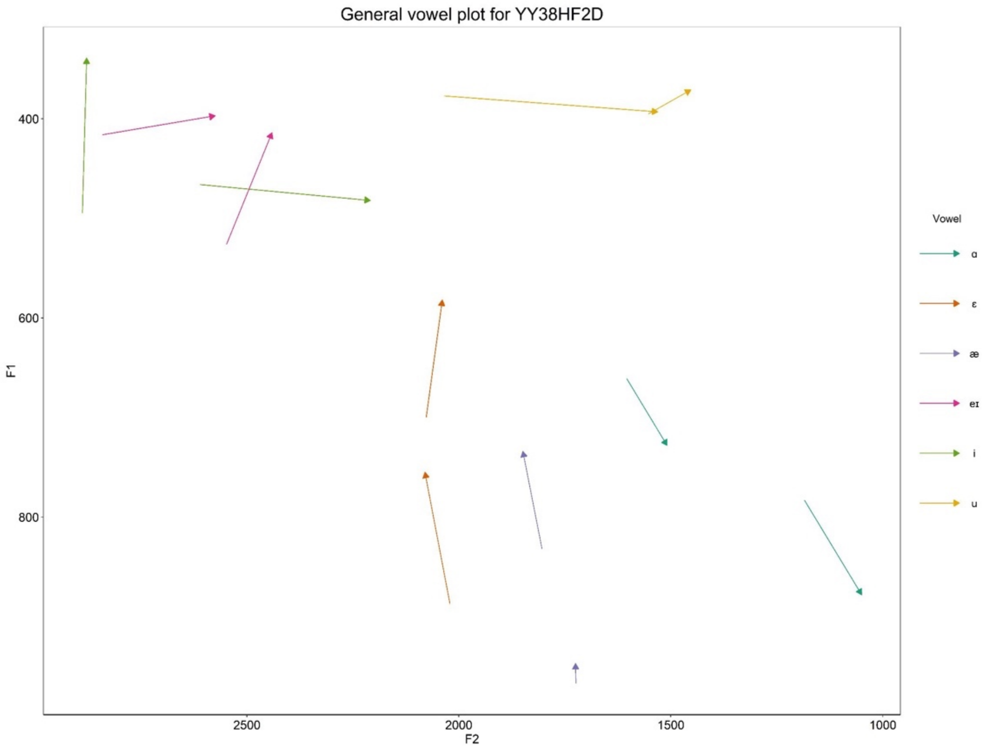


Figure 6: *General vowel plot for YY38HF2D.*

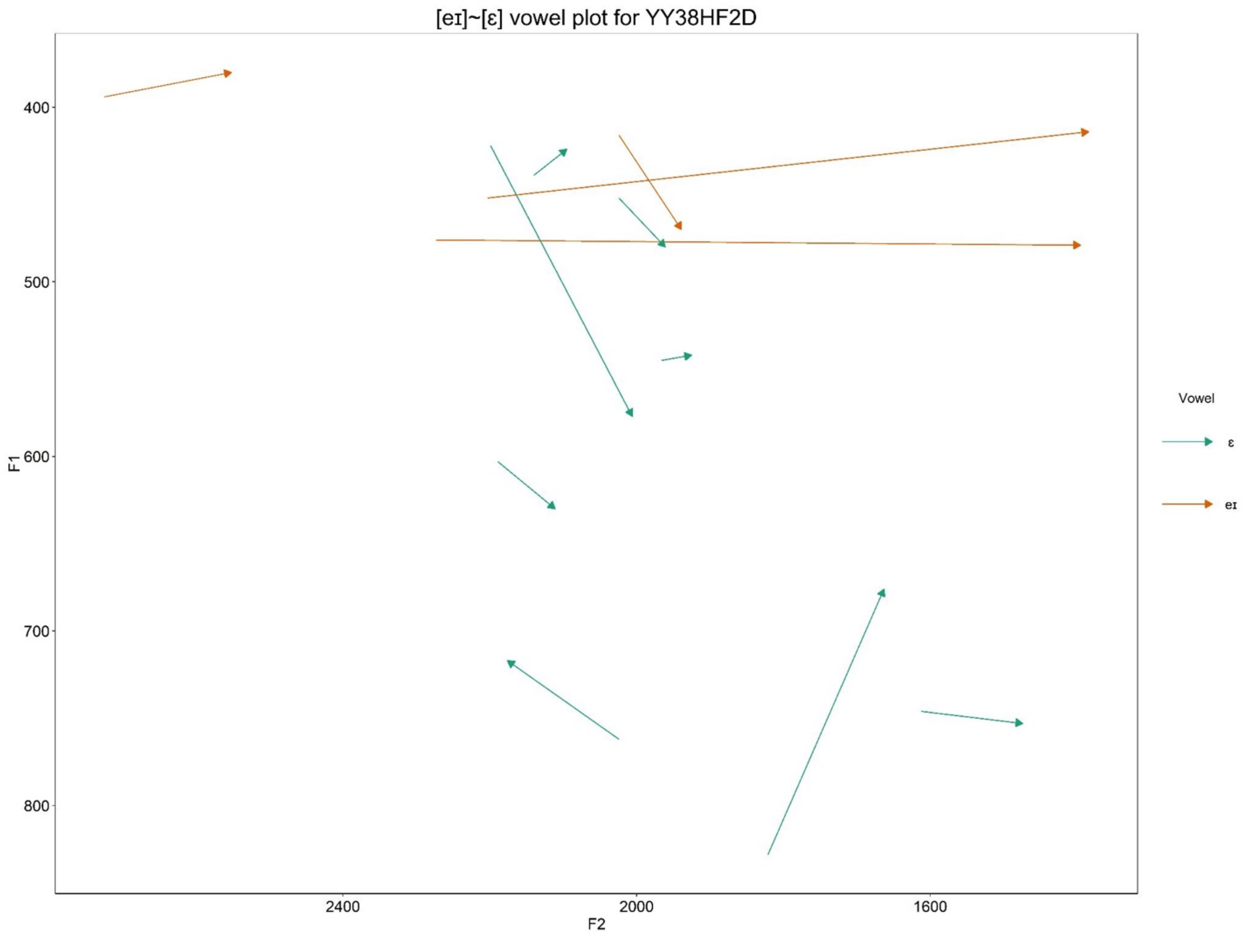


Figure 7: *[eɪ] ~ [ɛ] vowel plot for YY38HF2D in codeswitched sentences.*

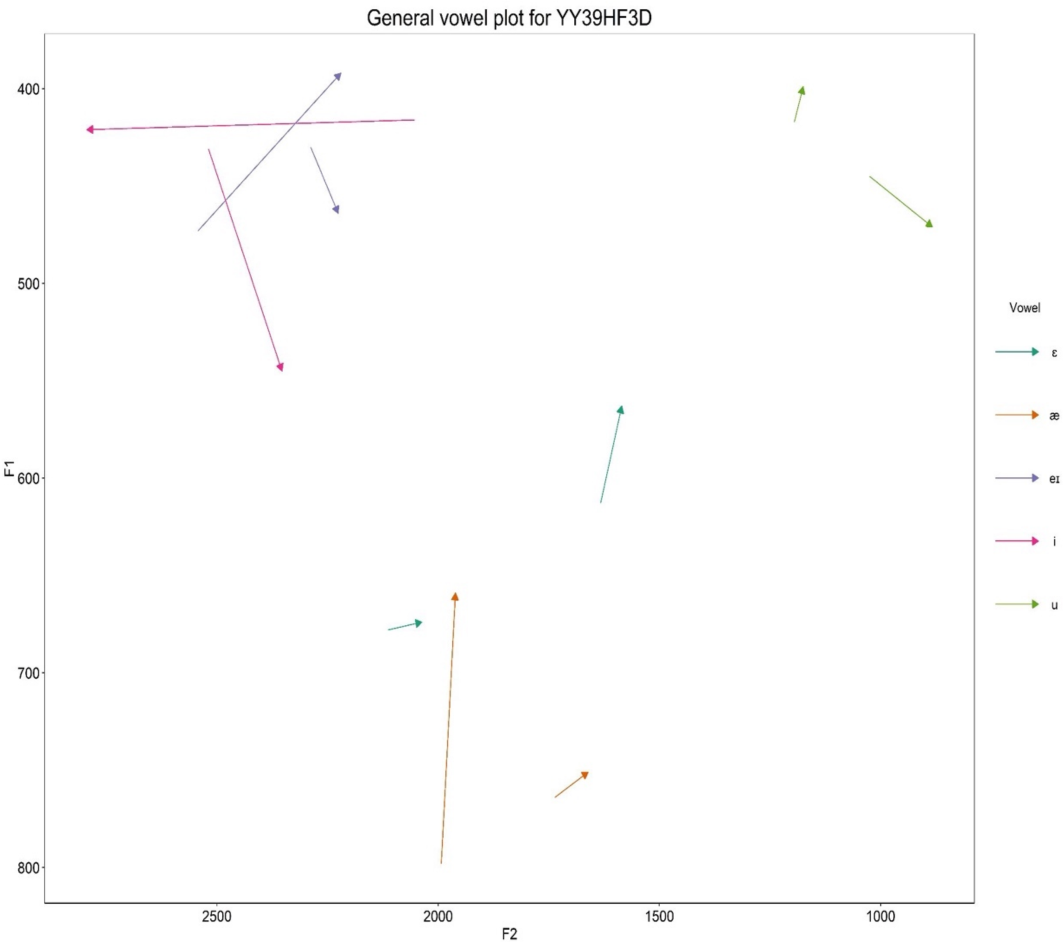


Figure 8: *General vowel plot for YY39HF3D.*

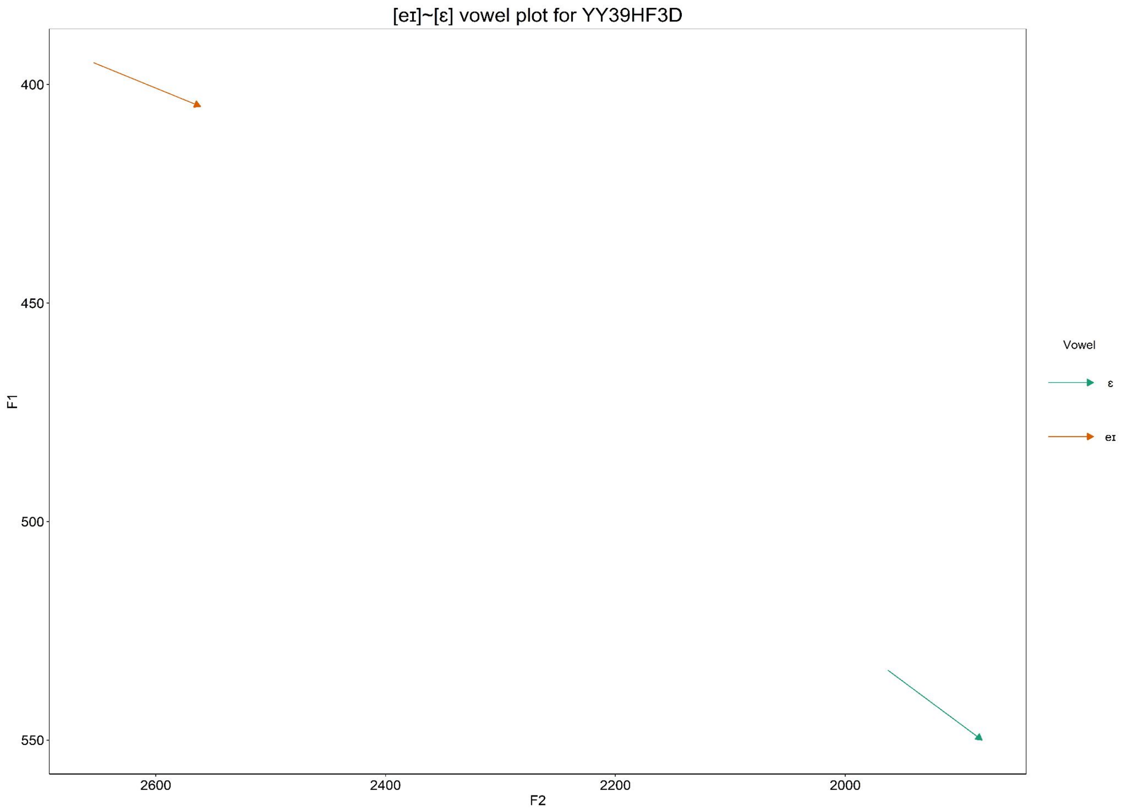


Figure 9: *[eɪ] ~ [ɛ] vowel plot for YY39HF3D in codeswitched sentences.*

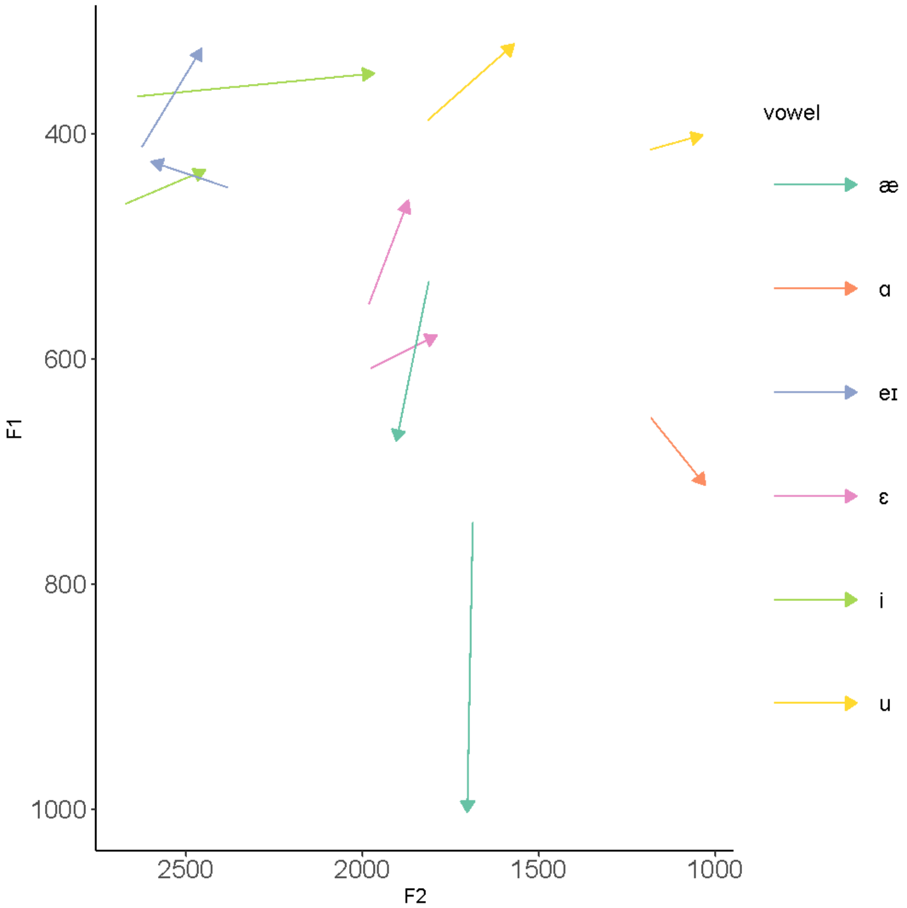


Figure 10: *General vowel plot for YY60HF2I.*

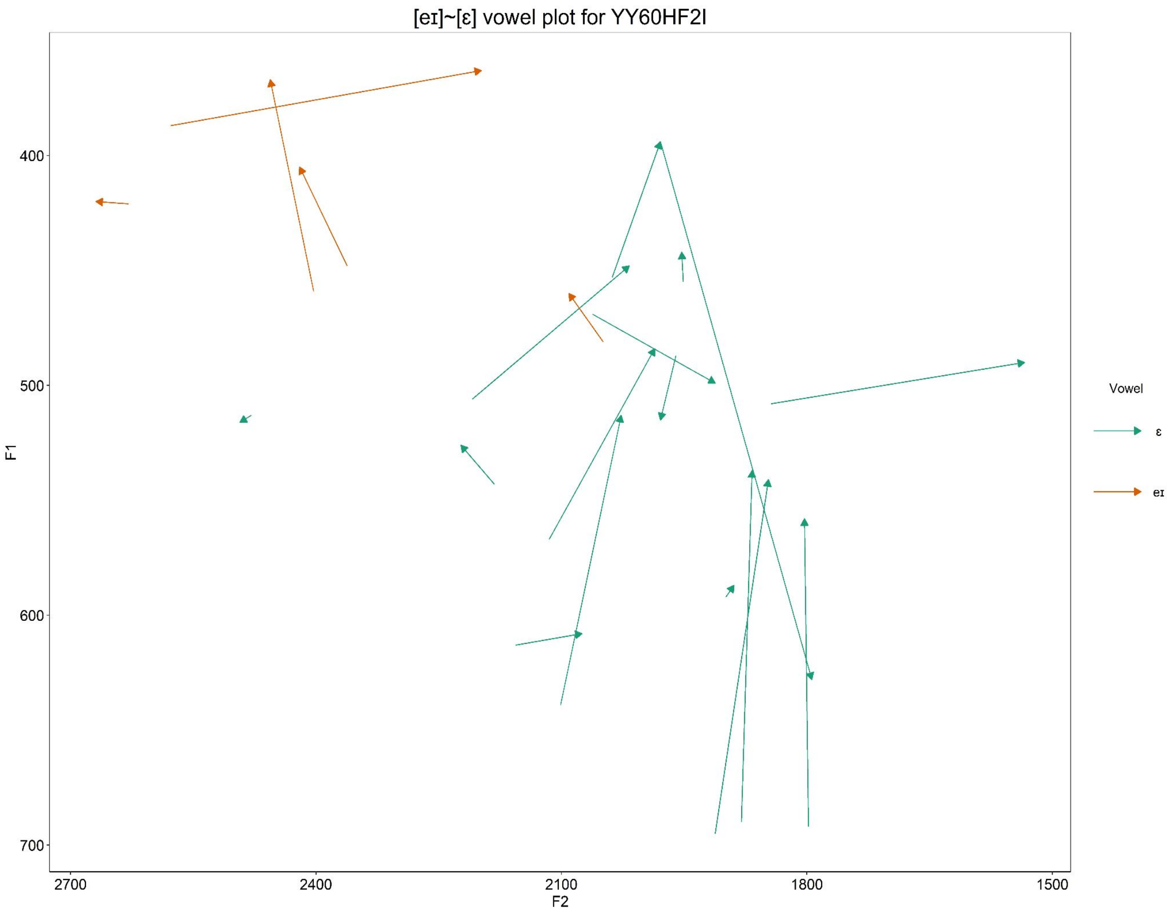


Figure 11: *[eɪ] ~ [ɛ] vowel plot for YY60HF2I in codeswitched sentences.*

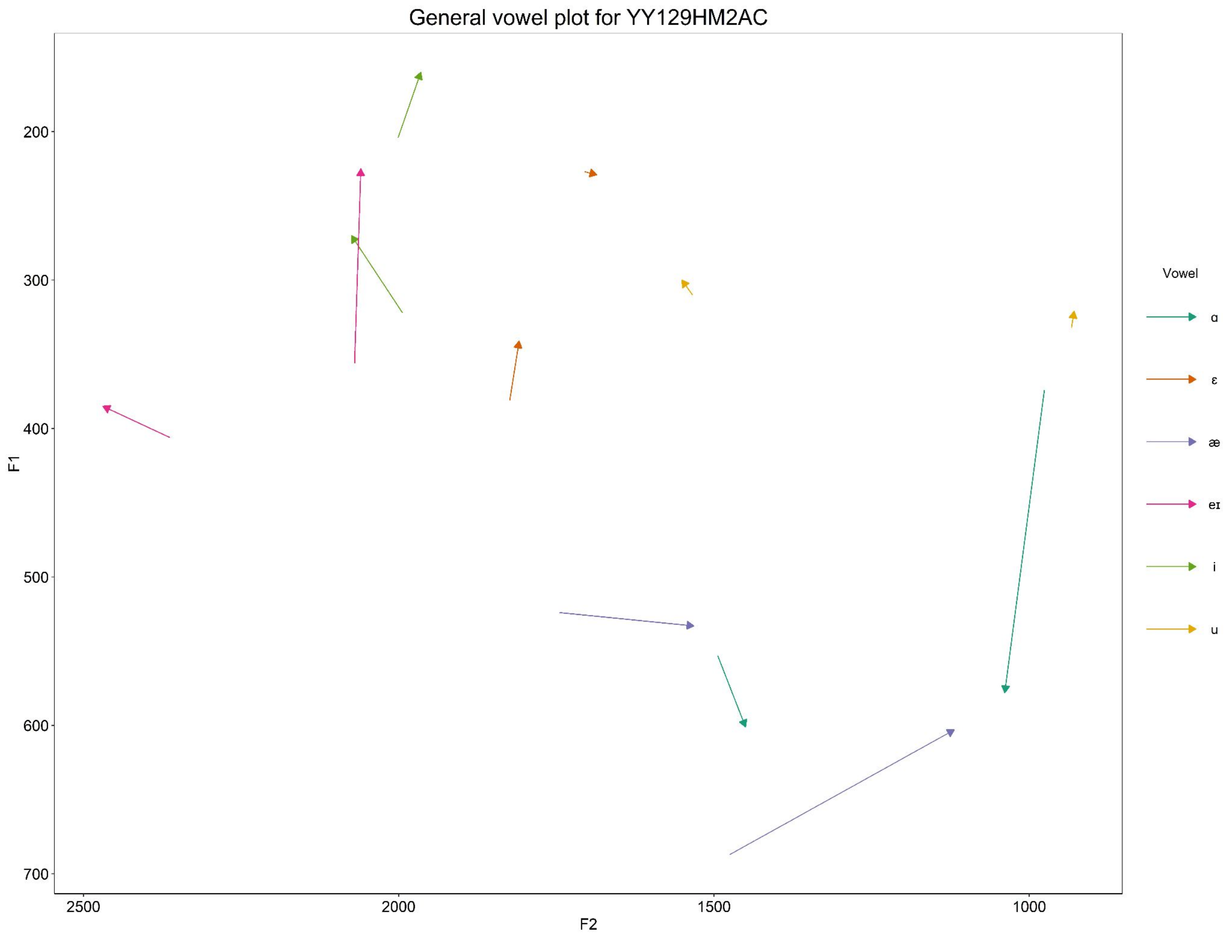


Figure 12: *General vowel plot for YY129HM2AC.*

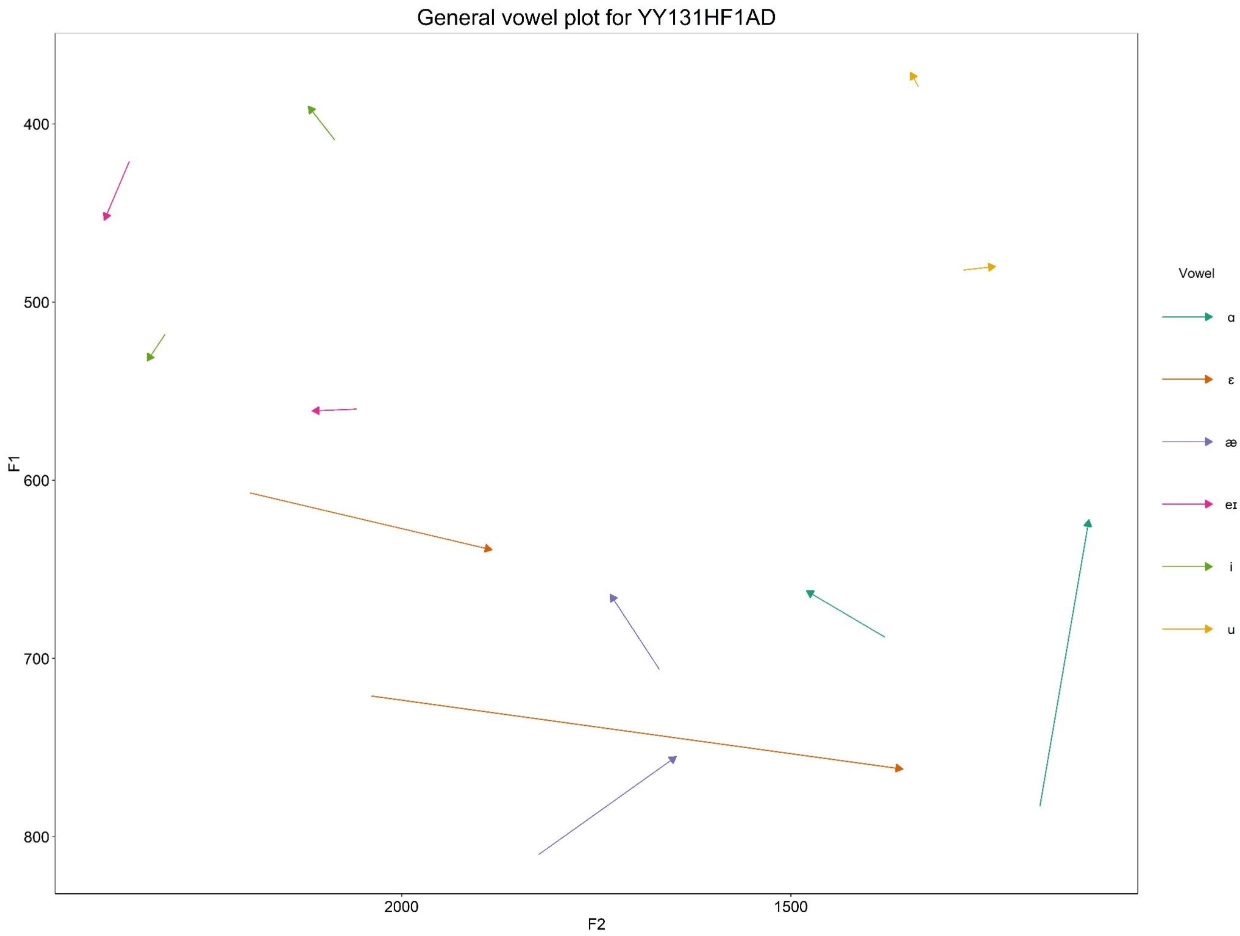


Figure 13: *General vowel plot for YY131HF1AD.*

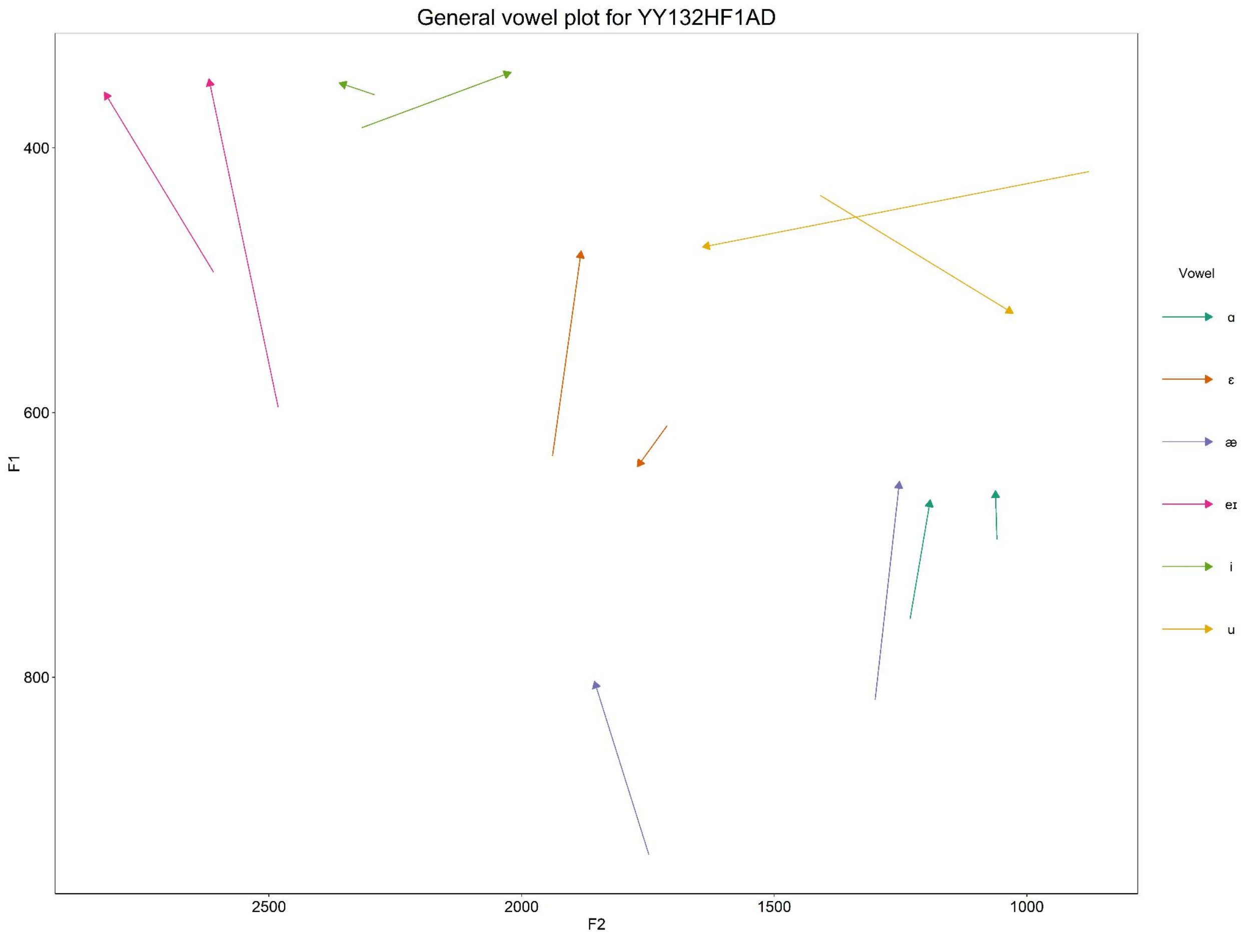


Figure 14: *General vowel plot for YY132HF1AD.*

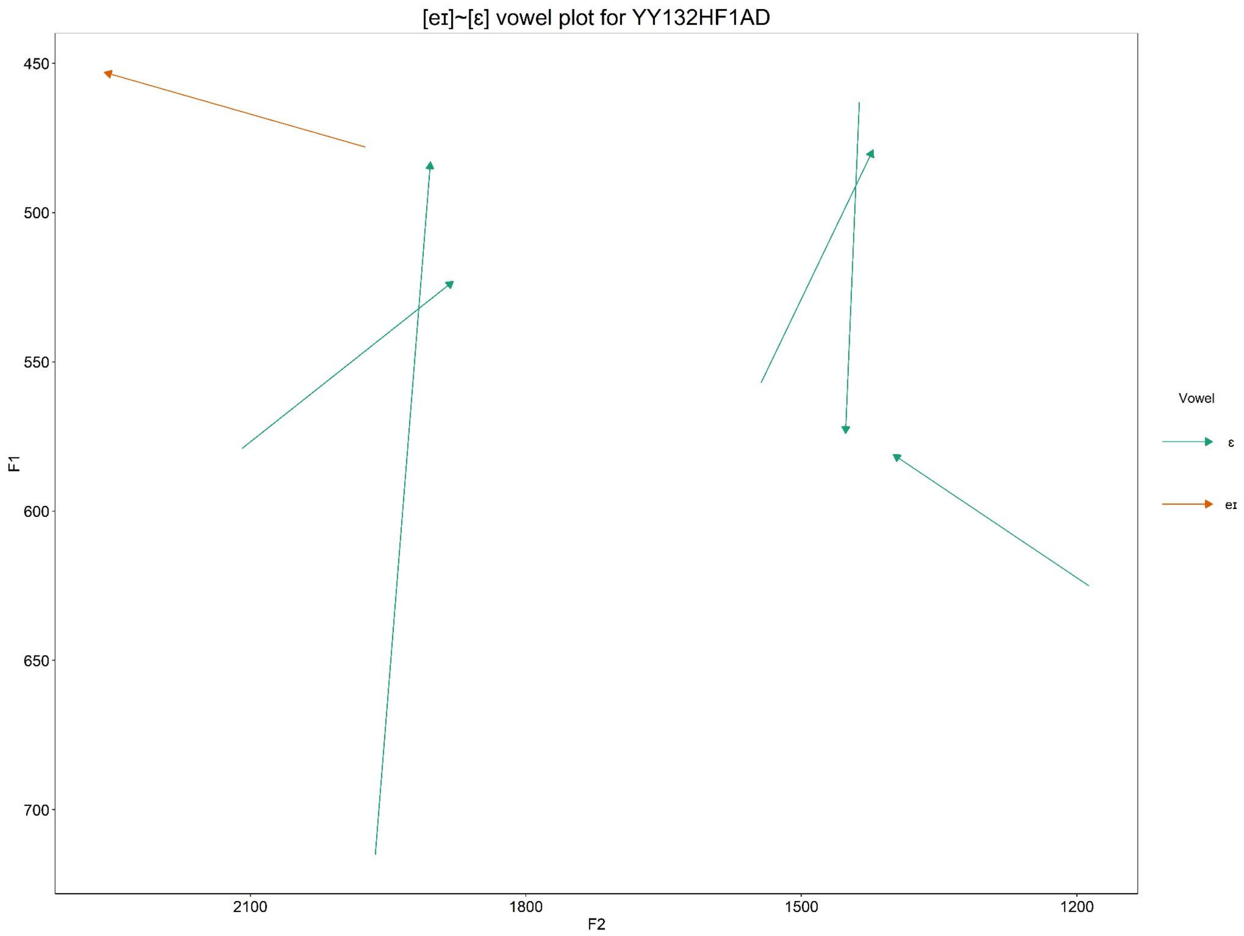


Figure 15: *[eɪ] ~ [ɛ] vowel plot for YY132HF1AD in codeswitched sentences.*

The Effects and Rates of Interpretive Listening in Time-Compressed Speech in Audiobooks

Wing Yin Ho
University of Edinburgh

Abstract. The present study employs an experimental design to investigate the rates and effects of interpretive listening in time-compressed audiobooks, more specifically if increased speech rates affected the emotive response of audiobooks compared to the normal speed, by how much, and whether there is a threshold for time-compressed speech that is sustained before the rate of interpretive listening is at a decline. Participants (N=50) that were fluent English speakers and had no hearing impairments were asked to rate their emotive response (using emotions: Happiness, sadness, anger, surprise, and fear) from a scale from 0 to 100 upon listening to twelve 30-second audiobook excerpts played at rates 1x, 1.5x, 2x, and a truncated condition. Participants also completed a multiple-choice comprehension question after each excerpt was played. The results found that there was a significant decrease in emotive response when the audiobook was played at 1.5x and 2x speed compared to the normal speed. The truncated condition did not have a significant difference in emotive response compared to the normal speed. The results presented also seem to suggest the existence of a threshold for the rate of interpretive listening. Furthermore, results stated that different time-compression rates did not have an effect on comprehension, although this could be due to the short length of each audiobook excerpt. With the ubiquitous audiobook and audio media content, more research is required to investigate individual emotions, develop a more systematic measure of interpretive listening and discover whether certain demographics have the same threshold for interpretive listening.

Keywords: interpretive listening; audiobooks; time-compressed speech

---

1 Introduction

There has been extensive research about the effects of the rate of speech on comprehension (Adank & Devin, 2010; Dupoux & Green, 1997; Carver, 1973; Foulke, 1966; Griffiths, 1992; Ritzhaupt et al., 2008; Wilson et al., 2018; Yang et al., 2020). However, a less investigated aspect of listening is interpretive listening, affective content that is derived from the tone of voice, inflection, and other prosodical information (Bodie & Fitch-Hauser, 2010). As there has been more interest in audiobooks and other media content, many audio speed playback functions are available. The current study aims to answer the question: What are the rates and effects of interpretive listening in time-compressed audiobooks?

The conflicting research on the effect of speech rates on interpretive listening diverge into three theories, it is that increased speech rates bring more emotive responses (LaBarbera & MacLachlan, 1979; Miller et al., 1979), less emotive responses (Apple et al., 1979; Moore et al., 1986; Schlinger et al., 1983), or have no effect on emotive responses overall (Adank & Devlin, 2010; Megehee, 2009; Wheeless, 1971). King et al. (1983) also suggested that the rate of interpretive listening in time-compressed speech is sustained until a certain threshold, before it declines sharply.

The current study presents participants with 30-second audiobook excerpts of 1x, 1.5x, 2x, and truncated condition. The participants are asked to rate the emotive content of the excerpts using each of the five emotions: Sadness, happiness, anger, surprise, and fear from a scale of 0 to 100. Using this design, the study aims to answer the hypotheses below:

(1) The 1.5x condition of audio will not have a significant difference in interpretive listening in audiobooks compared to the 1x condition.
(2) The 2x condition of audio will have a significant difference in interpretive listening in audiobooks compared to the 1x condition.
(3) The truncated condition of audio will reduce interpretive listening in audiobooks compared to the 1x condition.
(4) The comprehension skills will have a significant decrease across speed conditions compared to the 1x condition.
(5) All aspects of interpretive listening (happy, sad, anger, fear, and surprise) will decrease or increase uniformly.
(6) The manipulated conditions will have a worse effect with listeners aged 35 onwards compared to listeners aged 18-24 and aged 24-34.

2 Literature Review

Audiobooks have been a common medium to consume books in the recent decade. As of 2019, 1-in-5 Americans had listened to an audiobook in the previous year (Perrin, 2019). With the increasing popularity of podcasts, audiobooks, and other digital audio media, audio manipulation technology has improved immensely. Audio media consumption has become a highly personal experience, allowing users to adjust the speed of the audio without compromising the pitch of the audio, there is even a small community of users committed to speed listening, known as “podfasters” (Murrow, 2018). Most mobile applications containing audio media content have a function to increase playback speed such as Overcast (2021), an application that has a ‘smart speed’ function, which allows for the reduction of the length of long pauses in speech.

Time-compressed speech is the official term for the method of increasing the rate of audio, such as for the use of speed listening. The term refers to the altering of an audio recording in order that the full content of the audio is played in a shorter amount of time, typically while preserving its pitch (Barabasz, 1968). According to Audible, an audiobook streaming service, it is common to use time-compressed speech for audiobook consumption, with 5% of users listening to audiobooks at a speed of 1.5x and most users switching between normal speed and 1.5x speed (Hu, 2017).

2.1 Methods in which Audio is Time-Compressed

The main methods employed to compress speech are linear compression, selective sampling, and non-linear compression. Linear compression involves compression that is employed throughout the entire audio (Omoigui, 1999). Selective sampling is another method that removes silence as opposed to manipulating speech sounds (Arons, 1992). Non-linear compression comprises both linear compression and selective sampling techniques (He & Gupta, 2001). These different methods could lead to differences in the quality of comprehension depending on the method that the speech is compressed with.

More specific time-compression methods derived from these main methods are the speed changing method (Klumpp & Webster, 1961), the sampling method (Fairbanks et al., 1954), and pausing sampling (Megehee, 2009). The speed changing method involves increasing the rate of speech and frequency proportionally resulting in a high-pitch speech. Although the speed changing method may have varying degrees of effect on different age populations, older participants’ sensitivity to hearing can be subpar compared to younger participants (Hall & Mueller, 1997).

This is supported by several studies that reported a rapid decline in speech comprehension in older listeners when they encountered time-compressed speech. (Sticht and Gray, 1969; Tomasz & Nancy, 1996; Konkle et al., 1977).

The sampling method is more elaborate; the specific parts of the speech are removed at regular intervals in small segments. This preserves most of the sound of speech until larger segments of the speech are excluded. Pause sampling is similar to the sampling method, in which pauses are inserted or deleted between phrases and sentences without altering the speech sounds.

2.2 Time-Compressed Speech and its Effects on Comprehension and Recall

There has been a record of studies investigating comprehension and recall of information of audio at different speeds. Several studies suggest that there is a linear decline in comprehension and speech rate until it reaches the threshold at which the participant can follow the speech and comprehension declines dramatically (Carver, 1973; Foulke, 1966).

There has been varied findings pertaining to the threshold, with Carver (1973) reported it to be 150wpm as the threshold while Foulke (1966) reported 275wpm as the threshold. Wilson et al. (2018) postulated that there was a slight decrease in comprehension scores with time-compressed lectures. This is in contrast to the opposing study, which suggests that comprehension scores are significantly higher when passages are played at a slower rate than for passages at average and faster speech rates (Griffiths, 1992).

More recent studies, however, have concluded that the quality of comprehension is unaffected until the audio reaches a certain speed. Yang et al. (2020) stated that Chinese students were able to comprehend lecture content at 1.5x speed just as well as the normal speed. The participants of the study were also capable of recalling the content one week after the experiment. Another study (Ritzhaupt et al., 2008) showed that undergraduate students did not find a significant difference across the lecture recording speeds 1.0x, 1.4x, and 1.8x. In fact, the majority of students favoured the lecture content at 1.4x speed.

Studies by Adank and Devlin (2010) and Doupoux and Green (1997) are consistent with the recent findings that quality of comprehension remains unaffected until a certain speed threshold is reached. They concluded that this is because the brain is capable of accommodating different speeds of speech to a certain threshold. Their findings revealed that comprehension performance often improved with increased exposure to time-compressed speech, however, highly compressed speech required even more time for comprehension performance to improve.

2.3 Time-Compression by Method of Silence Removal

The method of reducing the duration of long silences, also known as pause sampling, has recently been used in applications such as Overcast (2021). Their ‘smart speed’ mode claims to reduce the duration of a podcast without distorting the original audio. If the silences are considerably reduced through pause sampling, however, this could affect the intonation, tone, stress, and rhythm of speech; the combination of these is known as the prosody (Schreiber, 1991). Prosody is an important factor in how listeners comprehend and interpret an utterance and can change the whole meaning of the speech despite the fact that the same words are produced (Lehiste, 1973).

Prosody also refers to crucial linguistic boundaries, which could be clauses, phrases or sentences made by pauses. Take for example **twenty-six** dollar bills versus twenty **six-dollar** bills; the former denotes a total of twenty-six dollars, whereas the latter indicates one hundred and twenty dollars. The varying length of the pause between the words six and dollar could change the amount being referred to. Intentional pauses that were placed for punctuation or effect could distort the meaning and effect of the utterance. Take another example:

(1) The friend yelled [pause] the passer-by is weird.
(2) The friend *[pause]* yelled the passer-by *[pause]* is weird.

The first example denotes the sentiments about the passer-by that the friend is expressing. The second example, on the contrary, indicates the stranger's expressed feelings toward the friend. Although the word stress and intonation would be different in the two examples, the alteration in the duration of the pauses could potentially confuse listeners.

MacGregor et al. (2010) investigated the purpose of silent pauses, a form of disfluency in speech. They discovered that listeners were more likely to recognise words that followed silent pauses. This demonstrates that the temporal delay contributes to the speech signal.

Silence could have also been used as a stylistic choice as a rhetorical device or to maintain the prosodic structure of an utterance. Therefore, silence or the lack thereof could affect comprehension and recall in listeners.

Miller and Licklider (1948) pioneered the comprehensive method for exploring temporal parameters associated with the perception of speech. They varied 'speech-time' ratio, altering speech and interjecting varying lengths of silence. They reported intelligibility was best when the speech-time ratio was reduced from 50% to 25%. However, when speech was manipulated, sometimes segments of the speech signal were removed which in turn affected the validity of the design. Huggins (1975) addressed the problem of the Miller and Licklider study (1948) and showed that intelligibility was optimal at silences at around <60ms and intelligibility was poor when the silences were longer at around >150ms.

A more recent study by Peelle and Davis (2012) discovered that time-compressed signals without insertions of silence were difficult to understand and that insertion of silence improved intelligibility, up until a certain range (20-120ms). The Pelle and Davis (2012) study is consistent with previous theories that the absence of silences could negatively affect comprehension.

2.4 Interpretive Listening

Bodie and Fitch-Hauser (2010) describe interpretive listening as the action of emotional or affective content that is derived from the tone of voice, inflection, and other prosodical information of a message. This theory is also supported by a few studies, Littlepage and Pineault (1981) indicated that vocalic cues are more useful in detecting deception compared to visual cues, and Leathers (1979) postulated when nonverbal and verbal signs contradict one another, listeners will more likely assume the nonverbal signs as the true feelings of the speaker.

King et al. (1983) revealed that interpretive listening scores were sustained until 1.6x time compression was introduced. Additionally, a time-compressed speech study by Skinner et al. (1999) concluded that fast speech rates could cause listeners to focus their attention more on forming global impressions than on specific facts. Interpretive listening is an important aspect of audiobook and general audio media consumption as most listeners listen to this form of media for enjoyment and experience. If compressing audio alters the interpretive listening aspect, then this could be an argument to not compress speech.

There has been conflicting research on the topic of the effects of speech rates on interpretive listening. The three main diverging conclusions are that increased speech rates either bring a positive effect, a negative effect, or a neutral effect towards interpretive listening. Yet, most of the studies come from an advertising standpoint and focus on persuasiveness and strictly positive emotions of a certain brand or advertisement.

2.5 How Interpretive Listening is Measured

The two notable listening comprehension tests that have measurements for interpretive listening are the Kentucky Comprehension Listening Test (Bostrom & Waldhart, 1983) as used in the King et al. study (1983) and the Watson-Barker Listening Test (Watson & Barker, 1983). Both provide a rudimentary measure of interpretive listening, requiring participants to listen to the dialogue with emotive content and evaluate its emotions.

These standardised tests provide a fairer assessment of listening comprehension skills, especially across individual studies. However, Fitch-Hauser and Hughes (1987) argue that the Kentucky Comprehension Listening Test (Bostrom & Waldhart, 1983) and the Watson-Barker Listening Test (Watson & Barker, 1983) have questionable external validity. This is due to the fact that each test was compared to a factor structure planned around the unit each test reported to measure. Villaume and Weaver (1996) provided more evidence that the two tests lacked external validity while conducting first and second order factor analyses on the tests. The interpretive listening aspect of the two listening tests is also described as a rudimentary measure of interpretive listening by Bodie and Fitch-Hauser (2010).

2.6 Emotions used to Evaluate Interpretive Listening

Psychologists have acknowledged the theory of basic emotions but the exact number of emotions and what the emotions are is up for debate. Ekman et al. (1969) indicated that there were six core human emotions: Happiness, sadness, fear, disgust, and surprise. While Plutchik (1982) identified eight primary human emotions: Anger, fear, sadness, disgust, surprise, anticipation, trust, and joy.

Plutchik (1982) also put forth the theory of the Wheel of Emotions. This means that primary human emotions are considered the basis of complex emotions, which have an array and degrees of the core emotions blended together. More recently Jack et al. (2014) have proposed that there are only four core emotions, being happy, angry, sad, and fear which is then detailed into the six emotions that Ekman et al. (1969) illustrated.

2.7 Theories about Time-Compressed Speech and its Effects on Interpretive Listening

LaBarbera and MacLachlan (1979) reported that participants reported the brand as more interesting and had high recall rates with the 1.3x time-compressed speech compared to the normal speed. The Miller et al. (1979) study also discovered that participants described increased speech rates as more persuasive than normal or decreased speech rates. However, Moore et al. (1986) criticised studies such as the LaBarbera and MacLachlan (1979) study for its questionable internal and external validity, as they used within-subject designs that assigned each group to a single-speed condition.

More recently, Ilie and Thompson (2006) theorised that faster, more high-pitched speech can be more arousing and intensifying, some participants claimed that this intensity is unpleasant and could negatively affect the audiobook experience. Although the study seems to propose a stronger emotive response, this could have arisen from the uncanniness of high-pitched speech, and not represent the heightening of emotions evoked within the speech message itself.

Nonetheless, contrasting results from the other side of the spectrum have also been concluded. Apple et al. (1979) state that participants found speech rates at 1.3x speed less emphatic and less truthful. Moore et al. (1986) found that time-compressed advertisements at speeds 1.3x and 1.6x captured less attention and evoked fewer cognitive responses to the radio advertisement. This could be because listeners may then decide that it is not worth the cognitive effort to maintain attention as it expends more cognitive load than normal rate speech.

Schlinger et al. (1983) observed that listeners considered increased speech rates at 1.2x less personal and more patronising, contrasting to the studies that indicate listeners preferring speakers with an increased speech rate. Although, it is important to note that Schlinger et al. (1983) investigated time-compressed audiovisual content which could pose a different effect than audio content alone.

There is also evidence of time-compressed speech having no effect on interpretive listening. Adank and Devlin (2010), concluded that brain plasticity aided in allowing listeners to adapt to rates of speech, which could suppose that interpretive listening remains unchanged by reasonable increased rates of speech. King et al. (1989) demonstrated sustained levels of interpretive listening rates until 1.6x of time-compressed speech. Schlinger et al. (1983) concluded that despite time-compression only having small effects on cognitive processing and post-viewing attitudes, as a result, listeners have fewer ideas in response to open-ended questions. This could inhibit positive and negative attitudes towards the advertised brand. Wheeless (1971) and Megehee (2009) have also concluded that time-compressed audio has made no impact on consumer buying intentions.

The current literature suggests a neutral or less emotive effect when speech rate is increased through time-compression methods. If the speech rate becomes less emotive its original emotive response is sustained until a certain speed threshold. This is because the studies in favour of increased rates of speech producing more emotive responses mostly focus on persuasiveness and only positive emotions. This is unlike audiobooks, which have various complex emotions that are highly individual to one another. Moore et al. (1986) also asserted that study designs such as Labarbera and MacLachlan (1979) have questionable internal and external validity.

2.8 Sleeper Effect

The sleeper effect, a phenomenon where listeners often remember the content more than the source over time and therefore tend to be influenced by the message content rather than the emotions arisen from the audio has been proposed by some studies (Hovland & Weiss, 1951; Pratkanis et al., 1988). This stipulates that while the time-compressed speech may initially be more persuasive, this may change as time passes. Skinner et al. (1999) found that when participants were asked to judge the product being advertised at different speeds, the majority rated the fast advertisement as significantly higher than the medium advertisement immediately after listening. However, when another group of participants rated the medium advertisement significantly higher than the fast ad when there was a ten-minute buffer between listening to the advertisement and rating the product. Moore et al. (1986) also postulated that improved general global impressions from increased speech rates may not sustain over time. This sleeper effect should be noted when interpreting results in our current study, as the methodology will not require participants to do a second survey.

2.9 Research Question

There has been substantial research surrounding speech rates and the effects it has on comprehension. A lesser investigated aspect of listening would be interpretive listening. Although there have been numerous studies pertaining to interpretive listening in advertising, the studies mainly focus on consumer intentions and persuasion. As audiobooks and podcasts become more frequent in everyday life, it would be crucial to further explore interpretive listening in its wider context, similarly to the King et al. (1983) study.

Since the rise of audiobooks and podcasts as a media genre, platforms such as Youtube (http://www.youtube.com), Audible (http://www.audible.com) , and Overcast (2021) provide speech rate increasing functions. There has been growing interest in the effects of these functions. Audiobooks, podcasts, and audio media are consumed for entertainment and enjoyment. By exploring whether the time-compressed speech in audio media could potentially affect interpretive listening, could lead to further research into methods of speech compressing with interpreting listening as a factor.

There have been few studies about the effects of time-compressed speech within the last ten years and especially few surrounding the effects of interpretive listening. A more recent perspective on interpretive listening and time-compressed speech would be beneficial with the ubiquitous nature of this media genre.

3 Methodology

3.1 Participants

A total of 61 participants were recruited through social media, word of mouth, and email. Prior to data collection, ethical approval was obtained by the Linguistics and English Language Ethics at the University of Edinburgh. Participants were required not to be hard of hearing, above the age of 18, and had high proficiency in English.

11 participants' results were removed due to incomplete surveys. After removing the incomplete surveys, the first version of the survey had 9 participants, the second version had 15 participants, the third version had 13 participants, and the fourth version had 13 participants. There is a difference in group sizes due to the Wix website (http://Wix.com) design used to lead the participants to different versions of the survey. The simple randomisation method has led to slightly uneven group sizes as some participants may have not finished the survey in its entirety or had click the website multiple times. The unequal sample sizes risk our statistical analysis to have less power compared to an equal sample size analysis. Although there is an even number of participants answering all speed conditions due to our pseudo-randomised order for each survey as outlined in the methodology.

14 outliers were identified by measuring the lower bound (2.5%) and the upper bound (97.5%) of the sum of emotions per participant and item 14 outliers. The outliers were trimmed prior to running the following analyses.

3.2 Survey Materials

The survey was hosted by Qualtrics (http://Qualtrics.com). A landing page created on Wix (http://Wix.com) was created to redirect participants to one of the four different versions of the survey.

3.3 Audiobook Excerpt Source

The audiobook excerpts were retrieved from a free domain audiobook website LibriVox (http://librixov.org). A wide range of genres was utilised, with the exception of thriller and horror as to not cause any physical or psychological harm to the participants.

Table 1: *The books used in the survey. The left-most column notes the book excerpt number that is coded later in the analysis.*

Audio	Genre	Book Title
1	Romance 1	First Love (Little Blue Book #1195) And Other Fascinating Stories of Spanish Life (Bazán, 2005)
2	Adventure 1	Around the World in Eighty Days (Verne, 1873)
3	Science-fiction 1	The End of All (Wheeler, 1891)
4	Fantasy 1	Alice's Adventures in Wonderland (Caroll, 1865)
5	Mystery 1	A Case of Identity (Doyle, 1981)
6	Romance 2	A Christmas Honeymoon (Mathews, 2015)
7	Adventure 2	The Adventures of Huckleberry Finn (Twain, 1884)
8	Science-fiction 2	The Invisible Man (Wells, 1897)
9	Fantasy 2	Peter Pan (Barrie, 1911)
10	Romance 3	A Sentimental Romance (Kuprin, 1969)
11	Mystery 2	A Scandal in Bohemia (Doyle, 1891)
12	Mystery 3	The Secret Garden (Burnett, 1991)

3. 4 Procedure

The survey, hosted on Qualtrics (http://qualtrics.com) required the participant to record their age, reading habits, audiobook listening habits, emotions experienced after listening to 12 audio excerpts at various speeds, and a comprehension question for each audio excerpt.

The audio excerpts were arranged in a pseudo-randomised order. The participants were required to listen to the audio excerpts carefully and only once. Participants were then required to record what emotions they experienced upon listening to the audio excerpt.

The audiobook excerpt was trimmed down to 30 seconds. The audio was then compressed and truncated using the Wavepad Audio Editor application (NHC Software, 2004). The audio was compressed by using the "speed change" function. For the 1.5x condition, the audio was reduced to 20 seconds, and for the 2x condition, the audio was reduced to 15 seconds. The audio pitch was not preserved through this process. A second of silence was inserted at the beginning of each audio to prepare the reader for the upcoming audio. Audio excerpts with the truncated condition were created by using the "Trim Silences" function. Silences that were softer than 24dB were trimmed and the relative threshold was set to the "Absolute level".

3.5 Dependent Variables

Emotions
A slider was given for each emotion (happy, sad, angry, fear, and surprise) per audiobook excerpt. The initial state of the slider was at 0 and ranged from 0 to 100.

Comprehension
This was a simple multiple-choice question out of two or three options.

3.6 Independent Variables

Speed
The speed conditions 1x, 1.5x, 2x, and truncated are arranged in a pseudo-random order, with it going along the order of an audio excerpt at 1x, 1.5x, 2x, and then truncated. It would then cycle back to the 1x condition and the order would repeat itself. This pattern was employed so that the participant will be accustomed to the speed. The participant, however, is not notified of the speed condition of the audio excerpts.

Item
The different audiobook excerpts (as illustrated in Table 1) are of varying genres.

Table 2: *The pseudo-randomised order that each audio is arranged in. In each of the version of the survey, the audio excerpt is only listened to once to avoid preconceived notions of the audio in question.*

	Version 1	**Version 2**	**Version 3**	**Version 4**
Question				
1	audio 1- 1x	audio 2 – 1x	audio 3 – 1x	audio 4 – 1x
2	audio 4 - 1.5x	audio 1 - 1.5x	audio 2 - 1.5x	audio 3 - 1.5 x
3	audio 3 - 2x	audio 4 - 2x	audio 1 - 2x	audio 2 - 2x
4	audio 2 - truncated	audio 3 - truncated	audio 4 - truncated	audio 1 - truncated
5	audio 5- 1x	audio 6 – 1x	audio 7 – 1x	audio 8 – 1x
6	audio 8 - 1.5x	audio 5 - 1.5x	audio 6 - 1.5x	audio 7 - 1.5 x
7	audio 7 - 2x	audio 8 - 2x	audio 5 - 2x	audio 6 - 2x
8	audio 6 - truncated	audio 7 - truncated	audio 8 - truncated	audio 5 - truncated
9	audio 9- 1x	audio 10 – 1x	audio 11 – 1x	audio 12 – 1x
10	audio 12 - 1.5x	audio 9 - 1.5x	audio 10 - 1.5x	audio 11 - 1.5 x
11	audio 11 - 2x	audio 12 - 2x	audio 9 - 2x	audio 10 - 2x
12	audio 10 - truncated	audio 11 - truncated	audio 12 - truncated	audio 9 - truncated

3.7 Interpretive Listening Recorded through Emotions

To record the interpretive listening that is best suited for the audiobook experience, basic human emotions were recorded. King et al. (1983) employ the Kentucky Listening Comprehension Test

(Bostrom & Waldhart, 1983). They similarly use emotions that are rated by the participants. Studies focusing on brand recall and persuasiveness requested participants to rate how persuasive and interesting the content was (LaBarbera & MacLachlan 1979; MacLachlan 1979).

The emotions utilised for this experiment were the four core emotions: Happy, sad, anger, fear as illustrated by Jack et al. (2014) with an additional emotion surprise. Surprise is added to the four core emotions as it can be considered a response derived from an unexpected situation. Manipulating the speed condition, which can be perceived as an unusual range of speed compared to normal natural speech, could cause an element of surprise. It would be helpful to note whether speed conditions would heighten or dampen the element of surprise.

3.8 Data Analysis

The data from Qualtrics (http://qualtrics.com) was downloaded into separate Microsoft Excel spreadsheets and organised into one single spreadsheet. The data was then summarised in RStudio version 1.4 (RStudio Team, 2021) using ggplot2 (Wickham, 2016).

This experimental design aims to discover whether there is a significant difference in the effects of interpretive listening across speed conditions. The hypotheses that were investigated were:

(1) The 1.5x condition of audio will not have a significant difference in interpretive listening in audiobooks compared to the 1x condition.
(2) The 2x condition of audio will have a significant difference in interpretive listening in audiobooks compared to the 1x condition.
(3) The truncated condition of audio will reduce interpretive listening in audiobooks compared to the 1x condition.
(4) The comprehension skills will have a significant decrease across speed conditions compared to the 1x condition.
(5) All aspects of interpretive listening (happy, sad, anger, fear, and surprise) will decrease or increase uniformly.
(6) The manipulated conditions will have a worse effect with listeners aged 35 onwards compared to listeners aged 18-24 and aged 24-34.

Linear mixed models using the Lme4 package (Bates et al., 2015) was used for the inferential statistics to question 1, 2 ,3, 5, and 6. Linear mixed models are utilised to allow for the fitting of the two random effects, the participants and the individual items. Question 4 was analysed using Logit regression using the glm function from the stats package (Rstudio Team, 2021). The logit regression is used instead of the linear mixed model because the variable correct is coding in binary (0 for incorrect, and 1 for correct). This type of regression is used for dichotomous outcome variables such as the data in question 4. The alpha level of 0.5 will be used to determine the statistical significance.

4 Results

4.1 Visualised Results

The mean of the Emotional Values across different speed conditions (as shown in Table 1) seem to show a slight decrease as the speed of the audio increases. The mean of the Truncated condition also seems to suggest that it is the most similar to the 1x condition compared to the other conditions. However, for more concrete conclusions, more thorough statistical analysis needs to be performed.

Emotional Value

All the five emotions (happiness, anger, sadness, fear, and surprise) are all summed up per audio and the overall mean is taken by participant.

Speed

The barplot from Table 3 suggests a slight decline in emotional values by speed conditions with the mean of 15.1 for the 1x condition, 12.0 for the 1.5x condition, 9.6 for the 2x condition. The barplot also seems to suggest that the truncated condition (with the mean of 13.3) was the most similar to the 1x condition. However, once again more extensive statistical analysis needs to be performed.

Table 3: *The mean and standard deviation of Emotional Value across different speed conditions.*

Speed	Mean	Standard Deviation	N
1x	15.1	22.9	750
1.5x	12.0	21.0	750
2x	9.6	18.2	750
Truncated	13.3	22.9	750

Emotions

More in-depth statistical analysis would be required to determine if specific emotions are affected by speed more profoundly if our statistical analysis shows a significant effect between speed and emotion.

Speed

More in depth statistical analysis would need to be performed as some items seem to have an overall effect on emotional value when the speed varies (for example, Item 3)

Item

According to Table 2, the emotions experienced by participants seem to vary significantly by item. This would be due to the fact that items had varying genres and would evoke separate feelings.

Figure 1: *The mean of emotional values of each emotion grouped by speed and separated by item.*

Emotion and Speed

According to Figure 1 Anger experienced by the participants were constant throughout the speed conditions. Happiness experienced by the participants also seem to be constant throughout speed conditions 1x, 1.5x, and 2x. The graph suggests that there is a spike in happiness experienced in the truncated condition. Surprise experienced by participants seemed to have a general decrease, with truncated experiencing more slightly surprise than 1x condition. Fear experienced by the participants was heightened slightly in the 1.5x condition and fell in the 2x condition. In the truncated condition, fear is experienced more by participants than the 1x condition.

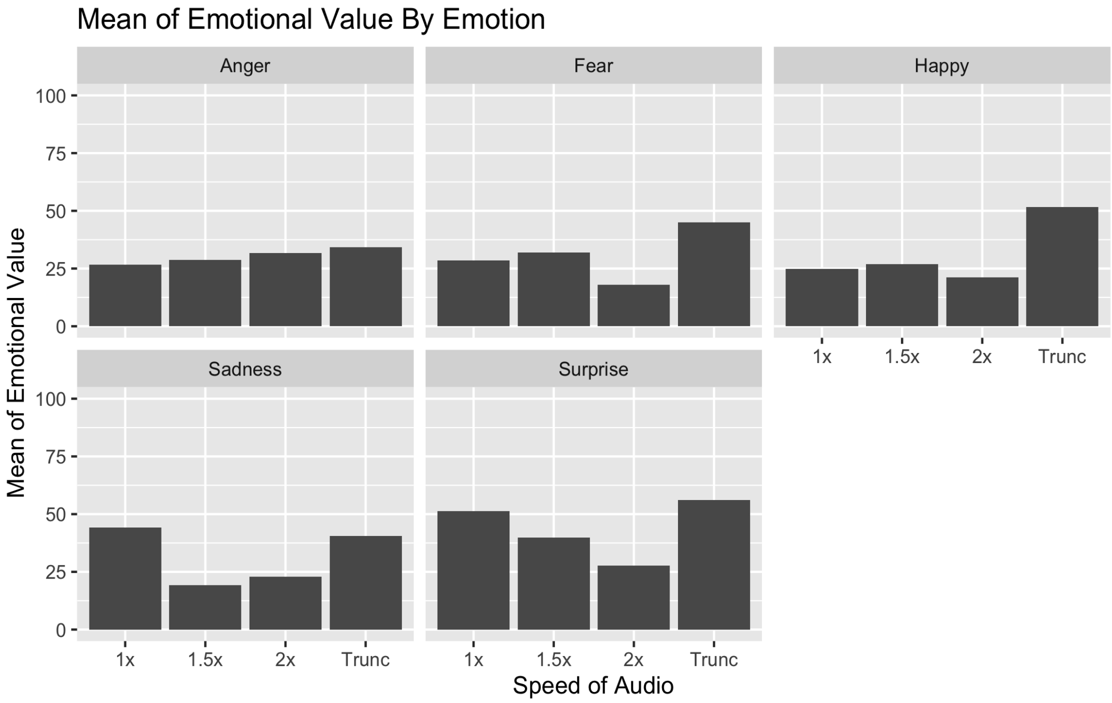


Figure 2: *Barplot of mean Emotional Values by Speed condition sorted into separate emotions.*

Comprehension and Speed

Figure 2 suggests that there is a slight decline in comprehension accuracy with speed conditions 1.5x and 2x. The truncated condition seems to not differ much from the 1x condition. However, it is important to note that the experimental design included audio clips that were 30 seconds long. Compared to other studies, this is a short duration and could have an effect on comprehension accuracy. There is a minute but linear decline in the speed conditions 1x (86.30%), 1.5x (84.41%), and 2x (80.14%). The comprehension results of the truncated condition (86.42%) seem to be quite similar compared to the 1x condition (86.30%).

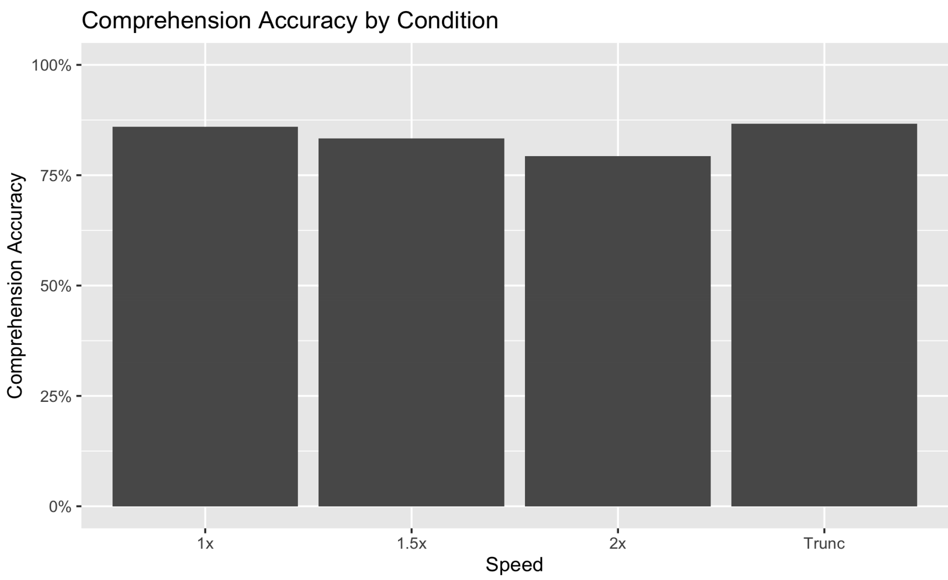


Figure 3: *Barplot of comprehension accuracy by speed condition.*

Reading Habits and Emotional Value

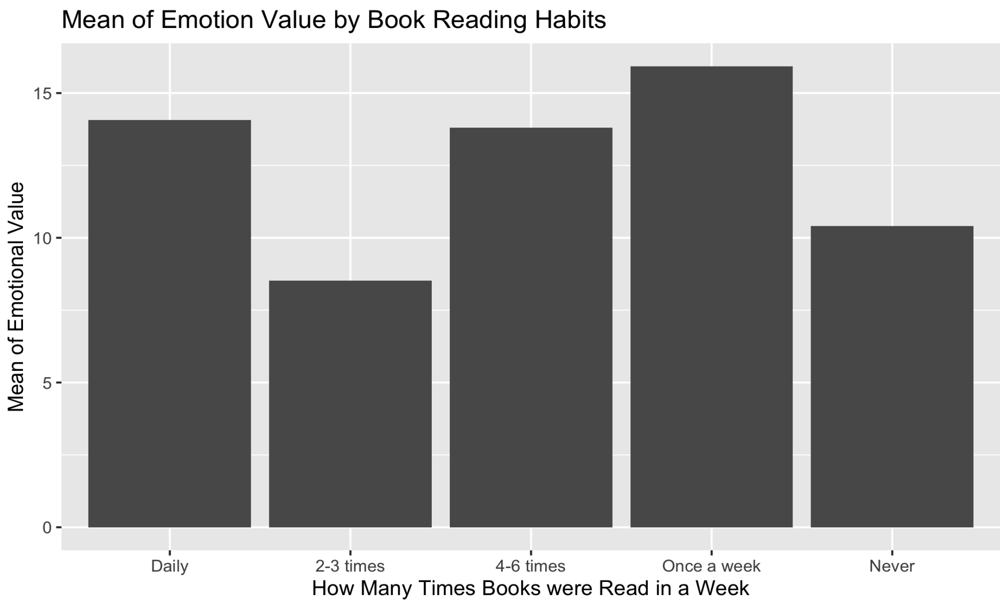


Figure 4: *Barplot of the sum of emotional values by book reading habits*

Audiobook Habit and Emotional Value

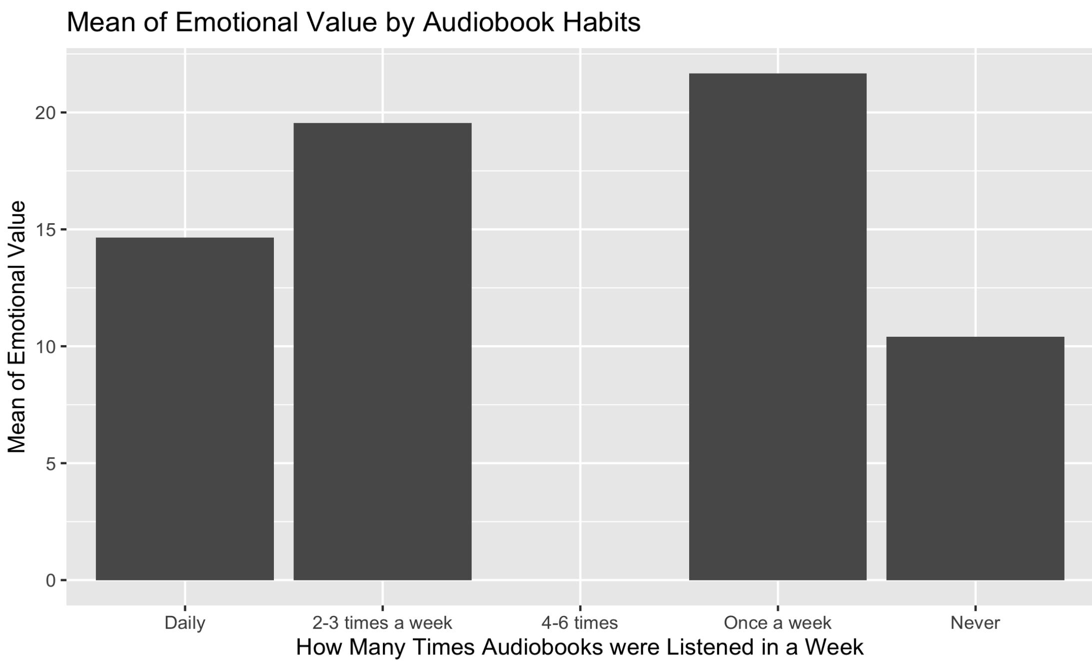


Figure 5: *Bar plot of the sum of emotional values by audiobook listening habits.*

4.2 Inferential Analysis

Five linear mixed models were performed using the Lme4 package (Bates et al., 2015). Speed was a fixed effect in each model. The participants and the item (the audio excerpt) were random effects in each model. P-values stated were achieved by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. Speed (without interaction term) was input as fixed effects into the model for the first four analyses. The random effects of the participants and items had intercepts.

4.3 Hypothesis One Results

A linear mixed model was used to test the first hypothesis stated: The 1.5x condition of the audio will not have a significant difference in interpretive listening in audiobooks compared to the 1x condition. The data was re-coded so that the sum of the five separate emotions were taken per item. As the hypothesis is investigating the effects of the 1x and 1.5x conditions only, a subset of only those condition variables were included in the analysis.

```
Data: dataQ1
Models:
modelQ1Empty: EmotionSum ~ 1 + (1 | Participant) + (1 | Item)
modelQ1: EmotionSum ~ Speed + (1 | Participant) + (1 | Item)
             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
modelQ1Empty    4 3148.4 3163.2 -1570.2   3140.4
modelQ1         5 3137.3 3155.8 -1563.6   3127.3 13.108  1   0.000294 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 6: *Linear mixed model of the sum of emotions and speed conditions 1x and 1.5x against an empty model.*

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: EmotionSum ~ Speed + (1 | Participant) + (1 | Item)
   Data: dataQ1

     AIC      BIC   logLik deviance df.resid
  3137.3   3155.8  -1563.6   3127.3      295

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.94884 -0.55302 -0.03099  0.49065  3.13251

Random effects:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept) 1444.7   38.01
 Item        (Intercept)  211.6   14.54
 Residual                1336.9   36.56
Number of obs: 300, groups:  Participant, 49; Item, 12

Fixed effects:
            Estimate Std. Error t value
(Intercept)   70.597      7.510   9.401
Speed1       -15.723      4.284  -3.670

Correlation of Fixed Effects:
       (Intr)
Speed1 -0.288
```

Figure 7: *Summary of regression model of the model with the interaction of the sum of emotions and speed conditions 1x and 1.5x.*

The null hypothesis that the 1.5x condition of the audio will not have a significant difference in interpretive listening in audiobooks compared to the 1x condition was rejected. The 1.5x speed condition affected emotional values ($\chi^2(1)$ =13.11, p=0.00029) lowering it by 15.70 ± 4.28 (standard errors).

4.4 Hypothesis Two Results

A linear mixed model was used to test the second hypothesis stated: the 2x condition of the audio will have a significant difference in interpretive listening in audiobooks compared to the 1x condition. The data was re-coded so that the sum of the five separate emotions were taken per item. As the hypothesis is investigating the effects of the 1x and 2x conditions only, a subset of only those condition variables were included in the analysis.

```
Data: dataQ2
Models:
modelQ2Empty: EmotionSum ~ 1 + (1 | Participant) + (1 | Item)
modelQ2: EmotionSum ~ Speed + (1 | Participant) + (1 | Item)
             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
modelQ2Empty    4 3087.7 3102.4 -1539.9   3079.7
modelQ2         5 3050.7 3069.1 -1520.3   3040.7 39.058  1  4.114e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 8: *Linear mixed model of sum of emotions and speed conditions 1x and 2x against an empty model.*

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: EmotionSum ~ Speed + (1 | Participant) + (1 | Item)
   Data: dataQ2

     AIC      BIC   logLik deviance df.resid
  3050.7   3069.1  -1520.3   3040.7      287

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.73729 -0.52603 -0.05332  0.42996  3.06399

Random effects:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept) 1456.4   38.16
 Item        (Intercept)  171.6   13.10
 Residual                1315.8   36.27
Number of obs: 292, groups:  Participant, 50; Item, 12

Fixed effects:
            Estimate Std. Error t value
(Intercept)   72.230      7.274    9.93
Speed2       -27.917      4.275   -6.53

Correlation of Fixed Effects:
       (Intr)
Speed2 -0.296
```

Figure 9: *Summary of regression model of the model with the interaction of the sum of emotions and speed conditions 1x and 2x.*

The null hypothesis that the 2x condition of the audio will not have an effect in interpretive listening in audiobooks compared to the 1x condition was rejected. 2x speed condition affected emotional values ($\chi^2(1) = 35.06$, $p = 0.00000000041$) lowering it by 27.91 ± 4.28 (standard errors).

4.5 Hypothesis Three Results

A linear mixed model was used to test the third hypothesis stated: The truncated condition of the audio will have a significant difference and reduce interpretive listening in audiobooks compared to the 1x condition. The data was re-coded so that the sum of the five separate emotions were taken per item. As the hypothesis is investigating the effects of the 1x and truncated conditions only, a subset of only those condition variables were included in the analysis.

```
Data: dataQ3
Models:
modelQ3Empty: EmotionSum ~ 1 + (1 | Participant) + (1 | Item)
modelQ3: EmotionSum ~ Speed + (1 | Participant) + (1 | Item)
             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)  
modelQ3Empty    4 3040.4 3055.1 -1516.2   3032.4                       
modelQ3         5 3039.3 3057.6 -1514.7   3029.3 3.1076  1    0.07793 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 10: *Linear mixed model of the sum of emotions and speed conditions 1x and truncated against an empty model.*

The null hypothesis that the truncated condition of the audio will not affect the interpretive listening in audiobooks compared to the 1x condition was not rejected. This suggests that the truncated condition did not affect emotional values ($\chi^2(1) = 3.10$, $p = 0.08$).

4.6 Hypothesis Four Results

A logit regression model was used to test the fourth hypothesis stated: the comprehension skills will have a significant decrease across speed conditions compared to the 1x condition. The data was re-coded this time compared to the first three analyses. The speed was re-coded from a categorical variable into a continuous variable, the time elapsed in each audio excerpt. Whether the participant answered the comprehension question correctly was coded in binary, 0 if it was not correct, and 1 if it was correct.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.88668    0.46887   1.891   0.0586 .
Speed        0.03323    0.01934   1.719   0.0857 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 509.42  on 585  degrees of freedom
Residual deviance: 506.53  on 584  degrees of freedom
AIC: 510.53

Number of Fisher Scoring iterations: 4
```

Figure 11: *Logit regression model of the speed conditions with the variable correct coded in binary (0 for incorrect, 1 for correct).*

The null hypothesis that the comprehension skills will not have an affect across speed conditions compared to the 1x condition. This suggests that speed conditions did not affect comprehension Pr(>|z|) = 0.086.

4.7 Hypothesis Five Results

A linear mixed model was used to test the fifth hypothesis stated: All aspects of interpretive listening (happiness, sadness, anger, fear, and surprise) will decrease or increase uniformly. Speed (interaction term) was input as fixed effects into the model. The random effects of the participants and items had intercepts.

```
Data: audio_long
Models:
model5: Value ~ Speed + Emotion + (1 | Participant) + (1 | Item)
interactionModel: Value ~ Speed * Emotion + (1 | Participant) + (1 | Item)
                 npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
model5             11 25646 25712 -12812    25624
interactionModel   23 25659 25796 -12806    25613 11.651 12     0.4741
```

Figure 12: *Linear mixed model of the interactions with emotions and speed against value compared against a simple model of emotions and speed with value.*

The null hypothesis that all aspects of interpretive listening (happiness, sadness, anger, fear, and surprise) will decrease or increase uniformly was not rejected. The chi-squared test suggests that there is no effect with the interaction ($\chi^2(12) = 11.65$, p = 0.47).

4.8 Post-hoc Analysis for Hypothesis Five

Five post hoc linear mixed models were used to investigate different emotions (happiness, sadness, anger, fear, and surprise) and whether it would be affected by different speed conditions. Speed (interaction term) was input as fixed effects into the model. The random effects of the participants and items had intercepts.

4.8.1 Happiness

```
Data: audioHappy
Models:
HappyEmptyModel: Happy ~ 1 + (1 | Participant) + (1 | Item)
HappyModel: Happy ~ Speed + (1 | Participant) + (1 | Item)
                npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
HappyEmptyModel    4 5028.6 5046.1 -2510.3   5020.6
HappyModel         7 5028.3 5058.9 -2507.1   5014.3 6.3555  3    0.09554 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 13: *Linear regression model of happiness and speed compared with its empty model.*

The chi-squared test shows that the null hypothesis that the speed conditions had no effect on the happiness emotional variable was not rejected ($\chi^2(3) = 6.36$, $p = 0.096$).

4.8.2 Sadness

```
Data: audioSad
Models:
SadEmptyModel: Sadness ~ 1 + (1 | Participant) + (1 | Item)
SadModel: Sadness ~ Speed + (1 | Participant) + (1 | Item)
              npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
SadEmptyModel    4 5037.1 5054.6 -2514.5   5029.1
SadModel         7 5010.7 5041.4 -2498.4   4996.7 32.315  3  4.493e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 14: *Linear regression model of sadness and speed compared with its empty model*

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Sadness ~ Speed + (1 | Participant) + (1 | Item)
   Data: audioSad

     AIC      BIC   logLik deviance df.resid
  5158.3   5189.0  -2572.1   5144.3      593

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.0797 -0.5544 -0.1410  0.2760  3.8606

Random effects:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept)  74.71    8.643
 Item        (Intercept)  57.42    7.577
 Residual                260.89   16.152
Number of obs: 600, groups:  Participant, 50; Item, 12

Fixed effects:
               Estimate Std. Error t value
(Intercept)      16.635      2.834   5.870
Speed1.5x        -7.445      1.875  -3.971
Speed2x         -10.619      1.872  -5.673
SpeedTruncated   -3.391      1.875  -1.809

Correlation of Fixed Effects:
            (Intr) Spd1.5 Sped2x
Speed1.5x   -0.331
Speed2x     -0.330  0.499
SpeedTrnctd -0.331  0.501  0.499
```

Figure 15: *Summary of the regression model for sadness with speed conditions.*

The null hypothesis that the speed conditions had no effect on the sadness emotional variable was rejected. The emotional value ($\chi^2(3) = 32.32$, $p = 0.00000045$) lowered by 6.61 ± 1.85 (standard errors) compared to the 1.5x condition, lowered by 10.54 ± 1.87 (standard errors) compared to the 2x condition, and lowered by 4.22 ±1.91 (standard errors) compared to the truncated condition.

4.8.3 Anger

```
Data: audioAngry
Models:
AngerEmptyModel: Anger ~ 1 + (1 | Participant) + (1 | Item)
AngerModel: Anger ~ Speed + (1 | Participant) + (1 | Item)
                npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
AngerEmptyModel    4 5089.1 5106.6 -2540.6   5081.1
AngerModel         7 5093.7 5124.4 -2539.9   5079.7 1.3673  3     0.7132
```

Figure 16: *Linear regression model of anger and speed compared with its empty model*

The chi-squared test shows that the null hypothesis that the speed conditions had no effect on the anger emotional variable was not rejected ($\chi^2(3) = 1.37$, $p = 0.71$).

4.8.4 Surprise

```
Data: audioSurprise
Models:
SurpriseEmptyModel: Surprise ~ 1 + (1 | Participant) + (1 | Item)
SurpriseModel: Surprise ~ Speed + (1 | Participant) + (1 | Item)
                   npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
SurpriseEmptyModel    4 5211.6 5229.1 -2601.8   5203.6
SurpriseModel         7 5206.0 5236.7 -2596.0   5192.0 11.604  3   0.008872 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 17: *Linear regression model of surprise and speed compared with its empty model*

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Surprise ~ Speed + (1 | Participant) + (1 | Item)
   Data: audioSurprise

     AIC      BIC   logLik deviance df.resid
  5340.6   5371.4  -2663.3   5326.6      593

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1201 -0.6015 -0.1136  0.3922  3.2684

Random effects:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept) 136.5    11.68
 Item        (Intercept) 100.2    10.01
 Residual                344.3    18.56
Number of obs: 600, groups:  Participant, 50; Item, 12

Fixed effects:
              Estimate Std. Error t value
(Intercept)     18.947      3.660   5.177
Speed1.5x       -2.132      2.154  -0.990
Speed2x         -5.055      2.151  -2.350
SpeedTruncated   1.271      2.154   0.590

Correlation of Fixed Effects:
            (Intr) Spd1.5 Sped2x
Speed1.5x   -0.294
Speed2x     -0.294  0.499
SpeedTrnctd -0.294  0.502  0.499
```

Figure 18: *Summary of the regression model for surprise with speed conditions.*

The null hypothesis that the speed conditions had no effect on the surprise emotional variable was rejected. The emotional value ($\chi^2(3) = 11.60$, $p = 0.0089$) was lowered by 2.83 ± 2.15 (standard errors) in the 1.5x condition, lowered by 5.28 ± 2.18 (standard errors) in the 2x condition, and increase by 1.71 ±2.22 (standard errors in the truncated condition.

4.8.5 Fear

```
Data: audioFear
Models:
FearEmptyModel: Fear ~ 1 + (1 | Participant) + (1 | Item)
FearModel: Fear ~ Speed + (1 | Participant) + (1 | Item)
               npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
FearEmptyModel    4 5019.3 5036.8 -2505.7   5011.3
FearModel         7 5010.3 5040.9 -2498.1   4996.3 15.05  3   0.001774 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 19: *Linear regression model of fear and speed compared with its empty model*

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: Fear ~ Speed + (1 | Participant) + (1 | Item)
   Data: audioFear

     AIC      BIC   logLik deviance df.resid
  5167.2   5198.0  -2576.6   5153.2      593

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2648 -0.5890 -0.0937  0.2716  4.4817

Random effects:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept) 115.46   10.745
 Item        (Intercept)  50.05    7.075
 Residual                257.60   16.050
Number of obs: 600, groups:  Participant, 50; Item, 12

Fixed effects:
               Estimate Std. Error t value
(Intercept)      13.837      2.865   4.829
Speed1.5x        -1.915      1.863  -1.028
Speed2x          -6.688      1.860  -3.596
SpeedTruncated   -2.550      1.863  -1.369

Correlation of Fixed Effects:
            (Intr) Spd1.5 Sped2x
Speed1.5x   -0.325
Speed2x     -0.325  0.499
SpeedTrnctd -0.325  0.501  0.499
```

Figure 20: *Summary of the regression model for fear with speed conditions.*

The null hypothesis that the speed conditions had no effect on the fear emotional variable was rejected. The emotional value ($\chi^2(3) = 15.05$, $p = 0.0018$) lowered by 2.29 ± 1.83 (standard errors) in the 1.5x condition, lowered by 6.87 ±1.85 (standard errors) in the 2x condition, and lowered by 1.54 ± 1.89 (standard errors) in the truncated condition.

4.9 Hypothesis Six Results

The sixth hypothesis proposed that the manipulated conditions will have a worse effect with listeners aged 35 and onwards compared to listeners aged 18-24 and aged 25-34. The dataset contained 50 participants of the 50 participants, 45 were aged 18-24, 3 were aged 25-34, 1 was aged 35-44, and 1 was aged 55-64. There were not enough participants of different age brackets to perform a statistical analysis.

5 Discussion

5.1 Hypothesis One

The results from the linear mixed model (Figure 6) suggests that there is a decrease in emotional response with the 1.5x condition compared to the 1x condition. The statistical analysis results are not consistent with our hypothesis that the 1.5x condition of the audio will not have a significant difference in interpretive listening in audiobooks compared to the 1x condition. The linear regression also states that when participants listened to the audio in the 1.5x condition there was a decline in emotional response.

The current results are consistent with the Moore et al. (1986) study that increased speech rates lead to fewer cognitive responses. Although they contradict the King et al. (1983) study, which states that interpretive listening levels are mostly sustained until it reaches the 1.6x condition. This could be because the study uses speed conditions 1.3x, 1.45x, and 1.6x. The study mentioned that in future studies more increments in speed condition are required to pinpoint the exact threshold for interpretive listening before interpretive listening levels decline significantly. Since the current experiment uses the increment 1.5x, which is in between 1.45x and 1.6x, this could suggest that the threshold could be slower than the 1.6x condition. However, it should also be noted that King et al. (1983) utilise a standardised comprehension test. The current design also differs in method of compression, source materials, test design, and length of audio.

Additionally, Ilie and Thompson (2006) stated that an increase of 1.26x speed in a speech reported more intensity in the speech. This is also contrary to our results; however, the study also elaborated that participants reported that the speech became unpleasant. This could negatively affect the audiobook experience, as the increase in intensity is not derived from the actual emotions from the speech but from the unpleasantness from the sped-up audio.

LaBarbara and MacLachlan (1979) concluded that brands were reported to be more interesting and caught more attention at the 1.3x condition. Although the sped-up audio caught more attention, it does not signify that it was more emotive, as this study was also focused on advertisements and brand image, which diverge from the current research aims.

5.2 Hypothesis Two

The statistical results from the linear mixed model (Figure 8) are in line with our hypothesis that the 2x condition will have a significant difference in interpretive listening in audiobooks compared to the 1x condition.

The results are in accordance with our first hypothesis, which already signify a decline in emotional responses in the 1.5x condition. The results of the current study fit the theory that there is a certain threshold for speed conditions before there is a sharp decline in emotional response, although King et al. (1983) states that this is any speed faster than 1.6x. There seems to be a 77.77% decline in emotional response in the 2x condition compared to the 1.5x condition.

5.3 Hypothesis Three

Our hypothesis that the truncated condition of the audio will reduce interpretive listening in audiobooks compared to the 1x condition is inconsistent with our statistical analysis (Figure 10).

This is inconsistent with the comprehension studies about truncated speech. MacGregor et al. (2010) position that silence could affect comprehension and recall. Huggins (1975) stated that silences that were less than 60ms favoured intelligibility. Pelle and Davis (2012) also reported that speech without the insertion of silences was difficult to comprehend and the insertion of silence improved comprehension. The method of truncation used in the current experimental survey only removed silences that were quieter than 24dB and only focused on the removal of longer segments of silence. The audiobook excerpts used also had varied recording equipment and could affect the process of silence trimming.

However, this would be consistent with the King et al. (1983) study that suggested the rate of interpretive listening was sustained for longer compared to comprehension. These results could be contrasting to our previous two hypotheses because our selective sampling method preserved the original pitch rate, as opposed to our other time-compression method which increased the pitch rate. As Ilie and Thompson (2006) posits, the higher pitched speech was often viewed as more unpleasant by participants.

5.4 Hypothesis Four

The hypothesis that comprehension skills will have a significant decrease across speed conditions compared to the 1x condition was not supported by our statistical analysis as shown in Figure 11.

Foulke (1966) and Carver (1973) proposed that comprehension is in linear decline until speech rate reaches a certain threshold. Although both state a different threshold, Foulke (1966) being 275wpm and Carver (1973) 150wpm. Our statistical analysis reveals that there is no significant difference in comprehension capabilities across different speed conditions. However, when we look at the barplot (Figure 3) we can see that there is a minute but linear decline in the speed conditions 1x, 1.5x, and 2x. Carver (1973) also stated that comprehension capabilities and speed thresholds were dependent on the individual. This could have posed an effect on the current results as there was a small sample size (n=50) for multiple items and conditions. Contrarily, Griffiths (1990) postulated that participants scored significantly higher when audio was played at a slower rate than at an average or faster rate. Although admitting that more needs to be achieved for the external validity of their experimental design.

The disparity in conclusions around comprehension skills and time-compressed speech could be because the current experiment utilises short audio excerpts, that are a maximum of 30 seconds long. Compared to most studies, the length of the original audio is shorter, which does not cause as much attention and memory fatigue to the listener. It could be possible that the short audio excerpts in the current study results are more attributable to short-term recall, which is also said to sustain up until a certain threshold (King et al., 1983).

The majority of studies agree that there is a threshold for comprehension and speed but have different indications to what the threshold is, with the exception of Ritzhaupt et al. (2008) that revealed no significant differences across 1x, 1.4x, and 1.8x conditions. Yang et al. (2020) stated that the threshold for comprehension was 1.5x for lecture content. While Wilson et al. (2018) indicated that the threshold was actually 1.6x to 1.7x for lecture content.

Increased exposure to time-compressed speech could allow the participants to accommodate to the speed condition (Adank & Devlin, 2010; Dupoux & Green, 1997). The current survey had a pseudo-randomised order that could have allowed for this accommodation. However, when the conditions cycled back from truncated to 1x condition, this could have interrupted the gradual exposure.

The comprehension results of the truncated condition are similar compared to the 1x condition. This is in contrast to the theory that participants were more likely to recognise words that followed silent pauses (Huggins, 1975; MacGregor et al., 2010; Pelle & Davis, 2012). However, the selective sampling method only removed the longer pauses and not every pause that was within the speech signal. This could posit that the removal of longer pauses could still preserve the quality of comprehension.

5.5 Hypothesis Five

The statistical analysis (Figure 12) suggests that all aspects of interpretive listening (happy, sad, anger, fear, and surprise) would decrease uniformly, which is in line with our hypothesis. Although our statistical analysis seems to suggest this, it could be possible that the core emotions selected for this study all have the same degree of decline in emotional value. Therefore, a post-hoc analysis was performed, inspecting each emotion and its emotional value per speed condition. This post-hoc analysis is important as it could provide a basis for what core emotions to use for future research regarding time-compression and interpretive listening.

From the post-hoc analysis, the emotions sadness, surprise, and fear show statistically significant results for the different speed conditions. This could suggest that the non-significant emotions happy and anger are more robust to time-compressed speech. This analysis would be consistent with the theory that participants found brands more interesting when with time-compressed speech (LaBarbera & MacLachlan 1979; MacLachlan 1979). Millet et al. (1976) also reported that in increased speech rates, messages were viewed as more persuasive. The suppression of negative emotions such as sadness and fear could be a potential reason why advertisements are reported as more favourable in time-compressed conditions.

Another interesting point is that both the emotions of fear and surprise are affected by time-compressed speech. These results could be evidenced to support the claim in the Jack et al. (2014) study which revealed that fear and surprise could instead be collapsed into one primary emotion. However, the post-hoc analysis needs to be interpreted with caution, as our primary statistical analysis revealed that the emotions decrease uniformly.

5.6 Hypothesis Six

There is insufficient data in age groups in our current experiment to illustrate any conclusion with the hypothesis that manipulated conditions would have a worse effect with listeners aged 35 onwards compared to listeners aged 18-24 and aged 24-34. Hall and Mueller (1997) indicated that when speed information were at higher frequencies, hearing perception can be poorer in older populations. Since our speech was manipulated in a way that resulted in a higher pitch, this could suggest that age would pose an effect on comprehension and interpretive listening across varying speed conditions in time-compressed speech, although further research would need to be performed.

5.7 Limitations of Study

The experiment provides participants with twelve audio clips that are arranged in a pseudo-random order. Studies have revealed that the amount of exposure you have to increased rates of speech could alter and improve comprehension (Adank & Devlin, 2010; Dupoux & Green, 1997). The pseudo-random order could potentially affect comprehension in this regard.

The method in which the audio is compressed could be a crucial factor in comprehension and interpretive listening. The audio is compressed without altering the pitch and therefore results in a higher pitch than the original audio. This could affect the results for older participants as studies show poorer hearing acuity amongst these participants (Hall & Mueller, 1997).

Ideally, this experiment would be performed in person in a quiet room. However, because of current restrictions due to the Covid-19 pandemic, this has been moved online. This could be a disadvantage to the data as it is not possible to monitor whether participants have listened to the audio more than once. On the other hand, the online survey method could strengthen the experiment's external validity as it mimics how most listeners consume audiobooks, in a place of their choice and on their own electronic devices. Since all the survey participants are anonymous, there is also no channels to invite them back to do a further study that explores recall abilities.

Since the experiment only provides English audio excerpts, the results and conclusions made may differ from other languages, especially tonal languages.

5.8 Further Research

Further study should focus on the effects and rates of time-compression with different methods of compression. Recall ability should also be investigated to further our understanding of this phenomenon. It may not always be the case that all audio clips have the same effect uniformly and the threshold amongst individual participants may also vary (Carver, 1973; Moore et al., 1986). Therefore, there could be further investigation into whether there is a pattern between demographics or groups that could have similar thresholds. There could also be specially designed audio that could further tease out what makes some audio more attractive when the rate of speech is increased.

6 Conclusion

There has been an increasing demand for speed playback functions in audiobooks and general audio media. As technology keeps up with the demands of the modern world, it is important to consider its ramifications. Linear mixed models and logit regression models were used in the current study to answer the question of what the effects and rates of interpretive listening in time-compressed speech in audiobooks were.

The first hypothesis explored whether the 1.5x speed condition had a significant difference in the rate of interpretive listening compared to the normal speed condition. The current results of the study do demonstrate a significant difference, which is in line with the Moore et al. (1986) study that suggested increased rates of speech would invoke less cognitive responses in participants. The second hypothesis inspected whether the 2x speed condition had a significant difference in the rate of interpretive listening compared to the normal speed condition. The current results once again indicate that there is a significant difference in interpretive listening between the two conditions. Additionally, the results of the first and second hypothesis also are consistent with the findings of the King et al. (1986) study, stating that the rate of interpretive listening is sustained moderately until a certain threshold. The third hypothesis investigated the effects of the truncated speed condition and whether it had a significant difference in the rate of interpretive listening compared to the normal speed condition. Our results concluded that there was no significant difference between the two conditions. This would be in line with the King et al. (1983) study that indicates the rate of interpretive listening is sustained for increased rates of speech compared to comprehension.

The fourth hypothesis examined whether the quality of comprehension was affected by the various speed conditions (1.5x, 2x, and truncation) compared to the normal speed condition. The results stated that there was no significant difference in the speed conditions and is contrary to the literature. Although descriptive data results seem to point towards the direction of the literature. The fifth hypothesis inspected the emotions happy, sad, anger, fear, and surprise separately, asking if all aspects of interpretive listening would decrease or increase uniformly with the speed conditions. Interestingly, among the five emotions inspected, the post-hoc analysis seems to reveal that the emotions sad, surprise, and fear have a more significant difference. The sixth hypothesis questioned if increased speech rates had a worse effect with older participants (aged 35 onwards) compared to listeners aged 18-24 and aged 24-34. This unfortunately had insufficient data to conclude the hypothesis.

The post-hoc analysis of the fifth hypothesis could be a potential area of future research, for example, different emotions from various theories could be used to design a thorough standardised interpretive listening test. More research also has to be performed for the rate of interpretive listening with increased rates of speech with older age groups.

7 References

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage,* 49(1), 1124-1132.

Apple, W., Streeter, L., & M. Krauss, R. (1979). Effects of Pitch and Speech Rate on Personal Attribution. *Journal of Personality and Social Psychology,* 37(May), 715-727.

Arment, Marco (2021). *Overcast* (Version 2021.2) [Mobile app]. Retrieved from *Apple App Store*: <https://apps.apple.com/us/app/overcast/id888422857>.

Arons, B. (1992). Techniques, Perception, and Applications of Time-Compressed Speech. *American Voice I/O Society,* Sep. 1992, 69-177.

Barabasz, A. F. (1968). A Study of recall and retention of accelerated lecture presentation. *Journal of Communication*, 18(3), 283-287.

Barrie, J. M. (1911). *Peter Pan.* Retrieved from *Libirvox*: <http://ia902707.us.archive.org/3/items/peterpan0801librivox/peterpan14barrie64kb.mp3>.

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, 67(1), 1–48.

Bazán, E. P. (2005). *First Love And Other Fascinating Stories of Spanish Life*. Retrieved from *Librivox*: <http://www.archive.org/download/lovestoriesvolume11910librivox/lovestories110various64kb.mp3>.

Bodie, G., & Fitch-Hauser, M. (2010). Quantitative Research in Listening: Explication and Overview. *Listening And Human Communication In The 21St Century*, 46-93.

Bostrom, R., French, R., Johnson-Laird, P., & Parshall, C. *1*. Uky.edu. Retrieved 24 April 2021, from *University of Kentuky*: <http://www.uky.edu/~bostrom/listfinal!.htm.>.

Burnett, F. H. (1911). *The Secret Garden.* Retrieved from *Librivox*: <http://ia800204.us.archive.org/15/items/secretgardenversion2librivox/secretgarden06burnett64kb.mp3>.

Caroll, L. (1865). *Alice's Adventures in Wonderful.* Retrieved from *Librivox*: <http://www.archive.org/download/alice_wonderland_0711_librivox/alice_09_carroll_64kb.mp3>.

Carver, R. P. (1973). Effects of increasing the rate of speech presentation upon comprehension. *Journal of Educational Psychology, 65*(1), 118-126.

David, E., & McDonald, H. (1956). A Note on Pitch-Synchronous Processing of Speech. *The Journal Of The Acoustical Society Of America*, *28*(7), 1261-1266.

Doyle, A. C. (1891). *A Scandal in Bohemia.* Retrieved from *Librivox:* <http://ia800304.us.archive.org/33/items/worlds_best_mystery_2_librivox/bestmysterystories2_07_hawthorne_64kb.mp3>.

Doyle, A. C. (1981). *A Case of Identity.* Retrieved from *Librivox:* <http://www.archive.org/download/worlds_best_mystery_2_librivox/bestmysterystories2_06_hawthorne_64kb.mp3>.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance,* 23(3), 914-927.

Eide, B. E. & Eide, F. F. (2011). *The dyslexic advantage: Unlocking the hidden potential of the dyslexic brain* [Audible audiobook]. Retrieved from *Amazon:* <http://www.amazon.com>.

Ekman, P., Sorenson, E., & Friesen, W. (1969). Pan-Cultural Elements in Facial Displays of Emotion. *Science*, 164(3875), 86-88.

Fairbanks, G., Everitt, W., & Jaeger, R. (1954). Experimental Phonetics. *University Of Illionois Press*.

Foulke, E. (1966). Comparison of comprehension of two forms of compressed speech. *Exceptional Children,* 33(3), 169-173.

Ghitza, O., & Greenberg, S. (2009). On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica*, 66(1-2), 113-126.

Griffiths, R. (1990). Speech rate and nns comprehension: A preliminary study in time-benefit analysis*. *Language Learning,* 40(3), 311-336.

Hall, J., & Mueller, G. (1997). Audiologist desk reference. *Audiologist Desk Reference, San Diego, CA: Singular Publishing*.

Hauser, M., & Hughes, A. (1987). A Factor Analytic Study of Four Listening Tests. *International Listening Association. Journal*, 1(1), 129-147.

He, L., & Gupta, A. (2001). Exploring benefits of non-linear time compression. *Proceedings Of The Ninth ACM International Conference On Multimedia - MULTIMEDIA '01*.

Hovland, C., & Weiss, W. (1951). The Influence of Source Credibility on Communication Effectiveness. Public Opinion Quartersly, 15(4), 635.

Hu, J. (2017, June 08). What's your 'x' rating? Retrieved 24/02/2021, from *Audible*: <https://www.audible.com/blog/the-listening-life/whats-your-x-rating>.

Huggins, A. (1975). Temporally segmented speech. *Perception & Psychophysics*, 18(2), 149-157.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception,* 23(4), 319-330.

Jack, R., Garrod, O., & Schyns, P. (2014). Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time. *Current Biology*, 24(2), 187-192.

Bostrom R. N. , Waldhart E. S. (1983). *Kentucky Comprehensive Listening Test.*

King, P. E., & Behnke, R. R. (1989). The effect of time-compressed speech on comprehensive, interpretive, and short-term listening. *Human Communication Research,* 15(3), 428-443.

Kisielius, J., & Sternthal, B. (1984). Detecting and explaining vividness effects in attitudinal judgments. *Journal of Marketing Research,* 21(1), 54.

Klumpp, R., & Webster, J. (1961). Intelligibility of Time-Compressed Speech. *The Journal Of The Acoustical Society Of America*, 33(3), 265-267.

Kuprin, A. I. (1969). *A Sentimental Romance.* Retrieved from *Librivox:* <http://www.archive.org/download/lovestoriesvolume_1_1910_librivox/lovestories1_15_various_64kb.mp3>.

LaBarbera, P., & MacLachlan, J. (1979). Time-Compressed speech in radio advertising. *Journal of Marketing,* 43(1), 30.

Leathers, D. (1979). The impact of multichannel message inconsistency on verbal and nonverbal decoding behaviors. *Communication Monographs*, 46(2), 88-100.

Lehiste, I. (1973). Phonetic Disambiguation of Syntactic Ambiguity. *The Journal of the Acoustical Society of America*, 53(1), 380-380.

Licklider, J., & Miller, G. (1948). The Intelligibility of Interrupted Speech. *The Journal of the Acoustical Society of America*, 20(4), 593-593.

Littlepage, G., & Pineault, M. (1981). Detection of Truthful and Deceptive Interpersonal Communications across Information Transmission Modes. *The Journal of Social Psychology*, 114(1), 57-68.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia,* 48(14), 3982-3992.

MacLachlan, J. (1979). What People Really Think of Fast Talkers. *Psychology Today,* 13(November), 113-117.

Mathews, F. A. (2015). *A Christmas Honeymoon.* Retrieved from *Librivox:* <http://www.archive.org/download/a_christmas_honeymoon_2012_librivox/christmashoneymoon_08_mathews_64kb.mp3>.

Megehee, C. M. (2009). Advertising time expansion, compression, and cognitive processing influences on consumer acceptance of message and brand. *Journal of Business Research*, *62*(4), 420-431.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology,* 34(4), 615-624.

Moore, D. L., Hausknecht, D., & Thamodaran, K. (1986). Time compression, Response opportunity, and persuasion. *Journal of Consumer Research,* 13(1), 85.

Murrow, L. (2018, September 29). Speed-Listening and the trouble WITH 'PODFASTERS'. Retrieved 26/02/2021, from *Wired*: <https://www.wired.com/story/stop-speed-listening-podcasts/>.

NHC Software. (2004). *Wavepad Audio Editor.* (Version 12.08) [Software]. Retrieved from *Apple App Store*: <https://apps.apple.com/us/app/wavepad-music-and-audio-editor/id395339564>.

Omoigui, N., He, L., Gupta, A., Grudin, J., & Sanocki, E. (1999). Time-compression: Systems Concerns, Usage, and Benefits. *Proceedings Of The SIGCHI Conference On Human Factors In Computing Systems The CHI Is The Limit - CHI '99*.

Pastore, R. S. (2009). The effects of diagrams and time-compressed instruction on learning and learners' perceptions of cognitive load. *Educational Technology Research and Development,* 58(5), 485-505.

Pelle, J., & Davis, M. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers In Psychology*, 3.

Perrin, A. (2021, February 22). One-in-five Americans now listen to audiobooks. Retrieved 24/02/2021, from *Pew Research:* <https://www.pewresearch.org/fact-tank/2019/09/25/one-in-five-americans-now-listen-to-audiobooks/>.

Plutchik, R. (1982). A psychoevolutionary theory of emotions. *Social Science Information*, 21(4-5), 529-553.

Pratkanis, A., Greenwald, A., Leippe, M., & Baumgardner, M. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead: Long live the sleeper effect. *Journal of Personality and Social Psychology*, 54(2), 203-218.

Ritzhaupt, A. D., Gomes, N. D., & Barron, A. E. (2008). The effects of time-compressed audio and verbal redundancy on learner performance and satisfaction. *Computers in Human Behavior,* 24(5), 2434-2445.

RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>.

Schlinger, M. J., Alwitt, L. F., McCarthy, K. E., & Green, L. (1983). Effects of time compression on attitudes and information processing. *Journal of Marketing,* 47(1), 79.

Schreiber, P. (1991). Understanding prosody's role in reading acquisition. *Theory Into Practice*, 30(3), 158-164.

Skinner, C. H., Robinson, D. H., Robinson, S. L., Sterling, H. E., & Goodman, M. A. (1999). Effects of Advertisement Speech rates on FEATURE recognition, and product and SPEAKER RATINGS. *International Journal of Listening,* 13(1), 97-110.

Twain, M. (1884). *The Adventures of Huckleberry Finn.* Retrieved from *Librivox:* <http://www.archive.org/download/huckleberry_mfs_librivox/huckleberry_finn_03_twain_64kb.mp3>.

Verne, J. (1873). *Around the World in Eighty Days*. Retrieved from *Librivox:* <http://ia802903.us.archive.org/11/items/around_world_rs_1005_librivox/aroundtheworldineightydays_06_verne_64kb.mp3>.

Villaume, W., & Weaver, III, J. (1996). A Factorial Approach to Establishing Reliable Listening Measures from the WBLT and the KCLT: Full Information Factor Analysis of Dichotomous Data. *International Journal of Listening*, 10(1), 1-20.

Watson, K., & Barker, L. (1983). *Watson-Barker Listening Test*. SPECTRA, Inc.

Wells, H. G. (1897). *The Invisible Man.* Retrieved from *Librivox:* <http://www.archive.org/download/invisibleman_1209_librivox_cb/invisibleman_08_wells_64kb.mp3>.

Wheeler, A. C. (1891). *The End of All.* Retrieved from *Librivox:* <http://ia801209.us.archive.org/9/items/short_science_fiction_collection_055_1607_librivox/ssf055_endofall_crinkle_jdb_64kb.mp3>.

Wheeless, L. R. (1971). Some effects of Time - Compressed speech on persuasion. *Journal of Broadcasting,* 15(4), 415-420.

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.

Wilson, K. E., Martin, L., Smilek, D., & Risko, E. F. (2018). The benefits and costs of speed watching video lectures. *Scholarship of Teaching and Learning in Psychology,* 4(4), 243-257.

Wright, P. (1981). Cognitive Responses to Mass Media Advocacy. *Cognitive Responses in Persuasion* (pp. 263-282). Hillsdale: Brock TC.

Yang, X., Lin, L., Wen, Y., Cheng, P., Yang, X., & An, Y. (2020). Time-Compressed Audio on Attention, Meditation, Cognitive Load, and Learning. *Educational Technology & Society,* 23(3), 16-26.

Konkle, D., Beasley, D., & Bess, F. (1977). Intelligibility of Time-Altered Speech in Relation to Chronological Aging. *Journal of Speech and Hearing Research*, 20(1), 108-115.

Sticht, T., & Gray, B. (1969). The intelligibility of time compressed words as a function of age and hearing loss. *J Speech Hear Res*, 12:443-446.

Tomasz, L., & Nancy, P. (1996). Comprehension of time-compressed speech: effects of age and speech complexity. *Journal of the American Academy of Audiology*, *7*(6), 447-457. Retrieved from *Audiology:* <https://www.audiology.org/sites/default/files/journal/JAAA070607.pdf>.

Acknowledgements

I would first like to thank my dissertation supervisor, Dr Chris Cummins for his encouragement and support throughout the entirety of my dissertation. I would also like to extend my gratitude towards the PPLS writing centre, which gave me valuable advice on how to make my dissertation more concise and clearer. I thank all the participants who took their time to take part in my research.

Additionally, I would like to thank my family for their encouragement and faith in me. For I would not currently be in Edinburgh, writing my dissertation and become the first in my family to receive a university education. Lastly, I extend my appreciation to my friends that have supplied home-cooked meals, small gifts, and a sympathetic ear throughout my dissertation writing process.

Evidence in Support of a Cognitive Bias for Cross-Category Harmony between the Verb Phrase and the Adpositional Phrase in the Absence of Surface-Level Patterns

Cliodhna Hughes
University of Edinburgh

Abstract. Most languages for which we have the data display syntactic harmony between the verb phrase and the adpositional phrase. Recent experimental work has indicated there may be a cognitive bias for harmony which contributes to this. These studies, however, all use the same objects for priming and testing stimuli. This leaves open the possibility of participants noticing surface-level patterns such as 'the saucepan is always gestured first', and using these in their responses, thus giving the appearance of a preference for syntactically harmonic patterns. This paper describes three experiments designed to establish whether there is a cognitive bias for harmony between the adpositional phrase and the verb phrase, when the possibility of using surface-level patterns is removed. Experiment 1 investigates the possibility of a baseline preference for adpositional phrase order in English-speaking participants, because it is the extent of their preference for particular adpositional orders which is manipulated in experiments 2 and 3. Results show no evidence for a baseline preference for either prepositions or postpositions. Experiment 2 finds that the experimental methods employed (silent gesture perception and artificial language learning) are sufficient to demonstrate a harmony effect when the elements in the priming and testing stimuli are the same. Experiment 3 reveals this preference for harmonic patterns is also present when there are no repeated elements in the priming and testing stimuli. This indicates that there is a bias for harmony between the verb phrase and the adpositional phrase which at least partially contributes to its cross-linguistic prevalence.

Keywords: cross-category syntactic harmony; cognitive bias; language evolution

---

1 Introduction

Languages differ widely from each other, yet there are a number of typological regularities which hold across languages (e.g., Greenberg, 1963). One such regularity is the tendency for dependents to consistently appear on one side of the head of the phrase. Languages which have this property are said to be *harmonic*. In this paper I investigate a particular type of harmony, specifically that between the verb phrase and the adpositional phrase. This is an instance of cross-category harmony, as it acts across phrase types. Previous work has suggested that harmony between the verb phrase and the adpositional phrase could be a product of the historical tendency for adpositions to derive from verb phrases or from genitives, in a process which preserves linear order (Moravcsik, 2010). Research using experimental methods, on the other hand, has found some evidence suggestive of a cognitive bias for cross-category harmony using the artificial language learning paradigm (Cook, 1988; Zhao and Fedzechkina, 2020; Wang et al., 2021). Zhao and Fedzechkina (2020) only find such evidence in the head-final direction, but this is likely due to the fact that there is a baseline preference for SOV for the type of events stimuli used in their experiments (Goldin-Meadow et al., 2008; Schouwstra and de Swart, 2014), which competes with the harmony bias, preventing the effect from being demonstrated in the prepositional condition. The studies in Cook (1988), Zhao and Fedzechkina (2020) and Wang et al. (2021) also fail to consider the possibility of a similar baseline adpositional order preference, either for prepositions or for postpositions, which may similarly be affecting their results. Moreover, in all three studies the items used in the test phase shared elements with those used in the training phase. It is therefore possible that participants used surface level rules in their responses, (e.g., ‘the word for table always comes last’) rather than a more general syntactic rule (e.g., ‘the head comes before the dependent’).

This paper describes three experiments which together aimed to establish whether there is evidence for a bias for cross-category harmony between the verb phrase and the adpositional phrase, on a syntactic rather than surface level. Experiment 1 of this paper uses silent gesture perception methods to demonstrate that there is no baseline preference for either prepositions or postpositions when testing native English speakers using this method. This is crucial for comparison with experiments 2 and 3, in which this preference is manipulated. Experiment 2 combines silent gesture perception with artificial language learning methods, in an extrapolation paradigm, finding that evidence of a bias for cross-category harmony can be found using these methods on native English speakers, when the objects in the training stimuli are the same as those in the testing stimuli. This replicates the findings of Wang et al. (2021), which used different methods. Finally, experiment 3 tests whether such evidence is still found using these methods when there is no overlap in the objects used for the training stimuli and testing stimuli. There is some evidence for a stronger harmony preference in experiment 2 than experiment 3, highlighting the importance of ensuring surface-level patterns are avoided when designing stimuli for such experiments. Nevertheless, a preference for harmonic patterns is still found in experiment 3, supporting the hypothesis that a cognitive bias for harmony between the verb phrase and the adpositional phrase exists.

2 Background

2.1 Word-Order Harmony

Word-order harmony is the phenomenon whereby dependents consistently appear on one side of the head within a language, across phrase types. Harmony is prevalent among the world's languages and is one of a number of statistical language universals first noted by Greenberg (1963). This paper deals specifically with harmony in relation to the verb phrase and the adpositional phrase. For our purposes, therefore, harmonic languages are those which have Verb-Object (VO) order and prepositions (head-initial), or which have Object-Verb (OV) order and postpositions (head-final). According to the World Atlas of Linguistic Structures (WALS), this instance of harmony is quite cross-linguistically robust, with only 56 languages showing an explicitly non-harmonic pattern, in a sample of 984 languages, as table 1 indicates (Dryer, 2013c).

Table 1: *Correlation between adposition type and relative orders of object and verb cross-linguistically. Languages with no dominant order for one or both of the phrase types in question, as well as languages with inpositions or no adpositions at all, have been excluded. Data from Dryer (2013c).*

	Prepositions	**Postpositions**
Object - Verb	14	472
Verb - Object	456	42

2.2 Explanations for Word-Order Harmony

A number of different explanations for the cross-linguistic prevalence of word-order harmony have been proposed. Broadly, these can be classified into two main categories:

(1) Cognitive explanations, which appeal to some aspect of human cognition contributing to the cross-linguistic tendency for harmony that we see;
(2) Historical explanations, which attribute harmony to the nature of historical processes of language change.

2.2.1 Cognitive Explanations

Cognitive explanations for cross-category harmony include that of the Principles and Parameters framework, which attributes harmony to a high-level parameter for head-directionality, which is set in the speaker's internal grammar only once per language acquired and therefore applies across all phrase types in the language (e.g., Chomsky, 2014). Hawkins (1990), on the other hand, proposes that harmony is due to the fact that harmonic syntactic structures are easier to process, because they minimise the number of words that need to be heard in order for the hearer to identify the immediate constituents of the sentence Moravcsik (2010).

Alternatively, Culbertson and Kirby (2016) argue that there is a domain-general cognitive bias for harmonic structures: specifically, a bias for simplicity, which has been suggested to drive a wide range of cognitive processes (Chater & Vitányi, 2003). Harmonic languages can be argued to be simpler than non-harmonic ones: for example, in harmonic languages, one rule describing head-directionality is sufficient to explain the word order in both the verb phrase and the adpositional phrase, whereas non-harmonic languages require two rules, and are thus more complex. This of course depends on the theoretical framework being used, however (Culbertson & Kirby, 2016).

Cognitive biases are believed to reveal themselves in the typological patterns that we see through a process of iterated learning, as each generation learns their language from the previous one. A number of computational simulations have indicated this to be the case, including that of Griffiths and Kalish (2007), who demonstrate that over several generations, the languages of people modelled as Bayesian learners directly reflect the learners' prior biases, when their learning strategy is modelled as sampling from the posterior distribution. Kirby et al. (2007) go further to demonstrate that, when the learners' strategy is modelled as one that involves choosing the language with the maximum a-posteriori probability, learners' prior biases are amplified in the resulting distribution of languages over the process of iterated learning. Moreover, Smith and Kirby (2008) argue that a maximum a-posteriori learning strategy has evolutionary advantages, and so is likely to be the strategy used by humans. Therefore, evidence for even a weak cognitive bias for harmony, would be sufficient to argue that this contributes to the cross-linguistic prominence of cross-category harmony.

Much experimental work has been undertaken investigating harmony within the noun phrase, with results supporting the influence of a cognitive bias on word order, acting upon the positions of the numeral, demonstrative and adjective in relation to the noun (e.g., Culbertson et al., 2012; Culbertson & Newport, 2015; Culbertson et al., 2020). Relatively little experimental work has investigated cross-category harmony, on the other hand. The few studies that have been carried out will be discussed in detail below.

Artificial Language Learning (ALL) is a method commonly used for investigating the impact of cognitive biases on language. It involves giving participants a miniature constructed language to learn, and then testing them on it. This method has been used widely to evaluate regularisation and extrapolation tendencies, and the learnability of certain language features (e.g., Culbertson et al., 2012; Yin and White, 2018; Culbertson & Adger, 2014). Three experimental studies have been carried out using the ALL method to investigate harmony between the verb phrase and adpositional phrase, specifically those of Cook (1988), Zhao and Fedzechkina (2020), and Wang et al. (2021).

Cook (1988) investigated harmony between the verb phrase, adpositional phrase and noun phrase, through a series of ALL experiments on school-aged children using an extrapolation paradigm. In one of these experiments, participants were taught the vocabulary of an artificial language (AL), and shown simple transitive sentences, either of the order Subject-Object-Verb (SOV) or Verb-Subject-Object (VSO). They were then asked to translate adpositional phrases from English into the AL, and thus were being tested on whether they extrapolated the head-directionality of the adpositional phrase from that of the verb phrase. Results showed that participants had an overall preference for postpositions, regardless of whether they were trained on SOV or VSO orders, but that this preference was significantly stronger for participants trained on SOV. This could indicate that there is a baseline preference for postpositions, which will be discussed further in Section 2.3.1. The difference in the extent of the preference could be due to a harmony bias increasing the preference for postpositions in the SOV condition and acting against the baseline preference in the VSO condition. However, because the lexical items used in the test phase were the same as those in the training phase, this pattern of results may instead be due to participants noticing surface-level patterns, such as ‘the word for tiger always comes first’, rather than syntactic ones. Moreover, Cook (1988) points out that the participants may have approached the experiment as a problem solving task, rather than a language learning one, and Zhao and Fedzechkina (2020) note that Cook's method of asking participants to translate sentences from English into the AL may have caused them to consciously consider English word order when approaching the task, and potentially to adopt a strategy of making the word order maximally different to English by choosing postpositions.

Zhao and Fedzechkina (2020) conducted an ALL regularisation experiment in which participants were first exposed to either prepositional phrases or postpositional phrases in the AL. They were then exposed to a set of verb phrases, 50% of which displayed VO order, and 50% OV order. At test, participants were shown images depicting events and asked to describe them in the AL. Participants in the easy lexical retrieval condition were given the vocabulary items to choose from when doing this, and those in the hard lexical retrieval condition were not. Results showed that participants exposed to postpositions in the hard lexical retrieval condition were more likely to produce OV structures at test, thus favouring harmonic patterns. Participants exposed to prepositions, however, did not have such a preference for harmonic structures, and participants in the easy lexical retrieval did not have a harmony preference when exposed to either adposition type. Looking only at the hard lexical retrieval condition, these results indicate that a cognitive bias cannot fully account for the typological patterns we see crosslinguistically, as there is only experimental evidence for it favouring harmony for one head direction. Zhao and Fedzechkina suggest that this may be because of a baseline preference for SOV orders, as suggested by the fact that it is the most common basic word order in WALS (Dryer, 2013b), and that experimental studies have shown it to be the most learnable (Tily et al., 2011). It is therefore possible that Zhao and Fedzechkina did not find evidence for a harmony bias in participants exposed to prepositions because the bias was interacting with this baseline preference for SOV order.

Wang et al. (2021), like Cook (1988), used an extrapolation paradigm to target harmony between the verb phrase and the adpositional phrase. Native speakers of English and of Chinese were trained on verb phrases in an AL either displaying OV or VO order, and then on the adpositional lexicon. At test they were asked to describe images depicting spatial relations in the AL. Results showed that both sets of participants had a strong preference for harmony in both the OV and the VO condition. This appears indicative of a cognitive bias for harmony, but, like the Cook (1988) study, the same lexical items were used in both the training and testing phases, meaning participants may have been utilising surface-level patterns.

Another method which is often employed in the investigation of language universals is the silent gesture paradigm. This paradigm was first used Goldin-Meadow et al. (2008) and has since been employed in many studies investigating word-order universals (e.g., Schouwstra & de Swart, 2014; Meir et al., 2010; Langus & Nespor, 2010). Typically, the method involves showing visual stimuli to hearing participants with no knowledge of any signed languages and asking them to convey what they see using only gesture. The order in which participants gesture the relevant elements of the picture (e.g., those representing the subject, object and verb of the corresponding descriptive sentence in English) is subsequently analysed. Silent gesture studies produce comparable results when employed on participants with a range of different native languages, including English, Spanish, Mandarin, Turkish (Goldin-Meadow et al., 2008), Irish, Russian, Tagalog (Futrell et al., 2015), and Italian (Langus & Nespor, 2010). This indicates that this method is relatively successful in inhibiting the effects of the word order of the speakers' native language, and therefore the similarities in the orders that they produce may be caused deeper shared preferences, such as those caused by cognitive biases.

Basic word order and, by extension, the verb phrase, have been the focus of recent silent gesture studies on language universals. Goldin-Meadow et al. (2008), conducted a series of experiments, one of which was a silent gesture production study, which aimed to investigate how people communicate when they are prevented from using language. They found that for basic transitive events, participants overwhelmingly communicated the elements of the stimuli in an order that corresponded to SOV. Schouwstra and de Swart (2014) noted that all stimuli in Goldin-Meadow et al. (2008) depicted extensional events, that is, those in which the direct object is manipulated, and exists independently of the event (e.g., verbs like *throw, poke, drop*). They then ran a silent gesture experiment which also tested participants on intensional events: events for which the direct object does not exist independently of the action, as for verbs like *paint, think about, dream of*.

They found that participants were indeed more likely to produce SOV gestured orders for extensional events, but for intensional events they were more likely to produce SVO. This demonstrates that word order preferences can be affected by semantics (Schouwstra & de Swart, 2014). As all of Zhao and Fedzechkina's (2020) verb phrase stimuli represented extensional events, this strengthens the argument that a baseline preference for SOV contributed to the lack of evidence for a harmony bias in the prepositional condition of their study.

2.2.2 Historical Explanations

For the type of harmony of interest in this paper, between the verb phrase and the adpositional phrase, a historical explanation has also been put forward. Specifically, there is a diachronic tendency for adpositions to derive from the verbs of verb phrases, or from possessum constituents of possessive phrases (which also participate in cross-category harmony), via a process which preserves linear order (Moravcsik, 2010). Since verbs and adpositions are the heads of verb phrases and adpositional phrases respectively, this would help to explain the prominence of syntactic harmony between these phrase-types cross-linguistically. This historical tendency has been observed in several unrelated languages, including Abkhaz, Basque, Bihari, Buriat, Kui (Bybee, 1988, p. 354), Mandarin (Li & Thompson, 1974), and English (Moravcsik, 2010), and thus certainly contributes to the cross-linguistic patterns of syntactic harmony we see. The question is whether this historical tendency is the sole reason for the harmony universal, or if cognitive biases also play a role.

2.3 The Present Experiments

This paper describes three experiments, which were designed to together establish whether evidence for a cognitive bias for harmony between the verb phrase and the adpositional phrase can be found. If such evidence is found, this would indicate that a cognitive bias for cross-category harmony contributes to its cross-linguistic prevalence, alongside the historical explanation presented above.

2.3.1 Experiment 1 Aims

Experiment 1 first aims to identify whether there is an overall baseline preference for either prepositions or postpositions, like the baseline preference for SOV basic word order for extensional events found by Goldin-Meadow et al. (2008), and to quantify the extent of such a preference. This is crucial for comparison with experiments 2 and 3 of this paper, which aim to manipulate this preference through priming participants with verb phrases of either OV or VO order. If an overall baseline preference for prepositions or postpositions exists, the preferences found in the results of experiments 2 and 3 should be interpreted in relation to this baseline preference, rather than just in relation to chance. The silent gesture paradigm is utilised here, as is typical for studies investigating such a baseline preference (e.g., Goldin-Meadow et al., 2008), because it minimises native language effects. Though such studies are typically done using silent gesture improvisation paradigms (e.g., Schouwstra & de Swart, 2014; Goldin-Meadow et al., 2008), here I opt for a silent gesture perception paradigm, as these are easier to implement online than improvisation studies, allowing more participants to be tested over a short period of time. Although it is a relatively new method, silent gesture production results have been shown to replicate using the silent gesture perception paradigm. Specifically, the perception results of Motamedi et al. (2021) replicate the production results of Schouwstra and de Swart (2014) for a difference in basic word order preferences based on semantics, and the perception results of Verhoef et al. (2016) replicate the production results of Padden et al. (2015) investigating the preferences for handling gestures in comparison with action gestures.

2.3.2 Experiment 2 Aims

Experiment 2 then aims to replicate the results of Wang et al. (2021) but uses silent gesture perception methods combined with ALL in the extrapolation paradigm, to establish whether these methods are sufficient to demonstrate a harmony bias when the objects used in the training stimuli are the same as those in the testing stimuli. This is required for comparison with experiment 3, in which the objects in the test stimuli are different to those in the priming stimuli. This is because, if a significant effect of condition is found in experiment 2 but not experiment 3, this would indicate that there is no harmony bias, and instead participants were just noticing surface-level patterns in experiment 2.

Participants in experiment 2 are first primed on either VO or OV verb phrase order, where in the VO condition, the stimuli exclusively represent intensional events, and in the OV condition the stimuli exclusively represent extensional events. This is to avoid the possibility of the differing baseline verb phrase order preferences found by Schouwstra and de Swart (2014) interfering with the strength of the prime in the two different conditions. Participants are then asked exactly the same questions as in experiment 1 to allow for direct comparison when evaluating whether the priming had an impact on their preferences for adpositional orders.

2.3.3 Experiment 3 Aims

Experiment 3 investigates whether the effect demonstrated in experiment 2 is the result of participants noticing surface-level patterns due to the repetition of objects in the training and testing stimuli, or whether there is evidence of a syntactic harmony bias at work. This is achieved by repeating experiment 2 but using different objects in the training stimuli to those in the testing stimuli.

3 Methodology

3.1 Experiment 1

3.1.1 Participants

140 monolingual adult native English speakers were tested for experiment 1. All were recruited and tested on Prolific and were paid the equivalent of £14.60/hr for their time. The experiment lasted roughly 3 minutes.

3.1.2 Materials

Nine drawings were made depicting adpositional relations between a human character and an inanimate object. Standard practice in silent gesture improvisation studies is to use a semantically-rich human character, such as a pirate or a chef (e.g. Schouwstra & de Swart, 2014), in order to encourage participants to express the subject in their gestures. In the present experiments a more generic representation of a person was used, as shown in figure 1, because the relative position of the subject is not relevant to the research questions. The objects in the stimuli were chosen to work semantically with the adpositions *in, under* and *in front of*, and to be easily gestured using a one-handed gesture. This is because using some one-handed gestures and some two-handed gestures for different elements of the sentence could affect the saliency of the different elements and thus impact the results. The objects chosen were a *teacup*, a *kettle* and a *saucepan*. The full list of stimuli is the following:

- [A person] in a teacup
- [A person] in a kettle
- [A person] in a saucepan
- [A person] in front of a teacup
- [A person] in front of a kettle
- [A person] in front of a saucepan
- [A person] under a teacup
- [A person] under a kettle
- [A person] under a saucepan

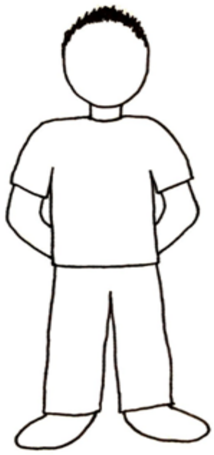


Figure 1: *Drawing of generic human used in image stimuli.*

Video stimuli were recorded of the experimenter gesturing interpretations of the image stimuli. Each video showed two distinct gestures, one for the object and one for the adposition, relying on the body being interpreted as the subject. For each stimulus, one video was made depicting the order adposition-object (i.e., a prepositional order), and one video depicting the order object-adposition (i.e., a postpositional order). All gestures were one-handed and involved movement. All gestures lasted 1 second each and all videos were exactly 4 seconds long in total.

3.1.3 Procedure

This and the following two experiments were coded in JsPsych and ran on the participant's web browser. After consenting to participate in the study, participants were instructed that they would see an image, alongside two videos of gesture sequences representing the image, and that they should watch both videos, before clicking on the button corresponding to the video which best represents the image. They then proceeded to the task, where the image appeared on the screen, and both videos played beneath it simultaneously side-by-side, as shown in figure 2. The videos looped until the participant had selected one, and the order of the videos (left/right) was randomised for each participant. Each participant received only one trial of this type in order to capture their immediate reaction to the stimulus. The stimulus they received was assigned randomly from the nine available.

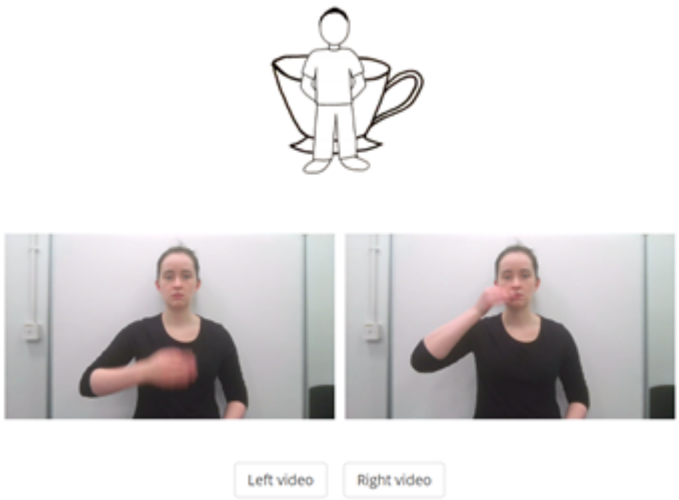


Figure 2: *Screenshot of a binary response question of experiment 1.*

Participants were then given a slider-response question, for which they were shown the same image and videos as in the previous question and asked to use the slider to indicate the strength of their preference for the video they chose, with the leftmost limit of the slider representing a strong preference for the video on the left, and the rightmost limit a strong preference for the video on the right. Figure 3 shows a screenshot of this trial type.

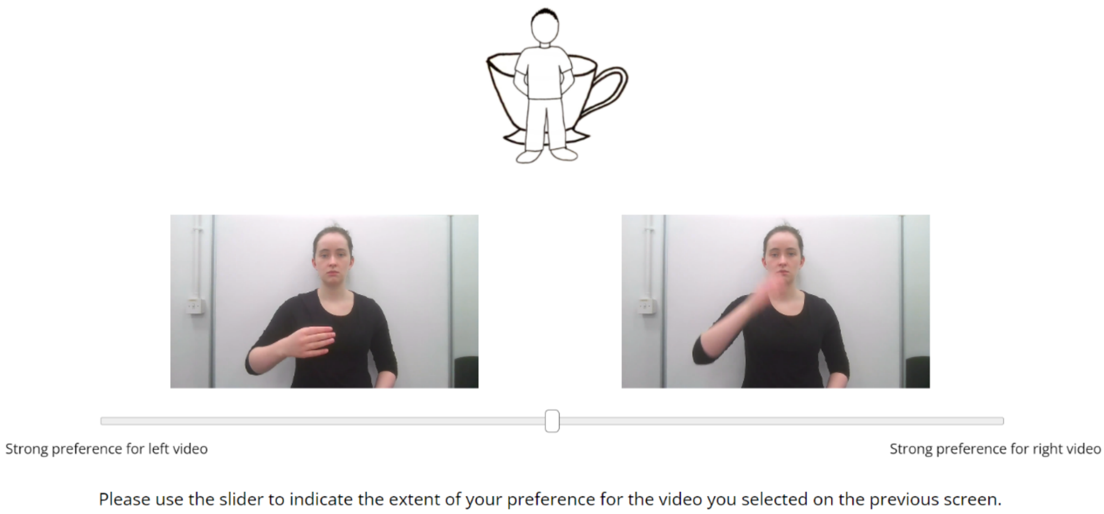


Figure 3: *Screenshot of a slider question of experiment 1.*

Participants were then once again shown the image stimulus from the previous two questions and asked to type a one-sentence description of the image in English.

3.2 Experiment 2

3.2.1 Participants

140 participants who were again monolingual adult native English speakers were recruited and tested on Prolific. Participants who took part in experiment 1 were prevented from participating in experiment 2. Experiment 2 lasted roughly 6 minutes and participants were paid the equivalent of £10.39/hr.

3.2.2 Materials

Nine drawings were made depicting intensional events and nine depicting extensional events. The objects in the images were the same as in experiment 1 and again the subject was a generic human character. The intensional verbs used were *paints, thinks of*, and *dreams of*, and the extensional verbs were *throws, drops*, and *pokes*. The full list of events stimuli is given in table Table 2.

Table 2: *Events stimuli for experiment 2.*

Extensional Events Stimuli	Intensional Events Stimuli
[A person] throws a teacup	[A person] paints a teacup
[A person] throws a saucepan	[A person] paints a saucepan
[A person] throws a kettle	[A person] paints a kettle
[A person] drops a teacup	[A person] thinks of a teacup
[A person] drops a saucepan	[A person] thinks of a saucepan
[A person] drops a kettle	[A person] thinks of a kettle
[A person] pokes a teacup	[A person] dreams of a teacup
[A person] pokes a saucepan	[A person] dreams of a saucepan
[A person] pokes a kettle	[A person] dreams of a kettle

Video stimuli were made in the same way as experiment 1. In this case, for each stimulus, one video depicted an OV gestured order, and the other a VO gestured order. The same adpositional stimuli from experiment 1 were used in experiment 2.

3.2.3 Procedure

Each participant was randomly assigned to either the VO condition or the OV condition. After consenting to participate in the study, they were instructed that they would first see some videos of gesture sequences representing images, and that they would be tested on these later so should pay attention. They proceeded to the training phase, where participants in the extensional condition saw all 9 extensional stimuli one by one, with the image presented at the top of the screen and the corresponding video for OV order underneath. The video played on a loop, but after it had played through once, a 'continue' button was made clickable, allowing the participant to proceed to the next stimulus. Those in the intensional condition saw the 9 intensional stimuli with the corresponding videos for VO order, presented in the same manner as in the extensional condition.

In the reinforcement phase, participants were tested on how well they had learned the gesture orders in the training phase. They were shown each of the 9 events stimuli again one by one, in a random order, along with both the OV and VO gesture videos for that stimulus. Participants were asked to click the button corresponding to the video which best fit with the AL they had just learned and were given feedback on their responses. This was to both reinforce their learning and to allow me to exclude participants from analyses if they did not learn the word order presented to them in the training phase.

Participants then proceeded to the critical test phase, where they were tested on the adpositional stimuli, with exactly the same questions and materials as in experiment 1. Like in experiment 1, each participant only saw one of the 9 adpositional stimuli.

3.3 Experiment 3

3.3.1 Participants

140 monolingual adult native English speakers were recruited and tested on Prolific for experiment 3. Participants who took part in experiments 1 or 2 were prevented from participating. Participants were paid the equivalent of £9.48/hr for their time. The experiment lasted around 6 minutes.

3.3.2 Materials

The same adpositional stimuli were used in experiment 3 as in experiments 1 and 2. Nine new extensional events stimuli and nine new intensional events stimuli were created for experiment 3. Here the objects were *hammer, spraycan* and *toothbrush*. The verbs were the same as in experiment 2. The full list of stimuli is in Table 3:

Table 3: *Events stimuli for experiment 3.*

Extensional Events Stimuli	Intensional Events Stimuli
[A person] throws a hammer	[A person] paints a hammer
[A person] throws a spraycan	[A person] paints a spraycan
[A person] throws a toothbrush	[A person] paints a toothbrush
[A person] drops a hammer	[A person] thinks of a hammer
[A person] drops a spraycan	[A person] thinks of a spraycan
[A person] drops a toothbrush	[A person] thinks of a toothbrush
[A person] pokes a hammer	[A person] dreams of a hammer
[A person] pokes a spraycan	[A person] dreams of a spraycan
[A person] pokes a toothbrush	[A person] dreams of a toothbrush

Image and video stimuli depicting these events were made in the same way as in experiment 1.

3.3.3 Procedure

The procedure for experiment 3 was identical to that of experiment 2, except that the stimuli for the training and reinforcement phases came from the set described in Section 3.3.2.

4 Results

4.1 Exclusion Criteria

In all experiments, participants were excluded from the analysis if they indicated in the slider response question that they preferred a different adpositional gesture video to the one they selected as their preferred video in the binary response question. For experiments 2 and 3, participants were also excluded if they gave more than two incorrect responses in the reinforcement phase. The number of participants excluded from the analysis for each reason in each experiment is listed in Table 4:

Table 4" *Reasons for participant exclusions in each experiment.*

Exclusion Criteria	Number excluded from experiment 1	Number excluded from experiment 2	Number excluded from experiment 3
Opposite preferences in binary response and slider questions	10	3	6
More than two incorrect answers in reinforcement phase	NA	5	7
Both of the above	NA	0	2
Total	10	8	15

4.2 Results of Experiment 1

4.2.2 Binary Response Data

I next analysed the responses to the critical questions in experiment 1, beginning with the binary response question, which asked participants to choose which gesture video they preferred to represent the image shown. Figure 4 shows the proportions of participants who preferred the prepositional gesture video in each condition of each experiment.

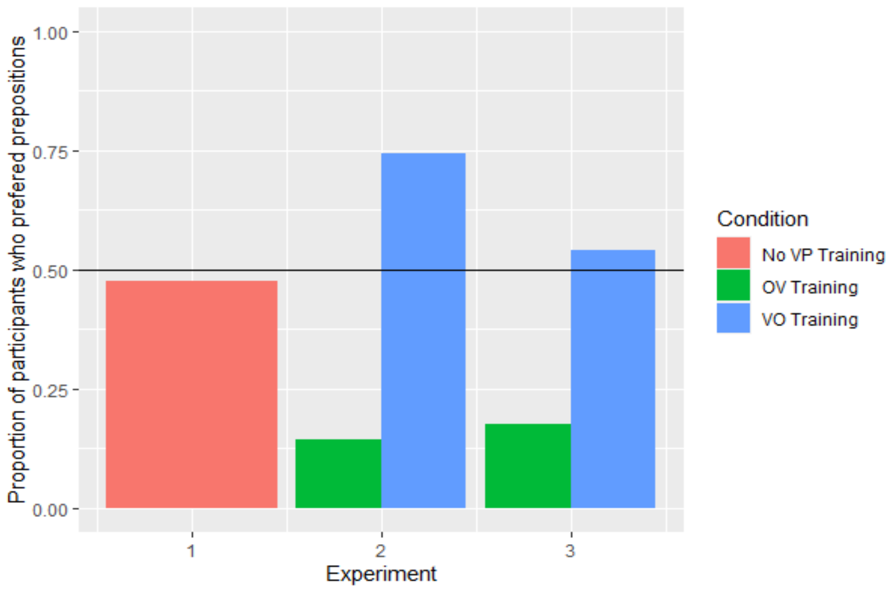


Figure 4: *Bar chart showing the proportions of participants who preferred the prepositional gesture video in the binary response question in each condition of each experiment, with the horizontal line representing chance performance.*

The proportion of participants who preferred the prepositional gesture video they were presented with in experiment 1 is 0.477, indicating a slight preference for postpositions. A logistic regression model was run on the experiment 1 data, using lme4 (Bates et al., 2015) in R (R Core team, 2019). The outcome variable was a binary variable indicating whether the participant preferred the prepositional gesture video or not. For all statistical tests in this paper the standard alpha level of 0.05 is adopted. The model returned a non-significant intercept estimate (β = -0.09, *se* = 0.18, *z* = -0.53, *p* = 0.60). There is therefore insufficient evidence to reject the null hypothesis that participants have no preference for either prepositional or postpositional gesture videos.

4.2.3 Slider Data

I next analysed participants' responses to the slider question in experiment 1, which asked them to give a more graded indication of their preference for the gesture video they chose using a slider. Figure 5 is a violin plot showing the results of this question for all 3 experiments, where 100 represents a strong preference for the prepositional gesture video and 0 a strong preference for the postpositional one.

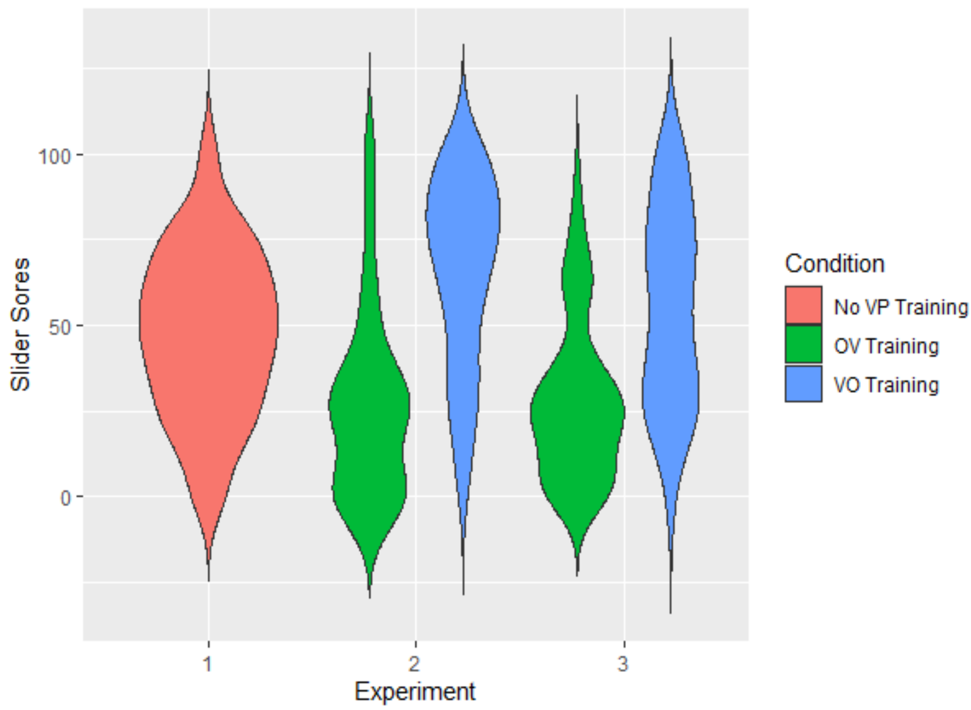


Figure 5: *Violin plot showing participants' slider rating for the gesture video they chose, for each condition in each experiment. 100 represents a strong preference for the prepositional gesture video and 0 a strong preference for the postpositional gesture video.*

For experiment 1, Figure 5 indicates a large number of participants had a very small preference for the video they chose, with many responses being centred around 50. After transforming the slider data to be in the range (0,1), a beta regression model was run on the experiment 1 data using mgcv (Wood, 2011) in R (R Core Team, 2019). The outcome variable was the participants' transformed slider scores for the adpositional videos. The model returned a non-significant intercept (β = -0.16, *se* = 0.12, *z* = -1.38, *p* = 0.17), indicating there is insufficient evidence to reject the null hypothesis that there is no preference for either prepositional or postpositional gesture videos, when preference is indicated using a slider.

Because for both the binary and slider data the prepositional preference was not significantly different from chance, there was no need for the following statistical tests to directly compare the results of experiment 1 with those of experiments 2 and 3.

4.3 Results of Experiments 2 and 3

4.3.1 Learning in Experiments 2 and 3

Before analysing the responses to the critical questions in experiments 2 and 3, I investigated how well participants learned the verb phrase order in these experiments (post exclusion), because if there is a difference between conditions this will affect how subsequent results are interpreted. Figure 6 shows the mean number of mistakes made by participants in each condition in the reinforcement phase of each of experiments 2 and 3, and indicates that more mistakes were made by participants in the OV condition in both experiments.

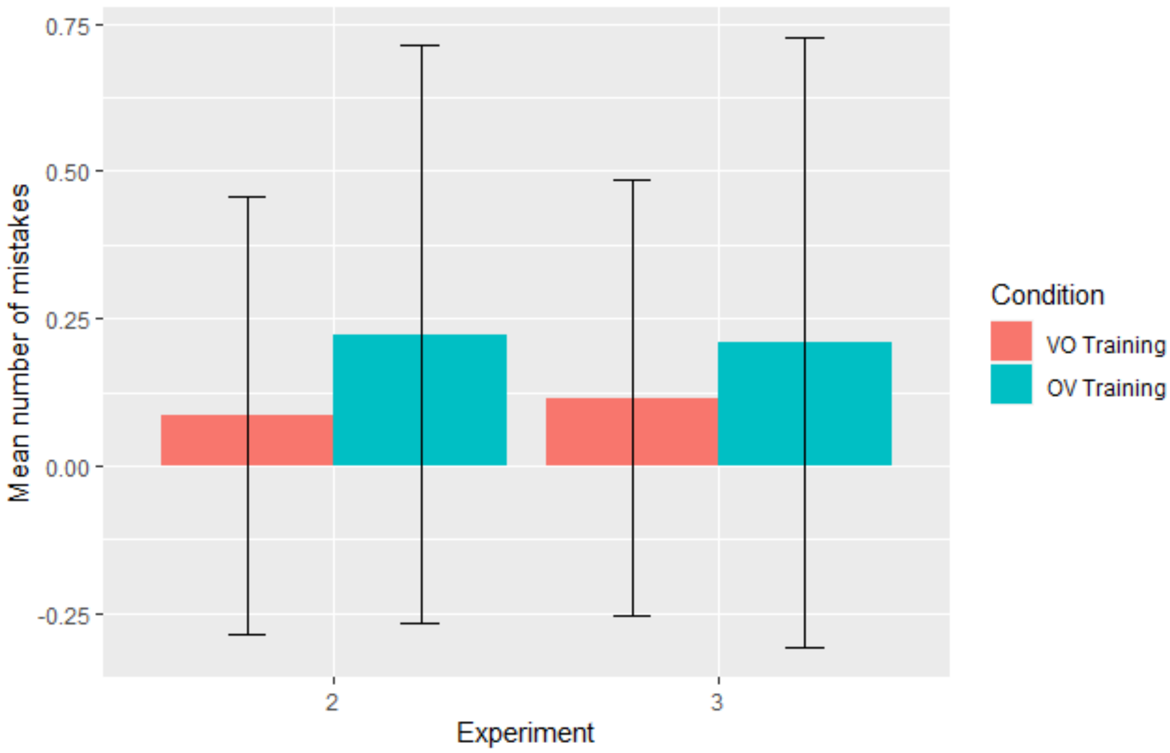


Figure 6: *Bar chart showing the average number of mistakes made by participants in verb phrase testing in each condition in experiments 2 and 3. Error bars show standard deviation.*

A linear model was run on these data to investigate whether this difference was significant. The outcome variable was the number of mistakes made by participants in the reinforcement phase of the experiment, and the fixed effects entered were experiment, condition and their interaction. For this and all following models, condition was sum-coded and experiment treatment-coded, with the reference levels being experiment 2 and the VO condition. This means the intercept estimate represents the outcome variable averaged over both conditions in experiment 2. Table 5 shows the results of the model.

Table 5: *Results of the linear model run on experiments 2 and 3 data with experiment, condition and their interaction as fixed effects, and the number of mistakes made by participants as the outcome variable. Condition was sum-coded, and experiment treatment-coded, and the reference levels were experiment 2 and the VO condition. * indicates significant results.*

Effect	Estimate	Standard Error	t-value	p-value
Intercept	0.153968	0.038229	4.028	7.46e-05*
Condition	-0.068254	0.038229	-1.785	0.0754
Experiment	0.008248	0.055113	0.150	0.8812
Condition: Experiment	0.020792	0.055113	0.377	0.7063

As Table 5 shows, none of the fixed effects were significant predictors of the number of mistakes made by participants. This indicates that the difference between conditions is not meaningful. Therefore, subsequent analyses are carried out without taking this into account.

4.3.2 Results of Experiment 2

For experiment 2, figure 4 clearly shows that harmonic patterns were preferred in the binary responses: only 14.3% of participants in the OV condition preferred the prepositional gesture video, compared with 74.3% in the VO condition. A logistic regression model was run on the experiment 2 binary response data, with the outcome variable being whether the participant preferred the prepositional gesture video or not. Condition was entered as a sum-coded fixed effect. The model results are given in table 6.

Table 6: *Results of logistic regression model run on experiment 2 data, with the binary response indicating the preference for adpositional gesture video as the outcome variable, and condition as a sum-coded fixed effect. * indicates significant results.*

Effect	Estimate (log odds)	Standard Error	z-value	p-value
Intercept	-0.3654	0.2261	-1.617	0.106
Condition	1.4263	0.2261	6.309	2.8e-10*

As Table 6 shows, again the intercept was non-significant, indicating we do not have sufficient evidence to reject the null hypothesis that, averaged across the two conditions, participants' preference for prepositional gesture videos was not underlyingly different from chance. Table 6 also reveals condition to be a significant predictor of preference for prepositional gesture videos: VO training leads to more prepositional choices than OV training. In other words, the majority of participants chose the adpositional phrase order that was harmonic with the verb phrase order they were trained on.

In terms of slider data, Figure 5 indicates that in experiment 2 the preference for prepositions in the VO condition is weaker than the preference for postpositions in the OV condition, as the participants' slider responses were slightly more dispersed in the VO than in the OV condition. A beta regression model was run on the experiment 2 data, with condition entered as a sum-coded fixed effect, and the outcome variable being the participants' transformed slider scores for the adpositional videos. The results of this model are summarised in Table 7.

Table 7: *Table showing results of beta regression model run on experiment 2 data, with slider responses as the outcome variable and condition as a sum-coded fixed effect. * indicates significant results.*

Effect	Estimate (log odds)	Standard Error	z-value	p-value
Intercept	-0.3465	0.1087	-3.189	0.00143*
Condition	1.0665	0.1087	9.815	<2e-16*

Here the intercept is significant so we have sufficient evidence to reject the null hypothesis that the average probability of a participant having a preference for prepositional gesture videos across both conditions in experiment 2 is not underlyingly different from chance, when preference is indicated by a slider response. The intercept estimate is negative, indicating that there is an overall postpositional preference in experiment 2. Condition is again a significant predictor of preference for prepositional gesture videos: participants in the VO condition preferred prepositions and those in the OV condition preferred postpositions. In other words participants showed a preference for harmonic patterns.

4.3.3 Results of Experiment 3

Figure 4 suggests that harmonic patterns were also preferred in experiment 3 when participants are asked to make a binary decision, though to a lesser extent than for experiment 2, with 17.7% of participants in the OV condition and 54.1% of participants in the VO condition preferring prepositions. A logistic regression model was run on the binary response data from experiment 3, with the same structure as the model for the binary response data from experiment 2. The model results are given in Table 8.

Table 8: *Table showing results of logistic regression model run on experiment 3 data, with binary response for adpositional preference as the outcome variable and condition as a sum-coded fixed effect. * indicates significant results.*

Effect	Estimate (log odds)	Standard Error	z-value	p-value
Intercept	-0.6848	0.2101	-3.260	0.00111*
Condition	0.8491	0.2101	4.042	5.3e-05*

Unlike experiments 1 and 2, for experiment 3, the model intercept for the binary response data was significant. As the intercept term is negative, this indicates there is a reliable preference for postpositions averaged across the two conditions. In this model, condition is again revealed as a significant predictor of preference for prepositional gesture videos: participants in the VO condition were significantly more likely to choose the prepositional gesture video than those in the OV condition.

Figure 5 suggests that, when indicating their preference using the slider, participants in the OV condition of experiment 3 again had a preference for postpositions on average, but the overall preference of participants in the VO condition is less clear. A beta regression model was run on the experiment 3 data with the same structure as that for the model run on the experiment 2 slider data. The results of the model are summarised in Table 9.

Table 9: *Results of beta regression model run on experiment 3 data, with slider response as the outcome variable and condition as a sum-coded fixed effect. * indicates significant results.*

Effect	Estimate (log odds)	Standard Error	z-value	p-value
Intercept	-0.3171	0.1163	-2.727	0.00639*
Condition	0.7443	0.1163	6.401	1.54e-10*

Here, the intercept estimate is significant and negative, providing sufficient evidence to reject the null hypothesis that the average probability of a participant having a preference for prepositional gesture videos across both conditions in experiment 3 is not underlyingly different from chance, when prepositional preference is indicated using a slider. This indicates that there is an overall postpositional preference in experiment 3. Like in all previous models, condition was a significant predictor, so participants in the VO condition were more likely to prefer the prepositional gesture video than participants in the OV condition.

4.3.4 Comparison Between Experiments 2 and 3

To evaluate whether there was a difference in the effect of condition on prepositional preference between experiments 2 and 3, or an effect of experiment, a logistic regression model was run on the binary response data from both experiments, with experiment entered as a treatment-coded fixed effect, and condition as a sum-coded fixed effect. The interaction between experiment and condition was also entered into the model as a fixed effect. The outcome variable was again whether the participant preferred the prepositional gesture video or not. The results are reported in table 10.

Table 10: *Results of logistic regression model run on the data from experiments 2 and 3, with the binary response as the outcome variable and condition, experiment and their interaction as fixed effects. Condition was sum-coded and experiment treatment-coded. * indicates significant results.*

Effect	Estimate (log odds)	Standard Error	z-value	p-value
Intercept	-0.3654	0.2261	-1.617	0.1060
Condition	1.4263	0.2261	6.309	2.8e-10*
Experiment	-0.3194	0.3086	-1.035	0.3007
Condition: Experiment	-0.5772	0.3086	-1.870	0.0614

Neither the main effect of experiment nor the interaction between experiment and condition were significant predictors of prepositional preference. There is thus insufficient evidence to reject the null hypothesis that there is a difference between experiments 2 and 3 in the preference for prepositions. There is also insufficient evidence to reject the null hypothesis that there is no difference in the extent of the effect of condition on prepositional preference between the two experiments.

A final beta regression model was run to evaluate the main effect of experiment and the difference between experiments 2 and 3 in the effect of condition on prepositional preference when prepositional preference was indicated using a slider. This model was run on the data from experiments 2 and 3, and experiment, condition, and their interaction were entered into the model as fixed effects. Experiment was treatment-coded and condition was sum-coded. The outcome variable was the participants' transformed slider scores for the adpositional videos. The model results are summarised in Table 11.

Table 11: *Results of beta regression model run on experiments 2 and 3 data, with slider response as the outcome variable and condition, experiment and their interaction as fixed effects. * indicates significant results.*

Effect	Estimate (log odds)	Standard Error	z-value	p-value
Intercept	-0.36194	0.10727	-3.374	3.68e-07*
Condition	1.14792	0.10727	10.702	<2e-16*
Experiment	0.06511	0.15918	-0.409	0.68252
Condition: Experiment	-0.46162	0.15918	-2.900	0.00373*

Condition was a significant predictor of prepositional preference, with participants in both experiments being more likely to choose the prepositional gesture video in the VO condition than the in the OV condition. The interaction between condition and experiment was also significant, but not the main effect of experiment. This indicates the prepositional preference in the VO condition in experiment 2 did not differ significantly from that of experiment 3, and similarly for the OV condition, however the difference between the prepositional preference in the VO condition and OV condition differed significantly across experiments. Specifically, the effect of condition is stronger in experiment 2 than in experiment 3. This interaction effect is illustrated by Figure 7, which shows the model's predictions for transformed slider scores.

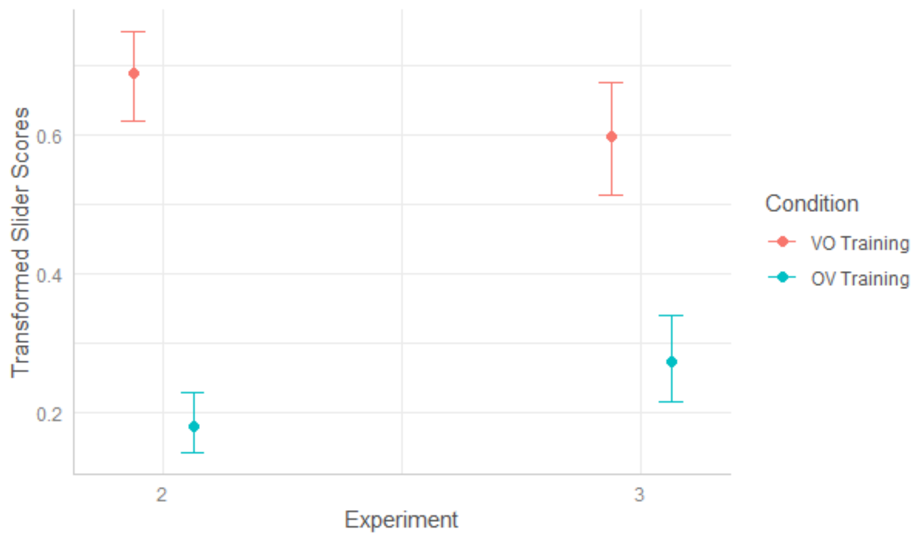


Figure 7: *Graph plotting the model's predicted transformed slider scores for each experiment and condition, to illustrate the significant interaction effect between experiment and condition. Here 1 indicates a strong preference for prepositions and 0 a strong preference for postpositions.*

4.4 Image Description

When asked to describe the adpositional image stimulus that they saw in English, the majority of participants in all three experiments correctly identified the object in the picture. For stimuli using the adpositions *in* and *in front of*, most participants also correctly identified the adpositional relation being targeted. This indicates that the images were clear enough for the purposes of the experiments. For the *under* stimuli, however, responses were more varied, with many participants giving descriptions like 'saucepan above head'. This indicates that this set of stimuli was not as well understood by the participants, and so maybe their responses were random. Just to be conservative, the above models were run again with all participants who received 'under' excluded and the results followed the same key patterns, indicating that this potential misunderstanding of stimuli did not have an effect on the results. It is worth noting that although the participants did not give the intended response to the image description question, this does not necessarily mean that they did not correctly interpret the gestures in the videos as representing the intended adpositions and nouns. Participants may have correctly interpreted the gestures as *under* and *saucepan*, and responded to the critical questions accordingly, but when asked to describe the image in English, they found that describing the relation as 'saucepan above head' felt more natural.

5 Discussion

The present experiments aimed to evaluate whether there is evidence for a cognitive bias for syntactic harmony between the verb phrase and the adpositional phrase, which may contribute to the patterns we see in language typology. I carried out three experiments for this purpose, combining ALL with silent gesture perception methods. In experiment 1, participants were asked whether they preferred prepositional or postpositional gesture videos, in order to gauge what the baseline preference for the order of elements in an adpositional phrase is, without any priming. In experiment 2, participants were primed with gesture videos depicting verb phrases, either of the order VO or OV, before being asked the same questions as in experiment 1. In experiment 2 the objects used in the priming stimuli were the same as those used in the testing stimuli, and so experiment 3 was devised to rule out the possibility that participants were using surface-level rules (such as 'the saucepan is always gestured first') to answer the critical questions. This was achieved by repeating experiment 2, but using different objects in the verb phrase stimuli than in the adpositional stimuli, to test whether participants generalise beyond the individual items to the entire class of nouns.

Taken together, the results indicate that there is no baseline preference for either prepositions or postpositions, and that there is evidence for a harmony bias, both when the training and testing stimuli contain the same elements (experiment 2), and when they do not (experiment 3). The evidence for harmony in the VO condition of experiment 3 did however seem less strong than in the other condition and experiment 2, as shown in figure 4, and for the slider data, the effect of condition was significantly stronger in experiment 2 than in experiment 3. This suggests that the surface-level patterns visible in experiment 2 partially contributed to the harmonic preferences observed there.

5.1 Baseline Adpositional Preference

Experiment 1 served its purpose in providing a reference to compare the results of experiments 2 and 3 to, even if the baseline preference it revealed was not significantly different from chance and was thus not needed for direct statistical comparison. However, the non-significant result in experiment 1 is surprising, since Cook's (1988) experiments point to an underlying preference for postpositions. Moreover, Gentner and Boroditsky (2009, p. 5) argue that relational terms require the entities that they link to be introduced first, thus explaining why SOV is preferred for extensional verbs. All adpositional stimuli in the present experiments involve adpositions used to describe spatial relations between nouns. Following Gentner and Boroditsky's (2009) logic, there should therefore be a baseline preference for postpositions, since they allow both nouns in the spatial relation to be introduced, before the relation between them is expressed. One possible reason such a preference was not observed here is that some participants may have been experiencing native language influence, as perhaps the silent gesture perception method is not as successful as the production method at inhibiting this. If this were the case, then participants' native language influence which causes them to prefer prepositions, would have been interacting with a baseline preference for postpositions, resulting in the overall proportion of participants choosing prepositions not differing significantly from 0.5. However, the results of Motamedi et al. (2021) do not support this hypothesis, as in their silent gesture perception study which tested native English-speaking participants' preferences for basic word order, no overall preference for SVO was found. It seems likely, therefore, that no baseline preference for one adpositional order or another exists. The typological evidence supports this: the number of languages with postpositions (577) in WALS is roughly the same as the number of languages with prepositions (511) (Dryer, 2013a). Additionally, the results of Cook (1988), which suggest an overall postpositional preference, could be due to methodological issues with their experiment as discussed in section 2.2.1. More specifically, translating from English to the AL may have caused the participants in Cook's (1988) study to adopt a strategy of making the word order maximally different from English and therefore choosing postpositions.

5.2 Harmony

The results of experiment 2 provide evidence that the methods utilised in the present experiments, namely ALL combined with silent gesture perception, were sufficient to demonstrate a harmony bias when the objects in the priming and testing stimuli are the same, thus replicating the main results of Wang et al. (2021).

The results of experiment 3 are clear in demonstrating a harmonic preference in participants in the OV condition, who preferred postpositions. However, the preference for prepositions of those in the VO condition was only slightly higher than chance. One possible reason for this is that the priming stimuli followed the pattern of English, their native language, and also did not contradict their natural preferences (i.e., VO order was used to represent intensional events). This means that there is a chance many participants did not register it as a prime and thus responded to the critical question as they would if they had not been primed, like in experiment 1. This leaves a question of why this would happen in experiment 3 but not experiment 2. One possibility is that in experiment 2, the surface-level patterns were so obvious that they were easily recognised and used by the participants in choosing their responses to the critical questions. In experiment 3, on the other hand, the analogies between the verb phrase stimuli and adpositional stimuli were less clear, so the participants could not rely upon such surface-level patterns and therefore responded as though they had not been primed at all, like in experiment 1.

Zhao and Fedzechkina (2020) similarly found evidence for harmony in the postpositional but not the prepositional condition, which may on the surface appear to be in conflict with the above hypothesis, since their test stimuli contained the same objects as their priming stimuli, like in experiment 2 of this paper. However, as discussed in Section 2.2.1, their results were likely instead caused by the fact that their priming stimuli were adpositional phrases and their test stimuli were extensional verb phrases, which have a baseline preference for SOV order (Schouwstra and de Swart, 2014; Goldin-Meadow et al., 2008; Tily et al., 2011).

It can still be argued therefore that experiment 3 provides evidence for a cognitive bias for syntactic harmony between the verb phrase and the adpositional phrase. Although the preference for prepositions in the VO condition is only slightly higher than chance, it still differs significantly from the preference for prepositions in the OV condition. This difference in preference according to condition is sufficient evidence for a bias which at least partially contributes to the typological patterns we see, in combination with the historical explanation described in section 2.2.2. Moreover, as explained in section 2.2.1, it is possible that even a weak bias for harmonic patterns is sufficient to explain the prevalence of cross-category harmony between the adpositional phrase and the verb phrase in the world's languages. This is because the process of iterated learning over generations amplifies weak biases in the population over generations, as indicated by the results of the computational simulations of Kirby et al. (2007) and Smith and Kirby (2008).

Finally, the fact that the effect of condition was significantly stronger in experiment 2 than experiment 3 for the slider data provides some evidence that some participants in experiment 2 were relying on surface-level similarities between the training and test stimuli when choosing their preferred adpositional gesture video. This highlights the importance of ensuring that stimuli for experiments investigating harmony are carefully designed, to avoid the possibility of surface-level patterns being utilised by participants.

5.3 Directions for Further Research

The present study provides the foundation for a number of possible future studies. For example, all adpositional stimuli used in these experiments denote locations. However, different adpositional types may have different baseline preferences. Some evidence for this possibility comes from Mandarin, where adpositions which are path designators (e.g., *cóng*, 'from'; *dào*, 'to') are prepositions, and those which denote locations (e.g., *shàng*, 'on'; *xià*, 'under') are postpositions (Whitman et al., 2013). This suggests that different adpositional phrase order preferences are possible for different semantic categories of adpositions, similar to what we see for extensional and intensional verb phrases (Schouwstra & de Swart, 2014). The results of experiment 1 indicating no baseline preference for adpositional phrase order are therefore not generalisable to all adposition types, and so future work could run similar experiments investigating baseline preferences for other semantic categories of adpositions.

English also displays cross-category harmony of the type investigated in this paper, and although it seems unlikely that there was native language influence at play in experiments 2 and 3, running the experiments on a population of speakers of a non-harmonic language would establish whether the key findings of this paper are robust.

I have proposed that the reduced prepositional preference in the VO condition of experiment 3 may be due to participants not being influenced by the prime in the VO condition. This is because the basic word order of their native language is VO, and there were no immediately obvious surface-level patterns between the priming stimuli and the test stimuli, like in experiment 2, so some participants responded as though they had not received any priming. This hypothesis could be directly tested by repeating experiment 3 on a population of speakers of an OV language. If the hypothesis is correct, then the participants should display a clear preference for harmony in the VO condition but not as strong a preference in the OV condition.

Finally, future research should also utilise the methods used here to investigate other types of cross-category harmony. Cross-category harmony describes a relationship between heads and dependents, but in harmony between the verb phrase and the adpositional phrase, the dependents of both phrase types are noun phrases. Thus, even if further experimentation proved that participants were using a harmony bias in experiment 3, and that the less obvious preference for prepositions in the VO condition is simply due to the priming stimuli following the pattern of their native language and therefore having no effect, this would not provide conclusive evidence for a harmony bias targeting the abstract categories of head and dependent. This is because the harmony preference could still be due to them noticing that the syntactic category of noun phrase comes in the same relative position in both the verb phrase and the adpositional phrase. Therefore, future research should work to evaluate whether the results described in this paper are generalisable to other types of cross-category harmony.

6 Conclusion

I have conducted three experiments utilising ALL and silent gesture perception methods which together aimed to establish whether there is evidence for a cognitive bias which contributes to the prevalence of harmony between the verb phrase and the adpositional phrase cross-linguistically. Experiment 1 demonstrated that, for native English speakers tested using silent gesture perception methods, there is no baseline preference for either prepositions or postpositions. Experiment 2 replicated results in Wang et al. (2021), thus demonstrating that ALL can be used in combination with silent gesture perception methods to show that when the priming and testing stimuli share elements, participants show a preference for harmonic patterns. In experiment 3, results showed that when there were no shared elements between the priming and testing stimuli, participants in the OV condition still showed a preference for harmony, but the preference for prepositions in the VO condition was only slightly above chance. I have argued that the reduced harmonic preference in the VO condition of experiment 3 may be due to the absence of immediately obvious surface-level patterns in combination with the fact that VO is the order in the participants' native language and so did not have the intended effects as a prime. Further experiments should be carried out to test this hypothesis, as well as to evaluate whether these findings are generalisable to other types of cross-category harmony. The fact that the effect of condition was stronger in experiment 2 than experiment 3 for the slider data indicated that the surface-level patterns visible in experiment 2 partially contributed to the significant preference for harmony observed. It is therefore important that stimuli of future experiments investigating biases for harmony using ALL methods are carefully designed to avoid this effect.

7 References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67 (1), 1–48.

Bybee, J. (1988). The diachronic dimension in explanation. In J. Hawkins (Ed.), *Explaining language universals* (pp. 350–379). Cambridge: Cambridge University Press.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences,* 7, 19–22.

Chomsky, N. (2014). *The Minimalist Program* (20th Ed.). Cambridge, Massachusetts: The MIT Press.

Cook, V. J. (1988). Language learners' extrapolation of word order in micro-artificial languages. *Language learning,* 38, 497–529.

Culbertson, J., & Adger, D. (2014). Language learners' privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences - PNAS,* 111*,* 5842–5847.

Culbertson, J., Franck, J., Braquet, G., Navarro, M. B., & Arnon, I. (2020). A learning bias for word order harmony: Evidence from speakers of nonharmonic languages. *Cognition,* 204.

Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in psychology,* 6, 1964–1964.

Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition,* 139, 71–82.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition,* 122, 306–329.

Dryer, M. S. (2013a). Order of adposition and noun phrase. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Retrieved from *Max Planck Institute for Evolutionary Anthropology*: <https://wals.info/chapter/85>.

Dryer, M. S. (2013b). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Retrieved from *Max Planck Institute for Evolutionary Anthropology*: <https://wals.info/chapter/81>.

Dryer, M. S. (2013c). Relationship between the order of object and verb and the order of adposition and noun phrase. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Retrieved from *Max Planck Institute for Evolutionary Anthropology*: <https://wals.info/chapter/95>.

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition,* 136, 215–221.

Gentner, D., & Boroditsky, L. (2009). Early acquisition of nouns and verbs: Evidence from Navajo. In V. Gathercole (Ed.), *Routes to language: Studies in honor of Melissa Bowerman* (pp. 5–32). New York: Psychology Press.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events non-verbally. *Proceedings of the National Academy of Sciences - PNAS,* 105, 9163–9168.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language,* 2, 73–113.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive science,* 31, 441–480.

Hawkins, J. (1990). A parsing theory of word order universals. *Linguistic inquiry,* 21, 223–261.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences - PNAS, 104*, 5241–5245.

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive psychology,* 60, 291–318.

Li, C. N., & Thompson, S. A. (1974). Coverbs in Mandarin Chinese: Verbs or prepositions? *Journal of Chinese linguistics,* 2, 257.

Meir, I., Lifshitz, A., Ilkbasaran, D., & Padden, C. (2010). The interaction of animacy and word order in human languages: A study of strategies in a novel communication task. In A. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The Evolution of Language* (pp. 455–456). World Scientific.

Moravcsik, E. A. (2010). Explaining language universals. In J. J. Song (Ed.), *The Oxford Handbook of Linguistic Typology* (pp. 69-89). Oxford: Oxford University Press.

Motamedi, Y., Wolters, L., Naegeli, D., Kirby, S., & Schouwstra, M. (2021). From improvisation to learning: How naturalness and systematicity shape language evolution.

Padden, C., Hwang, S.-O., Lepic, R., & Seegers, S. (2015). Tools for language: Patterned iconicity in sign language nouns and verbs. *Topics in cognitive science,* 7(1), 81–94.

R Core Team. (2019). *R: A language and environment for statistical computing.* Retrieved from *R Foundation for Statistical Computing*: <https://www.R-project.org>.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition,* 131, 431–436.

Smith, K., & Kirby, S. (2008). Cultural evolution: Implications for understanding the human language faculty and its evolution. *Philosophical Transactions. Biological sciences*, 363(1509), 3591-3603.

Tily, H., Frank, M., & Jaeger, F. (2011). The learnability of constructed languages reflects typological patterns. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 33.

Verhoef, T., Padden, C., & Kirby, S. (2016). Iconicity, naturalness and systematicity in the emergence of sign language structure. In S. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 11th international conference (Evolangx11)*. Retrieved from *The Evolution of Language*: <http://evolang.org/neworleans/papers/47.html>.

Wang, F., Kirby, S., & Culbertson, J. (2021). A cognitive bias for cross-category word order harmony. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 43.

Whitman, John, Paul, Waltraud, & Djamouri, Redouane. (2013). Postpositions vs Prepositions in Mandarin Chinese: The Articulation of Disharmony. In *Theoretical Approaches to Disharmonic Word Order* (p. Theoretical Approaches to Disharmonic Word Order, 2013). Oxford: Oxford University Press.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B),* 73(1), 3-36.

Yin, S. H., & White, J. (2018). Neutralization and homophony avoidance in phonological learning. *Cognition,* 179, 89–101.

Zhao, Y., & Fedzechkina, M. (2020). Learners' harmonic preferences are modulated by lexical retrieval difficulty. *Proceedings of the Linguistic Society of America,* 5, 708–721.

Exploring Gender and Sibilants: A Le/s/on From Transgender Speakers

Jaidan McLean
University of Oregon

Abstract. The variation of sibilant consonants is a well-studied topic in sociolinguistics, with previous studies showing variation across social axes of gender, sexuality, class, race, and region (e.g. Campbell-Kibler 2007; Levon 2007; Pharao et al. 2014; Podesva & Van Hofwegen 2015; Stuart- Smith 2007). More recently, scholars such as Zimman (2017) have raised questions about the social patterning of sibilants in regard to gender and speech production, given that the primary focus of gendered sibilant variation has been analyzed with only cisgender identity in mind. Using sibilant productions among transgender men and transmasculine people, Zimman (2017) discussed the social origins and implications of gender identity and gender expression on /s/ variation. The purpose of this study is to build on this recent research to explore how gender identity and expression influence sibilant production in a sample of transgender women and transfeminine speakers. As part of an undergraduate honours thesis, this study specifically attempts to address the following two questions regarding sibilant production, specifically /s/: What type of variation are present in the speech productions in the sample of speakers, particularly regarding the phonetic details of sibilants? How does speaker identity influence the patterns for these speech features? Sociolinguistic interviews are used to elicit natural sibilant production, then an acoustic measure and analysis is to be conducted, focusing on the frequency centre of gravity as characterised in sociophonetic research (Thomas 2011, Kendall and Fridland 2021).

Keywords: gender; sibilants; trans linguistics

---

1 Introduction

While the study of sibilant consonants is nothing new in sociophonetic research, there are new ways of thinking that are expanding social axes and hierarchies, so sociophonetics should adapt and explore these new ways of thinking. One of these new ideologies, and that which will be the topic of exploration in this thesis paper, is that the gender binary of man and woman are socially constructed ideals that many people challenge every day with their existence. People who are transgender, nonbinary, and otherwise gender nonconforming complicate this notion of gender as an innate binary system but rather a socially structured system.

These socially structured terms and labels are important to this thesis because they are the very aspects of social influence on language production analysed by scholars. How speakers identify and how they choose to present comes across in how they speak. Sibilant consonants, specifically /s/, have shown much interesting variation across social axes of gender, sexuality, class, race, and region in previous sociolinguistic research (e.g., Levon, 2007; Campbell-Kibler, 2011; Pharao et al., 2014; Podesva & Van Hofwegen, 2014). All of this social variation in how sibilants are produced makes these consonants highly salient for studying how social identities are tied to speech production. Similarly, scholars like Munson et al (2006) raise questions about speech style in relation to gender that will be explored in this study.

Before diving into the study more, I would like to establish some clarifying definitions to terminology that will be frequently used. As with anything related to social identities, no one label will be interpreted the same by every person and so in defining some of this jargon I am defining how it will be used in the context of this thesis. *Transgender* refers to a person whose gender identity differs from their assigned sex at birth – the shortened, more colloquial term ‘trans’ will be used. *Cisgender* refers to a person whose gender identity is the same as what sex they were assigned at birth – the shortened, more colloquial term 'cis' term will be used. *Nonbinary* (also written *non-binary*) refers to someone who does not identify as the sex they were assigned at birth, but they also do not necessarily identify in-line with the opposite sex. *Trans feminine* or *trans femme* is a descriptor used by those who usually were assigned male at birth but identify as trans or nonbinary and want to note that they are feminine without having to ascribe to being binarily a woman.

2 Background and Literature

The acoustic measurement focused on in this study is the spectral centre of gravity (COG), which measures for averaging high frequencies and intensity and is useful for examining sibilant production, as articulated in sociophonetic research (e.g., Thomas, 2011; Kendall & Fridland, 2021). COG variation has consistently been shown to be linked to gender and sexuality (e.g., Stuart-Smith, 2007; Munson et al, 2006), therefore this study focuses on COG for measuring in transgender speaker's /s/ production. Scholars have identified that the typical spectral centre of gravity frequency range for /s/ production among American English-speaking men and women are 4,000-7,000 Hz and 6,400-8,500 Hz, respectively (Avery & Liss 1996; Fuchs & Toda 2010; Flipsen et al. 1999; Nittrouer et al. 1989; Stuart-Smith 2007, 2019).

When it comes to gender and sibilants, there are three categories that seem to emerge in the literature for populations researched: men, women, and "those with complex profiles" (Hall-Lew, Moore, & Podesva 2021, p. 132). The third population have identities that are more "complex" because they do not align with the heteronormative research literature and are instead people who identify within the LGBTQ+ community either in gender, presentation, or sexuality.

Podesva and Van Hofwegen (2014) explore how heteronormative pressures from a person's regional politics can constrain the variation seen when it comes to how gender influences /s/ production. Evidence that social factors beyond gender identity, but how one chooses to express that identity, are linked to /s/ production are important for this study because it considers the relationship between the speaker's identity and their expression on stylistic choices and variation seen in data. Calder (2019) adds to this discussion of gender expression and sibilants by focusing on how visual presentation through drag in the LGBTQ scene correlates to more fronted-/s/ frequencies. They suggest there is a strong connection between visual presentation (i.e. expression) and linguistic performance, where acoustic differences of /s/ become part of the hair and makeup used in expressing one's fierceness in drag queen personae (59). This suggestion is linked to trans expression because there is a possibility, which this study aims to address, in a speaker's gender expression influencing language just as gender identity does.

In considering sexuality as a factor for influenced /s/ production, Podesva and Van Hofwegen's 2016 case study on LGBTQ speakers from Redding, California finds that /s/ productions by gay men and lesbians differ from that of straight men and women but do still fall within the established cisgender ranges previously discussed. Complicating the implication of these findings are other works by scholars like Munson et al (2006) who explored sibilants and sexuality but additionally speech style. In their study, they used read speech as opposed to spontaneous speech in order to control for the speech style external factor and focus on the social factor of sexuality on speech production and perception. Although my empirical study focuses on production and not perception, it is difficult to discuss gender expression without considering that they are expressing a certain way likely to be perceived in a certain way. I wanted to bring in the differences of speech style, spontaneous versus scripted (read), to further understand how gender, expression, and sexuality influence sibilant production.

Turning to literature on transgender people, novel questions relating to trans identities have been raise by scholars like Zimman (2012; 2017) regarding the social patterning of sibilants in the context of "those with complex profiles," after Zimman found the average COG ranges for transgender men and trans masculine speakers to be all over the board. These findings were significant because the speech was collected from trans masculine people in their first year or two of testosterone hormone therapy which physically lowers a person's vocal pitch (p. 994). This means that even though each speaker is experiencing their vocal pitch lower due to testosterone, the COG average frequencies patterned in-line with stylistic variation influenced by the trans masculine speaker's gender, expression, and sexuality. Zimman's (2017) average COG findings for trans masculine speakers are most relevant to this study because it begs the question about trans feminine speakers.

My study will focus on filling the sociophonetic research gap on trans feminine voices and finding if their social identities influence their /s/ production in similar ways. In exploring research gap, it is an important to consider a key difference in how hormone therapy, taking oestrogen specifically, affects trans feminine people's voices. There is no meaningful effect. Although the hormone replacement therapy aims to replace testosterone with oestrogen, after puberty the vocal tract has already been physically affected by the testosterone and oestrogen cannot reverse this (Hari Kumar et al., 2016). This means that for trans feminine people wanting to sound more feminine they have to consciously manipulate their voice, known as voice training. Further research is needed to understand if and how voice training for trans feminine people ties into sibilant production. According to Oates (2019), techniques used to increase voice perception includes lowering fundamental frequency, manipulating resonance, and breathiness. In this study, voice training will be considered particularly in regard to speech style.

When it comes to spontaneous versus scripted speech, I hypothesised that regardless of gender identity, expression, or sexuality speakers will exhibit a higher average COG in the reading tasks than in the interview. I am hypothesising this due to the increased attention to speech that will occur with reading aloud a short passage and word list, compared to the less attention paid to speech in conversation. In conjunction with this, the increased attention will likely lead to increased articulation in which /s/ production may move more front in the mouth, leading to a potentially higher frequency produced (Podesva & Van Hofwegen, 2014; Munson et al., 2007).

3 Research Questions

Inspired by the work of these other sociolinguists and sociophoneticians, particularly Zimman (2017), this study aims to answer two key questions about trans feminine people's sibilant production. Are the COG frequency ranges for speakers sampled stratified by gender identity, presentation, and/or sexual identity; if so, how? What are the /s/ frequency ranges for trans women and trans femme speakers sampled as in the context of spontaneous speech (interview) versus scripted speech (reading tasks)?

Based on the findings of previous scholars and their research, especially the findings of Zimman 2017, in answer to the first question I hypothesized that the COG frequency ranges will meaningfully vary in regard to gender identity, expression, and sexuality, where queer androgynous speakers will differ from more heteronormatively feminine speakers in both spontaneous and scripted speech. In other words, more typically presenting speakers who identify as trans women will produce higher average COG than less typically feminine presenting nonbinary trans femme speakers.

When it comes to spontaneous versus scripted speech, I hypothesised that regardless of gender identity, expression, or sexuality speakers will exhibit a higher average COG in the reading tasks than in the interview.

4 Methods

4.1 Collection

In order to elicit spontaneous /s/ production I conducted short sociolinguistic interviews, roughly 25 minutes each.[33] The sociolinguistic interviews included general metadata questions regarding the participant's age, hometown, and race/ethnicity, questions specifically designed to elicit /s/ production, and questions regarding the person's general experiences surrounding their gender identity, expression, and sexuality.

An outline of the interview questions asked is included in *Appendix A,* but an example of a question asked to specifically elicit /s/ in a spontaneous, conversational way was "What was your favourite subject in school?" or "Who was your favourite instructor from your classes?" where target /s/ words like *subject*, *school*, *instructor*, and *classes* would likely be repeated by the participant. More personal questions regarding the speaker's experiences in order for me to better understand their identity and expression include one like "Has there ever been a time in which you felt obligated to alter your expression or hide your gender identity?"

I interviewed and recorded a total of five speakers whose identities vary from nonbinary trans femme to "woman of transgender experience" to trans woman. Each speaker's race was white, with one speaker expressing Hispanic ethnicity, and they were from various US states: two from Oregon, two from Texas, and one from Hawai'i. The five trans women and non-binary trans femme people ranged in age from 18 to 26 years old and were from the University of Oregon community in response to recruitment emails and flyers. These emails and flyers were posted in on-campus LGBTQ-affiliated clubs and groups, as well as sent to organisers for trans women support groups such as the voice training program *Speak Up!*. The original intention of this study was to collect from as many trans feminine speakers as able to from the college community, however due to unprecedented times only five trans feminine people responded.

Using the same language and chosen pseudonym from the speakers who described themselves in the interviews, Table 1 shows an overview for each speaker's gender identity, expression, and sexuality. One thing that needs to be pointed out first and foremost is the unfortunate limitation to such a small study that I am unable to thoroughly test the sexual identity aspect of my study, simply because all of my speakers identified in some non-heterosexual way. Aside from that, it is interesting to see the similar yet still various aspects of each person's identity, such as Calypso and Onyx who identified their trans femme expression with androgynous, but also alternative style. Luna also expresses herself in an alternative, more gothic fashion, but notes that she is femininely doing so through lace and dresses. Similarly, Talia and Mira both express themselves in a "typical" more heteronormatively feminine way (e.g. painted nails, dresses).

Table 1: *Speaker Metadata.*

Speaker	Gender Identity	Expression	Sexuality
Calypso	non-binary trans femme	androgynous; alternative feminine	bisexual
Luna	trans woman	femininely; gothic	lesbian
Mira	woman of trans(gender) experience	typically feminine; painted nails, long hair, makeup	bisexual
Onyx	trans feminine	androgynous; punk style	lesbian
Talia	trans woman, trans femme, genderqueer	typically feminine; dresses	queer

[33] An exempt determination was made for the collection of this study by the University of Oregon's Institutional Review Board (STUDY00000426), because although I am working with people who are considered protected populations, all of my recordings and data have been anonymised either through direct audio editing or pseudonyms.

After the short interview, I then asked participants to read aloud a short passage and word list that I would use to identify variation in the scripted speech compared to the interview. These texts are provided in Appendix B and Appendix C and contain multiple words that occur in both the reading passage and word list.

4.2 Acoustic Analysis

The speech recordings were analysed acoustically using the popular free phonetics computer software package Praat (Boersma 2001) that allows for specific sounds, in this case /s/, to be isolated and measured for their COG. In addition to the isolation and extraction of /s/ produced by the trans feminine speakers, Praat was used to edit the audio files to remove any identifying information or parts of the interview requested to be removed.

Next, I went back through each isolated /s/ and labelled the segment with the word the /s/ was in, as well as annotated the label with a symbol indicating whether the /s/ was the first, second, or third /s/ in the word. See Figure 1 for an example with the word *successful* which has a word-initial /s/ noted with an exclamation point, a word-medial /s/ noted with a question mark, and a second word-medial /s/ noted with a tilde. Word-final /s/ productions are not measured because it has been argued that the assimilation that occurs with the following phonological environment can skew the COG of the /s/ when measure (Gunter et al 2021).

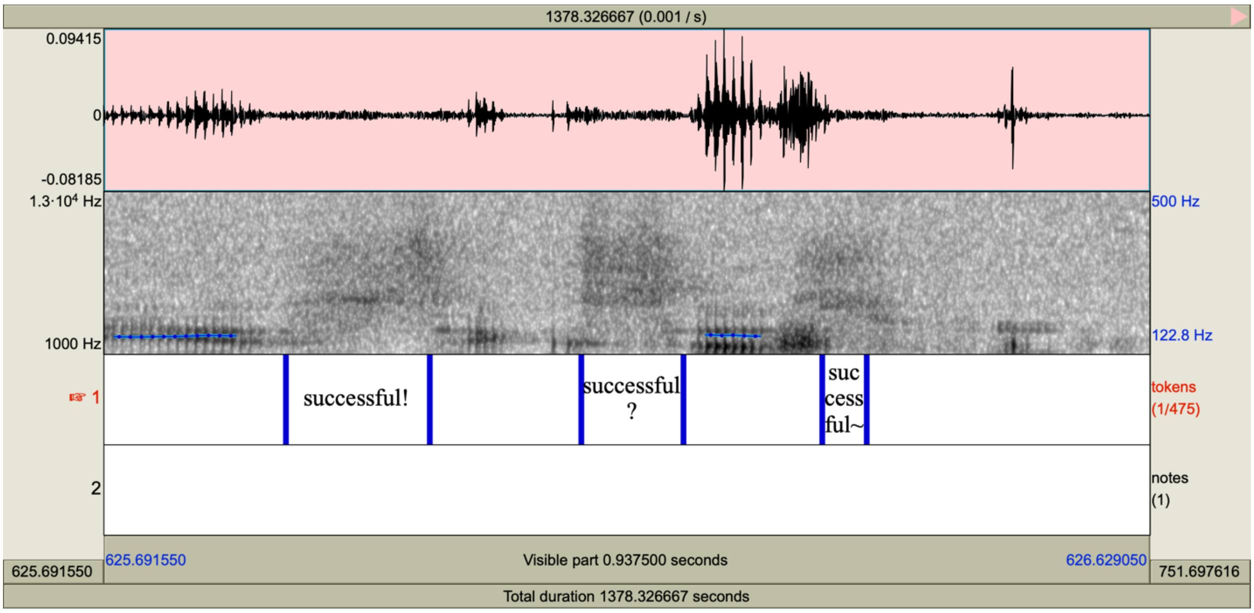


Figure 1: *'successful' parsed, labelled, and annotated for numerical position in Mira's interview at 625 seconds.*

Going through a listening one-by-one to each segment in order to label and annotate it allowed for second overpass of tokens where I could adjust the segmentation or add an additional annotation with an asterisk to indicate "don't count" when I got to extracting the sibilants.

Words that were deemed "don't count" tokens were ones in which there was audio interference with the spoken word, or where the /s/ was not realised fully as either word-initial or word-medial. For example, the word *just* was a frequent token that would orthographically has a word-medial /s/ before the /t/ but was often produced by my speakers as word-final with the /t/ dropped. For one participant in particular, *just* was primarily used as a filler in conjunction with *like* which meant the /t/ was almost always dropped (see Figure 2). In instances where I could not definitively decide if the /s/ was being produce word-final or not I discounted these too, especially if in instance of filler conjunctions *just like* to be consistent.

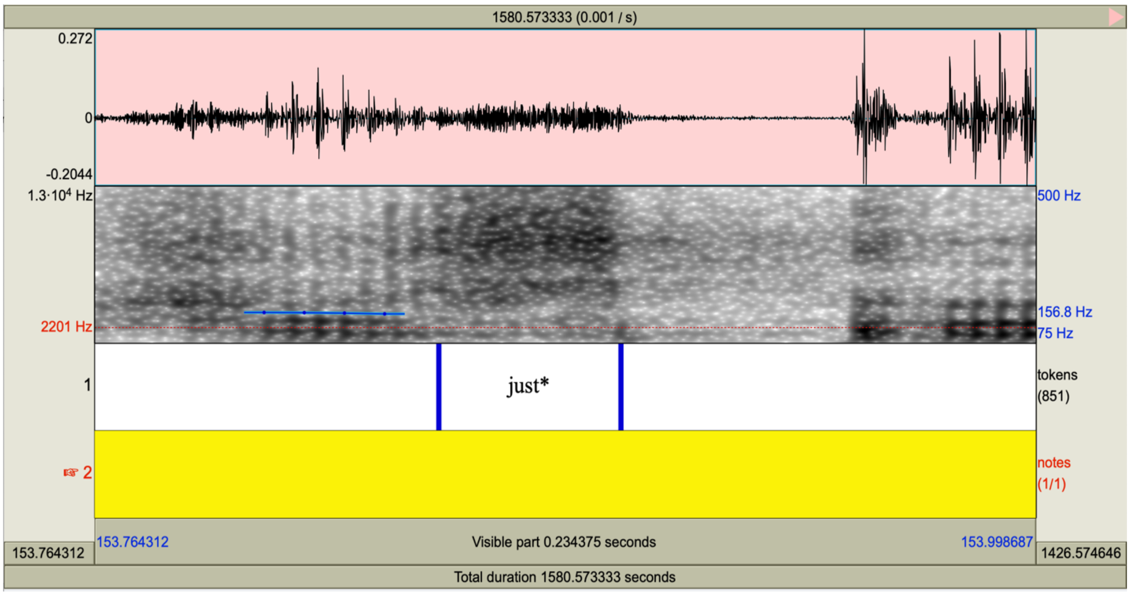


Figure 2: *'just' parsed, labelled, and annotated to be not counted in Luna's interview at 153 seconds.*

4.3 Statistical Analysis

After the labelling and initial annotating, the audio files were Hann pass filtered at 1,000-13,000 Hz to eliminate background noise and focus on the range in which the bulk of /s/ production is typically found (Kendall and Fridland 2021). The COG measurements were then extracted using the labelled sibilant extraction script in Praat, accessed from the Kendall and Fridland (2021) supplemental materials. These extracted measurements, specifically COG, were then analysed for variation and patterns influenced by speakers' gender identity and expression. These methods follow standard social and behavioural quantitative data analysis methods for sibilants (Kendall & Fridland 2021, ch. 3).

The linguistic factors that the COG measurements were organised and analysed by are sibilant's word position, phonological environment, and number of syllables in a token word. The word's position included whether the /s/ was initial or medial (e.g. *street* vs. *absolute*), but following if the /s/ was final, as mentioned with *just*, because of the messiness that becomes involved with word-final /s/ assimilation with the following word's sounds in a phrase (Gunter et al 2021). The environment included whether the /s/ was part of a consonant cluster or the single consonant preceding a vowel (e.g. *street* vs. *absolute*). In annotating for environment I realised that this was something I failed to control more for in choosing /s/ eliciting words for my word list, resulting in an uneven distribution of environmental features where /s/ occurred as a singleton preceding a vowel four times more than in a consonant cluster. This particular limitation and learning experience are discussed more in the *Results* section.

Once the average COG frequencies based on these factors were identified, I began comparing, contrasting, and overall analysing how the frequencies were stratified to explore my first hypothesis regarding the variation in /s/ based on gender identity, expression, and sexuality. Along with trying to find patterns in the linguistic factors of the data, I separated and organised the average COG frequencies by speech style: interview, reading passage, or word list. In answering how speaker's /s/ production varies in spontaneous speech versus scripted speech I wanted to further dissect the reading passage data from the word list data for the clearest picture. All of this data exploring was conducted using Microsoft Excel once I opened the extracted sibilants measurement text files into the spreadsheet. A mixed-effect linear regression was also conducted on the COG data, testing all of the factors of interest as main effects and including a random intercept for the speaker. Statistical patterns from the regression are included when relevant in the discussion of the results (Kuznetsova et al., 2017).

4.4 Interspeaker – Social Patterns

The COG frequency analyses conducted do indeed exhibit social patterns and trends among the speakers. Table 2 lists the overall average COG for /s/ productions for the interview speech, reading passage speech, and word list speech. Immediately it is evident that these speakers greatly range in their averages, with the overall range of these frequencies falling between about 3,100-6,800 Hz for interview (spontaneous) speech.

Table 2: *Average COG Frequencies by Speaker*

Speaker	Interview	Reading Passage	Word List
Calypso	3083.7	3542.66	3386.99
Onyx	4587.07	4251.05	4983.76
Mira	5233.57	5672.01	5988.37
Talia	6052.5	5423.87	6235.17
Luna	6822.99	6633.16	7207.37

I had hypothesised that the COG frequencies would meaningfully vary in regard to gender identity, expression, and sexuality. As already discussed, sexuality will not be a key component of my study after all due to my participant pool, but in addition I would like to note that it is tricky to divide expression from gender identity in this case because both nonbinary trans femme individuals also happen to both express themselves in alternative androgynous style. This alternative punk visual presentation is not necessarily the universal expression for nonbinary trans femme people and so my five-person study does not allow me to explore more heteronormatively feminine presenting nonbinary trans femmes.

The COG frequencies meaningfully varying in regard to gender identity does in fact seem to be true given the data in Table 2, where nonbinary trans femme speakers' (Calypso and Onyx) average COG ranged below 5,000 Hz and trans women's (Talia and Luna) averages ranged above 6,000 Hz. In other words, those identifying outside of the binary as nonbinary and "gender nihilist" (Onyx_int, 675 seconds) produce lower /s/ COG on average, while those who identify more with the binary as trans women produce higher /s/ COG on average aligning more with feminine expectations.

Interestingly, when asked her gender identity Mira in particular described in-line with the truth that trans woman are genuinely women that she was a woman but just "of trans experience" (Mira_int, 683-694 seconds). Interestingly, Mira's overall average interview speech COG was within the 5,000-6,000 Hz range between the nonbinary trans femme speakers and speakers identifying as unmarked trans women.

Overall, a meaningful and interpretable pattern is evident with the average COG frequencies linked being linked to gender. While this is importantly true for my five trans feminine speakers, I cannot make a generalising claim about this applying to all trans feminine speakers due to the small size.

4.5 Interspeaker – Linguistics Features

The data presented above do not suffice alone in allowing us to understand trans feminine people's sibilant production and how it's influenced by their gender identities and expressions. The linguistic features of /s/ position and environment were considered in order to account for potential linguistical influences that are resulting in the frequencies being seen (as opposed to it actually being influenced by gender and expression).

Continuing the focus first on the interview data, Figures 3 and 4 show the position and environment analyses organised by speaker. In Figure 3, it is evident that for every speaker word-initial positioned /s/ productions resulted in higher productions than /s/ in word-medial positions. The statistical model, considering all the data from the three speech styles, predicts that initial sibilants are 140 Hz higher in COG than non-initial sibilants and is statistically significant ($p < 0.01$). For environmental features, Figure 4 shows that /s/ produced in consonant clusters are on average lower in COG than /s/ produced as a single consonant preceding a vowel. Similarly the statistical model predicts sibilants produced preceding a vowel will be 300 Hz higher in COG than consonant cluster-occurring sibilants, and that this is statistically significant ($p < 0.001$).

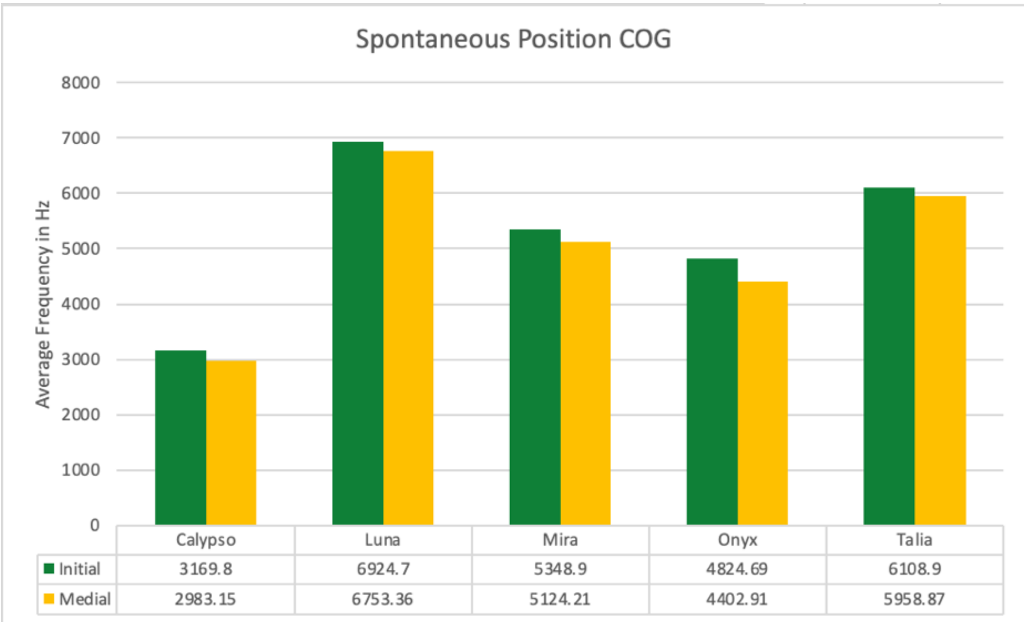


Figure 3: *COG for Interviews by Position*

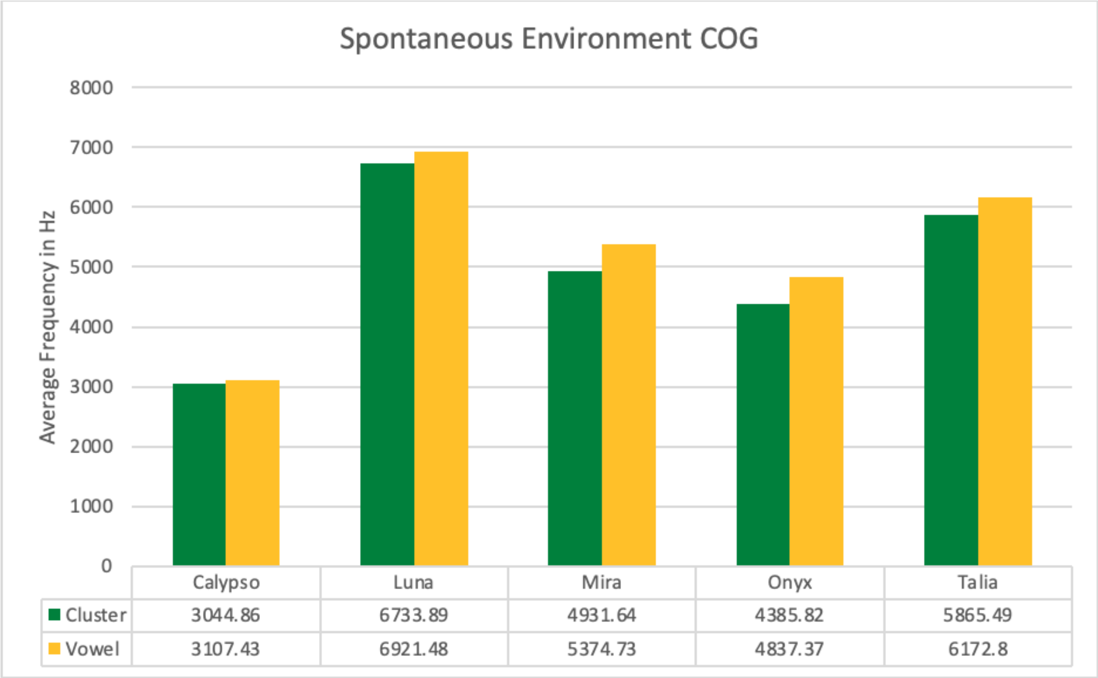


Figure 4: *COG for Interviews by Environment*

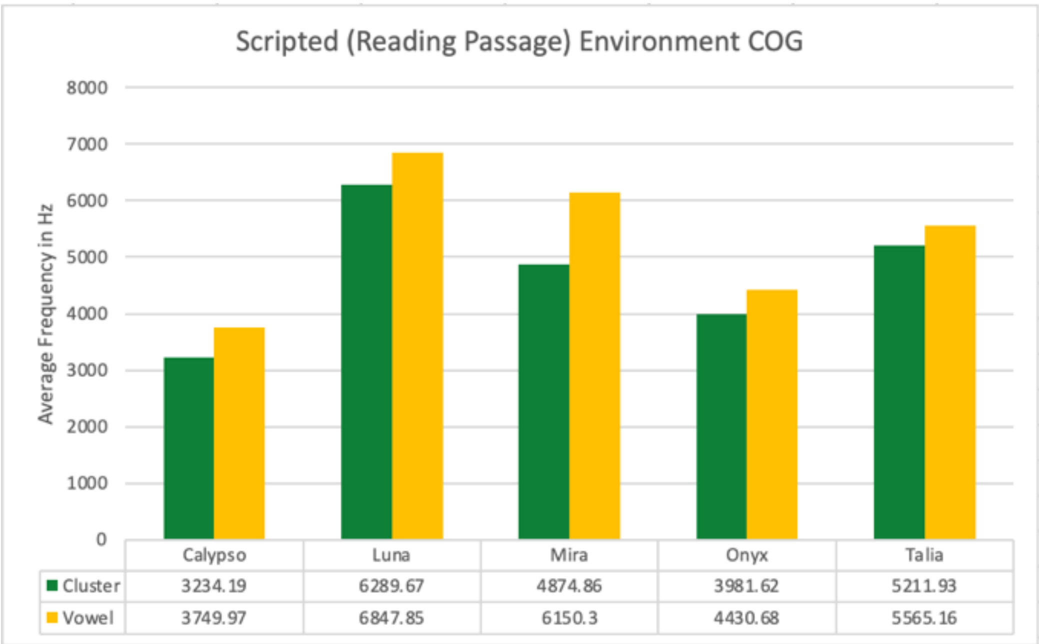
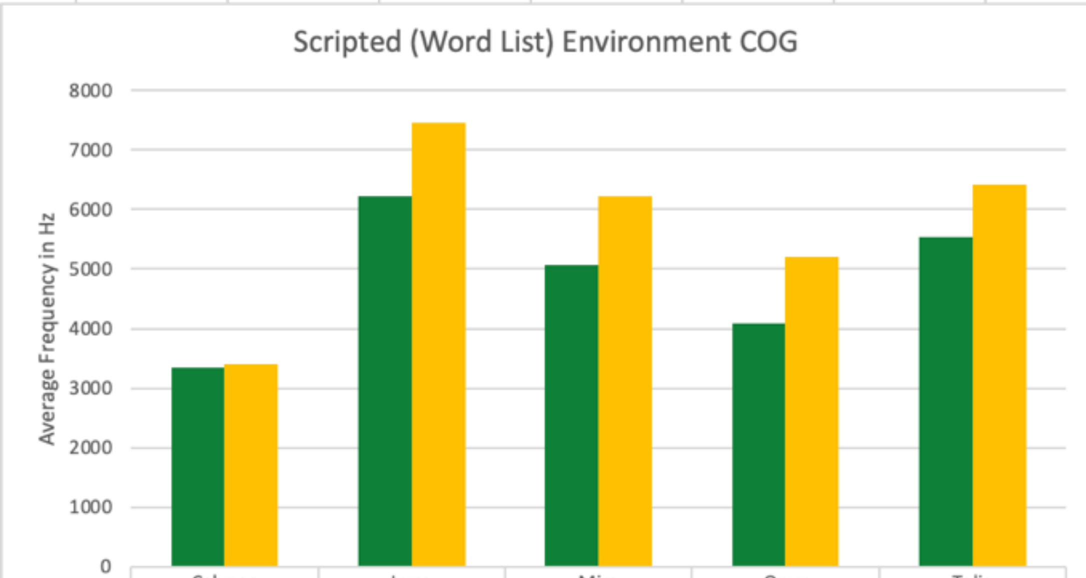


Figure 5*: COG for Reading Passage by Position*

	Calypso	Luna	Mira	Onyx	Talia
Initial	3783.77	6848.12	6041.89	4554.1	5539.81
Medial	3283.53	6382.4	5292.03	3922.76	5298.26

Figure 6: *COG for Reading Passage by Environment*



	Calypso	Luna	Mira	Onyx	Talia
Cluster	3339.1	6231.69	5057.88	4087.93	5540.91
Vowel	3398.96	7451.29	6231.11	5207.72	6408.74

Figure 7: *COG for Word List by Position*

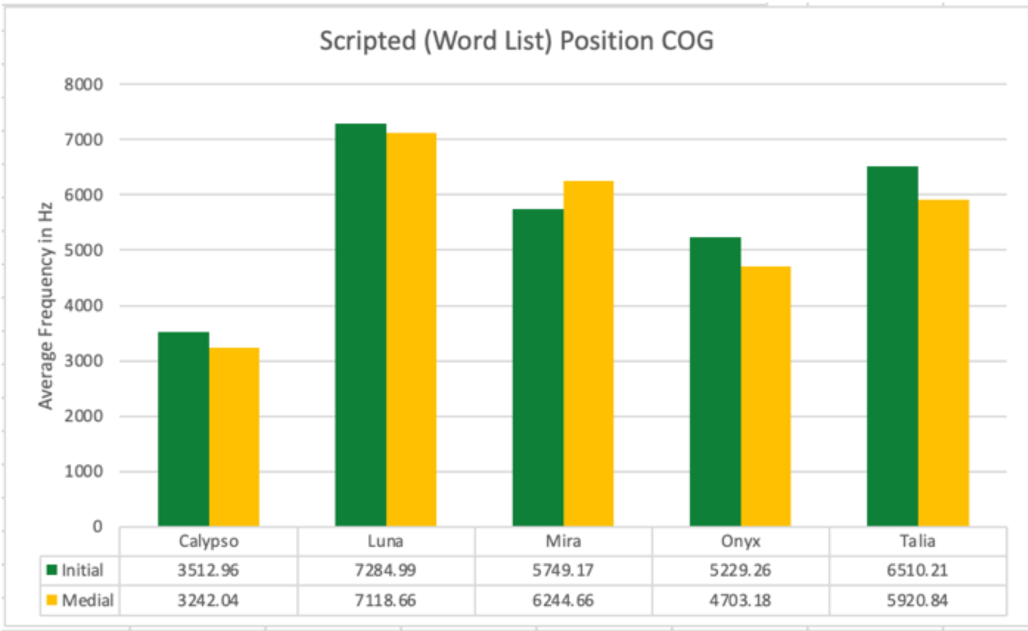


Figure 8: *COG for Word List by Environment*

Similarly with the scripted reading tasks of the passage and word list, Figures 5 through 8 show the same position and environment features as previously discussed. These figures show the same trend in position where word-initial /s/ result in higher COG on average than word-medial, and in environments where /s/ preceding a vowel result in higher COG on average than /s/ in consonant clusters. Each speaker exhibiting the same linguistic trends despite varying gender identities and expressions demonstrates that the variation between sibilant productions is indeed being socially motivated rather than linguistically motivated.

4.6 Intraspeaker – Spontaneous vs. Scripted Speech

While for gender identity my hypothesis was true, the same cannot be said for the spontaneous versus scripted speech analysis. I had hypothesised that regardless of gender identity or expression every speaker would produce higher average sibilant COG in the scripted over spontaneous speech. That hypothesis is only true for Calypso and Mira, as seen in Figure 9, in which both speakers exhibit lower average COG for /s/ productions in the interview than either of the reading tasks. However even still, both differ in terms of which speech style had the highest average COG, where Mira's patterned more along the lines of Luna, Onyx, and Talia who all produced the highest average COG frequencies in the word list task, but interestingly low averages for the reading passage.

Whereas in the previous results we saw an evident pattern for gender identity and average COG with Calypso and Onyx both identifying as nonbinary trans femme and producing lower average COG, we do not see this same gender identity trend for the data in Figure 9. Since Onyx shows a similar pattern as the trans women, Luna and Talia, it seems that gender identity may not meaningfully influence the /s/ production in the different styles. Although I had not expected this, such does support Zimman's (2012) argument about gender as stylistic bricolage discussed in the *Background and Literature Review*.

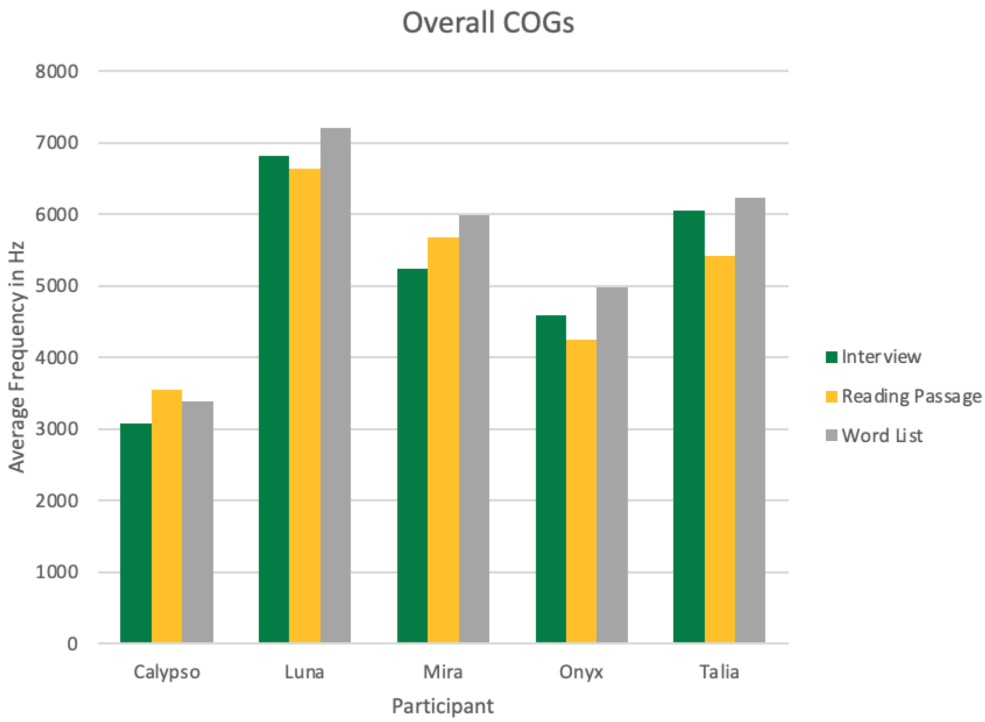


Figure 9: *Average COG Frequencies by Speaker*

If gender is not meaningfully influencing these differences, then there are two other potential influencers. One of these possibilities pertains to the words in the reading passage being within a phrase versus the words in the word list being in isolation. The attention to speech is going to intrinsically be higher with the word list since speakers are only attending to one word at a time, whereas in the reading passage there is less attention to individual words as there is to an overall segment or phrase. It is entirely possible that this is not true for every speaker and that Calypso and Mira could have actually been paying especial attention to the segments or phrases to where all of their speak was more fronted and therefore higher frequency.

In addition, I must revisit the limitation mentioned about annotating for environments where I found the uneven distribution of sibilants occurring in preceding vowel versus consonant clusters. Considering that in the previous section we see that /s/ preceding a vowel has a higher COG on average, another possibility for the word list showing much higher average COG than reading passage could be due to the uneven distribution influencing the averages found. Whether this is significant or not is something the definitely calls for further research.

Additionally, the simple statistical analysis indicates the difference between trans women having higher COG than trans femmes is not quite statistically significant ($p = 0.08$). However, the Wald statistics (from an analysis of variance function) does indicate that including gender identity significantly improves the statistical model's fit on the data ($p = 0.01$). In other words, with only five speakers it is hard to say anything concrete or generalizable about the social patterns from these data, but the fact that gender identity improves the model and is marginally significant as a factor in the model at all, supports the overall patterns identified.

5 Discussion and Conclusion

In an attempt to answer the questions of how sibilants pattern in relation to gender identity, expression, and sexuality this thesis paper raised interesting avenues to other important questions regarding transgender people's speech production.

So then what's the le/s/on? If only more questions were produced? Well that's just it: Research presented here showed that the trans feminine people sampled in this study pronounce their /s/ in-line with their gender identity and expression, not their assigned sex at birth. A large takeaway from this thesis is that these speakers maintained stylistic variations in their sibilant production that patterned with their gender identity and expression, similar to what other scholars like Zimman (2017) found for trans masculine people. It becomes evident when piecing small studies together like these that there is a larger, socially constructed, influence that that social identities have on a person's language production that goes beyond the normatively studied cisgender men and women.

Albeit a small empirical study, these similarities show that there is a need for further research on "those with complex profiles" such as trans people to create additional conclusions about gender and sibilants.

5.1 Further Research

As mentioned in the *Background and Literature Review* section, voice training is something this study considers but does not explore extensively.

In the interview, discussion about the participant's voice came up either naturally in conversation or through me directly asking if voice training was something they were interested in or had ever practised. Although there is always a sense of observer's paradox in studies like these, speakers expressed comfortability speaking with me, especially as another non-cisgender individual. Talia even offered to give me a sample of her feminine voice compared to what they had been using with me in conversation; "Right now I'm using my standard voice" referring to her non-hyperfeminised voice that she learned to produce using techniques from trans voice training (Talia_int, 1040-1043 seconds). She actually provides a 10-second "brief sample" of her speaking in her "non-standard" more feminized voice that could be analysed more closely (Talia_int, 1080-1090). While Talia's sample alone could not inform me of what voice training effects might be present in sibilant production, this is an aspect of trans women and trans feminine voices that needs sociophonetic attention.

Overall, this small empirical study as unveiled that transgender identities, particularly those of trans feminine people, need greater sociophonetic research in order to understand these "complex profiles" in the same way scholars sought to understand cisgender men and women's sibilants. Supporting the findings from trans masculine speakers in the Zimman (2017) study, trans feminine speakers' sibilant production in-line with their gender identity, demonstrating that language is influence by social structures like gender.

6 References

Avery, J. D., & Liss, J. M. (1996). Acoustic characteristics of less-masculine-sounding male speech. *The Journal of the Acoustical Society of America,* 99(6), 3738–3748.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International,* 5(9/10), 341–345.

Calder, J. (2019). The fierceness of fronted /s/: Linguistic rhematisation through visual presentation. *Language in Society*, 48, 31–64.

Campbell-Kibler, K. (2011). Intersecting Variables and Perceived Sexual Orientation in Men. *American Speech,* 86(1), 52–68.

Flipsen Jr, P., Shriberg, L., Weismer, G., Karlsson, H., & McSweeny, J. (1999). Acoustic characteristics of/s/in adolescents. *Journal of Speech, Language, and Hearing Research,* 42(3), 663–677.

Fuchs, S. and Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? *Turbulent sounds: An interdisciplinary guide,* 21, 281–392.

Hall-Lew, L., Moore, E., & Podesva, R. J. (Eds.). (2021). *Social Meaning and Linguistic Variation: Theorising the Third Wave.* Cambridge: Cambridge University Press.

Hari Kumar, K.V., Garg, A., Ajai Chandra, N.S., Singh, S.P., & Datta, R. (2016). Voice and endocrinology. *Indian Journal of Endocrinology and Metabolism,* 20(5), 590–594.

Gunter, K., Vaughn, C., and Kendall, T. (2021). Contextualising /s/ retraction: Sibilant variation and change in Washington D.C. AAL. Language Variation and Change. *Language Variation and Change,* 33(3), 331–357

Kendall, T. and Fridland, V. (2021). *Sociophonetics (Key topics in sociolinguistics).* Cambridge University Press.

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.

Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society,* 36(4), 533–554.

Munson, B., McDonald, E.C., DeBoe, N.L., & White, A.R. (2006). The acoustic and perceptual bases of judgements of women and men's sexual orientation from read speech. *Journal of Phonetics,* 34, 202–240.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R.S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research,* 32(1), 120–132.

Oates, Jennifer M. (2019). Evidence-Based Practice in Voice Training for Trans Women. In R. K. Adler, S. Hirsch, & J. Pickering (Eds.), *Voice and Communication Therapy for the Transgender and Gender Diverse Client: A Comprehensive Clinical Guide* (3) (pp.87–103). San Diego: Plural Publishing.

Pharao, N., Maegaard, M., Møller, J., Spindler, J., & Kristiansen, T. (2014). Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society,* 43(1), 1–31.

Podesva, R. & Van Hofwegen, J. (2014). How Conservatism and Normative Gender Constrain Variation in Inland California: The Case of /s/. *University of Pennsylvania Working Papers in Linguistics,* 20, 129-137.

Podesva, R. and Van Hofwegen, J. (2016). /s/exuality in smalltown California: Gender normativity and the acoustic realisation of /s/. In E. Levon & R. Beline Mendes (Eds.), *Language, sexuality, and power* (p.168–188). Oxford: Oxford University Press.

Stuart-Smith, J. (2007). Talkin' Jockney'? Variation and change in Glaswegian accent. *Journal of Sociolinguistics,* 11(2), 221–260.

Stuart-Smith, J., Sonderegger, M., Macdonald, R., Mielke, J., McAuliffe, M., & Thomas, E. (2019). Large-scale Acoustic Analysis of Dialectal and Social Factors in English /s/-retraction. International Congress of Phonetic Sciences (ICPhS 2019), Melbourne, Australia, 5-9 Aug 2019, 1273-1277.

Thomas, E. (2011). *Sociophonetics: An Introduction*. London: Palgrave Macmillan.

Zimman, L. (2012). *Voices in transition: Testosterone, transmasculinity, and the gendered voice among female-to-male transgender people*. [Unpublished dissertation]. University of Colorado.

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

7 Appendices

7.1 Appendix A

Outline of interview questions. This outline was used during the interview in which I asked the questions as the conversation went along, rather than having a more scripted dialogue to read from. I chose to do this to keep speech as spontaneous as possible, since there is always the consideration of the observer paradox.

● General Introduction o Name, pronouns, age o Year (if college student), occupation o Hometown/state, race/ethnicity
● School life o Schooling experience o Favourite school subject ▪ Is this your current/was your major or minor in college (if applicable) o Favourite instructor o Extracurriculars (currently or in the past)
● Personal life o Most proud of? ▪ Achievements, looks, personality, personal accomplishments o Could not be proud of yourself because of a social identity you held? o Gender identity, Sexuality ▪ Gender identity and/or sexuality to be an obstacle in your daily life or personal growth? Steppingstone? o Expression ▪ Fashion, lifestyle, hobbies ▪ Vocal training ▪ Instance where felt had to limit or change gender expression? Why?

7.2 Appendix B

Reading passage presented to participants to read aloud. *Italicised* words are target words that occur and are in the word list as well. The <u>underlined</u> words are additional /s/ token words extracted and measured for the reading passage data.

> "A few years ago, Kendall's family moved into an old, *secluded castle surrounded* by empty fields far *outside* of town. The home definitely had <u>some</u> <u>history</u> to it, <u>something</u> Kendall would often wonder about while they lied awake with *insomnia*, <u>staring</u> at cracks in the *ceiling*. In between the fields near Kendall's new home, there was an abandoned *cemetery* where Kendall often went to bring these wonderings to life. Kendall was an *artist* who was known for their *abstract* creations and unique colour theory, like with their <u>first</u> piece of a *salmon* colored *sailboat* trapped in a *strawberry*-pink *galaxy*. Although Kendall's family *swore* that no *sane* <u>person</u> would like their art, Kendall <u>persisted</u>. The *cemetery* and art became all the *medicine*

Kendall needed as they continued making creations that eventually many people came to love."

7.3 Appendix C

Word list presented to participants to read aloud after the reading passage. *Italicised* words are ones that also occur with those targeted in the reading passage. At the time of study these words were not italicised.

1.	*ceiling*	16.	monsoon
2.	*abstract*	17.	ceremony
3.	settlement	18.	taxi
4.	assembly	19.	sauna
5.	*cemetery*	*20.*	*outside*
6.	*surrounded*	21.	stream
7.	*strawberry*	22.	pencil
8.	missile	23.	seedling
9.	saddle	*24.*	*medicine*
10.	*artist*	*25.*	*swore*
11.	*sailboat*	*26.*	*insomnia*
12.	*castle*	27.	seminar
13.	*salmon*	28.	motorcycle
14.	*galaxy*	29.	currency
15.	*sane*	*30.*	*secluded*

A Diachronic Study of Place and Manner Assimilation in Eastern Canadian Inuit Apical-Initial Consonant Clusters

Leah Palmer
University of Cambridge

Abstract. Heterogenous consonant clusters in eastern dialects of the Inuit language have undergone regressive assimilation, with different dialects assimilating different types of consonant clusters. Previous research has focused on constructing a sociological explanation for assimilation of consonant clusters within the twentieth century. The assimilation of apical-initial consonant clusters, such as the /tk/ in *aŋ**atk**uķ*, 'shaman' is thought to have occurred sometime within the nineteenth century. However, the loss of apiC clusters is generally under-studied, with no clear timeline of, or explanation for, the change. In this study I use seven Inuit vocabularies from the early eighteenth to late nineteenth century to establish a more precise timeline of when apical-initial consonant clusters were lost in eastern dialects of Inuit, and to explain why they were lost much earlier than other consonant cluster types. I conclude that apical-initial consonant clusters started to assimilate in the Nunatsiavut dialect of Inuit spoken in Labrador by the year 1800. This early change was driven by the inherent phonetic properties of apical-initial consonant clusters that make them more prone to regressive assimilation. In this study I also show that, when analysed critically with focus on their background and social context, historical vocabularies created by non-native speakers can be a useful tool for studying change in languages with no historical written tradition.

Keywords: historical phonology; sound change; Inuit; polysynthesis

---

1 Literature Review

This section will cover the existing literature on the modern Inuit language, as well as an introduction to historical change in the Inuit language. I follow the conventions of Dorais (2010) in using the word 'Inuit' to refer to both the language and the indigenous people who speak the Inuit language. I will be avoiding the word 'Eskimo' unless quoting from historical sources, as the word is considered outdated and offensive.

1.2 The Inuit Language and its Phonology Today

The Inuit language is spoken by around 100,000 indigenous Inuit in an area spanning from the north coast of Alaska, across northern Canada and into Greenland. It is part of the Eskimo branch of the Eskimo-Aleut family, and is related to the Aleut languages of the Aleutian Islands, and the Yupik languages of Alaska and Siberia. It is split into dialects which form a dialect continuum from east to west over the North American Arctic. These form four major dialect groups; Alaskan Inupiat, western Canadian Inuktun, eastern Canadian Inuktitut, and Greenlandic Kalaallisut. The North American dialect groups are illustrated on the map below. These are the dialects I will focus on in this study.

Figure 1: *Map of Inuit dialect groups, based on Dorais (2010).*

These major dialect groups of Inuit are generally agreed upon, but the number of individual dialects is more difficult to discern. I will follow Dorais's (2010) classification of Inuit dialects, as his classification allows easy identification of where consonant clusters are allowed. According to Dorais's classification, there are sixteen dialects of Inuit, arranged into the dialect groups in Table 1. For maps showing where each dialect is spoken, see Dorais (2010), pages 12, 33 and 37.

Table 1: *Classification of Inuit dialects into dialect groups according to Dorais (2010, p28)*

Dialect group	Alaskan Inupiat	Western Canadian Inuktun	Eastern Canadian Inuktitut	Greenlandic Kalaallisut
Dialects	Bering Strait Qawiaraq Malimiutun North Slope	Siglitun Inuinnaqtun Natsilingmiutut	Kivalliq Aivilik North Baffin South Baffin Nunavik Nunatsiavut	West Greenlandic East Greenlandic Thule

Each dialect has the same three vowels: /i/, /a/ and /u/, all of which can be short or long. The consonant inventory varies between dialects, with western dialects having up to nineteen consonants. The Labrador dialect has the eleven consonants below, adapted from Dorais (1990), which are also found in all other dialects, except Alaskan Inupiat, where /s/ is realised as /h/.

Table 2: Consonants of the *Labrador dialect, adapted from Dorais (1990).*

	bilabial	apical	palatal	lateral	velar	uvular
stops	p	t			k	q
continuants	v	s	j	l	ɣ	R
nasals	m	n			ŋ	

A general tendency of Inuit is that dialects in the west, especially those that are part of the Alaskan Inupiat dialect group, tend to be more conservative than dialects spoken in the east, in the Inuktitut dialect group. The difference in conservativity between dialects is reflected in how the different dialects treat consonant clusters. The three westernmost dialects of Inuit, Bering Strait, Qawiaraq, and Malimiutun, have very free cluster formation. In these dialects, almost any consonant can combine with any other consonant to form a cluster (see Dorais (2010, p 58)). Dialects spoken in central Canada may allow heterogenous consonant clusters, but only clusters with the same manner of articulation. In eastern dialects, fewer types of consonant clusters are allowed. In the easternmost Canadian dialect, the Nunatsiavut dialect spoken in Labrador, no heterogenous consonant clusters are allowed. This pattern can be seen by comparing cognates in the Alaskan and Labrador dialects, in Table 3.

Table 3: *Cross-dialectal comparison of consonant clusters in modern Inuit dialects.*

Type of cluster	English gloss	Alaskan Inupiat Dictionary	Labrador Inuttut Dictionary
uvular-initial (uvuC)	Thigh	ḳu**ḳt**ugaḳ	Ku**tt**ugak
velar-initial (velC)	Dream	sinna**kt**uk	Sinna**t**omak
bilabial-initial (bilC)	Story	uni**pk**aaq	Uni**kk**atuak
apical-initial (apiC)	Needle	mi**tḳ**un	Mi**kK**utik

From these examples it is evident that consonant clusters in western Inuit dialects are equivalent to geminates in eastern dialects. In this study I will focus on clusters made of two stop consonants, because clusters with different manners of articulation are only found in the three westernmost dialects, but stop-stop clusters can be found in all dialects except the Nunatsiavut dialect. Table 4 illustrates exactly which types of stop-stop clusters are allowed in which dialects, with the dialects presented in approximate order of most western (at the top of the table) to most eastern (at the bottom).

Table 4: *Overview of which clusters are allowed in which dialects of Inuit (adapted from Dorais, 2010 p63).*

	Dialect	apiC	bilC	velC	uvuC
Most Western	Bering Strait	yes	yes	yes	yes
	Qawiaraq	yes	yes	yes	yes
	Malimiutun	yes	yes	yes	yes
	North Slope	yes	yes	yes	yes
	Siglitun	yes	yes	yes	yes
	Inuinnaqtun	yes	yes	yes	yes
	Natsilingmiutut	yes	yes	yes	yes
	Kivalliq	yes	yes	yes	yes
	Aivilik	no	yes	yes	yes
	North Baffin	no	no	yes	yes
	South Baffin	no	no	no	yes
	Nunavik	no	no	no	yes
	Nunatsiavut	no	no	no	no
	Thule	no	no	yes	yes
	West Greenlandic	no	no	no	yes
Most Eastern	East Greenlandic	no	no	no	yes

As seen in Table 4, uvuC clusters are allowed in the most dialects, then velC, then bilC. The consonant cluster type allowed in the fewest dialects is apiC. Dialects below the line, shaded in grey, do NOT allow apiC clusters in their modern forms. Of the dialects that do allow apiC clusters, all of them allow the apical stop /t/ followed by any other stop: so /tk/, /tp/ and /tq/ clusters. These are the clusters that this study will focus on. Overall, a general tendency in modern Inuit is that dialects that are more eastern are more innovative and allow fewer types of consonant clusters, while western dialects are more conservative and allow more types of consonant clusters. ApiC clusters are only permitted in the eight westernmost dialects of Inuit.

1.3 Historical Phonology of the Inuit Language

1.3.1 Historical phonology and consonant clusters in the Inuit language

According to Dorais (2010), Proto-Inuit allowed all four types of heterogenous stop-stop consonant clusters: uvuC, velC, bilC and apiC. In the past, these clusters were used across all Inuit dialects. The modern state of Inuit, where some dialects disallow certain consonant clusters, can be explained through a historic process of assimilation in these dialects. This process is outlined in Bobaljik (1996).

Firstly, Bobaljik discusses manner assimilation. Both voice and nasality assimilate in clusters in all dialects of Inuit, except the most western Bering Strait, Qawiaraq and Malimiutun dialects. In all other dialects, which have manner assimilation, only clusters with the same manner of articulation are permitted. Bobaljik also discusses regressive place assimilation. The eight westernmost dialects of Inuit (see Table 3) have no regressive place assimilation. They allow apiC, bilC, velC and uvuC clusters. The remaining eight dialects all have coronal regressive place assimilation. This is a process that applies both synchronically to underlying apiC clusters and diachronically to historical apiC clusters. The place of the second consonant in the cluster (the C2) spreads back to the apical consonant (the C1), so the apical consonant assimilates to the second consonant. All dialects that have regressive place assimilation also have manner assimilation, so the underlying and historical clusters become geminates, with the place of the second consonant in the underlying cluster. This can be seen in the word pair of

mitḳun 'needle' in Alaskan Inupiat versus ***MikKutik*** in the Labrador dialect. ***MikKutik*** has undergone coronal regressive place assimilation, so the apical C1 has assimilated to the place of the velar C2. The Aivilik dialect only has coronal regressive place assimilation. The North Baffin and Thule dialects have both coronal and bilabial regressive place assimilation, meaning both underlying apiC and bilC clusters assimilate to the place of articulation of the second consonant in the cluster. In these dialects, words such as ***unipkaaq*** 'story' that have a bilC cluster in Alaskan Inupiat, are realised as ***Unikkatuak*** with an assimilated cluster. The South Baffin, Nunavik, and Greenlandic dialects have coronal, bilabial and velar regressive place assimilation, meaning they only preserve uvuC clusters. Finally, the Nunatsiavut dialect has total place assimilation, so this dialect has no heterogenous consonant clusters.

Proto-Inuit allowed all kinds of heterogenous consonant clusters, so regressive place assimilation is an innovative process in eastern Inuit dialects. Dorais (2010) established an approximate time course of when each type of consonant cluster assimilated. Assimilation processes found in more Inuit dialects happened earlier in history. UvuC, velC and bilC clusters were still present in eastern Inuit dialects in the year 1900. UvuC clusters have only assimilated in the Nunatsiavut dialect, and this change took place recently, within the last generation of speakers. VelC clusters, which have assimilated in four dialects, were still present in the speech of older speakers in the 1970s. BilC clusters underwent regressive place assimilation around the 1950s. Overall, uvuC, velC and bilC clusters assimilated in eastern Inuit dialects throughout the mid to late twentieth century. The subject of this study will be the assimilation of apiC clusters. This assimilation is the most widespread in Inuit, with half of Dorais's dialects having no apiC clusters. However, unlike all the other cluster types, there is no clear timeline for the assimilation of apiC clusters. Dorais writes that they had disappeared by the mid-nineteenth century, but does not provide any more precise details on when they were lost in different dialects. This could be because there has been less examination of historical sources from before 1900 in the literature. In section 3, I will attempt to use historical sources to establish a more precise timeline for the loss of apiC consonant clusters, including when they were lost in which dialects.

It is possible to see the effects of regressive place assimilation as it progressed through Inuit dialects using historical sources recording the Inuit language. Until Europeans arrived in the Arctic, the Inuit language was maintained through oral tradition and was not written down. From the sixteenth to the nineteenth century, European colonialists, including explorers, traders, missionaries, encountered the need to communicate with the Inuit people, which led to wordlists and vocabularies of Inuit being written by outsiders to the language over a period of around 300 years. In section 3, I will also examine how useful historical sources written by outsiders to a language can be to studying the history of that language.

1.3.2 Explanations for regressive consonant cluster assimilation in the Inuit language

If Dorais is correct and apiC clusters had assimilated fully by the mid-nineteenth century, this would mean that regressive place assimilation in apiC clusters happened almost a full century before assimilation in other types of consonant clusters. Currently, no one has offered an explanation for why apiC clusters assimilated first, or why they assimilated so much earlier than other cluster types. There are various explanations proposed for why eastern Inuit dialects have lost consonant clusters, although it is rarer to find discussion on apiC consonant clusters specifically. Creider (1981) proposes that the loss of consonant clusters was innovated in Greenland, and in subsequent years diffused westward from Greenland into Labrador, then into more western dialects. This would align with the observation that consonant clusters were lost earlier, and more extensively, in eastern dialects than western ones. Other explanations rely on the general tendency for areas with more regressive assimilation in consonant clusters to have had longer contact with Europeans. European missionaries have been settling in Greenland and Labrador since the eighteenth century. Quebec has had a continuous presence of

Europeans since the mid-nineteenth century, and Inuit living on and around Baffin Island have had regular contact with whalers since the late nineteenth century. Contact with Europeans led to more Inuit speakers being bilingual, which could have led to phonological change in the Inuit language. For instance, Thomason and Kaufman (1988) wrote that bilingualism could cause structural borrowings into the substrate language. This could explain the spread in regressive assimilation, as areas with longer contact with Europeans had more bilingual speakers at the time of regressive assimilation.

Dorais (1985, 2010) suggests a sociolinguistic approach for general consonant cluster assimilation, however nothing specific for apiC clusters. He proposes that the system with more geminates and fewer heterogenous clusters was already present in female and child language at the time of European settlement. According to Dorais, there is some evidence for women and children using a different consonant system in historical written sources, and Nunavik elders have reported women and children using a different speech system in the past. Both pronunciations, with heterogenous clusters and geminates, coexisted in pre-European contact society. When Europeans arrived, life in Inuit society shifted to become more camp-based than nomadic, basing the society around camps where the women and children previously lived. Because of this shift in societal structure, the speech of women and children became more prestigious, and spread through the whole population. Dorais also offers an explanation for why this shift did not occur in Alaska; that is because English was the language of the new camps in Alaska, so Alaskan Inupiat was used only by those groups who remained nomadic, and it preserved more of its traditional pronunciation.

In summary, in the Inuit language, more conservative dialects in the west allow more heterogenous consonant clusters, and more innovative eastern dialects have more regressive assimilation. This current state of the language reflects a historical change towards regressive assimilation, which first affected apiC clusters in half of all Inuit dialects, then bilC clusters, then velC clusters, and most recently uvuC clusters in the Nunatsiavut dialect. Past authors have offered different explanations for the loss of consonant clusters in Inuit, including explanations rooted in bilingualism and social change. However, no author has proposed a specific explanation for why apiC clusters assimilated first, or why they assimilated long before other cluster types. In section 4 of this study, I will attempt to find an explanation as to explain why apiC clusters assimilated long before other cluster types in Inuit.

1.4 Research Questions

The three research questions that I aim to answer in this study are:

In Section 3:

Research Question 1: What was the exact time course of assimilation of apiC clusters in eastern Inuit dialects?

Research Question 2: To what extent can we use external historical sources to discover this?

In Section 4:

Research Question 3: Why did apiC clusters assimilate first, and long before other consonant clusters in Inuit?

2. Methodology

2.1 Sources Used in the Study

The assimilation of apiC clusters can be observed using historical sources of Inuit. For this study, I collected words with historical apiC clusters from seven sources on the Inuit language, dating from 1744 to 1893. Words from each source will be presented in tables. Information on exactly which words are included from which sources in their tables can be found in Table 4. To eliminate the possibility of transcription errors, I ensured that when a word was used from a historical source, an equivalent word could be found in a modern dictionary.

A full list of sources used for this study is given below. Grey cells indicate sources that have not been discussed before in any linguistic study of Inuit. To my knowledge, none of the sources used in this study appear in any previous studies of consonant cluster assimilation in Inuit.

Table 5: *Sources used in the study*

Date	Source name	Accessed via	Dialects in the source	What is presented from the source in this paper
1744	Dobbs: 'a Vocabulary of English and Eskima Words'	Online archive. Appendix to Dobbs, A. 1744. *An Account of the Countries Adjoining to Hudson's bay.* London: J. Robinson	Nunavik, Aivilik, Kivalliq and South Baffin	All words containing apiC clusters
1771	William Richardson: 'The Complete Inuttitut Vocabulary'	Full reprint in Stopp (2014).	Nunatsiavut	All words containing any consonant cluster
1833-1835	George Back: 'Notebook of Esquimaux Vocabulary'	I found and consulted this source in the Scott Polar Research Institute (SPRI) archive, Cambridge.	Natsilingmiutut	All words containing apiC clusters
1833	John Ross: 'Vocabulary of the English, Danish and Esquimeaux languages'	Online archive. Part of Ross, John. 1835. *Appendix to the Narrative of a Second Voyage in Search of a North-West Passage*. London: A.W Webster.	North and South Baffin, Natsilingmiutut	All words containing apiC clusters

1850	John Washington: 'Esquimeaux and English Vocabulary'	Online reprint by Forgotten Books, London	Split into three sections. Western dialects: Malimiutun Central dialects: Aivilik Eastern dialects: Nunatsiavut	All words containing apiC clusters
1851	Edward Adams: Vocabulary notebook	I found and consulted this in the SPRI archive.	Qawiaraq	All words containing apiC clusters, and a selection of other words to illustrate the discussion
1893	James Allan: Inuit wordlist, part of his journal	I found and consulted this in the SPRI archive	Locations mentioned in the journal are Carey Island and St John's, Newfoundland. Probably South Baffin, Labrador and Nunavik.	All words containing consonant clusters.

First, I created the maps in Figures 2 to 4 to provide a clearer picture of the approximate locations of each of the historical sources under investigation. I created these using a blank map of Canada sourced from mapchart.net and the Krita drawing software.

Figure 2: *Map indicating the approximate location of data collection sites of sources from western dialects.*

Figure 3: *Map indicating the approximate location of data collection sites of sources from eastern dialects.*

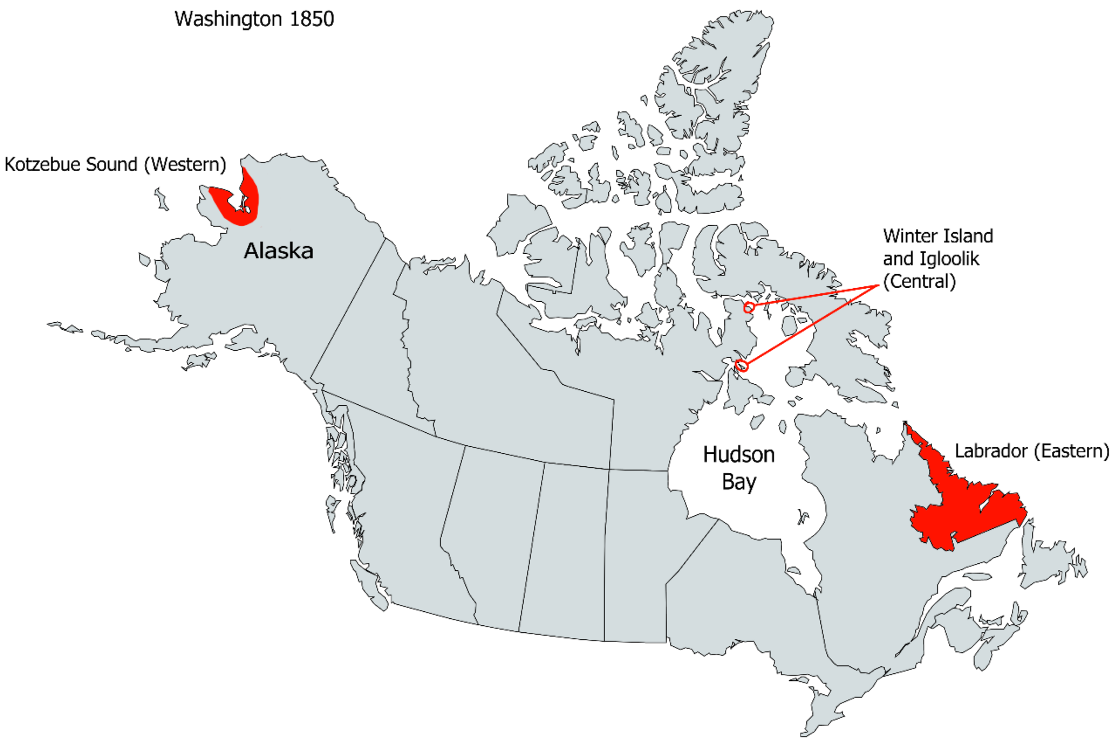


Figure 4: *Map indicating the approximate location of data collection sites from Washington (1850).*

2.2 Critical Evaluation of Sources

It is important to note that historical sources are not always reliable, since they were written by outsiders to the Inuit culture who were not native speakers of the language, who probably held certain prejudices about the language of indigenous peoples, and who had no formal linguistic training. However, as the Inuit language has no historical written tradition, these external sources are the only sources available. It is important to address the context of every historical source, to ensure that any conclusions made based on them reflect the actual state of the Inuit language. For each source used in this paper, I will address who wrote it down, and what their experience was with Inuit culture. I will attempt to establish where they sourced words from, and if they worked with a native speaker when compiling the word list.

To compare words from historical sources with their modern equivalents, and to ensure I use accurate transcriptions, I used two modern dictionaries of Inuit. These are the Labrador Inuttut Dictionary (LID), for modern usage in the eastern Labrador dialect, and the Alaskan Inupiat Dictionary (AID), for modern usage in western Alaskan dialects. The LID is an open-access online dictionary, with 4577 entries, compiled from multiple sources of the modern Nunatsiavut dialect, spoken in Labrador. It has been reviewed by native speakers of the Labrador dialect. As this dictionary is compiled from other, smaller sources, there are some inconsistencies in orthography. Most notably, it is stated on the LID webpage that a <K> is used to indicate a uvular sound. This would suggest that words like *SekKuk* 'knee' and *MikKutik* 'needle' contain a /kq/ cluster. However, other sources on the Labrador dialect, including Fortescue (1994), record uvular geminates where the LID writes <kK>, such as *siiqquq* for knee and *miqquti* for needle. Therefore, it is probably the case that in the LID <kK> represents a uvular geminate. The LID is still preferable for reference for this study, as it contains more entries and is presented in a more accessible format than Fortescue (1994). The Alaskan Inupiat Dictionary (AID) dates from 1970, and was collected from native speakers living in villages across Alaska. In the AID, <ḳ> is used for a uvular.

2.3 Methods in the study

The first aim of this study was to establish a precise time-course for the loss of apiC consonant clusters in eastern Inuit dialects. To achieve this, I searched for words that were common to multiple sources; that is, words that were recorded in both western and eastern dialects, and in the same dialect over time. Preliminary data from Dorais (2010) established that apiC clusters were certainly present in historical eastern dialects of Inuit in the sixteenth century. I first looked at sources from western dialects, including Back's notebook, Adams' wordlist, and Washington's western dialects. I found words that, in these sources, contained an apiC consonant cluster. If a word was only recorded in one source and was not found in modern dictionaries it was eliminated from the study in order to eliminate the possibility of transcription errors. In the second stage of the investigation, I established when the apiC clusters in these words assimilated in eastern dialects. I achieved this by looking at sources for eastern dialects over time and seeing when words that were previously recorded as having an apiC cluster, or had an apiC cluster in the sources on western dialects, started to be written with no apical. From this data, I drew more precise conclusions on the time course of the loss of apiC consonant clusters than have been made before in the literature.

3 Research Questions 1 and 2

3.1 Data Analysis

Data from historical sources is presented in tables, with the English gloss taken from the historical source. These tables also include the equivalent form of the word from modern dictionaries, for comparison. All orthography, capitalisations, diacritics, and hyphens in the historical data are taken directly from how the word was written in the historical source. Bold letters highlight apiC clusters. If the equivalent word in the modern dictionary column is greyed out, this indicates that the word appears to have a different etymology than the word in the historical source. Occasionally, an alternative English gloss is given in brackets, taken from a modern dictionary. Where a 'type of CC' is indicated, these have been taken from the AID, and may not reflect what was written in the historical source.

In section 3.1.1, I examine two sources from western dialects that maintain apiC consonant clusters to the present day. Examining these sources will demonstrate which words in Inuit historically contained apiC clusters, so it will be possible to examine these words in eastern dialects, which may or may not contain apiC clusters depending on what time they are from. In section 3.1.2 I examine, in chronological order, four sources on eastern dialects, to establish when apiC clusters were lost from these dialects.

3.1.1 Sources on Western Dialects

George Back's 'Notebook of Esquimaux vocabulary', 1833-1835, Scott Polar Research Institute archive.
George Back was an English naval officer from Stockport, who compiled an extensive notebook of Inuit vocabulary from 1833-1835 while exploring the Back River in western Canada. He was familiar with French and English. The notebook contains both vocabulary and Back's observations on Inuit grammar. Communication with the local people was important to Back, as the area he was exploring had never been accessed by Europeans, but was well known by the local Inuit. Using the map from Back's 1835 account of his land voyage, we can establish that he was likely communicating with speakers of the modern day Natsilingmiutut dialect, which still has apiC clusters today. Therefore, I expected to find apiC clusters in Back's notebook, which I did find. Back records twelve instances of apiC clusters, seen in Table 6, nearly all of which align with words from the AID where the apiC cluster is still visible today. Furthermore, Back's apiC clusters align with geminates from the LID, showing how apiC clusters in western dialects assimilate in modern eastern dialects.

There are some words of interest in Back's notebook. Firstly, 'Finger a' is written by Back as *Iik kiek.* This likely corresponds to 'little finger', which is written in the AID as *iki**tk**uraḳ*, with an apiC cluster. Back's geminate consonant in *Iik kiek* is probably either a transcription or perception error, as no apiC clusters are subject to regressive assimilation in the Natsilingmiutut dialect. Secondly, Back records an apiC cluster in the word ***mi**t-k**oo-shā,*** 'asbestos'. *Mi**t-k**oo-shā* may be related to the word for feather (*mi**tḳ**uḳ*, AID), due to the appearance and texture of asbestos. If so, it is a valid occurrence of an apiC cluster. Finally, Back's *ki**tk**ŏă,* 'weed tangle', is not in the AID, but may be recorded in its assimilated form in the LID as 'KikKuaK', seaweed.

Table 6: *Every word containing an apiC cluster from Back's Notebook of Esquimaux Vocabulary (1833-1835).*

English Gloss (from Back's notebook)	Back (1833-1835)	Alaskan Inupiat Dictionary	Labrador Inuttut Dictionary
Knee	si**t k**ŏ ă	si**tḳ**uḳ (kneecap)	SekKuk
Pot for cooking Pot stone	oo**t-k**oo-sĕek oo**t-k**oo seek seak	u**tk**isiḳ	Ukkusik (cooking pot)
Needle	Mi**tk**ote	mi**tḳ**un	MikKutik (sewing needle)
Finger a	Iik kiek	iki**tk**uraḳ (little finger)	IKikKuk (little finger)
Conjuror	Ang e**t k**ook	aŋa**tk**uḳ (shaman)	Angakkuk (someone with spiritual powers)
Beard	oo-mi**tk**ee	umŋich	Umik
Heavy it is	Oko-mā i**t p**oke	uḳumai**tch**uḳ	UKumaittuk
Dive he does	A**tk**ă moke	nakkaḳtuḳ a**tḳ**aḳtuḳ (descends)	AkKak (dive)
Asbestos	mi**t-k**oo-shā	mi**tḳ**uḳ (feather)	Ikisuittuk
Weed tangle	kitkŏă		KikKuak (seaweed)
Place Names			
Waye/Naye river	Oo**t-k**oo-sēĕk-sălĭk		
Repulse Bay	seā**tc**oke		

Overall, Back's notebook is a notable resource, which, due to remaining in a private collection until recently, has not been used before in any studies of Inuit. It gives evidence that the modern apiC clusters in western dialects were also present historically. Most importantly, it establishes which words certainly contained apiC clusters in historical western dialects, so that these same words may be examined in eastern dialects to see when the clusters assimilate.

Edward Adams's Vocabulary Notebook, Norton Sound, Alaska, 1851, Scott Polar Research Institute archive.

Adams's notebook is a 25-page long handwritten English-Inuit vocabulary, compiled during Adams's stay in Norton Sound, Alaska, in 1851. Adams was born in Suffolk and was in Alaska as part of a search for John Franklin and his crew who had gone missing in the Arctic in 1848. His notebook contains the wordlist as well as a written text with Adams's observations on Inuit grammar. In the notebook, Adams writes 'this is the dialect spoken by that tribe of Esquimaux inhabiting the eastern coast of the Sound', which corresponds to the modern day Qawiaraq dialect. The Qawiaraq dialect, as well as other dialects of Alaskan Inupiat, maintains apiC clusters in the modern day, so I expected to see apiC clusters in Adams's notebook.

Table 7: *A sample of words containing consonant clusters from Edward Adams's (1851) vocabulary notebook from Norton Sound, Alaska*

Type of CC	English Gloss	Edward Adams (1851)	Alaskan Inupiat Dictionary	Labrador Inuttut Dictionary
apiC	Knee	chu**s.k**uk	si**tḳ**uḳ	SekKuk
apiC	Head	Nu**sk**ok	nutchat (head hair)	NiaKuk
apiC	Needle	Mïng.a.kun	mi**tḳ**un	MikKutik
apiC	Finger (4th)	I.kush.li.kuk	iki**tk**uraḳ	IKikKuk
velC	Lie down	I.nòk.tüne	nallaḳtuḳ	Inillak
velC	Looking glass	Tek.ok.tüne	tautuktuḳ (see)	Tautuk
velC	Lose a	Tem.màk.tok		
bilC	Beard	Oom.gike	umŋich	Umik
	Nose	Inuk	Inuk (human being)	

In Table 7, it appears as though Adams uses quite unusual orthography, and may not be as accurate a transcriber as Back. He uses many diacritics, but it is unclear what they represent, and they are not used systematically. It is often difficult to corroborate Adams's word forms with word forms in modern dictionaries. Additionally, he makes some interpretation errors, such as his word for 'nose', which is written as *inuk.* In fact, this likely corresponds to 'human being', and may have been a result of his informant pointing to their nose, and Adams took this to mean the nose while his informant took it to mean the person.

Regarding consonant clusters, Adams uses inconsistent orthography. He records velar initial clusters in *I.nòk.tüne* 'lie down', *Tek.ok.tüne* 'looking glass', and *Tem.màk.tok* 'loose a'. *Tem.màk.tok* does not appear in the modern dictionaries, but appears to be a velar-final word with the suffix -tok added to the end. Adams also writes *Oom.gike* for 'beard', which contains a bilC cluster. Therefore, it does appear as though Adams records some velC and bilC clusters. Adam's record of apiC clusters is more inconsistent. His words *Mïng.a.kun*, 'needle' and *I.kush.li.kuk* 'little finger' do not contain their apiC clusters. On the other hand, *chu**s.k**uk* 'knee' and *nu**sk**ok* 'head hair' do contain apiC clusters, although the apical is transcribed as a fricative. This may be because /sk/ clusters are legal in Adam's native English, but /tk/ clusters are not, so he substituted the /t/ for the fricative /s/ due to influence from his own phonological system.

Adams's notebook serves as a reminder that historical sources cannot always be taken at face value. Even though Adams does not write stop-stop apiC clusters in his vocabulary, when the wider context of the notebook is taken into account, we see that Adams uses irregular orthography, writes words that are not found in other sources and occasionally makes unrelated transcription errors. Also, he does transcribe some apiC clusters with an apical fricative. Therefore, it remains very likely that stop-stop apiC clusters were present in 1850s Qawiaraq, but Adams did not transcribe them systematically.

3.1.2 Sources on Eastern Dialects

Arthur Dobbs's 'Eskima and English Vocabulary', 1744.

Dobbs was a British colonial official, born in Scotland, who spent time in Ireland and North Carolina. 'Eskima and English Vocabulary' is part of the appendices to Dobbs's 1744 account of his voyage across Hudson Bay, Canada. He compiled wordlists to communicate with local people and to trade, including this Inuit wordlist and the first known wordlist of a Dené language. The map in Dobbs's book shows that he travelled around both the east and west coasts of Hudson Bay. Therefore, the words in the 'English and Eskima Vocabulary' could have come from speakers of any of the Inuit dialects spoken in the region of Hudson Bay; the South Baffin, Nunavik, Aivilik and Kivalliq dialects. Of these, only the Kivalliq dialect still has apiC clusters today. Even though it was mostly compiled from eastern dialects that since lost apiC clusters, as Dobbs's vocabulary was compiled so early, we would expect to see apiC clusters within it.

Table 8: *Every word containing an apiC cluster from Dobbs (1744) 'Eskima and English Vocabulary'*

English Gloss	Dobbs (1744)	Alaskan Inupiat Dictionary	Labrador Inuttut Dictionary
Little finger	Lick-i**t-c**ock	iki**tk**uraḳ	iKikKuk
Knee	Sea**t-c**oke	si**tḳ**uḳ	sikkuk
Moon	Ta**t-c**ock	ta**tḳ**ik	takKik
Needle	Mi**d-c**oot	mi**tḳ**un	mikKutik
Whalebone	Shoo**t-c**ock	suḳḳaḳ (baleen)	sukKak

In Table 8, we can see that words in Dobbs's 'English and Eskima Vocabulary' do contain apiC clusters. From this, we can conclude that apiC clusters were lost in eastern dialects of Inuit sometime after 1744. The apiC cluster in *Shoot-cock,* 'whalebone' from Dobbs's vocabulary does not appear in the AID. This may be a case of transcription error, where the <t> was inserted, or a case of an apiC cluster that was lost even in modern western dialects. For more discussion of this, see section 3.1.4. Even if *shoot-cock* is a transcription error, there are enough apiC clusters in Dobbs's vocabulary to conclude that apiC clusters were certainly present in eastern dialects around Hudson Bay in the 1740s.

William Richardson's Inuttitut Vocabulary, from his journal, 177.

Richardson was a British naval officer who travelled along the coast of Labrador in 1771, where he collected a wordlist of the Nunatsiavut dialect in his journal. This wordlist was rediscovered and fully reprinted in a paper by Stopp (2014). According to Stopp, many of the words in Richardson's wordlist have equivalents in other Inuit lexicons, so he did achieve a good level of accuracy in his transcriptions. The purpose of Richardson's wordlist was probably to help establish positive relationships needed for trade with the local Inuit.

Richardson's wordlist contains thirteen words with a historical consonant cluster, seen in Table 8: four with uvuC clusters, one bilC cluster, five velC clusters and three apiC clusters. I excluded *Eccooma,* 'fire', from the following analysis as it contains a stop-nasal cluster in the AID, and it is unclear whether nasal manner assimilation had occurred in the Labrador dialect at this time. One of the velC clusters, in *amictuc* 'kettle', is not found in the AID, but appears to be a velar-final noun with the suffix -tok added.

Table 9: *Every word containing a consonant cluster from Richardson's (1771) Labrador Vocabulary*

Category of CC	**English Gloss**	**Richardson (1771)**	**Alaskan Inupiat Dictionary**	**Labrador Inuttut Dictionary**
uvuC	arrow	Cackchuc	ḳagruḳ	Katjuk
uvuC	otter	Pomiuctoo	pamiuḳtuuḳ	Pamiuttok
uvuC	whale	Auckbook	agviḳ	Apvik
uvuC	thigh	cookloak	ḳuḳtugaḳ	Kuttugak
uvuC	seal (bearded)	ogriuk	ugruk	Utjuk
uvuC	breeches	caligngai	ḳaḳliiḳ	
bilC	A European	Cobluanet	kavlut (eyebrow-origin of word for European)	Kallunak
velC	kettle	amictuc		
velC	fire	Eccooma	ikniḳ	Ikumak
velC	bow	petaioshic	pisiksi	Pititsik
velC	sword	Saviksoak	savik (knife)	Savitsuk
apiC	knee	chiookuck	si**tḳ**uḳ	SekKuk
apiC	needles	Muckett	mi**tḳ**un	MikKutik
apiC	moon	Tackuc	ta**tḳ**uk	TakKik

In general, Richardson is quite accurate at transcribing uvuC, bilC and velC consonant clusters. Excluding *Eccooma,* Richardson records at least eight out of ten possible uvuC, bilC and velC clusters. This could possibly be nine out of ten, if the <sh> in *petaioshic* is an attempt at recording a /ks/ cluster. The only word where Richardson certainly does not record a consonant cluster is *caligngai*, 'breeches'. The AID records *ḳaḳliiḳ* with a uvuC cluster where Richardson just writes <l>. Still, recording eight or nine out of ten possible clusters makes Richardson quite an accurate transcriber. It is surprising then, that out of the three words in his wordlist that contain a historical apiC cluster, Richardson never transcribes them with any evidence of an apical consonant. The historical clusters in his words for knee, needles and moon are all /tq/ clusters, but for each one Richardson only writes a <k> or <ck>. As Richardson is usually quite accurate at recording consonant clusters, this may be more significant than an ordinary transcription error. It could indicate that apiC clusters were starting to fall out of use in Labrador Inuit as early as the 1770s. An alternative explanation is that apiC clusters were still used in Labrador Inuit at this time, but were less salient than other types of consonant clusters, so were difficult to perceive for non-speakers or learners of the language like Richardson. If this is the case, the change towards regressive assimilation in apiC clusters probably happened with the next generation of Labrador Inuit speakers, who may, like Richardson, have found the apiC clusters almost inaudible. In either case, it shows that the change towards regressive assimilation in Labrador Inuit apiC clusters was almost certainly underway by 1800, and possibly slightly earlier.

John Ross's 'Vocabulary of the English, Danish and Esquimaux Languages', 1833.
John Ross was a Scottish naval officer who travelled to the Arctic in the 1830s, where he compiled this wordlist, published in the appendices to his book 'Narrative of a Second Voyage in search of a North-West Passage'. In the preface to the vocabulary, he writes 'this Vocabulary […] may be found useful to those who navigate Davis's straits and Baffin's bay'. From this, it is possible to infer that the words were probably collected from speakers of the North and South Baffin dialects, both of which lost apiC clusters. Also, it possibly contains words from the Natsilingmiutut dialect, which maintained apiC clusters, as Ross encountered speakers of Natsilingmiutut during his stay in Felix Harbour in early 1832.

It is also possible that Ross borrowed some words in his vocabulary from earlier published dictionaries, because in the preface he mentions using spelling conventions from a Danish-Inuit dictionary of Greenlandic, published in 1804. The purpose of Ross's vocabulary is linked with the purpose of the voyage; to communicate with local people in order to find a North-West passage.

The data from Ross (1833) suggests that, in the area where he collected his vocabulary, many apiC clusters had been lost by the early 1830s. This can be seen in the words in Table 9, all of which had lost their apiC cluster by the time that Ross recorded them. However, this is not a complete pattern. At least some words in his wordlist maintain evidence of an apiC cluster, such as *iterpok*, 'awake', where the <t> is evidence of a /tq/ cluster, still seen today in the AID *itḳumaruḳ.* Also, five words in Table 10 contain a historical apiC cluster that Ross writes as <rk>. I believe <r> is used by Ross in this context to transcribe a /t/, possibly because /r/ and /t/ are both apical sounds, so <rk> in these words represents a /tq/ cluster. It is not possible to say for certain that Ross's <r> represents a /t/ here, however it is true that there is a difference between the apiC clusters in Table 9, that Ross writes with just their second consonant so were probably assimilated, and the apiC clusters in Table 10 that Ross writes <rk>. This may be because some apiC clusters in Ross's wordlist are assimilated, and others are not, and he records these unassimilated clusters as <rk>.

Table 10: *Every word containing an assimilated apiC cluster from Ross (1833) 'Vocabulary of the English, Danish and Esquimaux Languages'*

English Gloss	Ross (1833)	Alaskan Inupiat Dictionary	Labrador Inuttut Dictionary
Conjuror or sorcerer	Angekok	aŋa**tk**uḳ (shaman)	Angakkuk (someone with spiritual powers)
Mittens (a pair)	Aketik	aa**tḳ**atik	akKikavak
Dive	Akkarpok	nakkaḳtuḳ a**tḳ**aḳtuḳ (descends)	AkKak
Decree, law	Pekkorsut	pi**tḳ**un	PikKujik (command)

Table 11: *Every word containing an unassimilated apiC cluster from Ross (1833).*

English Gloss	Ross (1833)	Alaskan Inupiat Dictionary	Labrador Inuttut Dictionary
Awake	I**t**erpok	i**tḳ**umaruḳ	IkKumak
Feather	Me**rk**ok	mi**tḳ**uḳ	MikKuk
Knee	Se**rk**ok	si**tḳ**uḳ	sikkuk
Hair (of animal)	Me**rk**ut	mi**tḳ**ut	MikKuk (human body hair)
Needle	me**rk**ut	mi**tḳ**un	mikkutik
Remember	E**rk**aïok	i**tḳ**aḳtuḳ	Ikkaumak

This mixed result, where some words maintain an apiC cluster and others do not, has two possible explanations. Firstly, Ross may have been speaking to speakers of one dialect as the change towards regressive assimilation was happening in that dialect. Some speakers, especially younger speakers, would use assimilated forms, and other speakers would use apiC clusters. If this is the case, it was probably with speakers of the South Baffin dialect, as the North Baffin dialect maintained clusters until later in the nineteenth century (seen in Washington's wordlist) and the Natsilingmiutut dialect still has apiC clusters today. Secondly, Ross may have been communicating with speakers of multiple dialects, some of which (the North Baffin and Natsilingmiutut dialects) still used apiC clusters, and others (the South Baffin dialect) did not. If he then compiled words from multiple dialects into the same vocabulary, it would result in a wordlist where some apiC clusters are still present, and others are not. In either scenario, the words in Table 10 where a historical apiC cluster is assimilated in Ross's vocabulary suggest that the South Baffin dialect had either fully lost or was starting to lose apiC clusters by the early 1830s.

Finally, seen in Table 12, Ross records eight words with <dl> clusters. Four are recorded as <vl> in the AID (section A, Table 12) and four are recorded as <l> in the AID (section B, Table 11). As these apiC clusters are not seen in modern western dialects, they could be a result of perception error by Ross. Alternatively, the words in Table 12 section B could be evidence of apiC clusters that have assimilated in even western dialects. This possibility is discussed in section 3.1.4.

Table 12: *Every cluster transcribed by Ross as <dl>.*

	English Gloss	**Ross (1833)**	**AID**
Section A: Clusters recorded as <dl> by Ross that the AID records as <vl>	Day	u**dl**ok	u**vl**uḳ
	Morning	u**dl**ak	u**vl**aaḳ
	Star	u**dl**oriak	u**vl**ugiaḳ
	Thumb	ku**dl**o	ku**vl**u
Section B: Clusters recorded as <dl> by Ross that the AID only records as <l>.	Unfortunate Unlucky	Pi**dl**uangitsok Pi**dl**uejtsok	pi**l**uktuḳ (is bad)
	Lamp	ko**dl**ek	ḳu**ll**iḳ
	Rainy	sie**dl**einarpok	sia**l**uktuḳ

John Washington's 'Eskimaux and English Vocabulary', 1850.

The most extensive vocabulary list in this study, Washington's 'Eskimaux and English Vocabulary' is a comparative dictionary of Inuit words in three dialects. The dialects in Washington's list are the eastern Labrador dialect (equivalent to modern day Nunatsiavut), the central dialects spoken on Winter Island and Igloolik (the modern day Aivilik and North Baffin dialects respectively), and the western dialect spoken in Kotzebue Sound (the modern day Malimiutun dialect). This dictionary was created to help English speakers communicate with Inuit to find information on the missing Franklin expedition. Washington spent much time in the Arctic and worked with a bilingual Inuit guide, Qalaherhuaq, to compile his vocabulary. A combination of the need to communicate with local people, and Washington and Qalaherhuaq's combined experience with the Inuit language, makes this an important, extensive, and accurate vocabulary.

Table 13: *Every word containing an apiC cluster from Washington (1850) Esquimaux and English Vocabulary (not formed by addition of a suffix)*

English Gloss	Washington (1850) Western Dialects	Alaskan Inupiat Dictionary	Washington (1850) Central Dialects	Washington (1850) Eastern Dialects	Labrador Inuttut Dictionary
Little finger	ikik-kok, eri**t**-**k**a-mak	iki**tk**uraḳ	erki**tk**oa	irkekok	iKikKuk
Kettle, copper		u**tk**isiḳ	U**t**-**k**u-sik (pot, earthen, native)	Uk-kusik	Ukkusik (cooking pot)
Conjuror, or sorcerer		aŋa**tk**uḳ (shaman)	Ang-e**t**-**k**uk	Ang-a-kok	Angakkuk (someone with spiritual powers)
Moon	ta**d**-**k**ak	ta**tḳ**ik	Ta**t**-**k**ak	Tak-kek	takKik
Gums		i**tḳ**iḳ	**It-kit-ka**	Kil-lak (probably 'palate')	ikkik
Knee	knab-lut	si**tḳ**uḳ	Si**t**-**k**o-a	Ser-ko	sikkuk
Asbestos		mi**tḳ**un (needle)	Mi**t**-**k**u-sha	u-ya-rak kau-ma-yok	miKkutik (needle)
Dive, he does		nakkaḳtuḳ a**tḳ**aḳtuḳ (descends)	**atk**a-mok	ak-ar-pok	AkKak (dive)
Tern, Arctic	mi**tk**uta-ila-k	mi**tḳ**utailaḳ	i-mi**t** **k**o-teil-lak		ImikKutailak
Fur	mi**t**-**k**ut	mi**tḳ**uk		aming-it merku-lingnik	
Needle	mik-kun (pin: me**t**-**k**un)	mi**tḳ**un	mi**t**-**k**ot	mer-kut	mikKutik
Trim a lamp, she does			ta**t**-**k**i-uk-pok	takkersorpok, tak-kut	
Weed, sea (tangle)			ki**t**-**k**o-a	ker-ko-yet	KikKuak

Table 14: *Words from Washington (1850) that appear to show an apiC cluster formed by a suffix*

English Gloss	Western Dialects	Alaskan Inupiat Dictionary	Central Dialects	Eastern Dialects	Labrador Inuttut Dictionary
Arrived, he is		tikittuḳ, tikitchuḳ	tik-ki**t-p**ok	tik-ki-lauk-pok	tikik
Clench his fist, he does		iḳit (fist)	ir-ki**t-p**ok	aggang-minik erking-awok	
Eyebrow	kablu-au, kablu-o**tk**a	kavlut	kablu**t-k**i	kab-lo	
Frost-bitten, it is		ḳiḳḳitigaa	ker-ki**t-p**ok	itse-mut piyau-wok	
Teeth	ki-u-tik, kau-ti**t-k**a	kigutit	ki-u-ti**t-k**a	kigutit	kigutik (tooth)
Heavy it is		uḳumaitchuḳ	oko-ma-i**t-p**ok	oko-mai-pok	ukumaittuk

From Table 13 it is evident that words in Washington's western and central dialects contain apiC clusters, but their equivalents in eastern dialects no longer contain apiC clusters by 1850. Also, Washington records six words that contain apiC clusters formed by a t-final root with the addition of a suffix. These can be found in Table 14. Although these exact forms are not found in the AID, they appear to be formed by a valid grammatical process, and these words fit the same pattern as other apiC clusters, where eastern dialects are assimilated but western and central are not.

Washington's transcriptions *ser-ko,* 'knee', *mer-ku,* 'needle' and *merkut* 'fur' are written with <rk> in eastern dialects where there was historically a /tk/ cluster. Unlike Ross, I do not believe Washington uses <rk> systematically to represent a /tk/ cluster, because there is evidence from Richardson's wordlist that the Labrador dialect had already lost apiC clusters by this time. Additionally, Washington uses <tk> for apiC clusters in central and western dialects, so he would use <tk>, not <rk>, if apiC clusters were also preserved in eastern dialects. Washington also records only three out of fifteen words in Eastern dialects with <rk>, so his use of <rk> is occasional, and may not reflect a general trend in the language. Instead, Washington's uses of <rk> could be taken directly from Ross's wordlist, where 'knee', 'needle', and 'fur' are written with an <rk> cluster. 'Needle' in western dialects, *mik-kun,* also appears to go against the general pattern, as it has no apiC cluster. However, other similar words such as *met-kun* 'pin' maintain their apiC cluster in western dialects, so this was probably a transcription error. The majority of words with a historical apiC cluster in eastern dialects have lost that apiC cluster in Washington's vocabulary. Therefore, it is possible to conclude that apiC clusters had disappeared from the Labrador dialect by 1850, which aligns with the earlier findings from Richardson suggesting they had in fact started to assimilate by 1800.

Additionally, there is a pattern in Washington's vocabulary that is unexpected given Dorais' assumption that apiC clusters had disappeared in all eastern dialects by the mid-nineteenth century. Washington's central dialects, spoken on Winter Island and Igloolik, maintain apiC clusters in this wordlist. This is unexpected since the dialect spoken near Winter Island today is the Aivilik dialect, and the dialect spoken in Igloolik is the North Baffin dialect, both of which no longer have apiC clusters. From this data it must be concluded that not all apiC clusters had assimilated by the mid-nineteenth century. In fact, apiC clusters in the Aivilik and North Baffin dialects were maintained into the second half of the nineteenth century.

Finally, in Table 15, there are some words in Washington's wordlist that contain an apiC cluster, but their equivalents in the AID only contain the assimilated version of this cluster. It is not possible to conclude if these are the result of a perception error by Washington, or if they are actual instances of apiC clusters that have since assimilated even in western dialects. Discussion of this is in section 3.1.4.

Table 15: *Words containing an apiC cluster from Washington (1850) where the AID only records an assimilated version*

English	Eastern	Central	Western	AID
Arm -below elbow	av-go-aub ata-ne	tai-yak-a-nak	ta**d-l**ik, ta-li-ak	talik
Beard or whiskers	u-ming-a, ul-luab aming-a	u-mi**t-k**i	u-mikh	umŋich
Five	te**d-l**i-ma, tel-li-met	te**d-l**i-ma		tallimat
Heel	kin-mik	kim-mi-ga	kai**t-m**ik	kimmik, kipmik, kikmik
Lamp	kol-lek	ku-**dl**i-uk	nai-nikht	naniḳ, ḳulliḳ
Old, he is	innu-tokak	it-tu**t-k**u-ak-pok		utuḳḳauruk
Golden plover		tu**dl**i ariu	tu**d-gl**iet, tu**d-l**ik	tullik
Puffin			a**t-p**ak	akpak

Overall, Washington's 1850 vocabulary is a very useful tool for historical Inuit phonology, with cross-dialectal comparisons, and being one of the first known sources with a bilingual native speaker contributing to it. From Washington's vocabulary, it is possible to conclude that apiC clusters remained in the Aivilik and North Baffin dialects until after 1850.

James Allan's Wordlist, 1893. Scott Polar Research Institute Archive.

James Allan, a sailor on a whaling ship, compiled an Inuit wordlist in his journal, which I found and consulted in the SPRI archive. No locations are mentioned in the wordlists, but in his journal, he mentions three different locations: Descent Harbour in the Davis Strait, Carey Island, and St James' in Newfoundland. Descent Harbour no longer exists and there is no information available online about its location, to my knowledge. Carey Island is in Hudson Bay, and St James' possibly refers to St John's, a major city in Newfoundland. It appears as though Allan sailed around the eastern Canadian Arctic on his whaling voyage, in the regions of the South Baffin, Nunavik and Nunatsiavut dialects. He does not credit the words to one location, so it is possible that they were collected from multiple places, or learned from another sailor or even copied form another wordlist, which were more widely available at this time.

Table 16: *Every word containing a consonant cluster from Allan's 1893 wordlist.*

Type of CC	English Gloss	James Allan (1893)	Labrador Inuttut Dictionary	Alaskan Inupiat Dictionary
velC	To steal	Ti**dl**ipook	Tillik	tigliktuḳtuḳ
velC	Yes	Ak-me-lak	Angik	aŋiḳtuḳ (says yes)
velC	Ship	mee-ak-twak	Umiatsuak (Battle ship)	umiaḳpak
velC	Sleep	sin-i-pa	sinik	siniktuḳ
velC	Deer	Tuktoo	Tuttuk (caribou)	tuttu
bilC	Day	oo**d-l**a-may	ulluk	uvluḳ
bilC	How many?	Ketch-ee-nay?	Katsinik	ḳapsinik, ḳavsinik
uvuC	Black whale	Akbin	Apvik (Sperm whale, bowhead whale)	Agvik (Bowhead whale)

In Table 16, there are two words containing a <dl> cluster: 'to steal' and 'day'. This is striking at first, as it suggests that apiC clusters were maintained until the 1890s. However, in the AID, 'to steal' is *tigliktuḳtuḳ,* with a velC cluster, and 'day' is *uvluḳ,* with a bilC cluster. In these cases, Adams appears to have written down an apiC cluster for what was probably a different type of cluster. As this wordlist was short, with only 48 entries, it did not contain any words with historical or underlying apiC clusters, so I did not draw conclusions about the loss of apiC clusters from it.

3.1.3 Non-native speaker (mis)perception: how accurate can wordlists written by outsiders be?

One issue with using external historical sources as evidence for sound change is that non-native speakers may perceive sounds differently than native speakers. This may be especially noticeable with perception of clusters that are illegal in a speaker's own language, such as apiC clusters by English speakers. Davidson and Shaw (2012) conducted a study with English speakers listening to non-native stop-stop clusters, including apical-velar clusters /tk/ and /dg/. This study focused on how well participants could discriminate between a word containing a non-native cluster and a variation of that cluster: either with prosthesis, schwa insertion between the consonants of the cluster, C1 deletion or C1 change. Participants in this experiment took part in a discrimination trial, where they listened to a non-word containing a non-native cluster and then the same word with one of the variations on that cluster listed above, and they had to judge if the two words were the same or different. Results were scored with a d prime value. A higher d prime signifies that participants were better at discriminating between the two forms. For apiC clusters, C1 deletion and vowel insertion had the lowest d prime value. This means that participants were the worst at noticing whether an apiC cluster had a missing C1 or a schwa inserted between the two consonants. However, confusion in C1 deletion trials was almost entirely due to /dg/ clusters. Participants could only discriminate between /dg/ and /g/ with 27% accuracy, but they could discriminate between /tk/ and /k/ with 89% accuracy.

This has two important implications for the current study. Firstly, it suggests that, if native English speakers do misperceive /tk/ clusters, it will be with schwa insertion, as /tək/. There is no evidence of schwa insertion in any of the clusters in the historical wordlists in this study, except perhaps *Mïng.a.kun* 'needle' in Adams' notebook. Secondly, it suggests that native English speakers can discriminate between /tk/ and /k/ 89% of the time. This means that, in the above wordlists, words written with a /k/ instead of a /tk/ will likely represent true cases of regressive assimilation, rather than misperception of the /t/. Still, overall accuracy in stop-stop clusters was lower than fricative-nasal, fricative-stop or stop-nasal clusters, and the lowest overall in the study. Also, Davidson and Shaw's study focused on onset clusters, while the apiC clusters in Inuit wordlists are intervocalic, which may result in a different pattern of perception.

The results of Davidson and Shaw's experiment are especially important for analysing Richardson's wordlist. It could be argued that apiC clusters had not started assimilating by Richardson's time, and instead he was misperceiving apiC clusters due to him not being a native speaker of Inuit. However, the most likely repair strategy used by English speakers for apiC clusters is schwa insertion, and Richardson does not use schwa insertion; he deletes the C1 of the cluster. This is consistent with regressive assimilation, rather than a typical repair strategy for /tk/ clusters. Additionally, Davidson and Shaw's experiment found that non-native velC clusters are often perceived as missing their C1. As established earlier, Richardson records most velC clusters with their C1 intact, so there is no evidence for Richardson systematically mis-transcribing clusters, even in velC clusters where C1 deletion is more likely. In summary, there is no evidence of Richardson using typical repair strategies in apiC or velC clusters, so he was likely recording the start of regressive assimilation in apiC clusters.

Overall, results from Davidson and Shaw suggest that English speakers do have difficulty perceiving non-native stop-stop sequences. However, if English speakers are going to misperceive /tk/ clusters, it will more likely be with schwa insertion, not with the first deletion or change of the first consonant. Also, not all the writers of these wordlists were completely inexperienced with the Inuit language; for instance, Qalaherhuaq, an Inuit man and native Inuit speaker, helped compile Washington's wordlist. Therefore, any case of systematic transcription of /tk/ as /k/ likely represents a case of true regressive assimilation.

3.1.4 Lexical Diffusion

Eight words from Washington's wordlist (Table 15), four from Ross's wordlist (Table 12 section B), and Dobb's *shoot-cock* 'whalebone', all contain apiC clusters where the AID records an assimilated cluster. This raises the possibility that some assimilation of apiC clusters has happened in even the most western dialects, and therefore, that the change may have spread by lexical diffusion. The term lexical diffusion was proposed by Chen and Wang (1975) and describes a proposed mechanism for the implementation of a sound change, whereby a sound change spreads lexeme by lexeme until it affects the whole lexicon. It is possible that regressive assimilation in apiC clusters spread by lexical diffusion, and the early effects of this process can be seen in western dialects that have assimilated some of the clusters observed in historical sources.

However, this study does not have enough data to definitively conclude that apiC clusters assimilated by lexical diffusion. The assimilated clusters recorded in the AID could be due to dialect mixing. It is easy to draw a solid line on a map separating western dialects with no assimilation from eastern dialects with assimilation, but in reality, groups of people move around and mix their dialects, which could be why some apparently assimilated clusters are seen in the AID. It could also be a case of systematic misperception by the historical writers. To conclude that lexical diffusion is active in Inuit cluster assimilation, more evidence is needed, perhaps from bilC, velC and uvuC clusters, for which there is more data from people with linguistic training. If evidence from these clusters shows a lexeme-by-lexeme spread of assimilation, then it is possible that apiC cluster assimilation also spread by lexical diffusion.

3.2 Summary of wordlist analysis

This study employed a close, critical examination of seven original English-Inuit vocabularies, none of which have used before in a study of Inuit consonant cluster assimilation. From this, I established a precise timeline for the loss of apiC consonant clusters, and for when regressive assimilation occurred in apiC clusters in specific eastern Canadian Inuit dialects, shown in Table 17. This is an improvement on earlier attempts to find when apiC clusters had assimilated, the most precise of which (Dorais 2010) simply stated that they had assimilated by the mid-nineteenth century, with no information on different dialects.

Table 17: *Timeline of the loss of apiC consonant clusters in eastern Canadian Inuit dialects.*

	Dialect	Time of apiC cluster assimilation	Source
More western	Aivilik	After 1850	Washington (1850)
	North Baffin	After 1850	Washington (1850)
	South Baffin	Around 1830	Ross (1833)
	Nunavik	Probably between 1800 and 1830	
More eastern	Nunatsiavut	By 1800	Richardson (1771)

Section 3 has answered research question 1 and 2 of the study by establishing a more precise timeline for the loss of apiC consonant clusters, through a critical examination of original historical sources. In Section 4, research question 3 is examined, using the above data to answer why apiC clusters were lost before other cluster types.

4 Research Question 3

I now discuss research question 3, and attempt to explain why apiC clusters were lost in Inuit dialects so early, and before all other cluster types. Section 4 will be split into two subsections:

4.1: Past approaches - why explanations of regressive assimilation by past authors may not be satisfactory to explain the new data.
4.2: An alternative explanation for the loss of apiC consonant clusters.

4.1 Past Approaches

In section 3, I established that apiC clusters in Canadian Inuit were first lost by a process of regressive assimilation in the Nunatsiavut or Labrador dialect, which probably started by the year 1800, significantly earlier than assimilation of other types of clusters. Considering this new data, there is a common issue with chronology between the previous accounts of consonant cluster loss. These accounts argue that consonant cluster loss diffused westward from Greenland, as in Crieder (1981), or that consonant cluster loss was brought about by changes to Inuit society due to contact with Europeans. The societal changes they describe as contributing towards the loss of consonant clusters happened after apiC clusters were already assimilated in some dialects. ApiC clusters were already being assimilated by the year 1800. According to Dorais (2010), significant travel between Inuit groups in Greenland and Canada was not common until the 1950s. Therefore, Creider's suggestion, that the loss of consonant clusters diffused westward from Greenland, is unlikely. Also, according to Dorais (2010), bilingualism was not well established in Inuit communities until the 20th century. Therefore, bilingualism cannot explain why apiC clusters started to assimilate in around 1800. It is true that there were bilingual Inuit before this time, but not a significant enough portion to cause large-scale language change. Dorais proposes sedentarisation as a major reason for the loss of consonant clusters, as sedentarisation caused the speech style used by women and children to become more common and prestigious in Inuit society. However, sedentarisation had not happened to a significant extent by the end of the 18th Century, when apiC clusters started to be lost. Although the transition from nomadic to sedentary life for Inuit communities was not an instantaneous process, Stuckenburger (2006) proposes that sedentarisation mostly came about in the 1960s, with the majority of sedentarisation happening long after apiC clusters started to assimilate around 1800.

Overall, the sociological phenomena used by past writers to explain the loss of consonant clusters in Inuit happen too late to explain the loss of apiC clusters. Furthermore, none of the above approaches explain why apiC clusters were lost before other types of clusters, or why they were lost so much earlier than other types of clusters.

4.2 An Alternative Explanation for the loss of apiC Consonant Clusters

4.2.1 Gestural overlap

Rather than a sociological explanation, which may not fit with the proposed timeline for the loss of apiC clusters, I propose that apiC clusters were lost before other types of clusters in eastern dialects of Inuit due to the inherent articulatory properties of apiC clusters. Byrd (1996) has found that apiC clusters have greater gestural overlap than other clusters, with the lingual gesture of the second consonant overlapping the apical consonant. This causes the apical consonant to be less easily perceived by the hearer, making apiC clusters generally more prone to regressive assimilation over time.

The unusual status of apiC clusters compared to other consonant cluster types was first proposed by Bailey (1969). Bailey proposed a universal principle that adjacent consonants in the order coronal-noncoronal form a marked cluster. These clusters are unstable and susceptible to change by regressive assimilation, expressed in the following rule:

$$[_{u}\text{cor}] \rightarrow [\text{-cor}] \;/\; \_\; [\text{-cor}]$$

This can be otherwise expressed as a rule that any coronal consonant will become non-coronal when immediately followed by a non-coronal consonant. Bailey's markedness proposal was based on English coronal assimilation, which is where an English coronal consonant will be produced with the place of articulation of the following consonant. For example, 'ten beans' is pronounced [tɛ**mb**iːnz]. Blust (1979) gave more evidence for Bailey's markedness proposal from Austronesian languages, all of which show coronal assimilation in clusters. Blust also proposed that the cross-linguistically marked status of coronal-initial clusters is due to a 'difficult articulatory environment'; speakers find apiC clusters difficult to articulate, so they rely on coarticulation techniques like regressive assimilation to make articulation easier. Overall, Bailey and Blust agree that coronal-initial clusters are more marked and more unstable than other cluster types.

Paradis and Prunet (1991) agree that coronal-initial clusters are marked, but offer a different explanation as to why they are marked than Blust's articulatory explanation. Working within a feature geometry model, Paradis and Prunet proposed that coronal consonants are underspecified for place. In this model, bilabial, velar and uvular consonants are represented as bilabial, uvular or velar in their mental representations, but coronal consonants have no place specification in their mental representation. Therefore, they assimilate easily in consonant clusters, as the other consonant can spread its underlying place specification to the coronal. However, there are some issues with the underspecification explanation. In Inuit, regressive assimilation can happen if the second consonant in the cluster is a coronal. This process can be seen, for instance, in the word 'fog', recorded as *tuck-took* in Dobbs's 1744 wordlist. The coronal consonant has spread its place of articulation to become the modern form *tattuk* (LID). In a feature geometry approach to underspecification, such as Archangeli (1988), segments that are underspecified for certain features cannot spread that feature. If coronals have no underlying place specification, then they should be unable to spread their place of articulation, so regressive assimilation in clusters where the C2 is a coronal would be impossible. Overall, there must be an explanation as to why coronal consonants assimilate more readily than other consonants in Inuit, without relying on underspecification.

Instead of underspecification, there may be an articulatory explanation as to why apiC clusters assimilated earlier in Inuit than other cluster types. A study by Byrd (1996) suggests that when apical consonants are the initial consonant in a cluster, they have a greater gestural overlap from the C2 than other types of C1 consonants. This means that the movement of the tongue for the C2 in an apiC cluster overlaps the movement of the tongue for the apical consonant. Because of this gestural overlap, the apical consonant in apiC clusters is more difficult to perceive than other types of C1, so apiC clusters are more prone to regressive assimilation than other cluster types. In Byrd's study, five English speakers took part in an electropalatography (EPG) experiment, where the movement of their tongue during consonant clusters was measured. The amount of overlap in the consonant cluster was measured in two ways; sequence overlap, which is the total amount that the gestures for C1 and C2 overlapped, and C1 overlap, which is how much the gesture for C2 overlapped the gesture for C1. Both types of overlap were always higher for apiC clusters than velC clusters. Averaged across all five speakers, sequence overlap for apiC clusters was 59%, while for velC it was only 46%. Mean C1 overlap for apiC clusters was even greater, at 87%, whereas velC clusters only had 53% overlap of their C1. Overall, in English, the second consonant in a consonant cluster overlaps the first consonant much more when the first consonant is apical. This could be an articulatory based explanation for why apiC clusters are more prone to regressive assimilation.

Patterns found in Byrd's study have been repeated in other studies on different languages, and using different methodology. For example, a review by Recasens (2018) found evidence for more gestural overlap in apiC clusters than other cluster types in Georgian, Moroccan, and Korean. Recasens (2018) also performed a similar study with Catalan speakers, this time using ultrasound data to measure gestural overlap. The speakers were measured saying apiC, bilC and velC clusters. Recasens concluded that gestural overlap of the C1 in a consonant cluster by the C2 was highest when the C1 was an apical, second highest when the C1 was a labial, and lowest when the C1 was a velar. This order of apicals > bilabials > velars matches the order in which Inuit consonant clusters assimilated. In summary, studies measuring tongue gestures in different consonant cluster types have reached the conclusion that the apical C1 in apiC clusters is highly overlapped by the C2, to a greater extent than the C1 is overlapped in other cluster types. This makes the apical consonant in apiC clusters difficult to perceive, while the C2 remains easily perceptible, which eventually may result in regressive assimilation.

In conclusion, a sociolinguistic approach as has been taken previously in the literature may explain later loss of consonant clusters in Inuit. However, the data from this study suggests that apiC clusters started to fall out of use in Inuit too early for an entirely sociolinguistic explanation to apply. Also, a sociolinguistic approach does not explain why apiC clusters specifically were lost so much earlier than other cluster types. The high degree of gestural overlap inherent in apiC clusters explains the early loss of apiC clusters without the need for a sociolinguistic explanation.

4.2.2 Frequency

Another factor that may have contributed to the loss of apiC clusters is their frequency. Jaeger and Hoole (2011) found that clusters in high-frequency words are more prone to regressive assimilation. This is because these words have greater lexical predictability, so there is less need for a speaker to articulate them carefully to be understood. Therefore, there is more gestural overlap in higher frequency words. Indeed, Jaeger and Hoole found that there was always more gestural overlap in clusters in high-frequency words than low-frequency words. If apiC clusters appeared in words with higher lexical frequency than other cluster types, this may contribute to them being assimilated earlier. I attempted to find whether words containing apiC clusters had an unusually high lexical frequency using the only corpus available of the Inuit language, the Nunavut Hansard Inuktitut-English Parallel Corpus. However, this corpus was unsuitable for this experiment. It is a corpus of modern Inuktitut, not historical Inuktitut, and it is specifically a corpus of governmental proceedings, so the word frequency taken from this corpus most likely does not reflect the frequency of words containing apiC clusters around 200 years ago, when they started to assimilate. In future, if more corpora of Inuit, especially historical Inuit, become available, then it may be possible to determine if lexical frequency played a role in the assimilation of apiC clusters.

5 Conclusions

5.1 Research Questions

In this study, I used seven historical English-Inuit wordlists which have not been examined before in the context of Inuit consonant cluster assimilation to answer these questions on historical Inuit phonology:

(1) What was the exact time course of assimilation of apiC clusters in eastern Inuit dialects?
(2) To what extent can we use external historical sources to discover this?
(3) Why did apiC clusters assimilate first, and long before other consonant clusters in Inuit?

For question one, using sources from online and physical archives, I confirmed that apiC clusters assimilated many years prior to assimilation in other consonant clusters. This study went further than other previous studies by establishing a precise time course for the assimilation of apiC clusters in different dialects of Inuit. I discovered that apiC clusters:

- Were in the process of assimilating by the year 1800 in the Nunatsiavut dialect;
- Were likely assimilating by around 1815 in the Nunavik dialect;
- Were assimilating by 1830 in the South Baffin dialect;
- Remained until after 1850 in the Aivilik and North Baffin dialects.

This makes the overall timeline for the assimilation of apiC clusters in eastern dialects of Inuit around sixty years from the early to mid-nineteenth century. Although this seems like a long time, it is not unexpected as there was little travel between groups of Inuit during this time, so any sound change would take some time to spread through the community. This proposed time course also means that there was a gap of around 150 years between when I propose the first assimilation of apiC clusters began, and when Dorais proposes that the first assimilation of bilC clusters began. When compared to the around 40 year gap between the loss of bilC clusters and velC clusters, and a similar gap between velC clusters and uvuC clusters, the assimilation process in apiC clusters takes place much earlier than other cluster types.

For question two, historical sources written by outsiders to the language can be useful, especially when they are the only option for studying the history of a language. However, it is important to be careful about using each source. Ensuring the words they have transcribed can be backed up by modern sources is vital. It is also important to establish a background for each source, as writers who have more experience with the Inuit culture, more of a drive to communicate effectively with the local people, or more time living in the Arctic alongside the Inuit, often make more accurate transcriptions. Overall though, with a critical analysis of each source, it has been possible to establish a timeline and explanation for the loss of apiC clusters in Inuit.

Finally, for question 3, later assimilation of consonant clusters in Inuit may have been due to changes in society brought about by European colonialism. However, for apiC clusters, the change actualised when many groups of Inuit were still quite isolated, and extensive contact with Europeans was still quite uncommon. Instead, apiC clusters were possibly lost so much earlier than other cluster types due the inherent articulatory properties of apiC clusters which make them more prone to regressive assimilation.

5.2 Significance of the Research

In the last centuries the Inuit language has come close to dying out, due to the Canadian government's enforcement of English and French as the only official Canadian languages throughout the 20th century, and the racism and forced re-education faced by Inuit speakers. Furthermore, as the language lacks a historical written tradition, the history of how the Inuit language sounded is not readily available. Work such as this study that aims to reconstruct the history of the Inuit language furthers our understanding of sound change processes in the Eskimo-Aleut family, and joins the efforts of many Inuit looking to preserve their language and culture. Additionally, I discussed the history of the Inuit language which has not been discussed before in the context of Inuit sound change, as well as sources from the Scott Polar Research Institute archive that have never been consulted for any study of the Inuit language. By recording and studying these new sources, I have contributed towards forming a more complete history of the Inuit language, as well as addressing the wider question in historical linguistics of how useful sources written by outsiders can be in studying the history of a language.

In the future, more research could be done into apiC clusters in the Nunavik dialect, which proved difficult to find sources for in English archives. Also, the sources on Inuit vocabulary that are new to this study could certainly be used as part of future studies into the history of the Inuit language. Throughout this study, I have given impressionistic thoughts on the accuracy of different sources, but this could also be done systematically. For this systematic approach, a native speaker of Inuit could examine the historical sources and record how many words from each source match with known words in their language.

6 References

6.1 Primary Sources

Adams, E. (1851, MS 1461/2). Vocabulary book, Norton Sound Inuit. Scott Polar Research Institute Archive. Cambridge: Cambridge University.

Allan, J. (1893, MS 1087/3). Arctic Journal. Scott Polar Research Institute Archive. Cambridge: Cambridge University.

Back, G. (1833-1835., MS 395/8). Inuit Vocabulary Notebook. Scott Polar Research Institute Archive, Cambridge: Cambridge University.

Back, G. (1835). *Narrative of the Arctic land expedition to the mouth of the Great Fish River.* Edmonton: M. G. Hurting.

Dobbs, A. (1744). A Vocabulary of English and Eskima Words. In A, Dobbs (Ed.). *An Account of the countries adjoining to Hudson's bay in the North-West part of America* (p. 206). London: J. Robinson.

Fortescue, M., Jacobson, S., & Kaplan, L. (1994). Comparative Eskimo dictionary with Aleut cognates. *Alaska Native Language Center Research Paper*, 9.

National Research Council of Canada. *The Nunavut Hansard Inuktitut-English Parallel Corpus.* Retrieved from: <http://www.inuktitutcomputing.ca/index.php>.

Pigott, P. (n.d.). Labrador Inuttut Dictionary. Retrieved from: <http://www.labradorvirtualmuseum.ca/english-inuttut.htm>.

Richardson, W. (1771). The Complete Inuttitut Vocabulary. Full reprint in M. Stopp. (2014). *The complete Inuttitut vocabulary collected by William Richardson ca.1765-1771.* Regional Language Studies Newfoundland, 25.

Ross, J. (1833). Vocabulary of the English, Danish and Esquimeaux languages. In J. Ross. (1835). *Appendix to the narrative of a second voyage in search of a North-West Passage.* London: A.W Webster.

Washington, J. (1850). *English and Eskimaux Vocabulary*. Reprint (2018), London: Forgotten Books.

Webster, D. H & W. Zibell. (1970). *Inupiat Eskimo Dictionary.* Fairbanks: University of Alaska.

6.2 Secondary Sources

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5(2), 183-207.

Bailey, C-J. N. (1969). A possible explanation for an assimilation curiosity. *University of Hawaii Working Papers in Linguistics* 1(5), 348-350.

Blust, R. (1979). Coronal-noncoronal consonant clusters: New evidence for markedness. *Lingua,* 47(2), 107-117.

Bobaljik, J. (1996). Assimilation in the Inuit Languages and the Place of the Uvular Nasal. *International Journal of American Linguistics,* 62(4), 323-350.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24(2), 209-244.

Crieder, C. (1981). Place of Articulation Assimilation and the Inuktitut Dialect Continuum. *Études/Inuit/Studies* 5, 91-100.

Davidson, L., & Shaw, J. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics,* 40(2), 234-248.

Dorais, L. (1985a). Regressive Consonant Assimilation in Eastern Inuktitut: Toward a Sociolinguistic Approach. *International Journal of American Linguistics*, 51(4), 390-393.

Dorais, L. (1990). The Canadian Inuit and their language. In Collis, D. (ed.) *Arctic Languages: An Awakening* (pp.185-291). Paris: UNESCO.

Dorais, L. (2010). *The language of the Inuit: Syntax, semantics, and society in the Arctic*. Montreal: McGill-Queen's University Press.

Jaeger, M. & Hoole, P. (2011). Articulatory factors influencing regressive place assimilation across word boundaries in German. *Journal of Phonetics*, 39(3), 413-428.

Paradis, C. & J-F. Prunet (Eds.). (1991). The special status of coronals: internal and external evidence. *Phonetics and Phonology* 2.

Recasens, D. (2018). *The Production of Consonant Clusters: Implications for phonology and sound change.* Berlin/Boston: De Gruyter.

Stopp, M. (2014). The complete Inuttitut vocabulary collected by William Richardson ca.1765-1771. *Regional Language Studies Newfoundland,* 25.

Stuckenberger, A. N. (2006). Sociality, temporality and locality in a contemporary Inuit community. *Études/Inuit/Studies*, 30(2), 95–111.

Thomason, S. G. & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics.* Berkeley, California: University of California Press.

Cross-Cultural Gestures: Differences in Type and Frequency between German-English Bilinguals

Eleanor Streatfield
University of Birmingham

Abstract. Previous cross-cultural and cross-linguistic studies argue that immersion in one language environment will affect bilinguals' unverbalised communication styles in both languages. Assimilated bilingual speakers (e.g. bilingual who grew up in a country speaking their L2), appear to exhibit similar frequency and space of gesture in both languages, where immigrant bilingual speakers do not (Efron, 1942). An assimilated bilingual speaker, or 'hybrid gesturer,' therefore, blends unverbalised communication styles together, indicating a transfer of gesture type and frequency from one language to another based on their predominant language environment. Investigating spontaneous gesture exhibited by bilingual German-English speakers, this study determined the effect of immersion in a country's predominant language on unverbalised communication styles, during descriptive tasks in a face-to-face conversation. Two different image-only comic strips were provided as prompts for a 5minute discussion in English and another in German. Co-speech gestures were transcribed, with results demonstrating a link between a language environment and subsequent gesture usage in that language. Participant B gestured more in her L1 overall; the language she was not immersed in growing up. Aziz and Nicoladis (2019) propose that bilingual speakers who are predominately immersed in their L2 tend to gesture more in their L1, as a method of compensating for the verbalised communication they are not surrounded by - and therefore may lack confidence in speaking. This participant also used a higher frequency of metaphoric gestures in German (L1) than English (L2) when explaining abstract concepts, which further reinforces this claim. Participant A produced a similar gesture style to previously studied immigrant bilinguals.

Keywords: spontaneous; gestures; German; English; bilingual; frequency

---

1 Introduction

This investigation explores the hypothesis that cross-cultural use of spontaneous gesture differs between bilingual German-English speakers, relative to the country they grew up in. The objective is to observe how distinct experiences of two cultures, where either one is surrounded by that country's dominant language, affects non-verbal communication whilst speaking either language. Both the participants are bilingual; therefore it was assessed how their different childhood environment has affected cross-cultural gesture through asking them to describe a comic strip during a face-to-face discussion, first in English and then German. The frequency of their spontaneous gesture use and number of each type (iconic, deictic, metaphoric, regulator) used within 30 seconds were recorded and analysed in conjunction with the research question. The outcome shows there is a contrast between the gesture usage in each language and suggests their predominant language environment influences how similarly or dissimilarly bilingual speakers gesture in each language.

2 Literature Review

Efron (1942) studied how assimilated groups of bilingual Italian-English and Yiddish-English speakers exhibited similar gesture usage (e.g., frequency, space) to each other despite their dissimilar heritages, where immigrant groups showed a contrasting difference in gesture. This presents foundational evidence in cross-cultural and cross-linguistic studies that the language you are predominantly immersed in affects unverbalised communication – regardless of heritage.

Furthermore, Pika, Nicoladis and Marentette (2006) propose the idea of a 'hybrid gesturer,' a bilingual person who blends two cultures' unverbalised communication styles together in both languages. It was suggested that the transfer of gesture from one language to another depended on the culture's typical use and frequency. Similarly, the study concluded that bilingual people gesture more, particularly iconic, and deictic, than (English) monolinguals. This provides data to support the research question, suggesting there could be a different style of gesture and rate between those whose predominant language growing up is German relative to English. However, it is also possible that gesture frequency in English will be lower, as studies have determined that it is typically a low-frequency gesture language (Ann Graham and Argyle, 1975). In addition to this, Nicoladis et al. (2018) established that European bilinguals produce slightly more gestures in their L1 than in English (L2), reinforcing a potential contrast in outcome of gesture rate during this investigation.

3 Methodology

3.1 Participants

Both participants in this investigation speak German as their L1 and English as their L2. Participant A (A) lived in Germany for 28 years and later immigrated to England. Participant B (B), conversely, grew up in England for 18 years; she has therefore immersed herself in the English language and culture, and could be considered a 'heritage learner' (Van Deusen-Scholl, 2003). The participants' bilingualism acts as a control factor to allow for more accurate analysis of how their home country (the language environment they grew up in) relates to their use of gesture in either language. Thus, their interactions during a description task answers the research question; one of them gesturing more than or differently to the other in either or both languages, for example, would provide evidence that the language and culture one grows up in influences how one gestures in other languages. There is an uncontrollable age difference between the two participants (mother [A] and daughter [B]), although age will not be a focus during this investigation. Other factors alike both identifying as female and having the same level of fluency in both languages were controlled, however.

3.2 Materials

For the topic of discussion during the recording, there were two different comic strips provided (with no speech and similar in length as a control variable), one for discussion in English and another in German. Both participants had a physical A4 paper copy of each prop, labelled A, B, C and D (see Appendix 3). This stimulated them to exhibit gesture use during descriptive tasks, such as events of the comics and any noticeable details about characters (e.g., relationships, appearance, emotions etc.)

Participants were encouraged to talk for approximately 5 minutes in English, then switch to the next two props and talk for 5 further minutes in German. The data was recorded at a table using a phone camera and microphone, perpendicular to the participants, who sat opposite each other to allow a clear view of gesture use and space from the torso upwards. Likewise, this allowed them to make eye contact, display fluid transitions between taking dominance over conversation and see hand gestures more clearly.

Once recording was complete, 30 seconds of the English conversation and then German were transcribed using annotation software ELAN v5.9. This aided the process of noting co-speech gestures and to count the frequency of gestures expressed by participants in both languages. The gesture types focused on were iconic; illustrating a physical object or movement; metaphoric; explaining an abstract concept; and deictic; moving attention towards referent in the real world, (Cochet and Vauclair, 2014; Novack and Goldin-Meadow, 2016). Regulators are not classed as gestures as such, however this study does group them amongst other gesture types studied as they are an interesting indicator towards dominance of and confidence during conversation.

3.3 Procedure

Due to restrictions during the COVID-19 pandemic in December 2020, I was unable to record the data myself. To overcome this issue, a simple instruction sheet (see Appendix 2) was provided, specifying what they should talk about and how to use the props. Ethical implications were carefully considered, especially during the conditions at the time of the study. To mitigate harm and stress, I thoroughly explained my investigation after approaching the participants and transparently disclosed where their data will be used and how it will be protected, as well as their anonymity and right to withdraw from the investigation should they wish to partake. Consent was obtained (see Appendix 1) after both participants read through, signed, and scanned through the consent form. The participants were not told the research question, however, to avoid inadvertent bias and prevent a sense of discomfort or overthinking their gesture use during the study. Participants' faces are anonymised in the example photos given in this paper to withhold confidentiality.

Similarly, I gave both participants adequate time to decide if they wanted to take part, and likewise to record and send the data across. This allowed them to feel in control of their own data and decisions, rather than feeling pressured to participate. Another positive side of this arrangement was that participants could record the data in a familiar and safe environment in their own living space. This proved to deliver a more relaxed conversation and hence more accurate use of gesture to analyse. The research was carried out in a quiet space (their dining room) to prevent noise distractions from affecting the results. Care was taken to minimise other factors which would disrupt the flow of conversation, such as phone notifications.

4 Results

After 30 seconds of each language were transcribed, it was evaluated that participant A, who grew up in Germany and immigrated to England, gestured more in the former language and less in English. In addition, she gestured significantly less than participant B, who gestured 8 times in German and 8 times in English within a 30 second period. Participant A only gestured once during the English conversation and 5 times in German.

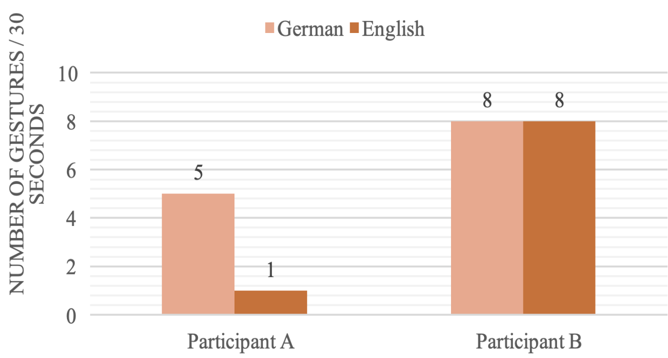


Figure 1: *Number of gestures per 30 seconds, in each language.*

Regarding the type of gesture expressed, there was a prominent difference between participants alongside the style most often displayed between them. For example, the only gestures used by A in German were regulators — and only one metaphoric gesture in English. Conversely, B, who gestured more frequently in both languages, used mostly metaphoric gestures in German and used a range of all four types in English — deictic being the most frequent.

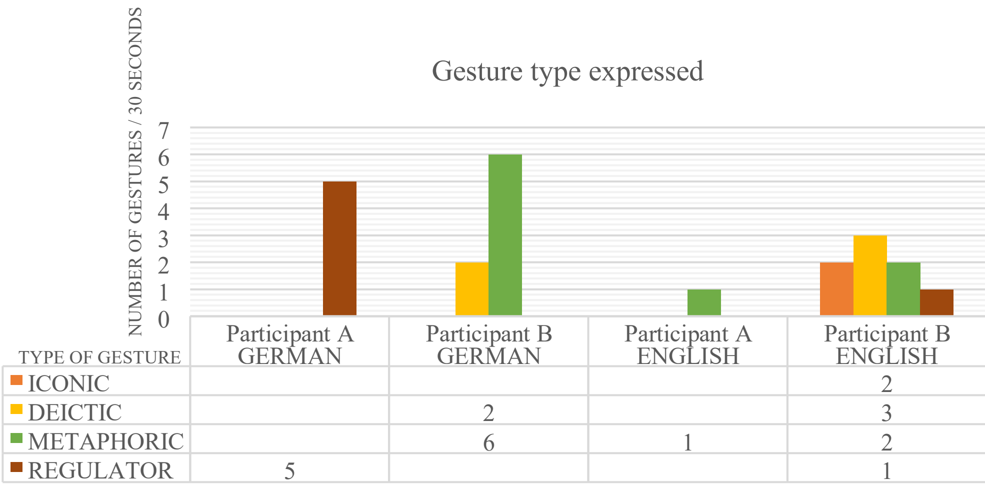


Figure 2: *Frequency of each type of gesture.*

5 Discussion

We can link together the frequency and type of gestures exhibited with the dynamics of the conversations to provide an answer to the research question. For example, the reason A tends to use more regulators in German could be due to B taking general dominance over both conversations. This is shown when A nodded frequently to show her understanding and held up her fingers (see Figure 3) to make B wait her turn in the conversation.

Figure 3: *"…and" (A)*

Similarly, we know that German is her native language and that she spent her first 28 years in Germany. Thus, it is possible that either she understands ideas more confidently in German *or* the culture tends to nod to show understanding more than England, reinforced by A not using regulators at all in English (nor B in German). Similar results were shown when B used more metaphoric gestures in German than English. A study by Aziz and Nicoladis (2019) indicated that bilingual people gestured more in the language they are not predominantly surrounded by; a form of trying to compensate for verbal communication as the language is not commonly spoken in English communities. Theoretically, B displays similar behaviour as she uses metaphoric gestures more in German when explaining concepts (see Figure 4) - perhaps because she grew up immersed in the English language and therefore lacks confidence speaking German. General frequency is similar in both languages, though this could explain the shift in type of gesture.

Figure 4: *"I think… the same…" (B) — example of a metaphoric gesture*

Moreover, there is a link between speech and gesture use yet a difference in this link between participants. For example, when participant A used a metaphoric gesture (throwing hands up in the air) in English, she says 'sorry to be picky'. The action of raising both hands with palms facing upwards directly correlates to the speech. It is worth noting this is the only gesture used by A in English, and the communicated idea is to defend her point which, again, could suggest a lack of confidence when speaking English.

Figure 5: *"Sorry to be picky" (A) — example of a metaphoric gesture*

Likewise, B displayed more referential gesture types than A, directly related to her speech. She used 2 iconic gestures in English (depicting a woman's short hair and throws hands to show idiom 'throwing out an idea'), 3 deictic gestures in English and 2 in German (pointing to the prop or pulling at clothing).

This reinforces previous evidence that European bilinguals use more iconic and deictic gestures in their L2 (in this case, English), which Azar, Backus and Özyürek (2019) explains as a method of organising discourse and reduce the 'cognitive load associated with being bilingual.' However, participant A did not display the same results, perhaps due to B dominating the conversation.

Figure 6: *"short brown bob" (B) – example of an iconic gesture*

Figure 7: *"I could describe what the mum looks like?" (B) – example of an iconic gesture*

If we view these results in relation to Efron (1942), the assimilated groups, which B is slightly comparable to, gestured more similarly than their immigrant counterparts, which A could be compared to. In other words, B has only lived in one cultural environment her entire life where A has lived in two, thus another potential reason why B gestures similarly in frequency in both languages where A does not, having grown up in one culture and immigrating to another.

6 Conclusion

There are undoubtedly limitations within this investigation, such as the inability to record the data first-hand, a small sample size and factors, such as age, that could not be controlled. However, future research could use the outcomes of this study as a baseline to conduct another with a larger sample size and therefore more confidently conclude whether a bilingual person's home country affects their gesture use, specifically during descriptive or chronicle storytelling tasks (Nicoladis *et al.*, 2018).

From the sample investigated, it can be concluded that there is a distinction in spontaneous gesture use and type, potentially caused by lack of confidence, 'rustiness' (Aziz and Nicoladis, 2019) whilst surrounded by one language only or factors such as one person dominating the conversation. Regardless, the evidence has answered the research question sufficiently and validates the hypothesis: there is a difference in gesture type and frequency dependent on the bilingual speaker's home country, despite the reasons behind this not being transparent.

7 References

Ann Graham, J. and Argyle, M. (1975) A Cross-Cultural Study of the Communication of Extra-Verbal Meaning by Gesture. *International Journal of Psychology*, 10(1), 57–67.

Azar, Z., Backus, A. and Özyürek, A. (2020) Language contact does not drive gesture transfer: Heritage speakers maintain language specific gesture patterns in each language. *Bilingualism*, 23(2), 414–428.

Aziz, J. R. and Nicoladis, E. (2019) My French is rusty: Proficiency and bilingual gesture use in a majority English community. *Bilingualism*, 22(4), 826–835.

Watterson, B. (January 1994) *Calvin and Hobbe. GoComics*. Retrieved on 26/05/2022, from *GoComics*: <https://www.gocomics.com/calvinandhobbes/1994/01/02>.

Watterson, B. (November 1995) *Calvin and Hobbe. GoComics*. Retrieved on 26/05/2022, from *GoComics*: <https://www.gocomics.com/calvinandhobbes/1995/11/19>.

Cochet, H. and Vauclair, J. (2014) Deictic gestures and symbolic gestures produced by adults in an experimental context: Hand shapes and hand preferences. *Laterality*, 19(3), 278–301.

Van Deusen-Scholl, N. (2003) Toward a Definition of Heritage Language: Sociopolitical and Pedagogical Considerations. *Identity, and Education*, 2(3), 211–230.

Efron, D. (1942) Gesture and Environment. *The ANNALS of the American Academy of Political and Social Science*, 220(1), 268–269.

Nicoladis, E. *et al.* (2018) Gesture frequency is linked to story-telling style: Evidence from bilinguals. *Language and Cognition*, 10(4), 661–664.

Novack, M. A. and Goldin-Meadow, S. (2016) Gesture as representational action: A paper about function. *Psychonomic Bulletin & Review 2016 24:3*, 24(3), 652–665.

Pika, S., Nicoladis, E. and Marentette, P. F. (2006) A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer? *Bilingualism*, 9(3), 319–327. do

8 Appendices

8.1 Appendix 1: Consent Form

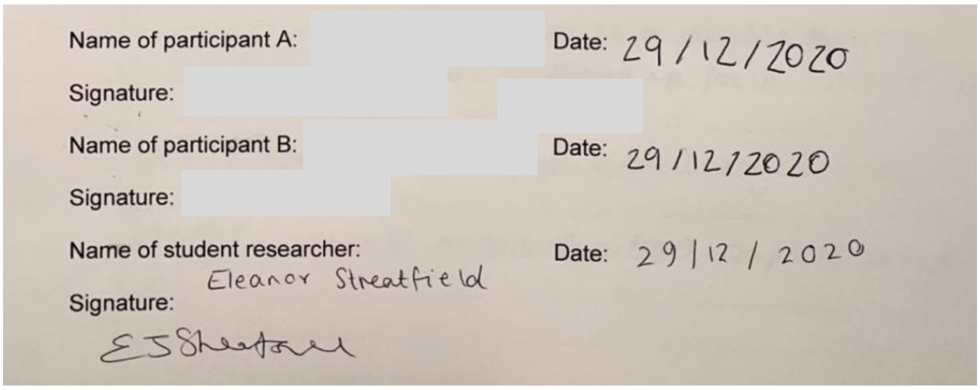
University of Birmingham
College of Arts and Law
Department of English Language & Linguistics
Information and Consent Form

I am a first-year student working towards an undergraduate BA degree in Modern Languages and English. As part of that degree I am taking a module that provides me with basic research skills, and which requires me to undertake a research project.

The information which you supply, or which may be collected as part of this research project will only be accessed by students involved in the project. No identifiable personal data will be published.

Statements of Understanding/Consent

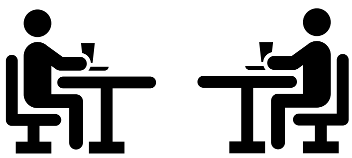
- I confirm that I have read and understand the participant information for this study. I have had the opportunity to ask questions if necessary and have had these answered satisfactorily.
- I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. If I withdraw, my data will be removed from the study and will be destroyed.
- I understand that my personal data will be processed for the purposes detailed above, in accordance with the Data Protection Act 1998.
- Based upon the above, I agree to take part in this study.


Name of participant A: Date: 29/12/2020
Signature:
Name of participant B: Date: 29/12/2020
Signature:
Name of student researcher: Eleanor Streatfield Date: 29/12/2020
Signature: EJStreatfield

8.2 Appendix 2: Participation Instruction Sheet

Participation in the language investigation

TASK:



1 Please sit opposite each other at a table with a stable camera (smartphone camera for example) positioned so it records your arms, torso, and face (see above for example) and can record any gestures you make, as well as hear everything you say.
2 Retrieve Prop A and B, a piece of A4 paper each which shows a comic strip without dialogue. Place down one of these in front of you, one in front of the other person.

3 Please discuss the events shown in the comic strip, the actions, the plot, and characters ***in English***. Try and be ***as descriptive and specific as possible***. Though it may feel obvious, try to talk about the most basic things, the personalities of the characters, how realistic the events are etc.
4 This will continue for approximately **5 minutes**. If you get stuck on things to say, try to ask each other questions.
5 Then, please retrieve prop C and D, another two sheets of A4 paper displaying a different comic strip.
6 Once again, discuss the comic strip: the actions, characters and what you think the plot is etc. only this time please discuss ***in German***. Try and be ***as descriptive and specific as possible.***
7 This will continue for approximately **5 minutes**. If you get stuck on things to say, try to ask each other questions again.
8 End of discussion. Thank you for participating!

8.3 Appendix 3: Props (A, B, C & D)

The cartoon images used within this experiment were accessed and printed with permission granted by Andrews McMeel licensing in the 'classroom usage statement' (available at: https://licensing.andrewsmcmeel.com/classroom-usage), as this investigation was carried out during my first year of university. The licence, however, does not extent to online usage or outside classroom use and cannot be presented on this paper. The weblinks to the comics are provided below :

> **Props A and B**: *Calvin and Hobbes by Bill Watterson for November 19, 1995 - GoComics* (no date) Available from: https://www.gocomics.com/calvinandhobbes/1995/11/19
>
> **Props C and D:** *Calvin and Hobbes by Bill Watterson for January 02, 1994 - GoComics* (no date) Available from: https://www.gocomics.com/calvinandhobbes/1994/01/02

Do Communication Pressures Drive Strategy Use in Iconic Productions?

Andrew Tobin
University of Edinburgh

Abstract. To iconically convey a handheld object, two strategies are common: handling, where the signer's handshape mimics the shape of a hand using the object, and instrument, where the signer's handshape mimics the shape of the object. Handling gestures are strongly preferred in the productions of hearing gesturers, while deaf signers use much more instrument in comparison. I hypothesised that this difference is due to greater communicative efficacy of the instrument strategy, leading to an increase in its frequency in the evolution of sign languages. In one experiment, I showed naïve hearing participants gestures using one of the two strategies, and tested for an effect of strategy on how successfully they guessed the sign's meaning. In a second experiment, I tested whether participants would increase their preference for instrument gestures when a communication pressure was introduced. The results found no significant effect in either experiment: In Experiment 1, strategy had no effect on how well participants guessed the intended meaning, suggesting no overall difference in communicative efficacy. In Experiment 2, a communication pressure did not affect strategy preference, suggesting communication does not drive choice of iconic strategy. While various limitations of the experiments are discussed, the results of the present study do not support communication pressures being a significant factor in the differences in iconic strategy usage between hearing gesturers and established sign languages.

Keywords: iconicity; signed languages; communication strategy

---

1 Introduction

In this paper I explore communication pressure as a possible explanation for differences in iconic strategies favoured by signers (whose languages are the outcome of generations of communicative pressures), and hearing non-signers using improvised gesture (whose strategies are not the result of such long-term adaptation to communicative pressures).

1.1 Iconicity

Arbitrariness has long been considered a defining feature of human language, with the vast majority of words taken to have no intrinsic connection between form and meaning (Hockett & Hockett, 1960). However, modern research into languages using the signed modality has revealed a set of fully functional languages with highly iconic lexicons (Sandler & Lillo-Martin, 2006), where the forms of many words are clearly motivated by properties of their referents. For example, about three quarters of the lexicon of American Sign Language (ASL) appears to be iconic, or to have iconic roots (Stokoe, 1965). Iconicity clearly constrains the range of forms we might expect for a sign, as for any given meaning the vast majority of possible forms will not be iconic. Despite this, there remains a wide variety of possible iconic forms for any single meaning, and many of the differences can be described in terms of iconic strategies.

1.1.1 Iconic Strategies

An iconic strategy is a certain method of conveying a meaning iconically, and the focus of this paper is on the iconic strategies used for manipulable objects (from here on, tools), such as a hammer, key, or comb. The most frequent iconic strategies for tools are the instrument strategy, where the signer's handshape imitates the shape of the object being referred to, and the handling strategy, where their handshape approximates the handshape used when handling the tool (Padden et al., 2013). Other less common strategies include tracing, where the outline of an object is 'drawn' in the air, and touch, where the hand touches a body part relevant to the object.

1.1.2 Patterning of Iconic Strategies

The choice of iconic strategy for a given word is not random, but has been shown to pattern with both the lexical category of the word, and its semantics, with this patterning also showing variation across languages. Padden et al. (2015) found that when describing tools, ASL signers primarily use the instrument strategy for nouns, and primarily use the handling strategy for verbs, even if the noun and verb involve the same object (e.g., the mop vs. to mop). Regarding semantics, Ortega & Özyürek (2020a) found a systematic connection between the iconic strategy used and the semantics of the thing being described, in the productions of silent gesturers (hearing participants prompted to communicate a meaning through gesture). For example, gesturers used the handling method over 75% of the time to convey tools, but only around 10% of the time to convey animate nouns, where they instead favoured the personification strategy. Between languages, preference for iconic strategies can vary drastically. While ASL signers use instrument about twice as often as handling for tool nouns, this pattern is reversed in New Zealand Sign Language (NZSL), where handling is used 67% of the time (Padden et al., 2013).

1.1.3 Iconic Strategies in Sign vs. Gesture

Despite this variation between sign languages, and across lexical categories, a consistent, unexplained, effect can be seen of native modality on iconic strategy used to convey tools. Specifically, hearing gesturers appear to use the handling strategy significantly more than sign language signers. Among the Al-Sayyid Bedouins of Israel for example, three quarters of non-signers use handling in silent gesture, compared to just a quarter of signers (Padden et al., 2013). In Ghana, both rural and urban non-signers prefer handling, while local sign languages have substantially more instrument signs (Edward, 2021). This pattern can be seen around the world: high rates of handling among hearing adults have been found among American (O'Reilly, 1995, p. 86), Dutch (Ortega & Özyürek, 2020a) and Mexican (Ortega & Özyürek, 2020b) gesturers. These contrast with comparatively lower handling preference found in ASL (Padden et al., 2013), Chatino SL (Hou, 2018), Dutch SL, Portuguese SL, and several African sign languages (Nyst et al., 2021).

For tool nouns, the lower handling rates in sign languages may be partly attributed to some languages exploiting the instrument strategy for nouns, with handling often reserved for verbs. However, even for verbs, the handling-preference is stronger for gesturers, with Padden et al. (2015) finding American gesturers using handling 91% of the time for verbs, compared to ASL signers who use it just 60% of the time.

Furthermore, there is evidence from emerging sign languages that the handling strategy drops in frequency over the course of a language's history. It has been argued that overall iconicity in sign languages declines over time as the language streamlines and becomes more systematic (Frishberg, 1975), however Pyers & Senghas (2020) demonstrate that this decline is not equally distributed across different iconic strategies. They compared the signs of first and third cohort speakers of Nicaraguan Sign Language, with the first cohort's signs taken to best reflect the language as it was in its earliest stage, and the third cohort's signs reflecting the influence of first language acquisition. They found that pantomime iconicity ('body-as-body', encompassing handling) diminished substantially by the third cohort, compared to perceptual iconicity ('body-as-object', encompassing instrument), which showed no significant change.

1.2 Research Question

This raises the question: why should established sign languages favour one iconic strategy significantly more than sign-naive gesturers when describing tools? Sign languages are known to have origins in the spontaneous creations of home signers, not unlike the spontaneous creations of silent gesture (Goldin-Meadow & Mylander, 1998), so a priori we may expect similar rates of strategies among signers and gesturers.

Various explanations have been suggested. Padden, et al. (2013, p. 303) speculate that a mix of instrument and handling forms emerges to expand the lexicons of sign language (compared to the primarily handling forms in silent gesture). Brentari et al. (2012, p. 16) link gesturers' high rates of handling with their lack of experience with the gesture-only medium, forcing them to fall back on handshapes they are familiar with (i.e., ones they actually use when handling the objects). Another possibility is that pressures affecting language change have an influence on iconic strategy. Various pressures are known to drive language change, for example pressures for language to be simple, expressive and easily processable (Slobin, 1977; Smith et al., 2012). In this paper I explore whether choice of iconic strategy is affected by linguistic pressures that are present in the evolution of sign languages, but are absent in the spontaneous productions of silent gesture.

Specifically, I hypothesise that iconic signs using the instrument strategy tend to be more effective at conveying a specific intended meaning than those using the handling strategy. Furthermore, I hypothesise that communication pressures present in sign language but not in silent gesture led to this more effective strategy increasing in frequency in sign languages. The higher rates of instrument in sign languages would therefore represent the cumulative output of generations of pressures to maximise communicative efficacy.

1.2.1 Instrument as the More Effective Strategy

There are a number of factors motivating the hypothesis that instrument handshapes would tend to be more communicatively effective. Firstly, it seems plausible that the set of distinct object shapes is greater than the set of distinct hand shapes while holding those objects (meaning handling handshapes would be more likely to be iconic of multiple different objects). Handshapes while holding an object are often very similar, for example a grasping gesture similar to the ASL C or O handshapes could easily represent the holding of a knife, hammer, or handbag, and a pinched thumb and index finger with the remaining fingers closed to the palm could likewise represent the holding of a piece of paper, a credit card, or a sewing needle (Sharma et al., 2019).

Furthermore, much of the distinctive potential in handshapes comes from extending different combinations of fingers, which is useful in instrument handshapes to emulate protruding features of an object (e.g., using the index and middle finger for the barrel of a gun, or the thumb and pinky for two ends of an old-fashioned telephone). In handling handshapes, this distinctive power is rarely useful, as handshapes while holding an object are rarely defined by distinctive finger protrusions. Eccarius (2008) found that sign language classifiers using instrument strategy tend to have more complex finger positions, and use a greater set of distinct finger positions, compared to classifiers using handling.[34]

Another potential constraint on the effectiveness of handling gestures is that many objects have a plurality of ways to be handled (e.g., a sheet of paper may be held in the hands, written on, folded, etc.), but only one shape. As a result, not only would a single handling gesture be likely to elicit multiple plausible semantic interpretations, but a single meaning could have multiple plausible handling gestures. With a weaker one-to-one link between form and meaning, handling handshapes are arguably a less effective iconic strategy.

1.2.2 The Efficacy of a Strategy and its Usage in Sign vs. Gesture

The second part of my hypothesis proposes that the higher communicative efficacy of instrument handshapes would provide a pathway to the attested higher frequency of instrument in sign languages, as compared to silent gesture. Specifically, I hypothesise that the pressure to communicate effectively present in the acquisition and transmission of sign languages would lead to the more communicatively effective strategy increasing in frequency.

Perlman et al. (2015) find, using a communication game, that pairs of interlocutors will settle on increasingly effective iconic (sound symbolic/onomatopoeic) words, when establishing a novel lexicon. Idiosyncratic, less easily interpretable iconic forms were ironed out as the pairs interacted and settled on a conventional form, leading to more effective communication. As well as interaction, the process of learning and transmission also favours more communicatively effective forms. In an artificial language learning experiment, Yin & White (2018) find that phonological rules that increase the amount of homophony in a language are acquired less successfully than those that do not increase homophony. If, as I hypothesise, handling handshapes have a more limited set of possibilities, and are more prone to multiple interpretations, this finding would fit well with the attested decrease in handling over emerging sign languages' histories. In spoken language, parallels can be seen for example in Mandarin, where disyllabic words have increased in frequency, as sound changes lead to an originally monosyllabic lexicon becoming too prone to homophony (Wen, 2012).

In line with this evidence, it follows that the lexicons of sign languages, which have experienced generations of interactive and learning pressures to be more communicatively effective, will have higher rates of the more effective iconic strategy. This contrasts with spontaneously-created silent gestures, which have experienced no transmission or learning pressures, and only involve fairly minimal pressure to communicate well. This is further supported by the fact that the few signs that silent gesturers do prefer to use instrument for are specifically those which are not spontaneously-created, but instead are drawn from conventionalised instrument forms in wider use. Padden et al. (2013) found the items eliciting the highest instrument frequencies from gesturers were scissors and telephones, both of which have conventionalised instrument forms (in the game rock-paper-scissors, and in the 'call me' gesture). The relatively high rate of instrument handshapes in conventionalised signs used in wider society (such as telephone, gun and all three gestures in rock-paper-scissors) provides further support for my hypothesis, considering they, like sign language signs, have emerged with the pressures of cultural transmission and communication.

1.3 Experiments

[34] Although Brentari et al. (2012) find the opposite pattern in terms of finger complexity among silent gesturers, with handling gestures actually using more complex finger arrangements than instrument. They attribute this to silent gesturers using more abstract forms borrowed from co-speech gesture when pushed to use instrument. Regardless, it seems plausible that the instrument strategy allows for a wider range of more complex iconic forms, even if this isn't utilised in silent gesture.

To explore my theory, I ran two experiments. The first experiment explored the hypothesis that use of instrument gestures is more communicatively effective, due to reduced risk of polysemy (one gesture iconically representing multiple meanings). To test this, I presented participants with either an instrument or handling gesture, and asked them what they thought it represented. I predicted that instrument gestures would elicit answers closer to the intended meaning.

The second experiment explored the hypothesis that the pressure to communicate effectively would lead to increased instrument use. To test this, another set of participants were shown both an instrument and handling video for a stated meaning, and asked which was most suitable for the meaning. Participants were assigned to one of two conditions, one with and one without a verbal communication pressure. I predicted that the communication pressure would lead to increased selection of instrument videos, in support of the hypothesis that instrument gestures will be favoured in a communicative context.

2. Experiment 1

The first experiment tested for an effect of iconic strategy (handling vs. instrument) on sign-naive participants' ability to guess a gesture's intended meaning.

2.1 Participants & Ethics

Participants for the first experiment were 491 English-speaking American adults, all of whom were recruited using Prolific (www.prolific.co). To avoid responses being influenced by the similarity of the gesture being viewed to some known word in a sign language, proficient users of sign languages were excluded. This led to the exclusion of data for 10 participants. Of the remaining participants, the majority were female (57%, 276/481) and the average age was 39. All participants granted consent before participating, and were paid £0.40, with the experiment lasting approximately 2 minutes.

Funding was supplied by the University of Edinburgh, and ethical approval was granted by the University of Edinburgh PPLS Research Ethics Committee.

2.2 Stimuli

The stimuli for both experiments were videos of handling and instrument gestures for 15 handheld objects (resulting in 30 videos in total). All gestures were performed by the same 21-year-old woman, in identical lighting. This section describes the process of selecting which objects to portray, and deciding how to portray them.

2.2.1 Object Choice

A number of factors went into deciding which handheld objects to portray in the videos. Ideally, I wanted to produce iconic gestures close to those used in both sign language and silent gesture, so words with attested forms in both mediums (for both instrument and handling strategies) were favoured. Sign language forms were found in a number of sources, including spreadthesign.com, handspeak.com, Kimmelman et al. (2018) and Padden et al. (2013). Silent gesture forms were found in Ortega & Özyürek (2020a), Padden et al. (2013), Padden et al. (2015), Edward (2021) and Ortega et al. (2019). Words for which a conventional iconic form is widely known were not used (such as phone, gun, and book), as the interpretation of these gestures would be strongly affected by participants' familiarity with them. Finally, only words for which reasonable and distinct iconic forms could be created were used.

For example, ‘cake’ was considered, but no instrument form was considered sufficiently iconic. Similarly, while ‘cup’ easily allows iconic instrument and handling forms, the two forms are not easily distinguished, so cup was not used.

The 15 handheld objects chosen were comb, drill, facemask, fork, hammer, key, knife, lipstick, paintbrush, saw, screwdriver, shovel, spoon, stapler and toothbrush. Data for one of these (saw) had to be excluded from some analyses due to issues with measuring semantic similarity, as most instances of ‘saw’ that defined its semantic vector were the past tense of ‘see’, not references to the tool.

2.2.2 Video Creation

When creating the videos, I tried to be faithful to the attested sign and gesture forms, however some alterations were made. For the most part, signed and gestured forms were very similar, meaning the videos were not influenced more by one than the other. Features in the attested signs/gestures that were not part of the handling or instrument iconic form were not included. For example, many signs/gestures for the word ‘key’ involved a second hand held up to represent a door, and this was not included in my videos.

To ensure that our videos were ‘minimal pairs’ between handling and instrument handshapes, each video for a given object had essentially the same hand location and movement, with the only difference being the handshape. No non-manual actions (e.g., facial expressions/mouthing) were included.

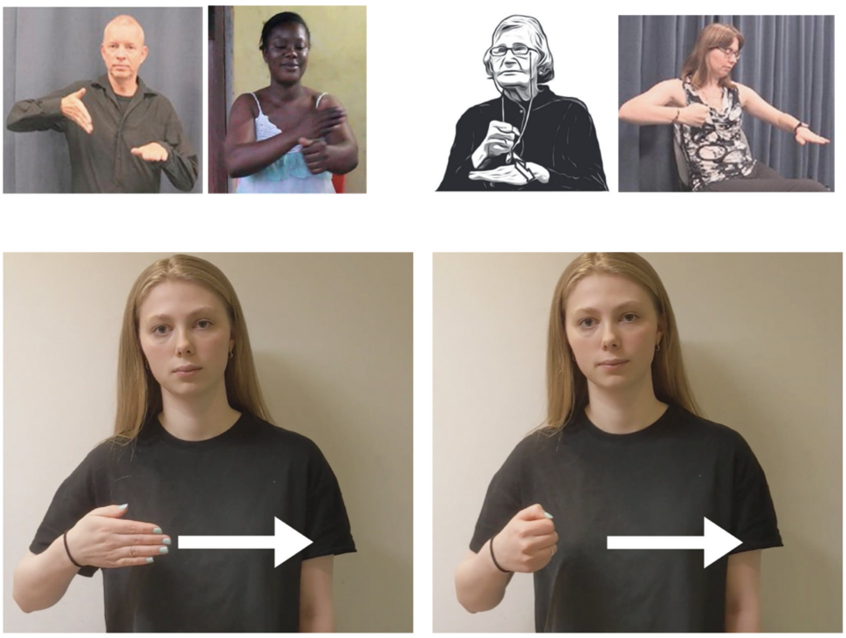


Figure 1: *Example of stimuli creation for the meaning ‘saw’; top row shows attested signs and gestures, bottom row shows the stimuli created for the present study. Top four examples, left-to-right, are from Ortega et al. (2019), Edward (2021), Padden et al. (2013), and Ortega et al. (2019).*

Each of the 491 participants were randomly assigned a single video (for example, a video of ‘hammer’ being gestured using the instrument strategy), and were directed to a Google Forms survey for that video. In the survey, the assigned video would play 3 times in a row, after which the following text would appear:

> *What object do you think she was trying to communicate? List any objects you can think of.*

After 7 seconds, the video would repeat again. Below, participants were able to input their answer in a text box.

2.4 Analysis

The primary analysis for the first experiment tested for an effect of iconic strategy on how close responses were to the intended meaning, with the hypothesis that instrument gesture videos would elicit closer responses.

2.4.1 Semantic Similarity

To ensure that a close but not perfect answer (for example 'chapstick' in response to a lipstick prompt) was considered more effective than a completely incorrect answer, a measure of semantic similarity was needed. For this, I used GloVe (Pennington et al., 2014), which generates semantic vectors for words, based on their occurrence in corpora. Specifically, I used the pre-trained Common Crawl (840B tokens) vector list.

Using R (R Core Team, 2013) and the lsa package (Leydesdorff, 2005), cosine similarity between two vectors was used to measure semantic similarity (s), giving values between -1 and 1. For example, compared against the meaning 'comb', a correct answer ('comb') returns a value of s=1, a close answer like 'brush' returns a value of s=0.592, and a semantically distant answer like 'chair' returns a value of s=0.175.

2.4.2 Data Cleaning

Before any analysis could be run, the raw response data, compiled into a single .csv, had to be cleaned, in order to work with the semantic similarity measure (which accepts only single words). This included many trivial changes, particularly removing articles and simplifying descriptions, such as changing 'a pair of scissors' to 'scissors', as well as some spelling corrections (e.g., 'ladel' → 'ladle'). The experiment prompt asked for the names of objects, so if a response described an object being used, this was changed to the name of the object on its own (e.g., 'put on a mask' → 'mask'). A number of responses had to be excluded altogether, as they did not refer to an object in any way (e.g., 'a welcome gesture', 'go away'). Finally, a very small number of object names had to be excluded due to not being present in the semantic vectors list, for example 'bubblewand'.

2.5.1 Similarity to Intended Meaning

To compare the similarity of guesses to the intended meanings of instrument and handling stimuli, we ran a mixed-effect regression analysis using R with lme4 (Bates et al., 2015). We tested for an effect of iconic strategy on the semantic similarity of participant responses to the target meaning, including random intercepts for participant, and random intercepts and slopes for the target meaning. The regression did not produce significant results (β= -0.075, SE = 0.052, t = -1.45, p = 0.17, baseline = handling), and therefore did not provide evidence for a communicative advantage of instrument gestures. In fact, handling gestures elicited responses slightly (but not significantly) closer to the target than instrument gestures (Fig. 2).

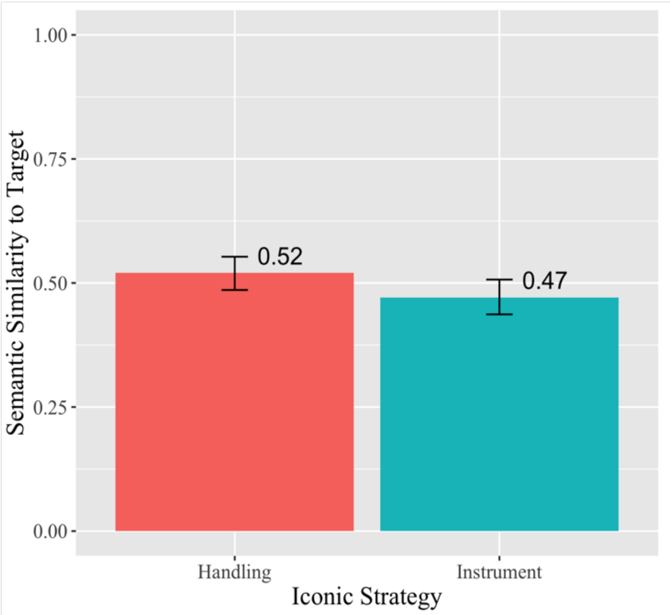


Figure 2: *Average semantic similarity of responses to the intended meaning, for handling (red, s=0.52) and instrument (blue, s=0.47) stimuli. Note that the y-axis begins at 0, although technically values below 0 are possible, however this is extremely rare (3/667 in my data).*

While these results do not support an overall advantage of either strategy, some words did show a significant difference between strategies. Responses to 'facemask' stimuli were significantly closer to the intended meaning when participants viewed the instrument strategy, while 'hammer' and 'toothbrush' performed significantly better with the handling strategy (Fig. 3).

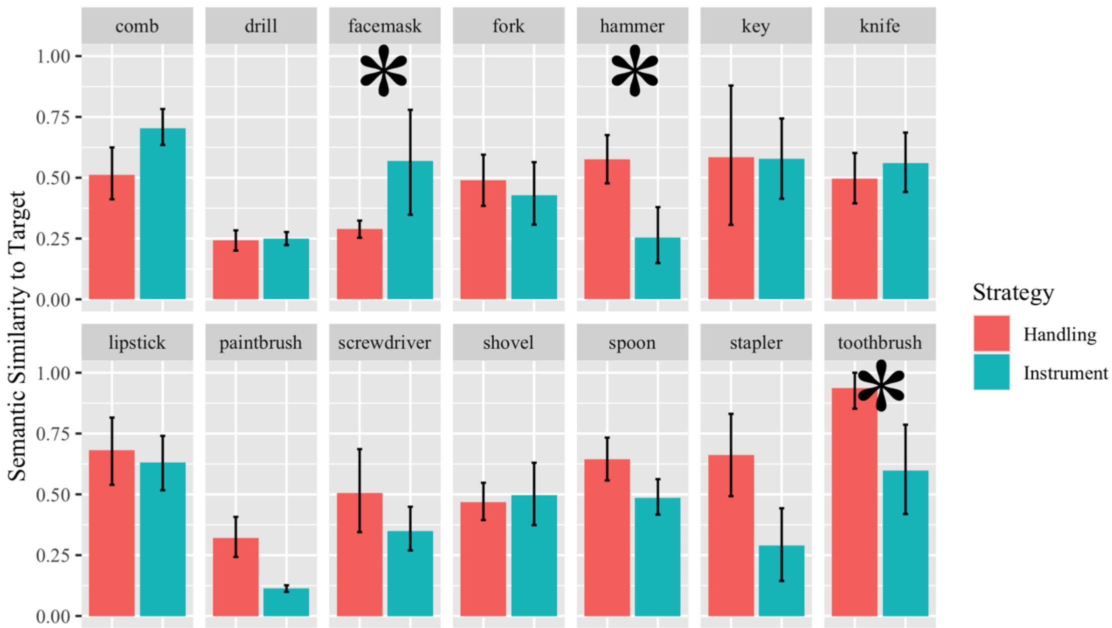


Figure 3: *Average semantic similarity of responses to intended meaning, comparing iconic strategies of stimuli. Each graph shows the results for a different intended meaning. Meanings with a significant difference between strategies (with Bonferroni correction; α=0.0036) are marked with an asterisk*

2.5.2 Inter-Participant Similarity

The analysis above tested whether instrument gestures elicited more accurate guesses. However, to test the effectiveness of the strategies in conveying a specific meaning, it's also worth looking at how similar participants' responses were to each other. To investigate this, I ran an exploratory analysis to measure the consistency between guesses, regardless of how close they were to the target meaning.

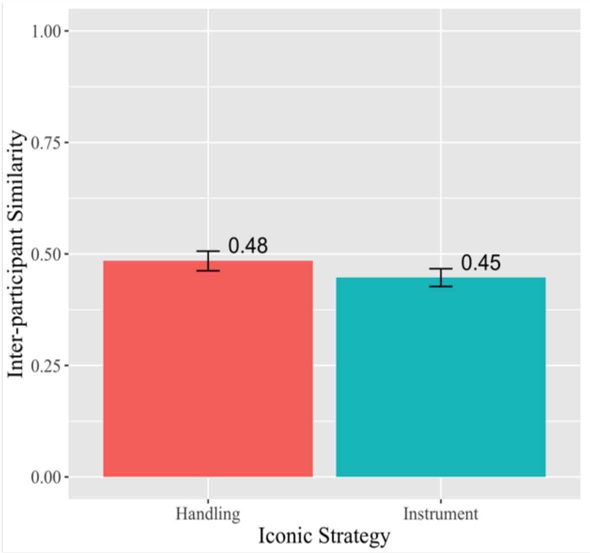


Figure 4: *No significant difference was found between strategies, in terms of similarity between participants' responses.*

For each response, I calculated the average semantic similarity to all other responses for that stimulus. On average, handling gestures elicited slightly more similar responses than instrument gestures (Fig. 4). However, a mixed-effect regression found no significant effect of strategy on inter-participant similarity (β= -0.057, SE = 0.05, t = -1.14, p = 0.272, baseline = handling), with random intercepts for participant, and random intercepts and slopes for the target meaning.

Several words showed a significant difference between strategies. For 'comb' and 'hammer', instrument stimuli prompted greater inter-participant similarity, while for 'shovel', 'stapler', and 'toothbrush', handling stimuli did so, demonstrating variation in which strategy prompts greater consistency in interpretation.

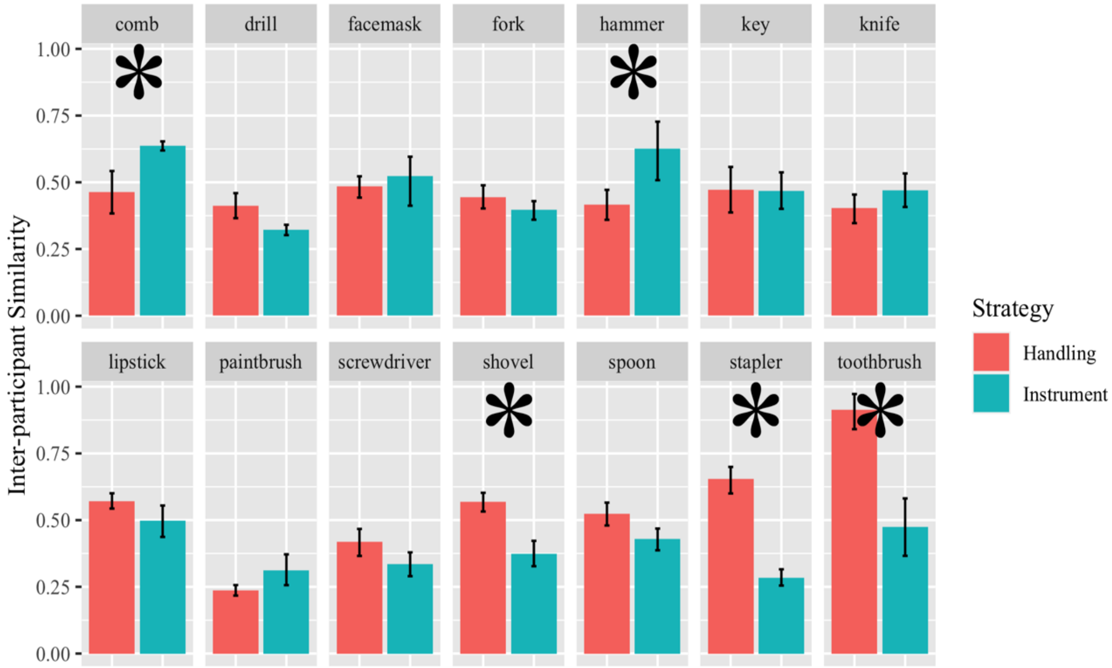


Figure 5: *Inter-participant similarity by word and iconic strategy. Meanings with a significant difference between strategies (with Bonferroni correction; α=0.0036) are marked with an asterisk.*

Notably, 'hammer' prompted responses significantly closer to the intended meaning when using the handling strategy, but prompted a significantly narrower range of responses with the instrument strategy. This appears to be due to the similarity of the instrument 'hammer' handshape to the typical 'telephone' gesture (Fig. 6), as 87.5% of respondents for this stimulus gave 'phone' or 'telephone' as a possible meaning.

Figure 6: *Comparison of the instrument 'hammer' stimulus (left) with a typical sign for 'telephone', in this case from Portuguese Sign Language (spreadthesign.com)*

2.5.3 Successful Communication

In actual communication, a moderately close-to-intended interpretation of a signal is arguably not much better than a semantically distant interpretation. For example, if one person is trying to communicate 'hammer', an interpretation of the signal as 'pen' (s = 0.3) is not much more use than an interpretation as 'coffee' (s = 0.16), as both would have failed to communicate the intended meaning. This motivates a further exploratory analysis into whether there was an effect of strategy on whether people were successful in guessing the intended meaning, and whether there was an effect on whether participants would be successful in communicating with each other.

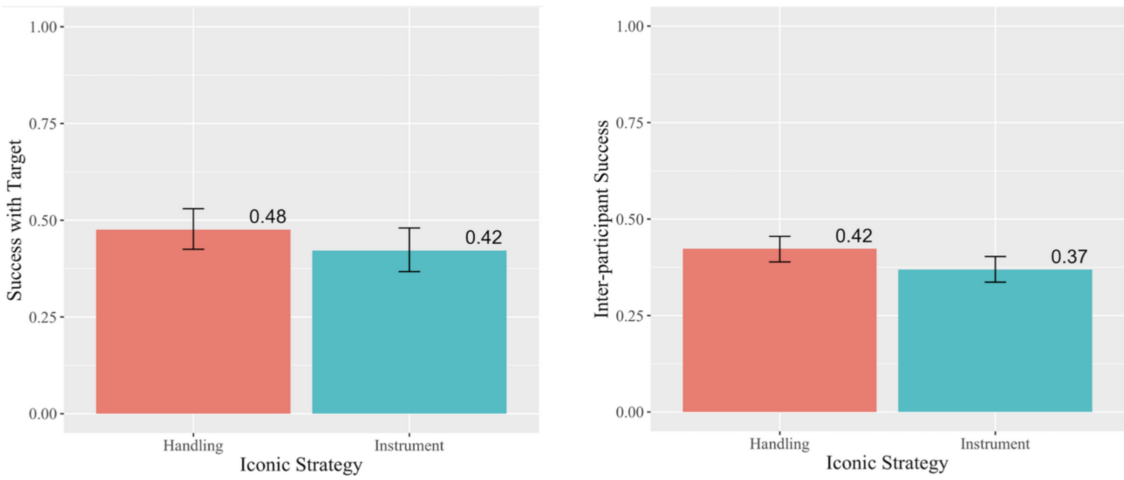


Figure 7: *Proportion of guesses successfully guessing the target meaning by strategy (left), and average chance of success between participant responses (right), where success is defined as s > 0.45.*

I defined successful communication as a pairing of meanings with a semantic similarity (s) above 0.45. For example, 'shed' - 'house' would be successful (0.47), as would 'mouse'-'rat' (0.62), while 'mouse'-'cat' would not be successful (0.43), nor would 'laptop'-'tree' (0.14). An alternative method, allowing success only for identical words, has the issue of not accepting synonyms or near synonyms (such as 'phone' and 'telephone'), and the 0.45 cut-off allows for a small amount of distance between meanings, where communication is unlikely to be seriously inhibited.

As with the average semantic distance measurements, no significant difference was found between conditions. A mixed effects logistic regression found no effect of strategy on whether participants gave a successful response, including a random intercept for participants, and random intercept and slope for the meanings being conveyed (β= 0.47, SE = 0.48, t = -0.99, p = 0.32, baseline = handling). Likewise, the same model found no effect on the chance of successful communication between participant responses (β= 0.10, SE = 0.08, t = -1.25, p = 0.23, baseline = handling).

2.5.4 Number of Responses

A final exploratory analysis tested for an effect of iconic strategy on the number of responses participants gave. If one strategy elicited fewer responses, it may suggest that strategy conveys a narrower range of plausible interpretations for any given person, even if there's diversity of interpretations between people.

However, a mixed effects Poisson regression found no effect of iconic strategy on the number of responses given, including random effects for stimulus meaning (β= 0.0083, SE = 0.081, z = 0.10, p = 0.92, baseline = handling).

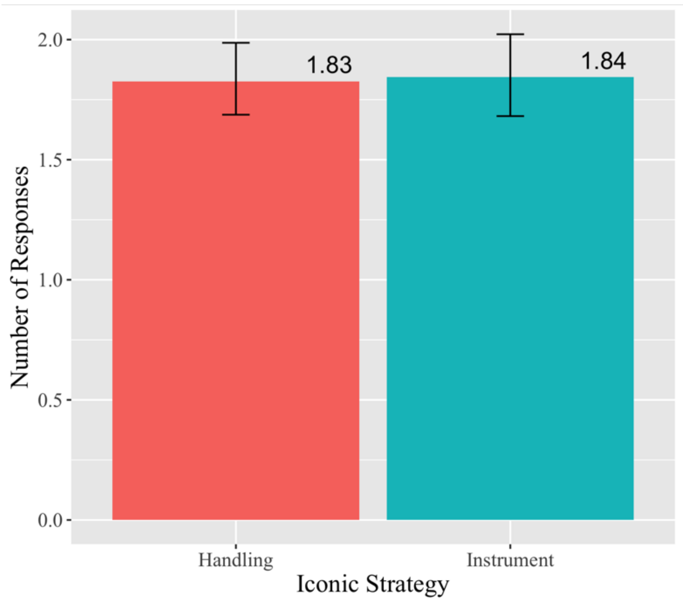


Figure 8: *Average number of responses given for handling stimuli (red) and instrument stimuli (blue).*

2.6 Discussion

The main conclusion from the analyses above is that the data does not provide evidence for a difference in communicative efficacy between the two iconic strategies. Guesses for instrument stimuli had an average semantic similarity of 0.47 with the target meaning, which is not significantly different from handling stimuli (0.52). Furthermore, exploratory analyses found that the chance of successful communication between participants was not significantly different between the two strategies, nor was the number of meanings participants proposed.

Does this suggest that iconic strategy is not relevant to how people interpret a gesture? After all, the instrument-handling pairs were essentially minimal pairs, with the same hand location and movement, differing only in handshape — could it be that the iconicity of location and movement were enough to make the difference in strategy negligible? Probably not. While it's possible the iconicity of strategy-neutral features of the gestures does increase the uniformity of their interpretations, the results do reveal differences in interpretation: for 21% (3/14) of the words, there was a significant difference between strategies in how close guesses were to the target, and for 36% (5/14), there was a significant difference in how semantically varied the responses were.

It seems possible then, that with a larger sample of objects, a significant overall effect of strategy might be revealed. However, the results in Padden et al. (2013) suggest we should not need a particularly large sample for a clear difference to be revealed (if one exists). Their study found a drastic difference in strategy usage between signers and gesturers, despite also using relatively few words (24), ten of which were also used in the present study.[35] If that substantial difference in usage is a result of a difference in communicative efficacy, we would expect to see that difference in efficacy in the present study, even with a fairly small sample of words.

[35] The overlapping words were fork, knife, scissors, saw, screwdriver, paintbrush, comb, hammer and lipstick.

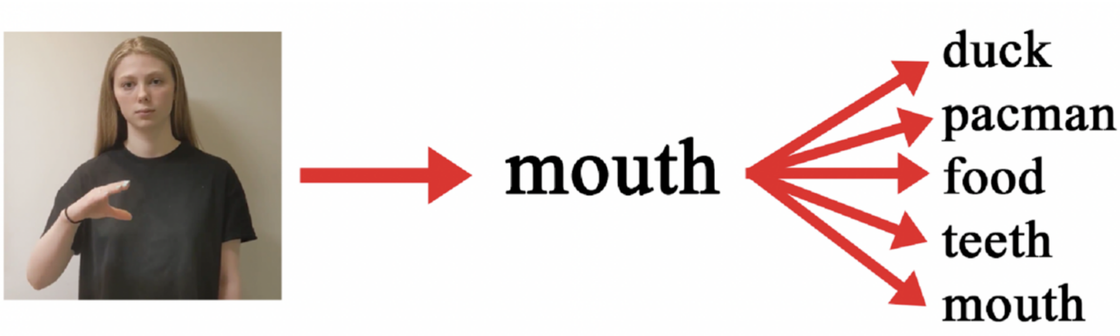


Figure 9: *A wide semantic range of responses (right) can be elicited from a single iconic interpretation (middle) of a gesture (left).*

Some interesting observations can be made from the results about the nature of iconic interpretations. The hypothesis that instrument would be more successful in Experiment 1 was largely based on the assumption that it would, on average, have a narrower range of possible interpretations. The example of 'hammer' though, shows that closeness of answers to the target does not necessarily imply a narrower range of interpretations: as mentioned before, most instrument answers for 'hammer' were incorrect, but had high internal success, due to the frequency of the 'telephone' interpretation. This motivated an exploration of inter-participant communicative success.

However, even when an instrument gesture is interpreted as embodying the same object by many participants, this does not necessarily result in high rates of inter-participant success, as the same interpretation may be extrapolated to a wide variety of responses (Fig. 9). For the instrument video for 'stapler', two thirds of responses seemed to interpret the gesture as embodying a mouth. Recognising the selective nature of iconicity (iconic signs will choose a salient element of the referent, as opposed to capturing the entire thing; Perniss et al., 2010), participants extrapolated a range of meanings from 'mouth', including things with a mouth (animals), things inside a mouth (teeth), and things that go in a mouth (food). This may partially explain why instrument gestures did not perform better than handling gestures: even if handling gestures are prone to a wider range of immediate iconic interpretations (since many things may be handled similarly), the same may be true for many instrument gestures due to a second layer of interpretation, where the object being embodied is assumed to represent a salient feature of a wide range of things.

3. Experiment 2

The second experiment explores the role of communication in driving choice of iconic strategy, specifically testing whether iconic strategy preference differs depending on whether participants are exposed to a prompt simulating a communication pressure.

3.1 Participants & Ethics

Experiment 2 involved 470 English-speaking American adults, recruited using Prolific. All participants granted consent before participating, after which 10 expressed proficiency in one or more sign languages and their data was excluded from the analyses. Of the remaining participants, 65% were female (298/460) and the average age was 34.

The experiment lasted approximately 1.5 minutes, and participants were paid £0.30. Funding was again supplied by the University of Edinburgh and ethical approval was granted by the University of Edinburgh PPLS Research Ethics Committee.

3.2 Stimuli

The same video stimuli as in Experiment 1 were used. These were handling and instrument strategies for the objects comb, drill, facemask, fork, hammer, key, knife, lipstick, paintbrush, saw, screwdriver, shovel, spoon, stapler and toothbrush. Unlike Experiment 1, data for the 'saw' stimuli was not excluded from the analyses, as the Experiment 2 analyses required no semantic similarity measure.

3.3 Procedure

Each participant was randomly assigned to one of the 15 objects (e.g., comb). The experiment was run on Google Forms, and at the beginning of the experiment participants were shown both videos for their assigned word (one video using the instrument strategy, and one using the handling strategy, with the order randomised for each participant). Participants were also randomly assigned to one of two conditions: a pressure condition, and a no pressure condition. After both videos had played, a verbal prompt appeared. In the no pressure condition, the prompt stated (here, for the object 'comb'):

> *Consider how you would use your hands to express the meaning "comb". Which video (A or B) best matches your own gesture?*

In the pressure condition, the prompt instead said:

> *You have just watched 2 videos (A and B). One of these will be shown to another participant, with the intention of communicating the meaning "comb". Which video do you think will have the best chance of successfully communicating this meaning?*

Participants were instructed to respond in an answer box below, typing 'a' to choose the first video they saw, and 'b' to choose the second. After 7 seconds, the videos would repeat, again followed by the appropriate prompt.

3.4 Analysis

The primary analysis tested for a significant effect of pressure condition (whether the prompt simulated a communication pressure or not) on preferred iconic strategy, with the hypothesis that communication pressure would increase preference for instrument gestures. Responses inconsistent with the instructions (neither 'a' nor 'b') were removed, bringing the number of valid responses to 456.

3.5 Results

3.5.1 Effect of Pressure Condition

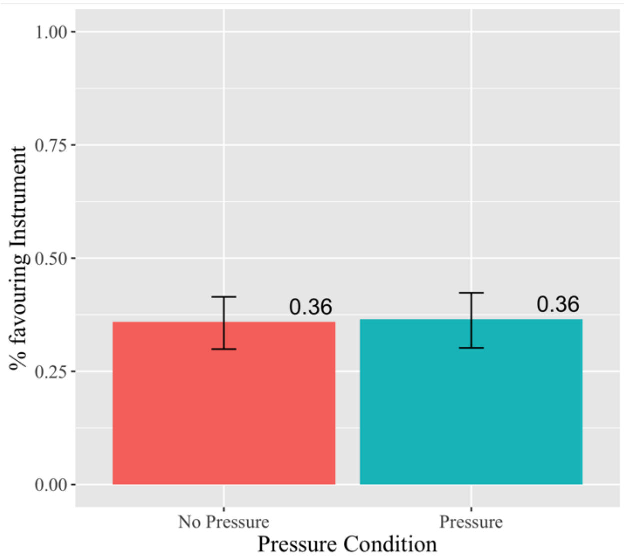


Figure 10: *Proportion of Experiment 2 participants favouring the instrument strategy for their word, with (blue) and without (red) a verbally-induced communication pressure.*

Using a linear mixed model with random intercepts and slopes for meaning, I found no effect of pressure condition on the iconic strategy participants preferred (β= 0.08, SE = 0.216, z = 0.35, p = 0.73, baseline = no pressure). 36.5% (81/222) of participants in the pressure condition favoured the instrument option, compared to 35.9% (84/234) in the no pressure condition. Furthermore, none of the 15 words displayed a significant difference (using Bonferroni correction on the critical value) between the two conditions. These results suggest that verbally inducing a communication pressure does not affect people's preferred iconic strategies. In both conditions, participants showed a significant preference for handling stimuli (binomial tests: $p < 0.0001$).

3.5.2 Variation Between Words

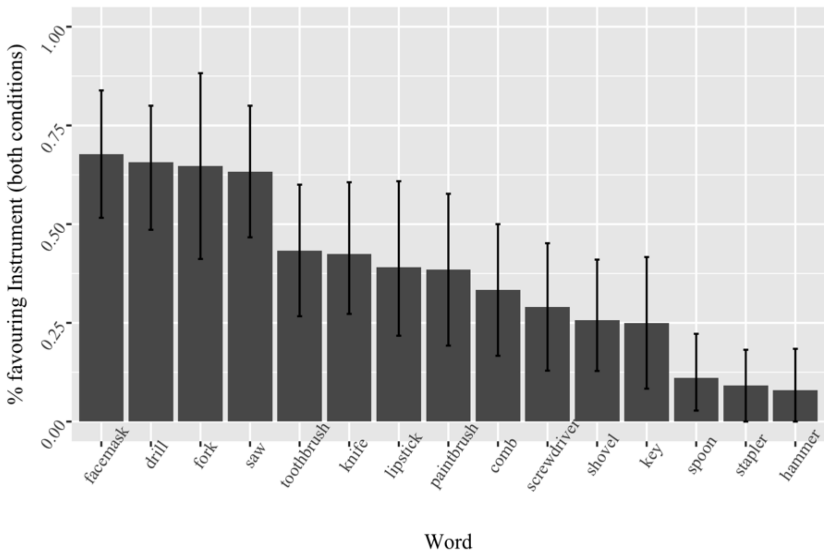


Figure 11: *Preference for the instrument strategy for each words (data from both pressure conditions combined).*

There was substantial variation between words. Across conditions[36], the proportion of participants favouring the instrument strategy ranged from 'hammer', with 8% to 'facemask, with 68%. The majority of words (11/15) were preferred in their handling form.

Interestingly, the words with the highest agreement ('spoon', 'stapler' and 'hammer', all overwhelmingly handling) also had comparatively high guessing success for the handling strategy in Experiment 1 — 69%, 80% and 51%, compared to an average of 48%. This is consistent with recent research (Fay et al., 2022) showing that the universality of an iconic gesture (how consistently people represent it in the same way) is associated with greater communicative success.

3.5.3 Comparison with Attested Strategy Frequencies

As an exploratory analysis, the strategy preferences for the different words were also compared with the strategies used in real life sign languages. Data on sign language strategy use was gathered from spreadthesign.com, a website providing a searchable database of signs for a wide range of sign languages. Strategy judgements were made by myself for 19 languages, making up all those languages which had a sign video for at least 80% of my 15 words.[37]

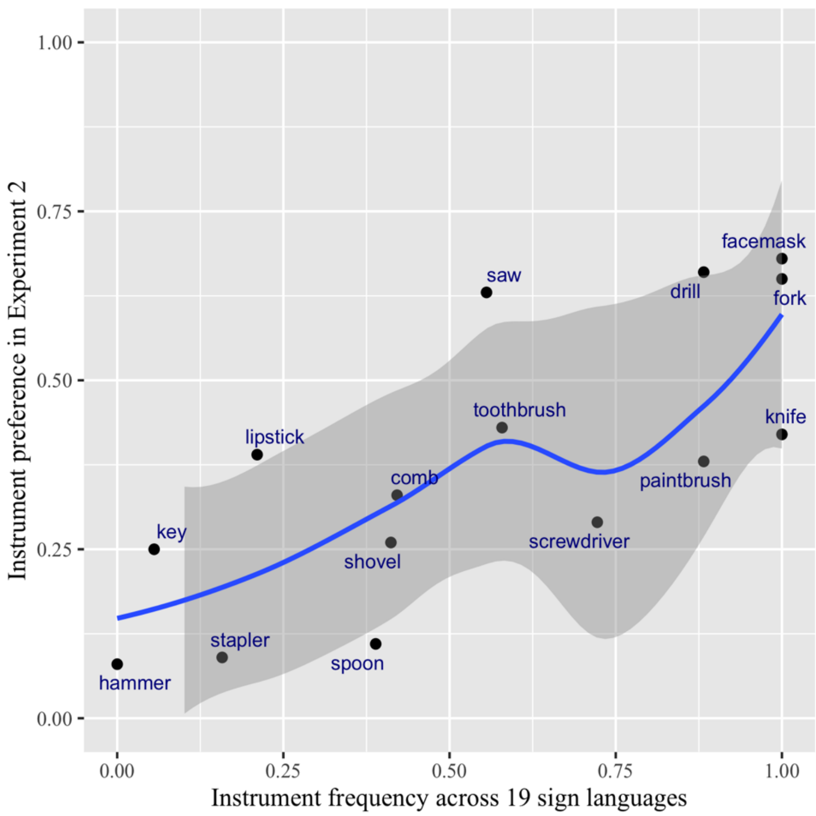


Figure 12: *Strategy preference among non-signers (Experiment 2, both conditions) correlates with strategy frequency in established sign languages. The blue line is a local regression line, with shaded 95% confidence interval.*

[36] Since no effect of condition was found for any of the words, this section looks at the combined data from both conditions.
[37] The 19 languages were the national sign languages of the UK, Spain, France, Germany, Lithuania, Russia, Poland, Iceland, Estonia, Austria, Latvia, Italy, Denmark, Croatia, Portugal, Turkey, Pakistan, China, and Argentina. The meaning "saw" had very few videos available, so was supplemented with videos for "handsaw".

Unsurprisingly, instrument use in the 19 sign languages was higher than instrument preference among the non-signers in Experiment 2. Excluding the 4% of signs where the iconic strategy was not clear, or was neither handling nor instrument, the 19 sign languages used instrument on average 50% of the time (ranging from 29% in German Sign Language to 71% in Estonian Sign Language). This is substantially more than the 36% instrument preference in Experiment 2. Looking at individual words, a moderate correlation can be seen (Fig. 12) between participant preference and real-life frequency, with increased instrument preference in Experiment 2 being associated with higher rates of instrument use in the sign languages sampled.

3.6 Discussion

There are three main findings from Experiment 2: participants prefer the handling strategy, a verbal communication pressure doesn't affect their preference, and preference varies substantially across different words, in a similar pattern to real sign languages.

The first finding is for the most part expected. Past research has found that non-signers tend to consider handling gestures to be more iconic than instrument gestures (Sehyr et al., 2017), so a general preference for handling gestures in a forced choice experiment is not surprising. It's worth noting that higher perceived iconicity does not necessarily mean a gesture is preferable in a communicative setting, as a highly iconic sign may also be iconic of several other objects, and thus be less useful for communication within an iconic lexicon (see Fig. 14 in General Discussion).

The second finding is that strategy preference did not differ between the two conditions (p=0.73). A major limitation on the present experiment is whether the verbal prompt was enough to simulate a true communication pressure. A variety of strategies have been use to induce a communication pressure in language experiments, with the gold standard being real communication between participants. For example, Tamariz et al. (2018) explored the effect of interaction on the emergence of sound symbolism, using separate communication and no communication chains. The communication chains involved direct interaction between participants, with successful communication rewarded by a point in the game. In Kirby et al. (2008), the pressure to communicate/be expressive was simulated by a 'filtering' process between iterated learning generations. If a participant in one generation assigned multiple meanings to the same signal, the next generation would only be trained on one of those signal-meaning combinations essentially simulating the need to distinguish different meanings when communicating.

In the Tamariz et al. (2018) example, we can be fairly confident in the simulation of real world communication, since there was actual interaction (even if the stakes were different to a typical communicative setting). The Kirby et al. (2008) example is a more contrived simulation, but still forces an effect (since polysemous signals are forcibly not transmitted). In the present study, however, effecting a communication pressure relied on participants interpreting the prompts as intended. In the no pressure condition I tried to elicit the strategy closest to participants' natural productions, while in the pressure condition I tried to elicit the strategy participants believed to be most effective. While it is possible that people's natural productions are the productions they consider most effective, the closeness of the two values (both 36% instrument) does suggest the difference in prompt may not be effective. An improvement to the present study could therefore make use of real productions and real communication, for example seeing if silent gesturers produce iconic strategies at different frequencies when they are communicating with another gesturer.

It remains possible, however, that the different prompts did work as intended, but there really is no difference between people's natural strategy and the strategy they consider most likely to communicate well. This could be because people naturally produce the strategy they consider most effective. Or, going the opposite way, people may assume the strategy they find most natural will also be the most natural for another person (an assumption that, more often than not, will be correct for a binary option), and thus consider it the most effective to communicate the meaning to that other person.

A final finding from Experiment 2 was a drastic difference in strategy preference between words, and an exploratory analysis found that strategy preference among the non-signers correlated with strategy preference in a sample of sign languages. A possible explanation for this is that, whatever the factors are that drive increased instrument use in sign languages, they are not strong enough to fully wash away the lexical patterning of iconic strategies of the sign languages' original spontaneous creations. Alternatively, the factors driving strategy usage over a sign language's evolution might be the same factors motivating people's decisions in Experiment 2 (for example, maximising iconicity or ease of production).

4 General Discussion

The signed modality of human language is notable for its exceptional ability to represent concepts iconically, and this paper aimed to explain why this ability is exploited in different ways by signers and gesturers. Specifically I aimed to explain past observations, especially in Padden et al. (2013), that deaf sign language users and hearing gesturers differ significantly in the iconic strategies they use to represent handheld objects. Gesturers by and large use the handling strategy, shaping their hand as though they were holding the object in question. In comparison, signers are much more likely to use the instrument strategy, shaping their hand to mimic the shape of the object.

To explain this, I hypothesised that instrument gestures are advantageous in a communicative setting due to a reduced risk of multiple iconic interpretations. Noting that use of the instrument strategy increases gradually against the handling strategy in emerging sign languages (Pyers & Senghas, 2020), I hypothesised that this communicative advantage leads to instrument gestures being favoured in a communicative setting, leading to progressively increased instrument usage in sign language communities.

The results of my experiments do not support these hypotheses. In Experiment 1, I found no effect of iconic strategy on people's success in guessing a gesture's meaning, and exploratory analyses also found no effect of iconic strategy on similarity between participant responses, nor on the number of meanings participants proposed. This suggests, at least for the objects tested, that instrument was not on the whole more (or less) communicatively effective. Similarly, Experiment 2 found no significant effect of communication pressure on the strategy participants favoured — when prompted to choose the strategy most like what they would produce themselves, they chose the instrument form at exactly the same rate (36%) as when prompted to choose the strategy most likely to convey the meaning effectively. This may suggest strategy usage is not driven by communicative pressures.

4.1 Potential Issues

There are a few issues that should be kept in mind when interpreting the results of these two experiments.

The first is the role of level of iconicity in driving people's decisions: while all of the gestures were iconic of their intended meanings, the spectrum nature of iconicity means some gestures may be considered more iconic than others (Dingemanse et al., 2015). Experiment 1 aimed to capture the (hypothesised) greater range of possible interpretations that handling gestures have by eliciting possible meanings for gestures of each strategy, but the usefulness of this method have been hindered by participants only responding with the most obvious, or most iconic option. For example (Fig. 13), suppose the concepts ‘comb’ and ‘hairbrush’ have identical handling forms, but distinct instrument forms. This would make handling a less effective strategy for conveying these tools. However, if participants are shown the handling and instrument strategies for ‘comb’ (A and B in Fig. 13), this difference in efficacy would not be captured if participants who viewed the handling gesture only respond with the form considered more iconic.

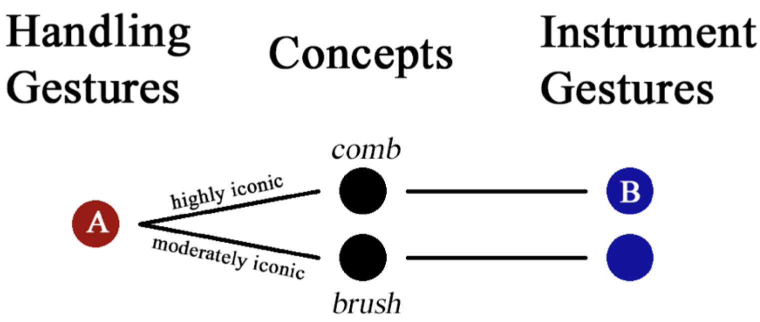


Figure 13: *Hypothetical gesture-meaning connections for ’comb’ and ’brush’. The common gesture for both ‘comb’ and ‘brush’ would make the handling strategy less effective, however if the handling stimulus only elicits the meaning it is most iconic of, Experiment 1 will not find a difference in efficacy between the strategies.*

To some extent this issue is resolved by allowing multiple responses, as participants were not totally constrained to only the most iconic/obvious option. However, the problem can still be seen, for example in the data for ‘hammer’ as described in Chapter 2. Despite the instrument form being iconic both of ‘hammer’ and ‘telephone’, the inter-participant similarity was very high (higher than the handling form), due to most people responding only with the latter interpretation.

A similar issue regarding iconicity arises in the second experiment, which aimed to elicit more communicatively effective forms when a communication pressure was present. Even if we assume the verbal prompt was enough to simulate a pressure, it is possible people reacted to the pressure in terms of iconicity, rather than communicative efficacy (Fig. 14). Non-signers typically consider handling gestures to be more iconic than instrument gestures (Sehyr et al., 2017), and in a forced choice format like Experiment 2, participants may select the most effective form based on which they deem to be more iconic, even if that form is at higher risk of misinterpretation due to also being iconic of other meanings.

A further drawback of Experiment 2 is that its format relies on the increase in instrument use being the result of conscious decisions, with participants being aware of the most effective strategy, and changing their strategy use in a communicative setting in response. However, it is also plausible that an increase in effective forms could be mediated through the language acquisition process, resulting from unconscious biases such as homophony avoidance (Yin & White, 2018). If this is the case, an experiment format that relies on the active decisions of participants may not be effective.

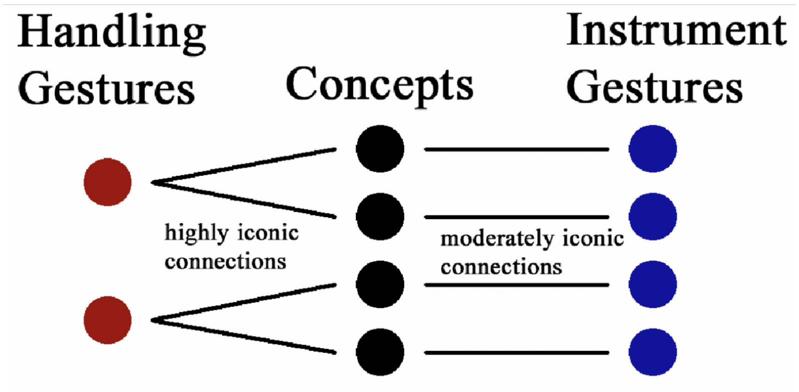


Figure 14: *Illustrating the difference between higher iconicity and higher communicative efficacy. To discuss the concepts (black circles), the handling gestures (red circles) may be more iconic than the instrument gestures (blue circles), but having the same iconic form for multiple meanings makes the handling strategy less effective, in terms of accurately communicating a meaning.*

Some alterations to the experiment methodologies could resolve the issues described above. To some extent the problems are a result of the role of the participants as interpreters (viewing videos and making judgements) rather than producers (creating their own signs), and production-based (silent gesture) experiments may therefore be more effective. In Experiment 1, evidence for weaker communicative efficacy in handling gestures may have been obscured by participants only suggesting the most iconic interpretations. A production-based alternative could involve participants creating their own iconic gestures for a selection of words, using either handling or instrument, to investigate whether the handling gestures they create are more similar to each other than instrument gestures. Similarly for Experiment 2, participants could be asked to produce gestures either on their own or in a real interactive context (i.e., with another participant). The presence of real communication would not only ensure an actual distinction between the two conditions is realised, but would also motivate participants to gesture in a way that is communicatively effective, not just highly iconic. Finally, to explore the effect of transmission (instead of just interaction) on strategy use, an iterated learning experiment could be run: teaching an initial generation of participants a selection of meaningful gestures, some of which have multiple meanings, would successive generations learning the gestures reduce this polysemy by making use of different iconic strategies?

4.2 Alternative Explanations

Regardless of some methodological concerns, the experiment results do not provide any evidence for my hypothesis, that a communicative advantage of the instrument strategy drives its higher use in sign languages. This motivates the present section, where I consider some alternative explanations for the different use of iconic strategies between signers and gesturers.

4.2.1 Instrument preference as a relic of home sign

One alternative explanation for the difference in iconic strategy usage between signers and gesturers is that the appropriate comparison is not between signers and adult gesturers, but rather between signers and child gesturers. Many, if not all, sign languages originate from the highly-iconic home signs spontaneously created by deaf children (Brentari & Coppola, 2013), with home sign gestures becoming conventionalised and fossilised in the lexicons of emerging sign languages. If children use more instrument signs than adults in their spontaneous productions, it might be expected that sign languages, whose lexicons were originally established by children, would also have more instrument form than the spontaneous gestures of hearing adults. What appears synchronically to be a difference caused by modality would diachronically reflect a difference caused by age.

Consistent with this explanation, Overton & Jackson (1973, p. 311) found a strong preference for ‘body part as object’ (i.e., instrument) forms in successful productions by 3-year-old gesturers (92% instrument), and this preference declined gradually with age, with only 31% of productions by 8-year-olds using instrument. However, recent studies have found the opposite effect among young signers. Ortega et al. (2017) found that when both an instrument and handling variant are possible, child signers use substantially more handling forms than adult signers. Confusingly, despite the apparently opposite results, both studies explain the change with age as resulting from cognitive development that allows the initially-dispreferred strategy to become better understood (Ortega et al., 2017, p. 95; Overton & Jackson 1973, pp. 313-314).

As well as a lack of clarity over which strategy is favoured by children, a further issue with this alternative explanation is its incompatibility with the gradual increase in instrument use compared to handling attested in Nicaraguan Sign Language’s history (Pyers & Senghas, 2020). If the instrument preference is a relic of home sign, we would expect it to be at its strongest in the earliest cohort of an emerging sign language (closest in time to the original home signers). Instead, the evidence from Nicaraguan Sign Languages suggests a gradual emergence of the instrument preference,

4.2.2 Instrument preference as a strategy-diversity preference

Another alternative explanation is that the higher rates of instrument strategy in sign languages does result from communication pressures, but this is not because instrument is the more effective strategy. Rather, a mix of strategies (using one strategy for some signs and another for some others) may be optimal in terms of maximising communication success and reducing the number of polysemous signs. Since the baseline in non-signers is a strong handling preference, an increase in diversity of strategies used will necessarily be realised as an increase in instrument usage. Padden et al. (2013, p. 298) noted that while handling was overwhelmingly preferred by gesturers, in the case of similar objects (such as spoon and fork), some gesturers used handling for one, and instrument for the other, suggesting strategy diversity can be employed when there is a need to distinguish similar objects.

It does seem to be true cross-linguistically that sign languages use a greater diversity of iconic strategies than gesturers. Kimmelman et al. (2018, p. 4233) compared iconic strategy use for tools in 19 sign languages. On average, the most common strategy in each sign language was used 61% of the time, much lower than the 79% handling usage by hearing gesturers in Padden et al. (2013).

On the other hand, this would not fully explain the differences found in Padden et al. (2013). While ASL strategies were more varied (65% instrument, 35% handling) than American gesturers (17% instrument, 83% handling), a pressure for strategy diversity would not explain the switch from handling dominance to instrument dominance. In the case of the Al-Sayyid Bedouins, strategy diversity was actually greater among gesturers.

4.2.3 Instrument preference due to resistance to erosion

Finally, perhaps the reason for signers’ instrument preference can be found by observing sign languages’ evolution. Pyers & Senghas (2020) found that both instrument and handling strategies decline in frequency over generations in emerging sign languages, but handling declines faster, leading to an increased instrument frequency relative to handling. If this decline involves handling signs being replaced with other iconic strategies, it would suggest issues with the strategy itself. However, if the decline is driven by handling signs being eroded phonologically until they are no longer iconic, it might suggest something about the handshapes used in handling signs makes them prone to erosion. More research into how signs change in emerging sign languages may be useful to answer this question. Alternatively, an iterated learning experiment could explore whether handling gestures are more prone to erosion (loss of iconicity) than instrument gestures, similar to past experiments on the loss of vocal (Edmiston et al., 2018) and pictographic (Caldwell & Smith, 2012) iconicity over generations.

5 Conclusion

In this paper I explored whether differences in iconic strategy use between hearing gesturers and deaf signers could be explained by differences in the communicative efficacy of the strategies. The first experiment found no significant difference between handling and instrument strategies, neither in terms of the semantic range of interpretations they elicited, nor in terms of how close interpretations were to the intended meaning. While some individual words showed a significant difference, the lack of an overall effect suggests the hypothesised difference in strategy efficacy may not exist, or may not be properly captured by the format of Experiment 1. The second experiment reaffirmed handling as the preferred strategy for non-signers, but found no evidence for preference being affected by a pressure to communicate. While these results do not support my theory, some features of the experiment design — such as the focus on gesture interpretation rather than production, and the reliance on a prompt-based communication pressure — may have concealed the hypothesised effects. An interaction between communication pressures and iconic strategy remains a possible, but here unproven, explanation for differences between signer and gesturer productions, warranting future research using altered methodologies, in particular making use of silent gesture, as well as genuine communication between participants.

6 References

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,* 67(1), 1–48.

Brentari, D., & Coppola, M. (2013). What sign language creation teaches us about language. Wiley Interdisciplinary Reviews: *Cognitive Science,* 4(2), 201-211.

Brentari, D., Coppola, M., Mazzoni, L., & Goldin-Meadow, S. (2012). When does a system become phonological? Handshape production in gesturers, signers, and homesigners. *Natural Language & Linguistic Theory*, 30(1), 1-31.

Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PLoS One* 7(8): e43807.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences,* 19(10), 603-615.

Eccarius, P. N. (2008). A constraint-based account of handshape contrast in sign languages. Purdue University Press.

Edward, M. (2021). Iconicity as a Pervasive Force in Language: Evidence from Ghanaian Sign Language and Adamorobe Sign Language [Unpublished doctoral dissertation]. University of Brighton.

Edmiston, P., Perlman, M., & Lupyan, G. (2018). Repeated imitation makes human vocalisations more word-like. *Proceedings of the Royal Society B: Biological Sciences,* 285(1874).

Fay, N., Walker, B., Ellison, T. M., Blundell, Z., De Kleine, N., Garde, M., ... & Goldin-Meadow, S. (2022). Gesture is the primary modality for language creation. *Proceedings of the Royal Society B*, 289(1970).

Frishberg, N. (1975). Arbitrariness and iconicity: historical change in American Sign Language. *Language*, 51(3), 696-719.

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, 391(6664), 279-281.

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American,* 203(3), 88-97.

Hou, L. (2018). Iconic Patterns in San Juan Quiahije Chatino Sign Language. *Sign Language Studies*, 18(4), 570-611.

Kimmelman, V., Klezovich, A., & Moroz, G. (2018). IPSL: A database of iconicity patterns in sign languages. Creation and use. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681-10686.

Leydesdorff, L. (2005). Similarity measures, author cocitation analysis, and information theory. *Journal of the American society for Information Science and Technology*, 56(7), 769-772.

Nyst, V., Morgado, M., Mac Hadjah, T., Nyarko, M., Martins, M., van der Mark, L., ... & Schu ̈ller, A. (2021). Object and handling handshapes in 11 sign languages: towards a typology of the iconic use of the hands. *Linguistic Typology.*

O'Reilly, A. W. (1995). Using representations: Comprehension and production of actions with imagined objects. *Child Development*, 66(4), 999-1010.

Ortega, G., & Özyürek, A. (2020a). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behaviour Research Methods*, 52(1), 51-67.

Ortega, G., & Özyürek, A. (2020b). Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures. *Language and Cognition*, 12(1), 84-113.

Ortega, G., Schiefner, A., & Özyürek, A. (2019). Hearing non-signers use their gestures to predict iconic form-meaning mappings at first exposure to signs. *Cognition*, 191.

Ortega, G., Sümer, B., & Özyürek, A. (2017). Type of iconicity matters in the vocabulary development of signing children. *Developmental Psychology*, 53(1), 89.

Overton, W. F., & Jackson, J. P. (1973). The representation of imagined objects in action sequences: A developmental study. *Child Development*, 309-314.

Padden, C., Hwang, S. O., Lepic, R., & Seegers, S. (2015). Tools for language: Patterned iconicity in sign language nouns and verbs. *Topics in Cognitive Science*, 7(1), 81-94.

Padden, C. A., Meir, I., Hwang, S. O., Lepic, R., Seegers, S., & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture*, 13(3), 287-308.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science*, 2(8), 150152.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology,* 1, 227.

Pyers, J., & Senghas, A. (2020). Lexical Iconicity is differentially favoured under transmission in a new sign language: The effect of type of iconicity. *Sign Language & Linguistics*, 23(1-2), 73-95.

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.

Sandler, W., & Lillo-Martin, D. (2006). Sign language and linguistic universals. Cambridge: Cambridge University Press.

Sehyr, Z. S., Fisher, D., & Emmorey, K. (2017). The influence of sign language experience on iconicity judgments. In *Psychonomic Society's 58th Annual Meeting.* Vancouver, Canada.

Sharma, A., Roo, J. S., & Steimle, J. (2019). Grasping microgestures: Eliciting single-hand microgestures for handheld objects. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Slobin, D. I. (1977). Language change in childhood and in history. *Language Learning and Thought*, 185-214.

Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35, No. 35).

Stokoe W. (1965). Dictionary of the American sign language based on scientific principles. Gallaudet College Press.

Tamariz, M., Roberts, S. G., Martínez, J. I., & Santiago, J. (2018). The interactive origin of iconicity. *Cognitive Science*, 42 (1), 334-349.

Wen, J. (2012). A statistical argument for the homophony avoidance approach to the disyllabification of Chinese. Singapore: National University of Singapore.

Yin, S. H., & White, J. (2018). Neutralisation and homophony avoidance in phonological learning. *Cognition*, 179, 89-101.

Acknowledgements

I would like to thank my supervisor, Simon Kirby, for his excellent advice, constant encouragement, and enthusiasm for my topic. Thank you also to Becca, for offering to be the model for my many hand sign videos (a kind gesture), and to the 951 people who participated in my experiments.

On the Meaning of the Preposition *of*: a Comparison with the French Preposition *de*

Chloé Vanrapenbusch
University of Sussex

Abstract. This research investigates the meaning of the English preposition of, in comparison with the French preposition de. Both of and de used to have a meaning 'away from' (OED, 2004; LGR, 2017), this was lost and replaced by the adverb off for the English preposition. The Oxford English Dictionary notes that multiple senses of *of* could have been influenced by French *de* and in the aftermath of the Norman Invasion. However, this research examines the present-day situation, through two corpora built for the purpose of this project. The corpora consist of online articles with topics that are recent for a more established analysis. The corpora are in French and English, whereby articles have been found in both languages, so no translation was required. These corpora were analysed by examining the frequency, n-grams, and patterns of use of the two prepositions in their various dictionary senses from the OED and Le Grand Robert. A few meanings were identified from the n-grams and the etymological background. Those meanings include the partitive expression where the French preposition de would switch to an 'article' to have this meaning in use, whereas the English preposition of remained a preposition and developed this meaning from the French *de*. The genitive case and possessive function are also investigated and give conclusive results that one preposition developed from the other. This data helped identify similarities and differences and concluded that the prepositions have converged in terms of meanings and functions but also diverged for some of them.

Keywords: corpus linguistics; French; prepositions

---

1 Introduction

1.1 Research Aims and Purpose

The aim of this study is to establish the most frequently used meaning of the English preposition *of* in comparison to the French preposition *de,* using data from corpora created for the purpose of this research. The corpora created are introduced as 'English corpus' and 'French corpus' and are accessible on the Sketch Engine software (Kilgariff, 2004). According to the Oxford English Dictionary (OED), the English preposition *of* has been influenced by the French preposition *de*. However, when looking closely at its meanings, both prepositions could have undertaken the same route and developed further. The OED 'seek[ed] to exhibit the main uses' (OED, 2004) of the English preposition *of*, and has done so by identifying around 110 meanings, obsolete ones included (Murphy, 2022). Similar to the English dictionary, Le Grand Robert (LGR) has found 59 meanings and functions for the preposition *de*. In fact, as both dictionaries categorise each meaning within sections and subsections making them similar in format, the Oxford English Dictionary and Le Grand Robert are the dictionaries used for this research. Dixon discussed that some of the preposition *of*'s meanings 'are thought to have developed in part through translation of French preposition *de*' (Dixon, 2022, p. 100). Thus, it could be a translation influence rather than both preposition going through the same process of acquiring similar meanings. However, due to the focus of this project to compare the meaning(s) instead of establishing a potential influence from the French *de*, the translation was not considered. The focus is then from a synchronic analysis, where this project is not looking to define the meaning of the preposition *de* but 'the concept that underlie [it] and the way people interpret [this] concept' in present-day (Landau, 2001, p. 153).

Thus, how does the English preposition *of* compare to the French preposition *de* in terms of their meanings and whether they have converged or diverged in written language in both languages.

1.2 Research Hypothesis

The prepositions *of* and *de* share similarities in their meanings, use in context and background information. Therefore, this research aims to establish what the most appropriate meaning of the English preposition *of* is, in comparison with the French preposition *de.* Even though it has been established that one meaning cannot be given to such a preposition.

1.3 Research Outline

Section 2 will introduce the background of each preposition, as well as past literatures. This will be done with a focus on the etymology and the grammar, with a chronological timeline of *of* and *de*. Section 3 introduces the methodology in a first place, followed by the data. In the latter, different data collection has been created, which are analysed and discussed in section 4, in relation to the background section. Finally, Section 5 concludes that the preposition *of* and *de* have similarities in their meanings and functions as well as some differences. This last section also outlines the limitations of this research as well as the next steps for further research.

2 Background Information and Literature

Only a few linguists have researched the meaning of the preposition *of* (Dixon, 2022; Murphy, 2022). Apart from searching in a dictionary for their meaning, grammar books define what their role is in a sentence, or how their meanings developed diachronically. The meaning of a preposition synchronically is not an area that has been well developed in linguistics. Indeed, when searching for the meaning of a preposition, the only responses given are either 'what is a preposition and how to use it' or the entry for the meaning of the word *preposition*. Consequently, researching in this area of small words could give another point of view on prepositions in linguistics and understanding how the preposition with the meaning 'away from' for *of* has changed to the meaning the most frequently used in recent written language. Dixon (2022) mentioned that 'each preposition has a basic meaning and a number of extensions from this, mostly of a more abstract nature.' (Dixon, 2022, p. 6). However, he also highlights the idea that the word is made of concrete things, consequently 'most prepositions have a basic meaning relating to space or to time' (Dixon, 2022, p. 6). Although his work is on establishing the meanings of the prepositions in general, Dixon gives a supportive analysis of each meaning for *of*, with for instance the possessive case, in a different perspective than the one given by grammar books for instance. Murphy (2022) also gives an overview on how the preposition *of* has acquired some of its meanings, giving a diachronic overview of the preposition *of*, how the preposition *of* has changed as well as how it continues to change. Indeed, the author mentions the meaning where the preposition also replaces the auxiliary verb *have*.

2.1 Etymological Explanations

This section outlines the etymology of the preposition *of* and *de*. This is an important factor to understand the meaning of the English *of* in comparison to the French *de* as 'the outer shape and meaning of a word may conceal a distinctly deeper or more relevant message' (Malkiel, 1993, p. 2).

The etymological analysis also gives insight into how each preposition changed diachronically, since most lexemes have 'the built-in capacity to migrate from one locus, or one speech community, to another' (Malkiel, 1993, p. 2). Therefore, it helps identify whether French and English have undertaken the same route etymologically, which can account for the recent meaning that can be found in the data (§3.2), which outlines the similarities between both prepositions. To have a chronological overview of the adpositions, a timeline has been created where the English *of* is above and the French *de* is below (Figure 1).

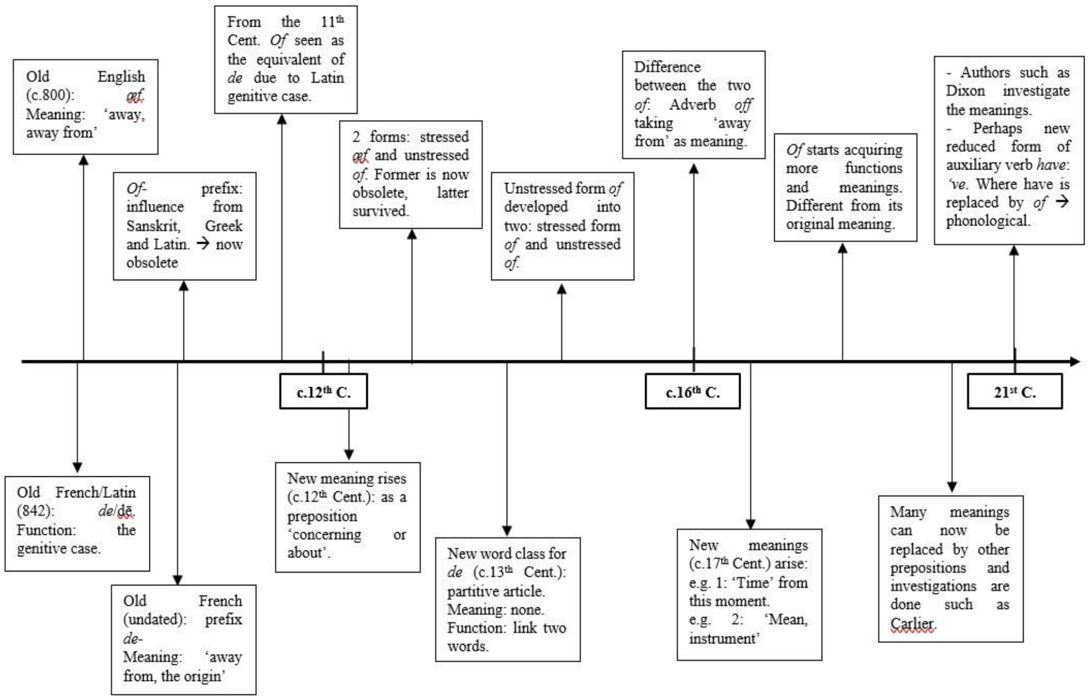


Figure 1: *Timeline of the preposition of (above) and de (below). Also in Appendix 4 (§8.1) for a bigger view.*

2.1.1 Preposition of

According to the Oxford English Dictionary, the preposition *of* used to be 'a low-stress variant of Old English *æf*' (OED, 2004) and has links to other languages such as Sanskrit, Latin, and Northern languages, as seen in table 1 and 2 (below). It's first meaning used to be 'away, away from' (OED, 2004), which has now been replaced by the adverb *off*.

Table 1: *Etymology of the preposition of-Northern Languages.*

Language	Preposition/Adverb/Equivalent
Old Frisian	*af*, *of,* prep.
Middle Dutch	*af,* prep. , *af*, *ave*, *of*, *aff*, *off,* adv.
Old Saxon	*af,* prep.
Old high German	*aba*, *abe*, *ab*, prep. & adv.
Old Icelandic	*af,* prep. & adv.
Norn (Shetland)	*ov,* prep. & adv.
Old Swedish	*af,* prep. & adv.
Danish	*af,* prep. & adv.
Gothic	*af,* prep.

Table 2: *Etymology of the preposition of-Indo-European Languages.*

Language	Preposition/Adverb/Equivalent
Sanskrit	*apa-* (prefix) meaning 'away''off'
Ancient Greek	*ἀπό* meaning 'off''away from'
Classical Latin	*ab* meaning 'off''away from'

Table 1 shows the first meanings the preposition *of* acquired, whereas table 2 shows the function *of* used to have taken from other languages. It was mainly a preposition and prepositional adverb that developed into two forms: a stressed form *æf* and an unstressed form *of* during the Middle English period as shown in the timeline. The former, now obsolete, was used for a 'few nominal compounds' (OED, 2004) whereas the latter survived and was used as a verbal particle and preposition. This unstressed form had two new forms by the end of the Middle English period: a preposition and an adverb, giving rise to a new stressed form of the word *of*. Both forms were written the same way during the Middle English period and up to the 17th Century (OED, 2004). It is only during this latter period that the forms differed where the stressed form with an adverbial function *of* became the adverb *off,* taking the original meaning of the preposition 'away, away from' (OED, 2004). Since the 18th Century, they have been used as two different words, whereby one is a preposition; *of,* and the other one has an adverbial function. Over the years, the preposition *of* has acquired a lot of meanings, for instance; 'of motion, direction, distance' (OED, 2004, I).

2.1.2 Preposition de

Le Grand Robert Dictionary (LGR) explains that the preposition *de* was first used around the year 842 (Le Grand Robert, 2022). The preposition comes from a Latin root and was used as a prefix to verbs to indicate 'the origin, movement from up and down and the meaning 'away from[38]' (Le Robert, 1976, p. 193) as well as 'concerning or about[39]' (Le Robert, 1976, p. 193). Moreover, Le Grand Robert explains that this word was used as a prefix to numerous verbs, marking the motion of going up and down, as seen in example 1 and 2.

[38] LeRobert Dictionary: 'indiquant l'origine, l'éloignement, le movement de haut en bas' (LeRobert, 1976, p. 193)

[39] LeRobert Dictionary: 'au sujet de' (LeRobert, 1976, p. 193)

(1) *Descendere*
de.PREF scendere
'go down'
(Le Robert, 1976, p. 193)

(2) scandere
'go up'
(Le Robert, 1976, p. 193)

More meanings to this affix were added such as the sense of 'an action that comes from an object', 'privation or remoteness' or even 'to reinforce'[40] the meaning of a word (Le Robert, 1976, p. 193), as presented in Table 3.

Table 3: *Latin examples of the preposition de with their translation.*

Meaning/function	Example
An action that comes from an object	Describere 'describe'
Privation or remoteness	Decapitare 'behead' 'decapitate'
To reinforce	Deperire 'die' 'to be lost'

As well as reinforcing verbs, the prefix *dē* also was used on 'a great number of adverbs'[41] (Le Robert, 1976, p. 193), this was especially the case in Latin, and can also be found in the French language, as examples 3 and 4 show. The prefix *de* has been used widely and has also spread in English as seen with example 5, however, this prefix in English does not have any relation with the preposition *of*, but with the prefix *de-* only.

(3) *de ex*
Fr. 'à partir de'
Eng. 'from'
(Le Robert, 1976, p. 193)

(4) *de foras*
Fr. 'dehors'
Eng. 'outside'
(Le Robert, 1976, p. 193)

(5) *deactivate*
'do the opposite of'
(Merriam-Webster, Undated)

Finally, the preposition *de* gave rise to a partitive article, which means it is empty of meaning. This article was first in use around the 13th Century (Le Grand Robert, 2022).

[40] LeRobert Dictionary: 'une action faite d'après un objet […] une privation ou un éloignement […] une simple renforcement' (LeRobert, 1976, p. 193)
[41] LeRobert: 'à renforcer un grand nombre d'adv.' (LeRobert, 1976, p. 193)

2.2 Grammatical Explanations

This section outlines the grammar of the prepositions *of* (§2.2.1) and *de* (§2.2.2). In English, a preposition can be described as a 'word that governs, and normally precedes, a noun or pronoun and which expresses the latter's relation to another word' (Pullum & Huddleston, 2002, p. 598), this can be seen in example 6, where the preposition is underlined. In French, it is described as an 'invariable word that established a subordination link between words or phrases'[42] (Grevisse, 1986, p. 1503). This is illustrated in example 7, where the prepositions are underlined. As well as having etymological similarities, the prepositions have the grammatical similarity of having to link and introduce new complements.

(6) Max sent a photograph to his parents
(Pullum & Huddleston, 2002, p. 598)

(7) Ma sœur est partie pour l'Afrique en avion
(Grevisse, 1986, p. 1503)

In both examples, the preposition introduces a new complement, where *to* introduces the noun phrase *his parents* and *pour* and *en* introduces the noun phrases *l'Afrique* and *avion* respectively. Analysing the grammar of each preposition is relevant to this research as they may express different meanings depending on the context. Dixon explains on his work on prepositions that 'which preposition is used in a particular circumstance, and which sense of it is appropriate, is determined by the meanings of the words it is combined with' (Dixon, 2022, p. 1). Therefore, looking into the grammar of both adpositions is essential for further understanding of what complements they can attach to and what meaning they give to their complement. Indeed, the main grammatical role of a preposition is 'to indicate the semantic role of its noun phrase' (Dixon, 2022, p. 24).

2.2.1 Preposition of

The preposition *of* acquired a great range of meanings and functions over the years, such as 'Of motion, direction, distance' (OED, 2004, I). The English language has what looks like a genitive case with the preposition *of*, as shown in example 8. However, the preposition *of* has lost most of its possessive's meaning since the clitic *'s* marks it as well, as seen in example 9.

(8) co-chair of the IPCC
(English corpus, 2022)
(9) the IPCC's co-chair

Furthermore, the preposition *of* often can be found next to, that is before or following, adjectives and verbs, as seen in examples 10 and 11, where the former is constructed verb + *of*, and the latter is *of* + adjective. However, it doesn't necessarily mean that *of* gives its semantic role to those complements.

(10) Children spoke of their concerns
(English corpus, 2022)

[42] 'un mot invariable qui établit un lien de subordination entre des mots ou des syntagmes' (Grevisse, 1986, p. 1503)

(11) We got ourselves in a lot of good attacking positions
(English corpus, 2022)

Thus, the preposition *of* doesn't necessarily link to a noun phrase but also to other parts of speech. In those instances, whereby the preposition follows or comes before the word are connected to a main meaning. They are mainly described as 'indicating something', where the preposition does not necessarily have a meaning and referent like a word such as 'castle' but would have a role that is to indicate a grammatical function within a sentence.

2.2.2 Preposition de

Le Grand Robert investigated the preposition from two points of view. The first category is called 'preposition with analysable meaning'[43] (Le Grand Robert, 2017) and the second one is introduced as 'the grammatical function prevailing the meaning; after a verb, an adjective or a noun'[44] (Le Grand Robert, 2017). Therefore, showing how the preposition has grammatical function such as 'after an adjective' in the subcategory 'to introduce the object of an action, the destination'[45] (Le Grand Robert, 2017), as seen with example 12, where *des* has been reduced from *de les*.

(12)	*La*	*meilleure*	*des*	*réponses*
	The	best	of	answer

'The best answer'
(French corpus, 2022)

It is another grammatical point to be noted; the different ways the preposition *de* can be written. Indeed, depending on the structure of the sentence and words that can be found around the adposition, *de* can differ orthographically. Table 4 shows the different ways *de* can be written, what they stand for, and an example taken from the French corpus.

[43] Le Grand Robert: 'preposition à sens analysable' (Le Grand Robert, 2017)

[44] Le Grand Robert: 'La function grammaticale primant le sens; après un verbe, un adjectif ou un nom' (Le Grand Robert, 2017)

[45] Le Grand Robert: 'Après un acjective' in the subcategory 'pour introduire l'objet d'une action, la destination' (Le Grand Robert, 2017)

Table 4: *Preposition de and its variants.*

Preposition/ Variant	What do they stand for?	Example
Du	Stands for *de le*	Glasgow du 31 octobre au 13 novembre *Glasgow de le 31 octobre au 13 novembre 'Glasgow from the 31st October to the 13th November'
D'	*De* is elided in front of a vowel or a silent /h/	L'idée d'une végétalisation * L'idée de une végétalisation 'The idea of a revegetation'
Des	Stands for *de les*	Au centre des négociations avec la Grande-Bretagne *Au centre de les négociations avec la Grande-Bretagne 'In the middle of the negotiations with Great Britain'

Finally, *de* isn't always a preposition but can also be employed as a partitive article, which is a word class whereby words are empty of meaning and merely connect two words. Grevisse describes it as 'nothing else than an indefinite article employed in front of a noun to indicate that it is nothing else than just an indefinite quantity'[46] (Grevisse, 1986, p. 911). Even though a preposition and a partitive article are similar in their use, which is by linking two complements, it is important to make the difference for the purpose of this research.

3 Methodology and Data

3.1 Methodology

The methodology chosen was corpus-based built for the purpose of this research. There are two corpora, one in French and one in English using the software Sketch Engine (Kilgarriff, 2004). Both corpora were built with written language articles respectively in French and English to avoid any translation therefore giving a more established analysis. The articles have three main topics: rugby, climate change and the fishing row conflict. These were chosen as they could be found in the news in both countries within the last few years. Providing instances where the meanings of the prepositions would be the most recent one since this research focuses on a synchronic point of view rather than diachronic. All articles were taken from French or English websites and built in the corpus software Sketch Engine (Kilgariff, 2004). An aim of 15,000 words per corpus was established at first but, as Table 4 shows, bigger corpora were realised, where the word count would be as similar as possible for both languages. Those corpora can be found within the software under 'French corpus' and 'English corpus', as they have been shared with University of Sussex users.

[46] Le Bon Usage: 'il n'est autre chose qu'un article indéfini employé devant un nom […] pour indiquer qu'il s'agit d'une quantité indéfinie' (Grevisse, 1986, p. 911)

Table 5: *Word count of each corpus.*

Language	Word count
English	23,137
French	22,376

Within each corpus, the preposition *of* and *de* were identified, providing a total of 546 instances for the former and 2,051 instances for the latter. However, the software does not consider the instances where *de* could be a partitive article. Therefore, a part of the French corpus has been sampled and analysed to have an estimate of the partitive article present in the corpus (see Section 3.2.2). To account for the difference between the two prepositions, looking into the clitic *'s* was also very important. Indeed, the possessive case could be the reason why there are less instances of *of* than *de*, as there could be more *'s* than *of* in terms of the possessive case. This can be seen into more details in the Section 3.2.3. Moreover, n-grams were created using the tool on the Sketch Engine software. This allowed to find the most used sequences within each corpus with the preposition *de* and *of* (please refer to Section 3.2.4). Identifying the n-grams also allowed to investigate how the dictionaries are relevant for this research. Indeed, dictionaries are one of the primary resources to research meanings for this research.

3.2 Data

3.2.1 Frequency of the Prepositions

The use of prepositions in the corpora is very different, as seen in table 5, between the two languages. Indeed, the English language uses almost twice more prepositions than the French language. This could be due to another part of speech called 'article' in French, where *'le, la, les'* are definite articles and *'un, une, des'* are indefinite articles. Therefore, French has more than one word class for some function words where English would mainly have them as prepositional.

Table 6: *Frequency and items of overall prepositions present in the corpora.*

	English Corpus	French Corpus
Frequency	3,262	4,123
Items/KWIC	61	33

To understand further the use of each preposition, the frequency of each has been established where we can see a clear difference between the use of the two words (Table 7). The French preposition is used 4 times more than the English preposition. However, *de* can also be a partitive article. Therefore, the frequency could be wrong and is investigated in Section 3.2.2. On the other hand, *of* does not change category and remains a preposition no matter what the context is, thus the frequency is accurate.

Table 7: *Frequency of each preposition in the corpora.*

	English *of*	French *de*
Frequency	546	2,051

3.2.2 *Identifying the Partitive Article de*

As previously cited, the French word *de* can be found in different part of speech (POS) such as preposition and partitive article. The software Sketch Engine does not allow to search for partitive article, or any other article, and will automatically put any non-prepositional word into the category 'other'. When looking at the frequency of *de* specifically, the results didn't seem conclusive. Table 8 shows the frequency of *de* and the categories it has been assigned, where one shows the frequency found in table 7 and the 'other' category shows instances that aren't prepositional.

Table 8: *Frequency of the preposition de and the different POS.*

	Frequency *de*
Preposition	2,051
Other	4

To confirm those results from the software are correct, 2 random samples of 200 instances (representing 10% each of all occurrences) were investigated to recover occurrences of the non-prepositional *de*.

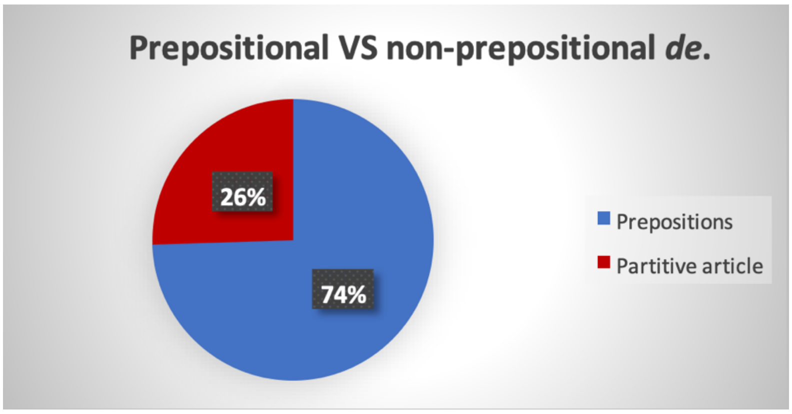


Figure 2: *Prepositional VS non-prepositional de from sample 1.*

The results were different than the one suggested by Sketch Engine, where within the 200 instances of the first sample, 26% were not prepositional, as shown in Figure 2. This means the preposition *de* is not as shown in table 8 but would be estimated at 1,518 instances.

The second sample gave slightly different results, as seen in Figure 3, where 22% instances were not prepositional. However, it also confirms that the Sketch Engine results are inaccurate.

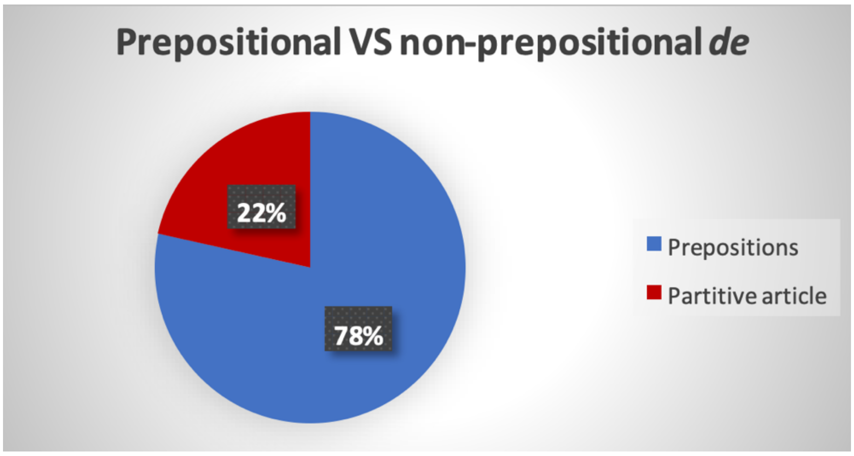


Figure 3: *Prepositional vs non-prepositional de form, sample 2.*

The mean of both samples is 24%, so an estimation of 1,559 instances is prepositional with *de*, against 492 instances that would be qualified as partitive article. To confirm those results a final sample of 400 instances was investigated (all data analysed can be found in the appendix §8). The results were conclusive, as seen in Figure 4, where the average instances of non-prepositional *de* is 24%. Thus, confirming the previous two samples' mean is accurate and that the Sketch Engine software is wrong in identifying 'other' as a category for *de*.

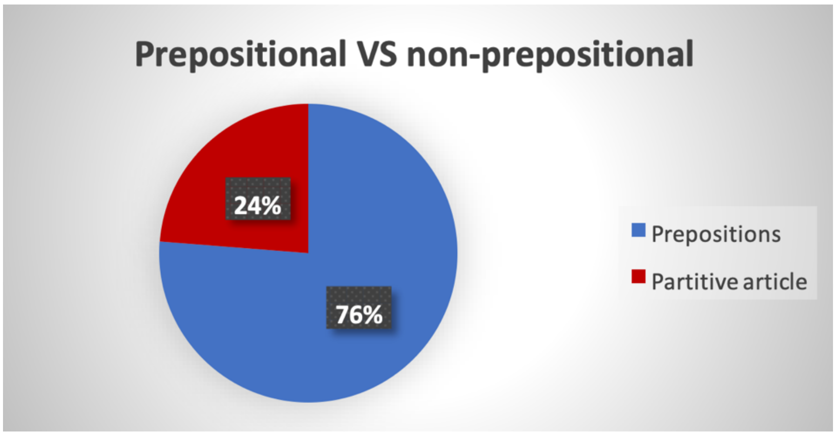


Figure 4: *Prepositional vs non-prepositional de from sample 3.*

The partitive article is the word class that the preposition *de* can be when it is found next to a noun that cannot be counted, or is a noun mentioned as a whole (Le Grand Robert, 2017), as seen in example 13, where '*visibilité'* is a feminine noun that is uncountable.

(13) *Et donner un peu de visibilité aux pêcheurs*
'And give more visibility to the fisherman'
(French corpus, 2022)

Another way to identify partitive article would be where the preposition is found next to an article, as seen with example 14. Indeed, those *de* become partitive article due to not linking two content words. However, this is not always the case.

(14) *Cette annonce intervient à deux jours de la date-butoir*
‘This announcement comes two days before the deadline’
(French corpus, 2022)

Partitive articles can also be found before an adjective, if this one ‘is part of the noun’[47] (Grevisse, 2009, p. 109), as seen with example 15, where the adjective ‘*bonne’* is found following *de* and cannot be disconnected from its noun ‘*foi’*.

(15) *[…] en espérant que Jersey soit de bonne foi*
‘Hoping that Jersey is of good faith’
(French corpus, 2022)

Identifying the partitive article within the French corpus was essential to understand the gap between *of* and *de* in their use within the corpus. Finally, instances within the corpus where an adjective marking the possessive form follows *de* are part of the corpus, as the latter remains the same, and the former gives its function to the preposition, further investigation is done on the possessive case in Section 3.2.3.

3.2.3 The Clitic ‘s and the Possessive Case

The clitic *‘s* acts as the possessive case the same way *of* would, which could explain why *of* has less instances in the corpus than the French *de*. Investigating the clitic within the corpus was required to understand the gap between the two prepositions, and to check whether *of* had kept the genitive case as one of its meanings. However, the results were not as expected where the clitic *‘s* occurred only 3 times as seen in Figure 5. Out of the 73 instances *‘s* appears in the corpus; 70 are the auxiliary verb ‘be’ contracted. Thus, the preposition *of* is used more than the clitic *‘s* in terms of the possessive case.

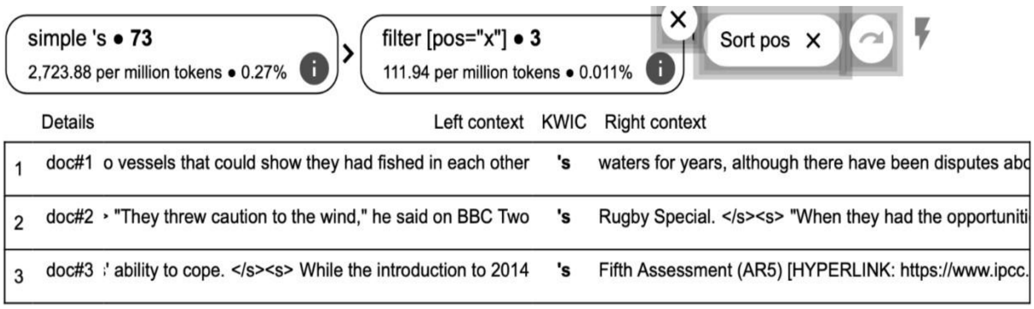


Figure 5: *Occurrence of the clitic ‘s in the English corpus.*

To confirm these results are the only ones that should be expected, investigating the number of clitics *‘s,* within a bigger corpus was required, where the corpus used to compare is English Web 2020 on Sketch Engine.

[47] Le Petit Grevisse: ‘quand l’adjectif fait corps avec le nom’ (Grevisse, 2009, p. 109)

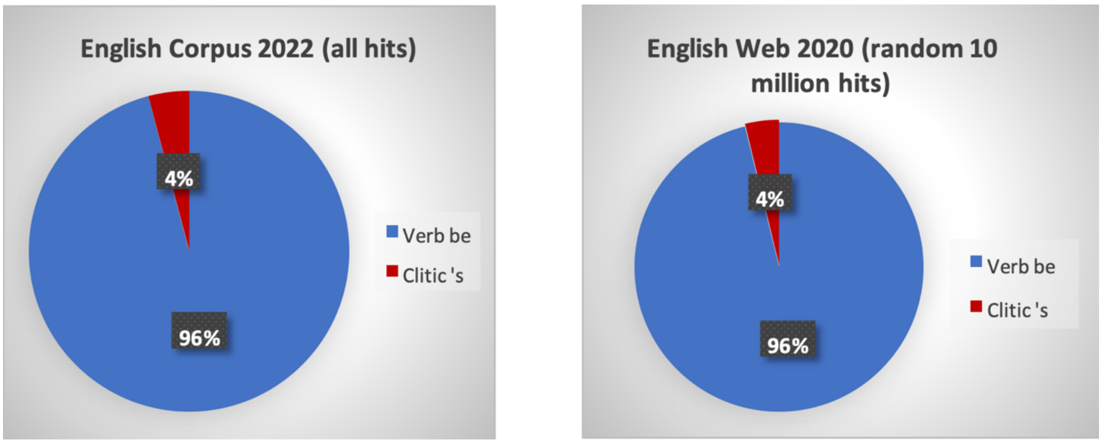


Figure 6: *Clitic 's in English corpus 2022 (on the right as 'a') and in the English Web 2020 (on the left 'b').*

As seen with Figure 6b, the results are the same as 6a, thus validating the results from the corpus built for the purpose of this project are accurate.

3.2.4 N-grams

Checking for n-grams within the corpora is relevant to find the most frequently used *of* and *de* in a certain context thereby helping with identifying the meaning that is the most used in written language. With the help of n-grams, it will be possible to link the most used instances to an entry of the dictionary; this should help identifying the meaning of the preposition *of* in comparison with the preposition *de*. To choose the appropriate N-grams, that is whether it would be constituted of 2 to 6-grams, checking for the instances where repeated, was required. The final choice was 5 or 6-grams as it seemed the most appropriate choice since more similarities in occurrences were present, as seen in Figures 7 for *of* and 8 for *de*.

	Word	Frequency ?	
1	the end of the century	3	•••
2	by the end of this	3	•••
3	to dozens of French boats	3	•••
4	the end of this century	3	•••
5	by the end of the	3	•••
6	the rub of the green	3	•••
7	at the end of the	3	•••
8	licences to dozens of French	3	•••
9	brief window of time to	2	•••
10	had a history of fishing	2	•••

	Word	Frequency ?	
1	by the end of the century	3	•••
2	by the end of this century	3	•••
3	licences to dozens of French boats	3	•••
4	boats had a history of fishing	2	•••
5	by the latest threats of sanctions	2	•••
6	close to an act of war	2	•••
7	deny fishing licences to dozens of	2	•••
8	coastal waters of the UK and	2	•••
9	brief window of time to avoid	2	•••
10	a brief window of time to	2	•••

Figure 7: *N-grams (5/6) for the preposition of from the English corpus. Where the image on the left is 5-grams (a) and on the right 6-grams (b).*

	Word	Frequency ?
1	dans les eaux de Jersey	9
2	de la Baie de Granville	7
3	britanniques à condition de pouvoir	5
4	à condition de pouvoir prouver	5
5	de pouvoir prouver qu' ils	5
6	eaux britanniques à condition de	5
7	gouvernement de l' île anglo-normande	5
8	condition de pouvoir prouver qu'	5
9	pêcher dans les eaux de	5
10	la ministre de la mer	5

	Word	Frequency ?
1	condition de pouvoir prouver qu' ils	5
2	britanniques à condition de pouvoir prouver	5
3	eaux britanniques à condition de pouvoir	5
4	à condition de pouvoir prouver qu'	5
5	de pouvoir prouver qu' ils y	5
6	accord de commerce et de coopération	4
7	le gouvernement de l' île anglo-normande	4
8	du comité régional des pêches de	4
9	comité régional des pêches de Normandie	4
10	répondre à toutes les exigences de	3

Figure 8: *N-grams (5/6) for the preposition de from the French corpus. Where the image on the left is 5-gram (a) and on the right is 6-gram (b).*

These n-grams for the preposition *of* have been checked, where chosen instances that are found in both 5 and 6-grams have been identified to be the meaning to investigate with the partitive expression and the possessive case. Indeed, those two last meanings have been chosen as they can be found in both languages etymologically. The same process as *of* was also applied on the preposition *de*, and the chosen n-grams can be found in Table 9. To choose the appropriate n-grams, any instances where the prepositions were at the end were removed as we are looking to have a complement after the preposition. For the English, it became obvious which one would have to be an analysed meaning, as seen in Table 9 with 'the end of the century'. The second one was chosen because of the verb following the preposition *of*. In contrast, for the preposition *de,* it was a more complicated process due to the numerous repetitions of the same sentence, for instance; '*à condition de pouvoir prouver'*, making this one an n-gram to investigate in relation to *of*, which is whether both prepositions have the same meanings in the most frequently used context. The second one was chosen as the preposition wasn't found at the end of the sentence but had a relation to the partitive meaning, like one of the n-gram for the preposition *of*.

Table 9: *Chosen n-grams to identify the meanings.*

English n-gram	Meaning	French n-gram	Meaning
'the end of the century'	X. Expressing possession or being possessed. 33. c. belonging to a time, as existing or taking place in it.	'gouvernement de l'ile anglonormande'	Partitive article. 1. Devant un nom concret
'boats had a history of fishing'	IX. in partitive expressions, indicating things or thing which of a part is expressed by the preceding words. 28. b. Preceded by a noun.	'à condition de pouvoir prouver'	LGR: II. La fonction grammaticale. 4. Devant un infinitif. Devant un infinitif sujet 'In front of an infinitive subject'

The table is constructed where two n-grams have been selected, the meanings are given from the OED and LGR, and the n-grams present some similarities in meaning and function between the two languages. Therefore, suggesting the prepositions aren't used in the same context at the same frequency.

3.2.5 Purpose of the dictionaries for this research

Using dictionaries for the purpose of this research was essential since the etymological explanations were drawn from those. They also provide a better understanding of words in general, thus providing greater chances of assessing the meanings and functions of the prepositions correctly. Dictionaries also allow for connection between words; the OED makes a few assumptions to the preposition *of* being like *de,* but it also provides with examples where *of* could be replaced with another function word. As does the French dictionary Le Grand Robert. Moreover, as the definition states a *dictionary* has a purpose to 'explain or translate, usually in alphabetical order, the words of a language or languages' (OED, 2010). Also, the definition includes what it should for each word which is 'spelling, an explanation of its meanings, and often other information, such as pronunciation, etymology, synonyms, equivalents in other languages, and illustrative examples' (OED, 2010). Therefore, the use of dictionaries for this research project was needed as most of these categories are analysed and discussed.

4 Analysis and Discussion

So far, we have seen that the adpositions *of* and *de* have carried similar meanings since they were first in use when they had the prefix function indicating the meaning 'away from'. They also have a partitive expression and possessive sense in common, found to be the most used in a certain context from the corpora. Therefore, looking further into those two meanings is important to establish whether both prepositions use them the same way. The last section uses the n-grams to investigate further the meanings and establish whether it is possible to choose one appropriate meaning for the preposition *of*, in comparison with the preposition *de*.

4.1 The Partitive Function

It was an important step to identify the partitive article within the French corpus, to check whether it carried the same meaning as the English preposition *of.* When it is employed as the entry from the OED suggest 'In partitive expressions, indicating things or a thing of which a part is expressed by the preceding words' (OED, 2004, IX). On the other hand, French *de* is part of the word class called 'partitive article', as previously introduced, and represents around 26% of instances in the French corpus. The fact that both prepositions present a partitive sense since Old English for *of* and Old French for *de* questions whether the two prepositions have undertaken the same route. Indeed, examples 16 and 17 show the partitive function with both the preposition *of* and partitive article *de* respectively. Where the former means 'preceded by a word of number or quantity' (OED, 2004, IX.28.a), and the latter 'preceding a concrete noun'[48] (Le Grand Robert, 2017).

(16) One of the broadcasters
(English corpus, 2022)

(17) *Et donner un peu de visibilité aux pêcheurs*
'Giving a little bit of visibility to the fishermen'
(French corpus, 2022)

The idea that both prepositions share this function suggests one of these words could have had a process of influence via either translation, etymological process, or an influence from either language. However, as previously cited in sSection 2.1.2, English *of* only had a Classical Latin root, using the meaning of a word from the Latin word *ab*. The OED suggest that *of* may have 'render[ed] Latin *ex* or *de*' (OED, 2004), thus the partitive function for this preposition is not due to an etymological influence. But it may be an Old English function that already existed before some of the meanings it has, which are now more in use. The fact that the OED is giving details on where *of* could be used where Latin *ex* and *de* could be found, shows a connection between *of* and *de*. The first use of the partitive function is found in the OED as being approximately during the Old English period, to be more precise it is necessary to look at first example to give a better estimation. In fact, the example the OED uses is taken from the Baedae Historica Ecclesiastica Gentis Anglorum, which was dated to be written around 731. This suggests the partitive function may have looked different before the preposition *of* acquired its meaning, as does the French *de* since it acquired the 'partitive article' around the 8th Century. Therefore, it seems that both *de* and *of* as partitive expression came across at the same time. Consequently, checking whether synchronically *of* is still used predominantly the same way *de* is used, will show whether the partitive function is a similarity between both prepositions.

[48] Le Grand Robert: 'devant un nom concret' (Le Grand Robert, 2017)

	Lemma	Frequency	Relative ?	% of conc. ?
1	one	20	746.27	3.66 %
2	part	19	708.96	3.48 %
3	end	15	559.70	2.75 %
4	%	15	559.70	2.75 %
5	out	14	522.39	2.56 %
6	number	10	373.13	1.83 %
7	amount	8	298.51	1.47 %
8	level	8	298.51	1.47 %
9	because	7	261.19	1.28 %
10	half	6	223.88	1.10 %
11	loss	5	186.57	0.92 %
12	sort	5	186.57	0.92 %
13	range	5	186.57	0.92 %
14	lot	5	186.57	0.92 %
15	area	5	186.57	0.92 %
16	impact	5	186.57	0.92 %
17	ownership	4	149.25	0.73 %
18	risk	4	149.25	0.73 %
19	dozen	4	149.25	0.73 %
20	issue	4	149.25	0.73 %

Figure 9: *Frequency of the lemmas for the preposition of.*

As seen in Figure 9, the frequency of each lemma is mainly formed with quantifiers such as *one* (1) or *dozen* (19) and even the symbol *%* (4), suggesting the partitive function is significantly used in written language rather than any other meanings or function. Indeed, the percentage symbol is always linked to a numeral, thus a quantity. Since these quantifiers are found preceding the preposition *of*, they cannot be linked to the entry VII. 21. a. (Fig. 10). Where this one does not have the same meaning as the partitive as it is linked to the main category 'indicating the material or substance of which is made or consists' (OED, 2004, VII). Hence, the quantifiers for this entry would be linked to physical things.

a. Connecting two nouns, of which the former is a collective term, a quantitative or numeral word, or the name of something having component parts, and the latter is the substance or elements of which this consists.

Figure 10: *Oxford English Dictionary entry for VVII.21.a*

On the other hand, the partitive article has quantifiers linked to anything, as seen with example 18, where the quantifiers are found preceding the preposition that links to a noun phrase.

(18) One of the drawbacks to these types of jobs
(English corpus, 2022)

Moreover, Dixon explains that the preposition *of* can be used with a quantifier but those can also be employed alone as seen with examples 19, where it describes one of the numerous ministers from a set of ministers (19a) and the other example indicates any minister (19b) (Dixon, 2022, p. 112).

(19) a. She asked one of her ministers

b. She asked one minister
(English corpus, 2022)

Moreover, the author draws from the explanations on the quantifier meaning that 'the distinction between senses is never water-tight' (Dixon, 2022, p. 113). Whereby some sentences may have a certain meaning due to the complement preceding the preposition but can also have a different sense depending on the complement following *of* (Dixon, 2022, p. 113). Hence, example 19 can also suggest a possessive sense in this context since it is 'one of her ministers', whereby *her* gives a possessive sense to the complement *ministers* which is attached to the preposition *of*. This instance is part of the partitive function like the entry VII. 32 suggests, whereby *of* is 'followed by a noun in the genitive case or a possessive pronoun' (OED, 2004). This is the case with example 19, suggesting the partitive function may also be part of the possessive entry. However, the partitive case was originally used but then was 'subsequently used instead of the simple possessive where this would be awkward or ambiguous, or as equivalent to an appositive phrase' (OED, 2004). Therefore, the possessive and the partitive are connected and perhaps also linked to the preposition *de*. In accordance with the word class partitive article, it is used in instances where the possessive case cannot be applied as the preposition *de* but the requirement for the partitive article is essential to link two complements (Grevisse, 1986, p. 901). As seen in Section 3.2.2, the partitive article in French consists of roughly 26% of the total instances of *de* that can be found in the French corpus. This is probably because French always needs a function word to link content words, especially nouns (Murphy, 2022). This is evident as there are only 546 instances of *of* against 2,051 occurrences of *de*. The fact that out of 2,051 instances 26% are the partitive article, shows that the preposition *of* does not need to be present in places where French would require *de*. Thus, the partitive expressions for *of* is used in places where the partitive article *de* is required as well as places where *of* may not be used. Moreover, Keun-Young explains that 'English partitive noun phrases are marked by the preposition *of*' (2007, p. 127). This can be compared to the French *de* and other ways of it being written (§2.2.2), where it is also the only partitive article in the French language. This suggests both prepositions share a common most frequently used sense and function within a sentence in terms of the partitive expression, as seen with example 20, where the preposition *de* can be translated as *of*. Therefore, showing the partitive expression is used the same way in either language.

(20) *[…] en espérant que Jersey soit de bonne foi*
'Hoping that Jersey is of good faith'
(French corpus, 2022)

This translation of this example can be linked to a meaning from the OED; 'without preceding partitive word, forming the complement of a verb, or the predicate after *be*' (OED, 2004). The preposition *of* has similar senses and uses as the preposition *de*, indicating that they have taken the same route in terms of the partitive expression.

4.2 The Possessive Case

The possessive case with the preposition *of* and *de* can be marked differently. As seen in Section 3.2.3, the possessive case in English can also be marked with the clitic *'s*, for instance, example 19a represents the possessive case with the preposition and 19b represents the same sentence with the clitic *'s*. The meaning remains the same, as well as acceptable, 'however this is by no means always so' (Dixon, 2022, p. 108) that the preposition *of* can be replaced with the clitic *'s* as seen in example 21.

(21) a. Co-chair of the IPCC
b. The IPCC's co-chair
(English corpus, 2022)

(22) It also ended the Bay of Granville agreement
(English corpus, 2022)

This last example (22) cannot be written as 'Granville's Bay', as the use of the clitic 'is generally preferred when R[the referent] has human (or higher animate) reference' (Dixon, 2022, p. 108). Therefore, this could be a reason why the clitic *'s* is only present in 4% (against 96% for the verb *be*), as those 3 instances aren't linked to an animate reference. Suggesting, the possessive case is mostly expressed with the preposition *of* rather than the clitic *'s*.

Moreover, the preposition *de* can also express the possessive case. This is shown either with *de* on its own, but it can also be found preceding a possessor such as *mon*, *ma*. The entries from Le Grand Robert express senses for the possessive case under the two different main categories, as explained in Section 2.2.2. Giving the preposition *de* more than one way of indicating this meaning.

(23) *De son côté*
'On his side'
(French corpus, 2022)

The first one, as seen in example 23, expresses the meaning 'belonging, dependant'[49], where *son* is used to indicate the possessive and *de* remains a preposition. The second entry for the preposition *of* can be found under the main category 'the grammatical function prevailing the meaning; after a verb, an adjective or a noun'[50] (Le Grand Robert, 2017), and under the subcategory 'After a pronoun (possessive)'[51] (Le Grand Robert, 2017). This last meaning will not be used, although it expresses the possessive case, it is proper to spoken language, unlike the data of this research which is based on written language. No data for this entry of the dictionary could be found in the French corpus, showing that *de* as a possessor for the spoken language remains as such. Grevisse also explains how the possessive with *de* was first mainly used in places of animate reference only such as nobility or with titles. This can be linked to the meaning of *of* as 'belonging to a place, as deriving a title from it, or as its lord, ruler, owner, etc.' (OED, 2004), where the OED also explains that this sense was rare up to the 11th Century. It became more and more popular 'when it became the regular equivalent of French *de*' (OED, 2004). Showing here that the prepositions had already acquired this meaning before they 'crossed paths.' Therefore, the French *de* did not influence the English *of* but *de* reinforce *of*'s meaning and they became equivalent to each other through translation processes, or as Murphy and the OED suggests through the Norman Conquest (Murphy, 2022. OED, 2004). They describe the preposition *de* as 'undeletable, indefatigable, and undeniably in demand' (Murphy, 2022), suggesting that the Norman Invasion is the reason why the Latin genitive case can also be found in English. Thus, the main meaning 'away from' was replaced with adverb *off* and the new meanings of the preposition *of* were acquired through a process of language change. Neither Le Grand Robert nor the OED suggests that *of* could have given some of its meanings to the French *de*. However, as previously cited, they share the meaning 'away from' whether it is as a preposition or a prefix. One comes from the Germanic languages; *of*, and *de* acquired it from Latin, showing how the prepositions did not have a link when they first appeared.

[49] 'Appartenance, dépendance' (Le Grand Robert, 2017)
[50] Le Grand Robert: 'La fonction grammaticale primant le sens; après un verbe, un adjectif ou un nom' (Le Grand Robert, 2017)
[51] Après un pronom (possessif)' (Le Grand Robert, 2017)

The possessive case like many other meanings was acquired before a French invasion or influence but was reinforced when French *de* came around into use in England. Because the preposition *of* with a possessive sense is very broad (10 meanings), this sense would be hard to categorise as anything else than what the OED is suggesting ‘expressing possession or being possessed’ (OED, 2004). Each entry goes into detail to which context *of* can be found in a sentence. Whereas French *de*, proposes only 1 meaning for written language, according to Le Grand Robert. This one meaning does not mention possession but ‘belonging’ (Le Grand Robert, 2017). This could suggest that the preposition *de* has only put forward the different ways the preposition *of* can be used. Indeed, the former tends to be followed with a possessor to mark the possessive case whereas the latter can be replaced with the clitic but can also stand on its own to mark the possessive case.

4.3 Choosing an appropriate meaning for *of*

So far, the data has shown that the preposition *of* can acquire more than one meaning and that it can be compared to the French preposition *de*. There is also the fact that *of* was not influenced by *de* but rather has developed its meanings from the preposition *de*, as seen with both meanings previously investigated. Dixon also described the preposition *of* as developing some of its meaning ‘in part through translation of French preposition *de*’ (Dixon, 2022, p. 100). Indeed, both prepositions came around the same time and Norman invasion probably added to the meanings of the preposition *of* to change or develop further. The idea that Dixon suggests a development of *of*’s meaning to be partly done through translation shows an influence from *de* could have been the case. N-grams have provided somewhat non-conclusive results to what meaning is the most used in a certain context. But this was expected since one of the editors of the Oxford English Dictionary stated that *of* is ‘probably the most difficult of the preposition, to the most concise possible treatment of which many weeks of work and some twenty columns of space, with 955 quotations, have to be given.’ (Murray, c1905). This was not the case for *de*, even during the 17th century, where the meaning found consisted of the genitive case related to Latin, and was used as an adverb (Le Robert, 2022). In fact, the preposition *de* having so little in terms of meaning at the time could suggests that *de* also developed meaning through translation from *of*. Identifying the right meaning for the n-grams selected was not an easy task since the preposition *of* could be linked to more than one subcategory in certain contexts. However, after further investigation and confirmation, it became clear that the main meanings for each n-gram chosen was the appropriate one. The difficulty of choosing one meaning for a preposition is probably because it ‘can’t be achieved with a standard substitutable definition’ (Atkins, 2008, p. 447). Indeed, Atkins raises the fact that such small words that are thought to have a grammatical function ‘do not really ‘mean’ anything’ (Atkins, 2008, p. 447). The author also adds that it is ‘our job to explain their function in the sentence and the contribution they make to its overall meaning’ (Atkins, 2008, p. 448). However, when a preposition, such as *of*, has more meanings than some concrete words and has changed from its original meaning, it becomes difficult to establish one meaning, and easier to link it to a grammatical function. Most dictionaries will present it as a complement to a noun phrase or as a function word to indicate something (Collins, Merriam-Webster), this is perhaps due to learners of English needing to have a clear and concise definition. The Oxford English Dictionary doesn’t present it as such but start straight away with a first category that gives a meaning, as seen in figure 11.

I. Of motion, direction, distance.

Figure 11: *First entry for the preposition of in the OED.*

Figure 11 may show the first meaning given to the preposition *of* but it does not mean that this entry is the most frequently used meaning, as seen with the n-grams, where the most frequently used meanings seem to be the possessive case and the partitive expressions (table 10). The preposition *de*, on the other hand, has the partitive article as one of the main meaning the most frequently used. The other one is part of a 'grammatical function' in front of concrete noun (Le Grand Robert, 2017). Thus, showing here again the prepositions are used the same way, in written language.

Table 10: *Selected n-grams that are the most frequently used in the corpora.*

English n-gram	Meaning	French n-gram	Meaning
'the end of the century'	X. Expressing possession or being possessed. 33. c. belonging to a time, as existing or taking place in it.	'gouvernement de l'ile anglonormande'	Partitive article. 1. Devant un nom concret
'boats had a history of fishing'	IX. in partitive expressions, indicating things or thing which of a part is expressed by the preceding words. 28. b. Preceded by a noun.	'à condition de pouvoir prouver'	LGR: II. La fonction grammaticale. 4. Devant un infinitif. Devant un infinitif sujet 'In front of an infinitive subject'

Moreover, the genitive case could be a reason why both prepositions have similarities, as this one used to 'express primarily the relationship of source or possession' (Tyler & Evans, 2003, p. 209). This case can be found in Old Saxon as well as Old French, where both are often mentioned under entries of the OED for the preposition *of*. However, as the assumptions from the OED suggest, the possessive case developed further due to crosslinguistic influence where it 'caused the gradual extension of *of* to all uses in which Old English had the genitive' (OED, 2004, X). *Of* was slowly changing from a few meanings to acquiring a lot more due to the development of the genitive case (Latin *de*, for instance) on other meanings. So, if the possessive case was not the only meaning associated to the genitive case, the partitive expressions may have also been influenced by this case. Indeed, as previously shown '*Of* may have render Latin *ex* or *de*' (OED, 2004, IX. 28. A), can be found below the entry for the partitive function, added to this information is 'Old English more commonly had the genitive case' (OED, 2004). Showing how the genitive case was perhaps the main meaning of the preposition *of* and can still be in use in instance where it is meant to give the sense of 'source' or 'possession'. The genitive case can be seen associated with those two meanings, as seen with example 24, where 24a represents the possessive case and 24b represents the partitive expression. However, due to having meaning developed from the genitive case as being 'possessive' and 'partitive', they cannot be called genitive.

(24) a. Fisherman of Hauts-de-France
(English corpus, 2022)
b. All of his family must be so proud of him
(English corpus, 2022)

The prepositions *of* and *de* have similar meanings in the most frequently used instances in certain context. They share the genitive case as a common ground where both developed or reinforce their meaning from it, and as for the preposition *of*, it seems that it developed further when in contact with the preposition *de*.

5 Conclusion

5.1 Conclusion

To conclude, this research investigated the meaning of the English preposition *of*, in comparison with the French preposition *de*. It was found that they share similarities in terms of the partitive expression, however, French has a word class for this function called 'article' whereas English includes this meaning within the preposition's meanings. Moreover, both prepositions share a meaning of possession, that arose from the genitive case. This case was found to come from the Latin, where both developed further from it. The preposition *of* developed and reinforced its meanings regarding the possessive case from language contact with the French preposition *de*.

They have differences in terms of the n-grams and etymologically, where they don't have the same root language, and they also aren't used at the same frequency as seen with the corpora analysis.

Finally, the main similarities are that they both shared a first meaning of being 'away from', they used to be prefix, and they still have the main function a preposition has, which is to introduce a new complement.

5.2 Limitations

The difficulty encountered for this project was the lack of research on establishing the meaning of a preposition or researching a preposition as whole from a semantic point of view. Due to financial costs, the purchase of the etymological dictionary of French words was not possible and therefore the information on this were limited to an online source at a reduced cost.

Moreover, to not take into consideration the partitive function in the data in the English corpus as well as the possessive case caused the analysis to be more difficult. Investigating the partitive *de* was relevant but could have benefited with doing the same for *of*. This is also the case for the possessive case where both prepositions could have benefited from an investigation.

5.3 Next Steps for Further Research

For further research, I would recommend researching the meaning of the preposition from a translation point of view, that is, perhaps with asking bilinguals (French/English) to translate a set of sentences in either language. Looking into the instances where the partitive expressions occur for the preposition *of* could bring more conclusive results than just the few samples for *de*. Perhaps another area to develop this research would be to focus on spoken language. Indeed, this project focuses on written language only. Finally, an investigation done diachronically rather than just synchronically could evaluate this research from another view.

6 References

Atkins, B. T. S. & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Dixon, R. M. W. (2022). *English Prepositions: Their Meanings and Uses*. Oxford: Oxford University Press.

Grevisse, M. & Lits, M. (Eds.). 2009. *Le Petit Grevisse*. Louvain-la-Neuve: De boeck supérieur.

Grevisse, M. & Goosse, A. (Eds.). 1986. *Chapitre VII: La préposition.* In *Le Bon Usage: Grammaire Française*. Paris-Gembloux: Editions Duculot.

Keun Young, S. (2007). *Partitive Structure*. Retrieved: 06/05/2022, from *ProQuest*: <https://www.proquest.com/docview/304852773/fulltextPDF/B2AC62682C614EC8PQ/1?accountid=14182>.

Kilgariff, A. Rychlý, R., Smrž, P. , Tugwell, D. (2004). *The Sketch Engine.* Retrieved: 27/10/2012, from *Sketch Engine*: <http://www.sketchengine.eu>.

Landau, S. I. (2001). *Dictionaries: the art and craft of lexicography*. Cambridge: Cambridge University Press.

Le Grand Robert. (2017). *Preposition 'de'*. Retrieved: 20/04/2022, from *le Grand Robert*: <https://grandrobert.lerobert.com/>.

Malkiel, Y. (1993). *Etymology*. Cambridge: Cambridge University Press.

Murphy, L. (2021). *Small Words* [Unpublished master's dissertation]. University of Sussex.

Murray, J., A., H. (c.1905). A new English Dictionary on Historical Principles. Volume VII. Retrieved 05/05/2022, from *OED*: <http://public.oed.com/wp-content/uploads/Volume-VII-O-Ozyat.pdf>.

Pullum, G. & Huddleston, R. K. (2002). Prepositions and preposition phrases. In G. Pullum & R. Huddleston (Eds.), *The Cambridge Grammar of the English Language* (pp. 597-661). Cambridge: Cambridge University Press.

Robert, P., Rey, A. & Morvan, D. (1976*). Dictionnaire Etymologique de la Langue Française*. Paris: Dictionnaires Le Robert.

Simpson, J. & Weiner, E. (2004). *Oxford English Dictionary*. Oxford, England: Oxford University Press.

Tyler, A. & Evans, V. (2003). *The Semantic of English Prepositions: Spatial Scenes, Embodied Meaning, and Cognition.* Cambridge: Cambridge University Press.

7 Appendices

7.1 Appendix 1: The Partitive article

7.1.1 Sample 2

A page of sample 2 is given here to show an example of how the identification was done.

29/04/2022, 18:33

Concordance | Sketch Engine

SKETCH ENGINE

SAMPLE
2

200

6

simple de • **2,055**
78,214.2 per million tokens • 7.8%

shuffle • **2,055**
78,214.2 per million tokens • 7.8%

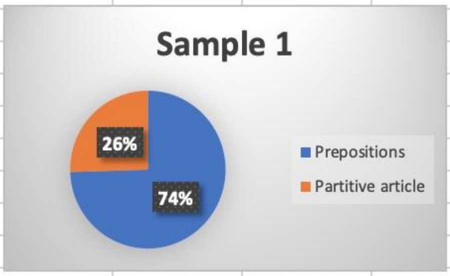
Sample 201 • **201**
7,650.15 per million tokens • 0.77%

		Details	Left context	KWIC	Right context
P	1	doc#0	<s> Cotentin : comment anticiper les impacts	du	changement climatique sur les huîtres et les moules ?
P	2	doc#0	illes d'huîtres vides réduites en poudre dans les zones	d'	élevage. </s><s> Cet article vous a été utile ? </s><s>
AP	3	doc#1	lité de vos villes et marques favorites. </s><s> La part	de	la France dans la lutte contre le changement climatique
P	4	doc#1	tre le changement climatique en question Paris, du fait	de	sa responsabilité historique dans le réchauffement, dev
AP	5	doc#1	l'une réalité simple : les pays développés ont le primat	de	la lutte contre le changement climatique, dont ils sont l
P	6	doc#1	ıla l'obligerait ainsi à abaisser ses émissions de 168 %	d'	ici à 2030 par rapport à 1990. </s><s> Un objectif total
P	7	doc#1	En marge du processus officiel, 33 coalitions de pays,	d'	entreprises ou d'institutions ont en outre été lancées po
P	8	doc#1	›cessus officiel, 33 coalitions de pays, d'entreprises ou	d'	institutions ont en outre été lancées pour réduire les én
P	9	doc#1	e huit EPR2 supplémentaires, un projet qui fera l'objet	d'	études. </s><s> L'EPR 2 est un projet de version << o
AP	10	doc#1	e << plus simple à construire >>, bénéficiant d'un effet	de	série (construction par paires) et de préfabrications en
P	11	doc#1	; raisonnables >>. </s><s> Tout projet de construction	d'	EPR 2 devra notamment recevoir l'autorisation de l'Aut
P	12	doc#1	n débat public sur le projet et le dépôt par EDF auprès	du	gouvernement d'une demande d'autorisation de créatio
P	13	doc#2	st pas encore terminée. </s><s> Pêche dans les eaux	de	Jersey : les Normands mieux lotis que les Bretons, ma
P	14	doc#2	ste insuffisant Le bras de fer continue pour l'attribution	des	licences de pêches dans les eaux britanniques. </s><s
P	15	doc#2	</s><s> Ce mercredi 29 septembre, le gouvernement	de	Jersey a délivré un certain nombre de licences aux ma
P	16	doc#2	dernières attributions de licences par le gouvernement	de	Jersey. </s><s> Les marins pêcheurs normands ont ol
P	17	doc#2	n a une série de bateaux en attente qui doivent fournir	des	preuves de présence dans les eaux de Jersey. </s><s
AP	18	doc#2	ıze milles nautiques. </s><s> Dimitri Rogoff, président	du	comité régional des pêches de Normandie. </s><s> D
P	19	doc#2	; normands qui ajoute "c'est compliqué pour les moins	de	douze mètres où on est pas sur des positions satellitai
P	20	doc#2	aisceaux de preuves". </s><s> L'horizon bouché pour	des	pêcheurs manchois Ils sont quelques uns dans le port
P	21	doc#2	> A quelques kilomètres plus au nord sur la côte ouest	du	Cotentin, Gilles Musard a eu plus de chance. </s><s>
P	22	doc#2	icences des pêcheurs de l'Union européenne (et donc	de	Granville) La pêche est l'un des sujets épineux dans le
AP	23	doc#2	envenu pour les pêcheurs français et notamment ceux	de	la Baie de Granville. </s><s> Trois mois de plus. </s>
P	24	doc#2	ır les pêcheurs français et notamment ceux de la Baie	de	Granville. </s><s> Trois mois de plus. </s><s> De quo
P	25	doc#2	›rd sur la pêche dans les eaux de l'île anglo-normande	de	Jersey. </s><s> Le gouvernement français et l'Union e
AP	26	doc#2	/s><s> Leur interlocuteur outre-manche a donné gage	de	bonne volonté dans un dossier qui envenime les relatio

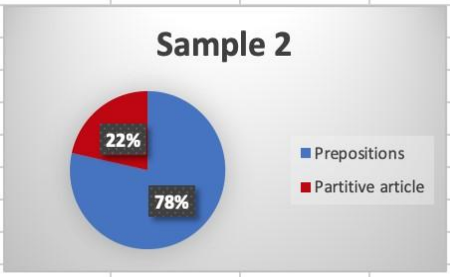
7.1.2 *Sample Calculations (Excel)*

SAMPLE PARTITIVE ARTICLE

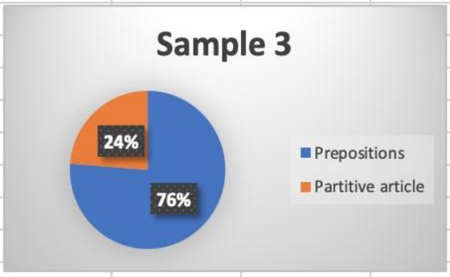
Sample data 1/200	Instances
Prepositions	149
Partitive article	51



sample data 2/200	Instances
Prepositions	157
Partitive article	43



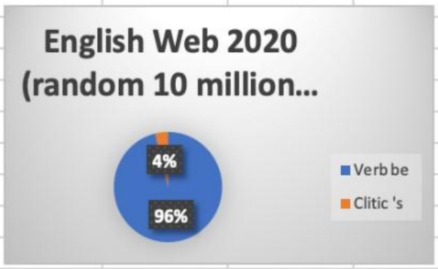
sample data 3/400	Instances
Prepositions	305
Partitive article	95



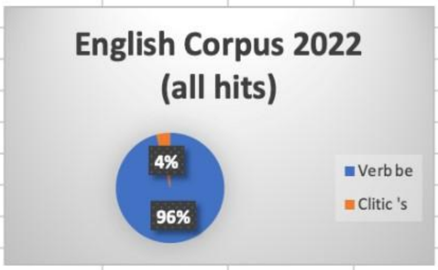
7.2 Appendix 2: Clitic *'s* and Possessive Case

CORPUS CLITIC/VERB *BE*

Clitic *'s*	English Web 2020 (random 10 million hits)
Verb *be*	9,620,580
Clitic *'s*	379,420



Clitic *'s*	English Corpus 2022 (all hits)
Verb *be*	70
Clitic *'s*	3

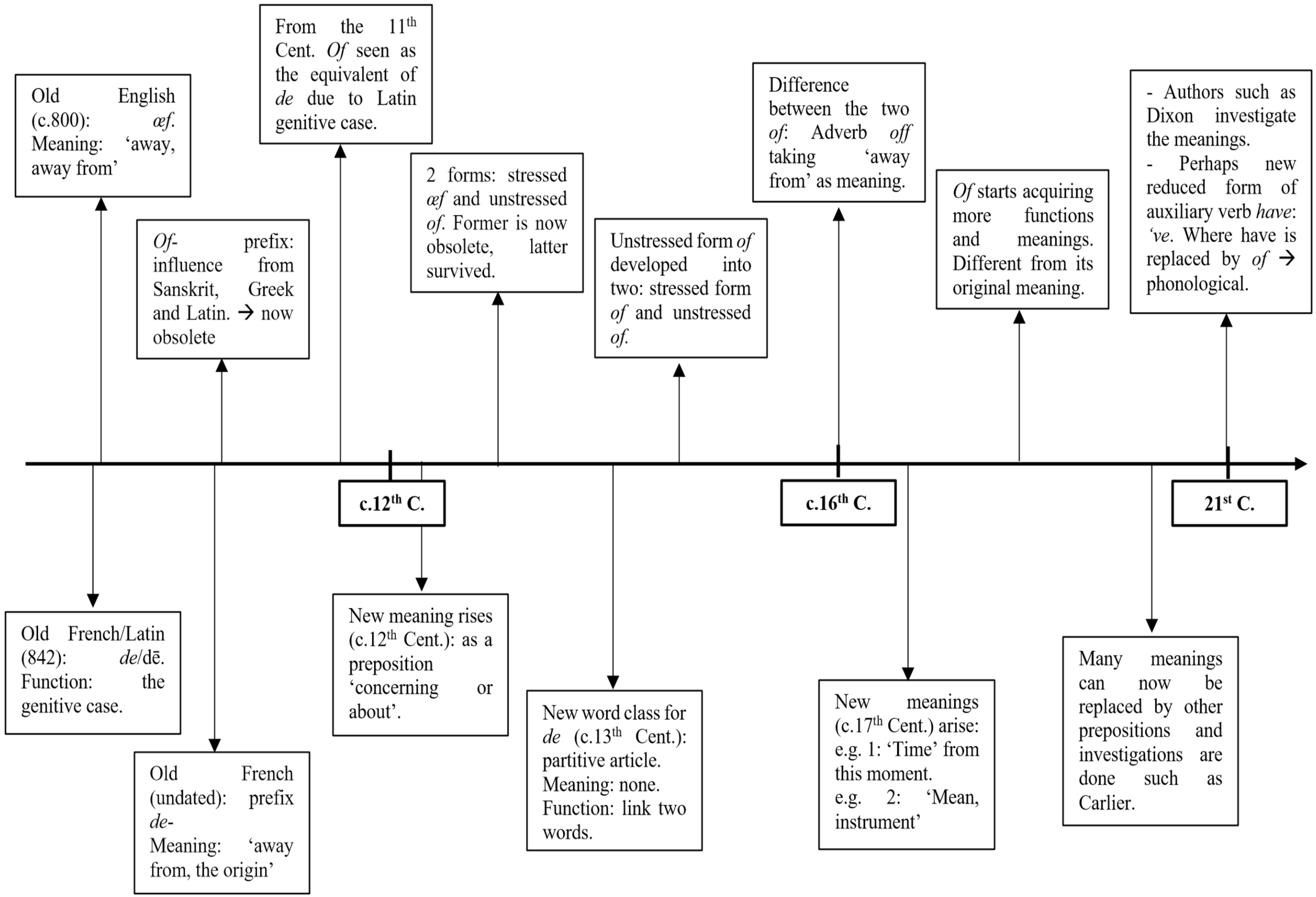


7.3 Appendix 3: Websites of the Articles

French	English
Actu.fr	BBC News
Lemonde.fr	BBC Sports
Franceinfo	Climatechangenews.com
Liberation	France24
Lefigaro Tf1info.fr	The Guardian

7.4 Appendix 4: Timeline in a Bigger View

Timeline of the prepositions *of* and *de*. (Dates have been estimated for the preposition *de* depending on the example given from Le Grand Robert dictionary).



Old English (c.800): *æf*. Meaning: 'away, away from'
Of- prefix: influence from Sanskrit, Greek and Latin. → now obsolete
From the 11th Cent. *Of* seen as the equivalent of *de* due to Latin genitive case.
2 forms: stressed *æf* and unstressed *of*. Former is now obsolete, latter survived.
Unstressed form of *of* developed into two: stressed form *of* and unstressed *of*.
Difference between the two *of*: Adverb *off* taking 'away from' as meaning.
Of starts acquiring more functions and meanings. Different from its original meaning.
- Authors such as Dixon investigate the meanings.
- Perhaps new reduced form of auxiliary verb *have*: *'ve*. Where have is replaced by *of* → phonological.
c.12th C.
c.16th C.
21st C.
Old French/Latin (842): *de*/dē. Function: the genitive case.
Old French (undated): prefix *de-* Meaning: 'away from, the origin'
New meaning rises (c.12th Cent.): as a preposition 'concerning or about'.
New word class for *de* (c.13th Cent.): partitive article. Meaning: none. Function: link two words.
New meanings (c.17th Cent.) arise: e.g. 1: 'Time' from this moment. e.g. 2: 'Mean, instrument'
Many meanings can now be replaced by other prepositions and investigations are done such as Carlier.

Acknowledgements

Throughout the year, I have received significant help and advice for this research project. I would particularly like to thank Lynne Murphy for the idea and putting me in the right direction, her constant support and advice whether it was in our dissertation meetings, online or in person. I would also like to thank Freya Stott, Ella Nelson, and Becky Hunt for being amazing dissertation peers, support and all the reviewing they have given me. It has been amazing working with them, and I count myself lucky to have been part of this dissertation group.

Many thanks also to Melanie Green for her advice and tips throughout the year.

I would like to thank Lynn Day and Gina Chapman for all the proof-reading and recommendations. Finally, thank you to Amber Chapman, Myriam and Alain Vanrapenbusch, Elodie Vanrapenbusch and Guillaume Herbomel as well as Mathilde and Jean-Baptiste Zolla for their non-stop support and encouragement.

Thank you to everyone for believing in me.

SECTION B

Squibs and Write-ups

Coordinating Theories of Coordinate Structures: Evidence from Thai

Ariane Branigan
The University of Edinburgh

Abstract. Since Ross' seminal work on coordination in 1967, a great deal of attention has focused on trying to explain the patterns we see in what can – and cannot – be extracted from these structures. Attempts to explain them through a lens of either syntax (e.g., Ross, 1967, and Gazdar, 1981) or discourse (e.g., Lakoff, 1986) alone have proved to be problematic. There is, however, a promising new theory developed by Altshuler and Truswell (in press) unifies both perspectives; unfortunately, it suffers from a lack of cross-linguistic data. I will provide cross-linguistic evidence for this theory by taking evidence from Thai to show that the authors' predictions around the status of adjuncts as weak islands and topical material can account for patterns of grammaticality and ungrammaticality when extracting from coordinate structures.

Keywords: syntax; coordination; Thai

Coordination shows a great deal of cross-linguistic variation, both in how it is presented syntactically, and how it contributes to discourse coherence (Ramm, 2011). Accordingly, there has been a growing body of research dedicated to documenting these cross-linguistic differences (Mauri, 2008, p. 6). Although these studies' focus has widened (notably Haspelmath, 2004), the majority of analyses are still restricted to Indo-European languages (cf. Ramm, 2011; Mauri, 2008). In this essay, I take a tentative and provisional step towards addressing this shortcoming by examining coordination in Thai; as Thai has received little attention in English-language coordination literature, I approach the phenomenon through the lens of various influential works on the subject. I first examine the syntax of Thai coordinate structures and demonstrate that they match the extraction patterns outlined by both Ross (1967) and Lakoff (1986). Next, I examine the discourse-coherence functions of *lae* 'and' in Thai, showing that they are congruent with Txurruka's (2003) discourse-based assessment of 'and'. I then use the evidence gathered in the previous sections to demonstrate that Thai provides cross-linguistic support for Altshuler and Truswell's (2022) predictions around the syntactic and discourse-coherence criteria for extraction from coordinate structures. Finally, I provide a brief overview of Thai Serial Verb Constructions (SVCs) and their discourse functions and suggest directions for future research which encompasses both coordination and SVCs.

Coordination-related syntactic research often focuses on the extractability of syntactic elements from within coordinate structures; very little has been written about Thai in this regard. In English, extractability is tested using *wh*-movement — although Thai is a *wh*-in-situ language, it still has A' movement in the form of topicalization (Ruangjaroon, 2007, p. 121). This provides a parallel diagnostic for determining the acceptability of extraction (Chomsky, 1977, p. 94). I will now examine how Thai conforms to two different models of extraction from coordinate structures - Ross (1967) and Lakoff (1986) — based on topicalization grammaticality judgements from native Thai speakers[52].

Ross (1967, p. 161) introduces the Coordinate Structure Constraint (CSC), which stipulates that you cannot move any element in a conjunct, nor the conjunct itself, out of a coordinate structure. Previous research suggests that Thai obeys the first of these generalisations; no elements in a conjunct can be topicalized:

[52] Unless otherwise indicated, all Thai examples are my own.

(1) *Dek tham kanbaan lae duu thiiwii kriang*
Child do homework and watch TV CLASS.
nan phroomkan
that simultaneously
'The child did his homework and watched that TV at the same time.'

(2) **Thiiwii khriang nan nae dek tham kanbaan lae*
TV CLASS. that TOP child do homework and
duu phroomkan
watch simultaneously
'That TV, the child did his homework and watched _____ at the same time.'
(Wilawan, 1993, p. 61)

Acceptability judgements suggest that Thai also obeys the second constraint (moving conjuncts out of a coordinate structure):

(3) **Sofa taw nan khea wang keaxi wi rahwang*
Sofa CLASS. that he put chair some between
toa bang taw lae
between table some and
'That sofa, he put a chair between some tables and ______.'

Ross (1967, p. 174) also notes an exception to the CSC, which he terms Across The Board (ATB) movement: extraction from conjuncts is acceptable only if the same element is extracted from every conjunct. Thai fulfils this prediction, too:

(4) *Mamuang luk nan chan chorp lae khun gliad*
Mango CLASS. that I like and you hate
'That mango, I like ______ and you hate ______.'

Since Ross' (1976) PhD, however, further research has suggested that ATB is not the only exception to the CSC. Lakoff (1986, PAGE NUMBER) notes three scenarios where extraction from individual conjuncts is acceptable; this prompts him to suggest that there is no purely syntactic CSC. The first of these exceptions, which he names *Type A*, depicts "an expected course of events" and requires extraction from the final conjunct:

What did he go to the store and buy _____?

The second type of extraction pattern, *Type B,* describes an event where conventional expectations are violated. It does *not* require extraction from the final conjunct:

How much can you drink ______ and still stay sober?

In *Type C*, the final pattern, the first clause describes an action and the second describes its result. This also only requires extraction from the first conjunct:

That's the stuff that the guys in the Caucasus drink ______ and live to be a hundred.

Thai follows these patterns, with all three types resulting in grammatical extraction:

(5) **Type A**

Mamuang		*luk*	*nan*	*chan*	*bai*	*dalaat*	*lae*	*suu*
maa								
Mango	CLASS.	that	I	go	market	and	buy	PAST

'That mango, I went to the market and bought ______.'

(6) **Type B**

Beiir	*nan*	*chan*	*duum*	*lae*	*mai*	*mao*
Beer	that	I	drink	and	not	drunk

'That beer, I drank ______ and still remained sober.'

(7) **Type C**

Kaijiew	*nan*	*chan*	*gin*	*laew*	*lae*	*ruesuk*	*im*
Omelette	that	I	ate	PAST	and	feel	full

'That omelette, I ate ______ and feel full.'

This suggests that Thai is similar to English in terms of extractability from coordinate structures. Having given an overview of the syntactic background, I will now turn to the semantics of coordination and show that Txurruka's (2003) analysis of English 'and' matches the semantics of Thai *lae*.

Txurruka (2003) suggests that clause-linking 'and' indicates the presence of a coordinating discourse relation, which she terms a Coordinator. She argues that the discourse relations Narration and Result are Coordinators:

Narration:
Mary put on her tutu and pruned the apple tree.
Result:
I drank coffee after lunch and couldn't sleep the whole night.

She also touches on the role of Subordinators, another set of relations which build discourse differently to Coordinators; this contrast is best described as follows:

> '[…] assuming we have two constituents, labeled α and β for which R1(α,β) is already established… If you can attach ɤ to α, then R1 is Subord. If you can attach only to β, R1 is Coord.' (Ascher and Vieu, 2005, p. 600, paraphrased)

Txurruka argues that, because these Subordinators are structurally incompatible with Coordinators, 'and' cannot introduce Subordinator discourse relations like Explanation and Elaboration. Her argument focuses on English; however, *lae* seems to license the same Coordinators, and block the same Subordinators, as 'and':

(8) **Narration**

Chan	*bai*	*dalaat*	*lae*	*suu*	*mamuang*
I go	market	and	buy	mango	

'I went to the market and bought mangoes'

(9) **Result**

Chan	*duuem*	*gafae*	*lae*	*mai*	*lap*	*khun*	*nan*
I drank	coffee	and	NEG	sleep	night	that	

'I drank coffee and couldn't sleep that night'

(10) **Explanation**

??*Nong lom long lae kao luen bon phun piiyak*
Nong fall down and she slip on floor wet
'Nong fell and she slipped on the wet floor'

(11) **Elaboration**
??*Muuea wan chan gin aahaan aroy lae bai Burger King*
Yesterday I eat food tasty and go Burger King
'Yesterday I ate tasty food and went to Burger King[53].

Thai coordination is clearly similar to English in terms of both its syntax and its function in discourse coherence. With this wider context established, we can now turn to Altshuler and Truswell (2022), who provide an explanation for the extraction patterns and discourse relations outlined by the three works mentioned previously (along with numerous others).

The authors refute a purely syntactic- or discourse-based account of coordination extraction. From a syntactic perspective, they point to Lakoff's (1986) examples as exceptions to the element condition of the CSC which are "conditioned by discourse", i.e., not syntactically motivated (Altshuler and Truswell, 2022, p. 46). However, they also note that Lakoff's (1986) patterns of extraction are not solely determined by discourse relations, either; for example, they identify the following Type A sentence which, counter to Lakoff's assertion, does not require extraction from the final conjunct:

'Which knife did Lizzie take _____ and hack the steak to pieces?' (Brown, 2017, p.21)

Instead, the authors suggest an alternative explanation which unifies syntax- and discourse-based accounts. Firstly, they derive a rule for determining whether a language allows extraction from a non-initial conjunct, using the syntactic structure of coordination proposed by Munn (1993) in which non-initial conjuncts are adjuncts of the initial conjunct. Accordingly, they argue that extractability from a non-initial conjunct is determined by the island status of adjuncts in the language under discussion (Altshuler and Truswell, 2022, p. 139; Truswell, 2021, p. 1).

Altshuler and Truswell (2022, pp. 136, 290) note that so far, this pattern has only been observed in Western European languages such as English and French (Truswell, 2021, p. 1; Postal, 1998, p. 76). However, it seems to extend to Thai as well - firstly, Thai adjuncts are weak islands:

(12) *Khon nan Nong bai Angrit phuey pachern*
Person that Nong go England to confront
'That person, Nong went to England to confront ______.'

Therefore, because Thai adjuncts are weak islands, we should also be able to extract from non-initial Thai conjuncts. This prediction is borne out, as we have already seen in (5), repeated below:

(13) *Mamuang luk nan chan bai dalaat lae suu*
Mango CLASS. that I go market and buy
'That mango, I went to the market and bought ______.'

Extraction from non-initial conjuncts is clearly possible in Thai - we now turn to the rest of Altshuler and Truswell's argument to understand how discourse determines actual extraction patterns. Building off Txurruka's (2003) work, the authors identify three classes of discourse relations expressed by 'and':

[53] One of my informants notes that this sentence "makes you sound very greedy".

Class 1, which requires a common topic (Continuation, Narration, Background →), Class 2, which requires a common theme (Parallel), and Class 3, which does not require a common topic or theme (Result) (Altshuler and Truswell, 2022, p. 261-7; Truswell, 2021, p. 3). They argue that the ability to extract from a conjunct is determined by these classes: ATB is possible for all three, extraction from initial conjuncts is possible for Classes 1 and 3, and extraction from non-initial conjuncts is possible for Class 1 assuming the language's adjuncts are weak islands, and the extracted material is topical (Altshuler and Truswell, 2022, p. 261-7; Truswell, 2021, p.3). Examples of each are paraphrased from Altshuler and Truswell (2022):

Class 1

ATB: What did Ava [write _____] and [discuss _____ with Teia]?
Initial: What city did Mary [go to____] and [buy a painting]?
Final: What painting did Mary [go to NYC] and [buy____]?

Class 2

ATB: What book did [John buy ___] and [Bill read ___]?
Initial: ?? What book did John [buy ____] and [read a magazine]?
Final: ?? What magazine did John [buy a book] and [read ____]?

Class 3

ATB: What did you [hear ____] and [think of ____]?
Initial: What's the stuff the guys in the Caucasus [drink ____] and [live to be 100]?
Final: ?? Who did Kharms [hear a news story] and [think of ____]?

Before we test these claims in Thai, we should first confirm whether *lae* can introduce the discourse relations which comprise the three classes mentioned above. We have already shown through evaluating Txurruka's (2003) work that *lae* introduces Narration and Result discourse relations, so we only need to check Continuation, Background →, and Parallel. Again, Thai speakers judge these sentences to hold the same discourse relations as English:

(14) **Continuation**

Nong sai suuea-koed lae Nok gin mamuang
Nong put on coat and Nok eat mango
'Nong put on her coat, and Nok ate mango.' (simultaneously)

(15) **Background**

Nong yuu gap Nok lae aan nangsuu hai kao
Nong live with Nok and read book to her
'Nong lives with Nok, and reads a book to her.'

(16) **Parallel**

Nong hen Nok lae Ponrawee hen Nok duuay
Nong see Nok and Ponrawee see Nok also
'Nong saw Nok, and Ponrawee saw Nok too.'

We have seen that Thai behaves like English both in terms of adjunct extraction, and which discourse relations can be introduced by *lae*. By extension, we would expect Thai to meet the rest of Alshuler and Truswell's (2022) predictions around extractability based on the three classes described above. These predictions are indeed met:

Class 1

(17) ATB: *Reung nan Nong khian lae pruksa Nok*

Book that Nong write and consult Nok
'That book, Nong wrote ______ and discussed ______ with Nok.'

(18) Initial: *Muueang nan Nong bai tii yuu lae suu papwad*
City that Nong go visit and buy painting
'That city, Nong visited ______ and bought a painting.'

(19) Final: *Papwad nan Nong bai Krungthep lae suu*
Painting that Nong go Bangkok and buy
'That painting, Nong went to Bangkok and bought ______.'

Class 2

(20) ATB: *Nangsuu lem nan Nong suu lae Nok aan*
Book CLASS. that Nong bought and Nok read
'That book, Nong bought ______ and Nok read ______.'(

(21) Initial: ??*Nangsuu lem nan Nong suu lae aan nitayasar lem*
Book CLASS. that Nong bought and Nok read magazine CLASS.
'That book, Nong bought ______ and read a magazine.'

(22) Final: ??*Nitayasar lem nan Nong suu nangsuu lae aan*
Magazine CLASS. that Nong bought book and read
'That magazine, Nong bought a book and read ______.'

Class 3

(23) ATB: *Ai-dim nan chan leum sai nai tu-chae lae kor leuy gin mai dai*
Ice cream that I forgot put in freezer and therefore eat NEG can
'That ice cream, I forgot to put in the freezer ____ and therefore can't eat ____.'

(24) Initial: *Kaijiew nan chan gin laew lae ruesuk im*
Omelette that I ate PAST and feel full
'That omelette, I ate ______ and feel full.'

(25) Final: ??*Khon nan chan diyin khao lae kor leuy nukthung*
Person that I hear news story and therefore thought of
'That person, I heard a news story and thought of ______.'

Our investigation of Thai lends valuable cross-linguistic credence to Altshuler and Truswell's theory. Not only does it provide syntactic evidence that there is indeed a link between the weak-island status of adjuncts and extractability from initial conjuncts, but it also shows that the same coordinating discourse relations result in the same extraction criteria and patterns.

However, Thai coordination is not restricted to *lae*; like many other Southeast Asian languages, it also has Serial Verb Constructions (SVCs), which are sequences of multiple verbs not overtly linked by coordination or subordination (Aikhenvald, 2006:1,8). Below is an example of a Thai SVC:

(26) *Khao bai suu maa gin*
He go buy come eat

'He went to buy something and brought it back to eat.'
(Smyth, 2002, p.81)

The syntactic structure of Thai SVCs does not seem to match that of coordination (cf. Wilawan, 1992; Thepkanjana, 1986). However, from a discourse perspective, SVCs fulfil a similar function to *lae* - all examples are taken from Diller (2006), although the emphasis on the relations is my own:

(27) **Narration**

Yip	*duu*
Pick up	look

'She picked it up, and looked at it'

(28) **Result**

Gin	*im*
Eat	full

'They ate, and became full'

(29) **Background**

Yuun	*phuud*
Stand	speak

'He speaks while standing'

Of particular interest is the pragmatic and syntactic constraints which determine when each construction can be used; although there is acknowledgement in the literature that the same sentence can have different meanings depending on whether it is presented as an SVC or coordinate structure, to the best of my knowledge there is no research on how this applies to Thai specifically (Aikhenvald, 2006, p. 6). Cursory investigation suggests there is great potential for further research; for example, despite Background → being available for both, there seem to be instances where only SVCs are appropriate:

(30)

Nong	*yuun*	(??*lae*)	*takon*	*sai*	*Nok*
Nong	stand	(and)	shout	at	Nok

'Nong stood and shouted at Nok'

More research into this relationship could also contribute to our understanding of how speakers perceive event structure cross-linguistically; existing literature suggests that SVCs generally portray more "tightly sequenced" events than coordinate structures, but there has been no work examining this difference empirically (Diller, 2006, p. 164).

Through the lens of various influential works, I have provided a preliminary summary of the syntax and discourse of Thai coordination. In particular, I have shown that Thai coordination bears remarkable similarities to English; in addition to having the same extraction patterns outlined by Ross (1967) and Lakoff (1986), *lae* also introduces the same discourse relations suggested by Txuruka (2003) for 'and'. Furthermore, I have demonstrated that Thai supports the theory proposed by Altshuler and Truswell (2022) regarding adjunct- and topic-based extraction patterns; this serves to emphasise the importance of cross-linguistic research when formulating and testing linguistic theories. Finally, I have provided suggestions for future research which could delineate the similarities and differences between coordination and SVCs. Although this essay has given a much-needed overview of coordination in Thai, it is clear that a great deal more work is needed to rectify the gap in the literature both for Thai, and non-Indo-European languages in general.

References

Aikhenvald, A. Y. (2006). *Serial Verb Constructions in Typological Perspective.* In A. Y. Aikhenvald and R. M. W. Dixon (Eds.), *Serial Verb Constructions: A Cross-Linguistic Typology* (pp. 1-68). Oxford: Oxford University Press.

Altshuler, D. and Truswell, R. (2022). *Coordination and the Syntax-Discourse Interface* (Vol. 12). Oxford: Oxford University Press.

Ascher, N. and Vieu, L. (2005). Subordinating and coordinating discourse relations. *Lingua*, 115(4), 591-610.

Brown, Jessica (2017). *Heads and Adjuncts: An Experimental Study of Subextraction from Participials and Coordination in English, German and Norwegian*. [Unpublished doctoral dissertation]. University of Cambridge.

Chomsky, N. (1977). *On wh-movement*. In P. Culicover, T. Wasow, and A. Akmajian (Eds.), *Formal syntax* (pp. 71-132). New York: Academic.

Diller, A. (2006). Thai *Serial Verbs: Cohesion and Culture*. In A. Y. Aikhenvald and R. M. W. Dixon (Eds.), *Serial Verb Constructions: A Cross-Linguistic Typology*. Oxford: Oxford University Press.

Haspelmath, M. (2004). *Coordinating Constructions*. Amsterdam: John Benjamins Publishing Company.

Lakoff, G. (1986). *Frame Semantic Control of the Coordinate Structure Constraint*. Chiacago: Chicago Linguistic Society.

Mauri, C. (2008). *Coordination relations in the languages of Europe and beyond*. Berlin, New York: De Gruyter Mouton.

Munn, A. (1993). *Topics in the Syntax and Semantics of Coordinate Structures*. Unpublished doctoral dissertation]. University of Maryland.

Postal, P. M. (1998). *Three Investigations of Extraction*. Massachusetts: MIT Press.

Ramm, W. (2011). *Discourse-structural salience from a cross-linguistic perspective: Coordination and its contribution to discourse*. In C. Chiarcos, B. Claus, & M. Grabski (Eds.), *Salience: Multidisciplinary Perspectives on its Function in Discourse* (pp. 143-172). Berlin, New York: De Gruyter Mouton.

Ross, J. R. (1967). *Constraints on Variables in Syntax*. [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.

Ruangjaroon, S. (2007). Thai Wh-expressions at the Left Edge of the Clause: Contrastive and Identificational Wh-clefts. *Concentric: Studies in Linguistics*, 33(2), 121-157.

Smyth, D. (2002). *Thai: An Essential Grammar*. London: Routledge.

Thepkanjana, K. (1986). *Serial Verb Constructions in Thai*. [Unpublished doctoral dissertation]. University of Michigan.

Truswell, R. (2021). Unpublished Class Handout. University of Edinburgh.

Txurruka, I. G. (2003). *The Natural Language Conjunction "and"*. Linguistics and Philosophy, 26(3), 255-285.

Wilawan, S. (1992). *A reanalysis of so-called serial verb constructions in Thai, Khmer, Mandarin Chinese, and Yoruba*. [Unpublished doctoral dissertation]. University of Hawaii.

Acknowledgements

I would like to extend sincere thanks to my Thai informants - Nong Branigan, Ponrawee Prasertsom, and Tanae Cheurthai - for their invaluable contribution in the form of (many) acceptability judgements.

Indecisiveness: The Acceptability of 'It Depends' as a Response to *Or*-Questions

Aaliyah Bullen
Swarthmore College

Abstract. Conversation between even the most effective interlocutors can contain points of uncertainty and indecisiveness, but we propose there are some situations in which this indecisiveness can be unacceptable. In this paper, we defend the use of *it depends* as a sign of indecisiveness. We standardise ineffectiveness as conversational moves that lead to the disobedience of conventions regarding talk exchanges. The conventions in our discussion will primarily include the Gricean maxims of quantity and quality. Through this discussion, we will investigate and identify specific contexts during which *it depends* is acceptable and unacceptable. We reason that when an interlocutor responds with it depends, they are expressing an uncertainty caused by the variation of their answer depending on the situation. To better understand the meaning conveyed when an interlocutor responds with it depends, we can look at the entailments of the response. Entailments will be defended through conducted cancellation tests. From these cancelation tests, we can conclude that *it depends* entails that there are situations in which both given options are possible. We consider this a reflection of what we loosely define as uncertainty and indecisiveness and will use these entailments as a formalisation of this property for the paper.

Keywords: acceptability judgements; Gricean maxims

1 Introduction

Conversation between even the most effective interlocutors can contain points of uncertainty and indecisiveness, but we propose there are some situations in which this indecisiveness can be unacceptable. An example of this is shown in (1), an adjustment to a common conversation between a store cashier and a paying customer:

(1) (a) Do you want to use either cash or credit?
(b) It depends.

In this paper, we defend the use of *it depends* as a sign of indecisiveness and explore how and when speaker B's response in (1) can be semantically anomalous. We standardise ineffectiveness as conversational moves that lead to the disobedience of conventions regarding talk exchanges. The conventions in our discussion will primarily include the Gricean maxims of quantity and quality. Through this discussion, we will investigate and identify specific contexts during which *it depends* is acceptable and unacceptable.

2 Formalising 'It Depends'

We reason that when an interlocutor responds with *it depends,* they are expressing an uncertainty caused by the variation of their answer depending on the situation. To better understand the meaning conveyed when an interlocutor responds with *it depends*, we can look at the entailments of the response. Entailments will be defended through conducted cancellation tests.

We can draw an example of the meaning from the conversation introduced in (1) with no creed to the acceptability or otherwise of the response. Statements (2) and (3), show below, are generated entailments of Speaker B's response:

(2) There are situations where I want to use cash.
(3) There are situations where I want to use credit.

To conduct the cancelation test, we generated sentences (2a) and (3a), which are conjunctions of Speaker B's response and the negation of the entailments in (2) and (3), respectively:

(2) There are situations where I want to use cash.
(2a) #It depends and there are no situations where I want to use cash.
(3) There are situations where I want to use credit.
(3a) #It depends and there are no situations where I want to use credit.

We claim that (2a) and (3a) are contradictions. For (2a), if there are no situations where Speaker B wants to use cash, saying i*t depends* would be untruthful and extraneous information. An accurate response would be to say they want to use credit as there are no cases where they want to use cash. Similarly, for (2b), if there are no situations where Speaker B wants to use credit, saying *it depends* would create a contradiction. Speaker B's response depends on nothing if they only want one of the two options, so saying *it depends* would be nonsensical.

From these cancelation tests, we can conclude that *it depends* entails there are situations in which both given options are possible. We consider this a reflection of what we loosely define as uncertainty and indecisiveness, and will use these entailments as a formalisation of this property for the remainder of the paper.

All of the instigating statements to be discussed will be in the form of binary or-questions. By providing two contrasting options to be accounted for in the response, we believe tracking the response's entailments will intuitively follow those established from conversation (1), the cash or credit example.

3 When 'It Depends' is Acceptable

In our investigation, we identified two situations where *it depends* is an acceptable response: (1) giving preferences and (2) describing socially-spaced states of being. We do not suggest that these situations are all-inclusive and only intend to contribute a base understanding of when *it depends* is acceptable. For each given category, we will delve into a specific example conversation that displays the acceptability in our conjecture.

Acceptability will be measured according to their cooperation with Grice's maxims of quality and quantity. To elaborate on this, our method involves labelling a response as acceptable when it, and its entailments, provide accurate information that is neither extraneous nor insufficient. Accordance with the Gricean maxims is an important measure of acceptability because it allows the listener to make assumptions in line with the cooperative principle and continue the conversation with suitable information.

3.1 Giving Nonurgent Preferences

This category involves situations where *it depends* is in response to a prompt for a speaker to give their preference. *Like*, *want*, *prefer*, and other verbs of a similar vein are common indications of an instigating question to give preferences. The second key descriptor of this situation category is that the preference to be given is nonurgent. This lack of urgency prevents a violation of the quantity maxim solely due to how quickly the information is needed. Violations of quantity in urgent situations will be discussed further in our upcoming discussion of when *it depends* is unacceptable.

Our example conversation in (4) surrounds Speaker A and B who just began an hour-long, casual meeting in a coffee bar to get to know each other better. The dialogue and the entailments of Speaker B's response, shown in (5) and (6), are listed below:

(4) (a) Do you like either dogs or cats?
(b) It depends.
(5) There are situations where I like dogs.
(6) There are situations where I like cats.

We maintain our view that Speaker B's response is suitable in this context as it satisfies Grice's maxims of quality and quantity. Quality is sufficiently met because as far as Speaker A knows, there are situations where Speaker B likes dogs and situations where Speaker B likes cats. If there were any faults in the quality of this information, it would not be due to the use of *it depends* as a response. The maxim of quantity is satisfied because Speaker B conveys they like dogs and cats, just in different situations. This information helps Speaker A's desire to learn more about Speaker B at a level sufficient for casual conversation, and allows the conversation to continue as intended.

3.2 Nonurgent Societally-Spaced States of Being

This category involves situations where *it depends* is in response to a question about the responder's state of being, so how they are in the moment. More specifically, these questions all exist in a relative, societally-spaced spectrum. For instance, questions regarding race, maturity, emotional state, and gender could all give rise to an acceptable *it depends*. Additionally, these situations should be nonurgent in the manner we defined previously.

Our first example conversation in (7) is set in a restaurant bar where Speaker A is trying to get to know more about Speaker B, a potential romantic interest seated near them. The dialogue and the entailments of Speaker B's response, shown in (8) and (9), are listed below:

(7) (a) Are you a boy or a girl?
(b) It depends.
(8) There are situations where I am a boy.
(9) There are situations where I am a girl.

Working with the generally-accepted perception that gender exists on a spectrum, we consider Speaker B's response to be effective. Speaker B adheres to the maxim of quality because depending how Speaker A wishes to define what a boy or a girl is, Speaker B could fall in different categories as they exist somewhere on the spectrum that is not clearly identifiable. We argue that this response also follows the maxim of quantity because Speaker B presents the information that Speaker A needs to know for building their view of Speaker B as a potential partner. Some may find it preferable that Speaker B gave one specific choice in the binary of options given, but that would be a strong violation of the maxim of quality. Quality is especially important when responding to questions as Speaker B is responsible for providing new information to Speaker A. In these less urgent situations, the maxim of quality can take precedence with little consequence because there is sufficient time for Speaker A to gain more clarity later in the conversation. But either way, *it depends* would be an acceptable response as it does not interfere with the continuation of the conversation.

Our second example conversation in (10) takes place in the same restaurant bar with the same speakers. Now, we have the additional information that Speaker B is 18. Again, the dialogue and the entailments of Speaker B's response, shown in (11) and (12), are listed below:

(10) (a) Are you an adult or a child?
(b) It depends.
(11) There are situations where I am an adult.
(12) There are situations where I am a child.

We consider Speaker B's response to be semantically acceptable because it abides to the maxims of quantity and quality. The maxim of quality is met because whether Speaker B is an adult or a child is dependent on which metric Speaker A wishes to use to define adult maturity. It could be that Speaker A is using 18, the legal adult age in America. It could also be that Speaker A is using the legal age of drinking which can be anywhere from 16 to 21 depending on the area. By this, saying *it depends* can accurately express Speaker B's state of being. The maxim of quantity is also met because Speaker B provides enough information given the question asked. Speaker B can be an adult or a child depending on why the Speaker is asking, so the question is sufficiently answered. Some may think a response with singular choice, either adult or child, would have been a better response, but expressing that the answer varies is suitable for the situation. Speaker B's response would prompt Speaker A to expand on the question and Speaker B could give a new response that equally satisfies quantity and quality. We established previously that this luxury of prioritising quality over quantity is ideal in these less urgent situations, but what about when the information is vital?

4 When 'It Depends' is Unacceptable

We identified two categories of situations during which *it depends* is an unacceptable response: (1) states of being and (2) urgent situations. Much like with the acceptable categories, we do not intend for this sampling to address every situation and examples will be given to illustrate the main points.

The ineffectiveness of *it depends* as a response will be described based whether it defies the maxim of quality or the maxim of quantity.

4.1 States of Being

This category also involves situations where *it depends* is in response to a question about the responder's state of being, so how they are in the moment. In contrast to the acceptable state of being category, in this category, the situations are not ones that are socially-spaced, and do not exist on a spectrum. All of the instigating questions in this category stem from the desire to know something about the responder's state of being in the moment. This could include things about an activity they are doing or where they physically reside in a moment. The issues with *it depends* in these situations primarily because it violates the quality maxim.

Our example conversation in (13) surrounds Speaker A and B who are on the phone. Speaker B was at a show and was complaining that they could not see the stage properly in the moment. Curious as to why Speaker B could not see properly, Speaker A posed a question to Speaker B regarding their state of being. The dialogue and the entailments of Speaker B's response, shown in (14) and (15), are listed below:

(13) (a) Are you sitting or standing?
(b) #It depends.
(14) There are situations where I am sitting now.
(15) There are situations where I am standing now.

This response is unacceptable as it violates the maxims of quality. It is not possible for Speaker B to be sitting in some situations and standing in some situations when they can only be doing one of the two. There are no sensible situations where *it depends* would be a true statement.

Consider the ineffectiveness of Speaker B's response in (13) in contrast to the conversation in (16) which would fall under the *Giving Nonurgent Preferences* acceptable category.

(16) (a) Do you want to sit or stand?
(b) It depends.

Speaker B's response in (16) does not violate the maxim of quality because it is possible for Speaker B to have the desire to stand and to have the desire to sit, whereas in (13), a fairly similar conversation, it is clearly a violation of quality.

4.2 Urgent Situations

This category involves situations where the question posed needs a clear answer because the conversation is space in a short period of time or because the answer is vital and needed to continue to a new action. We claim that *it depends* as a response in this category of situations is unsuitable primarily due to issues with the maxim of quantity. Note that this urgency exists in contrast to the nonurgent acceptable situations.

For our first example, we return to the conversation between a cashier and a customer from the introductions. The conversation is given in (1) and the entailments of Speaker B's response are given in (2) and (3), shown below:

(1) (a) Do you want to use either cash or credit?
(b) #It depends.
(2) There are situations where I want to use cash.
(3) There are situations where I want to use credit.

We consider this response to be unacceptable because Speaker B's response does not hold against the maxim of quantity. In this situation, Speaker A (the cashier) needs Speaker B's response to change the setting on their cashier monitor. Having an uncertain response like *it depends* is highly inconvenient for Speaker A as they need a clear response to continue. Given the fact that the length of the interaction between a cashier and a customer is meant to be short, there is a sense of urgency that makes *it depends* even more ineffective in this situation.

Our second example conversation in (17) is a conversation between a movie ticket booth worker and a theatre patron trying to buy a ticket for a PG movie. The booth worker, or Speaker A, asks the patron, or Speaker B, a question to help with the purchase. Note that Speaker B is 18. The dialogue and the entailments of Speaker B's response, in (18) and (19) are listed below:

(17) (a) Are you an adult or a child?
(b) #It depends.
(18) There are situations where I am an adult.
(19) There are situations where I am a child.

Speaker B's response in this example can be considered to be unacceptable because it, again, violates the maxim of quantity. For the purpose of the ticket sale, Speaker A only wants to know how Speaker B should be charged. By saying *it depends*, Speaker B is giving extraneous information about their maturity. This problem is heightened as this exchange is not meant to last for an extended period of time and the ticket booth worker has other work to tend to. For the restaurant bar conversation in (10), this same dialogue was acceptable because of ambiguity surrounding the socially-spaced spectrum of adulthood and because there was time for the conversation to extend. This is not the case in this conversation.

5 Conclusion and Improvements

After looking at the function of *it depends* in several contexts, we were successfully able to identify two situations where the indecisiveness surrounding *it depends* is acceptable and two situations where it is not acceptable. These identifications were primarily understood through Grice's maxims of quality and quantity. With this we satisfied our goal of contributing to an understanding of what *it depends* means and when indecisiveness harms conversational goals. Future research should elect to look further into the role of Grice's maxims of relation and manner as well. We selected to focus on quality and quantity because they elicited the most variation between our examples.

Another major improvement to this study would be to extend the investigation of uncertainty to statements beyond *it depends*. We are unable to make any broad sweeping claims about uncertainty or indecisiveness in these situations as we only looked at the impact of one statement. We hope this investigation will inspire further work in the area of *it depends* and uncertainty.

For Want Of: Evidence for the 'Deprivative' From Australian Languages

Grace Ephraums and Patrick Das
Monash University

Abstract. Within Australian languages, nominal modification is widespread (e. g. Dench & Evans, 1988; Dixon, 2002; Nordlinger, 1998; Simpson, in press). We contend crosslinguistic criteria for an underlooked nominal modifier, the DEPRIVATIVE, based on evidence from five Australian languages: Yankunytjatjara (Goddard, 1983), Gooniyandi (McGregor, 1990), Wanyjirra (Senge, 2015), Yuwaalaraay and Gamilaraay (Giacon, 2014). Despite being identified in at least these languages, the deprivative has received little comparative attention. Despite some recognition by Nose (2006), it receives little mention in existing studies of Australian negation (e. g. Dixon, 2002, p. 81; Phillips, in press). Hence, our study is the first to comparatively analyse the DEPRIVATIVE. From our initial analysis, we propose the following minimal criteria: (1) Syntactic: a) A nominal modifier (i.e., a morpheme that can modify nominals). (2) Semantic: a) Encodes LACK - absence of the nominal referent it modifies ([food-DEPRIV] encodes the absence of food) and; b) Encodes REASON - the LACK motivates some action ([food-DEPRIV] in a construction with the main verb [die] encodes death due to absence of food). Similar nominal modifiers encoding LACK have received study: the privative (Phillips, in press; Saulwick, 1996), caritive (Kozhanov, 2019; Oskolskaya et al., 2020; Rudnitskaya, 2020) and absessive (Hamari, 2011; Miestamo et al., 2015). However, the DEPRIVATIVE is unique in also encoding REASON. We seek to semantically and syntactically distinguish the deprivative from other established categories. By establishing a comparative definition, we aim to facilitate future study of this under-examined linguistic category, and inform the broader typological understanding of nominal negation.

Keywords: morphosyntax; typology; deprivative; negation; Australia

---

Abbreviations

Table i: *Abbreviations appearing in the data that are not included in the Leipzig Glossing Rules.*

AUG	**Augmented**	**(Giacon, 2014; Senge, 2015)**
DIM	Diminutive	(Giacon, 2014)
EV	Evitative	(Harvey, 2011)
FU	Future	(Harvey, 2011)
IGNOR	Ignorative	(Giacon, 2014)
IO	Indirect Object	(Harvey, 2011)
INCHO	Inchoative Verbaliser	(Goddard, 1983)
MOV	Continuous-Moving Suffix	(Giacon, 2014)
PD	Personal Declension	(Giacon, 2014)

PRES	Present Verb Suffix	(Goddard, 1983)
RDP	Reduplication	(Senge, 2015)
REAL	Realis	(Senge, 2015)

1 Introduction

The present study provides a preliminary description for an underlooked negation category, the DEPRIVATIVE, based on evidence from eight Australian languages. Despite being identified in several descriptions of individual languages, the deprivative has received little recognition in existing studies of negation strategies in Australian languages (e.g., Dixon, 2002, p. 81-85; Phillips, in press). We present an overview and preliminary analysis of a group of negation methods in Australian languages that we have called the "deprivative". We contend that these may be considered a category and aim to develop a robust comparative criterion for further typological study.

First, we describe the background of negation strategies as they have been described, with particular attention to Australia and a related negation market, the privative. Next, we will familiarise you with our research agenda and the methodology we used to conduct the study, surveying 20 Australian languages, sampled for variety. Lastly, we discuss our results and findings, that eight of those 20 languages display deprivative functions, and we consider its expression as a construction or as a monomorphemic nominal marker, before concluding and making recommendations for future study.

2 Background

The typological study of negation strategies across Australian languages has received reasonable but incomplete attention in Australianist literature. While many grammars of individual languages describe negation methods in detail, there is a small number of studies that consider the behaviour of negation across larger samples of Australian languages.

The earliest concerted effort is Dixon (2002, p. 81-85), who offers a survey of Australian negation, and posits a taxonomy of four points at which Australian languages "prototypically" encode negation in their grammars. In other words, Dixon claims that Australian languages generally have strategies for performing these types of negation. These are described by Dixon as follows: (1) a complete clause consisting of 'an interjection'; (2) a negator of non-imperative clauses ('not'); (c) a negator of imperative clauses ('don't'); (d) a derivational privative suffix.

The fourth of these, the privative marker, has subsequently received the most dedicated attention in Australianist literature (Ephraums, 2021; Phillips, in press) and is considered a defining feature of Australian languages. It is generally described in relation to its positive counterpart, the proprietive/comitative (Blake, 1987; Dixon, 1976; Saulwick, 1996). This pair of markers often features in the literature as the subject of an ongoing discussion surrounding their status as either derivational (e.g. Dench & Evans, 1988) or inflectional (e.g. Nordlinger, 1998) morphemes.

Most recently, Phillips (in press) offers a dedicated nonexhaustive overview of negation methods found in Australian languages, with a focus on their grammatical development. Phillips focuses on four broad negation strategies: (1) standard clausal negation of assertions; (2) negative imperatives/prohibitives; (3) nominal negation; (4) lexical negative verbal stems in several languages. Phillips recognises Australian negation as an underdeveloped area, and flags several phenomena in need of typological study.

3 Research Agenda

Having established the previous work in this area, this next section now outlines the goals of our present study. To do so, we first define the formal and functional criteria for our search, and then we describe our methodology for conducting the present research. Overall, our objective is to search for grammatical encodings of a particular cluster of meanings: that we call the deprivative, and motivate similarities and differences from available theory.

The deprivative for our purposes can be said to have two semantic components: LACK – indicating an absence of an entity, and WANT – indicating that the absence of the entity has caused an effect on the participant. In our data then, we aim to characterise the nature of LACK and WANT, further specify them, and interrogate possible grammaticalization pathways for them.

We collected our data by conducting a survey of 20 Australian languages (10 Pama-Nyungan and 10 Non-Pama-Nyungan), listed in the table below, a cross between a sample of convenience and a sample of variety. The sources we considered for the study were all reference grammars. Our sampling methodology rests on an adaptation of the Diversity Value method (Miestamo et. al, 2016). This means out of the pool of well-described languages (for our purposes, this meant languages with a reference grammar published after 1970), we only select one language per 'mother node', that is, the subgroup that it is most immediately classified under. Since languages that share a mother node have shared innovations, by maximising the number of mother nodes, we are in some way maximising the variety in our sample. However, our method doesn't weigh higher levels as more diverse as it should, as the classification of Australian languages can still be very contentious (Koch, 2014). At best, it provides a method of ensuring more variety than a typical sample of convenience with no regard for genetic bias. Our sample also suffers from an unavoidable bibliographic bias, due to the availability of sources and only using reference grammars at the present stage of study.

Table 1: *Survey*

Language	Mother Node	Classification	Source
Arrernte	Aranda	PN	Green 1994
Bilinarra	Ngumpin	PN	Meakins, Nordlinger, 2014
Panyjima	Ngayarta	PN	Dench, 1981
Djinang	Northern Yolngu	PN	Waters, 1989
Warlpiri	Yapa	PN	Simpson, 1983
Yuwaalaraay/Gamilaraay	Central-NSW	PN	Giacon, 2014
Muruwari	Muruwaric	PN	Oates, 1988
Diyari	Western Karnic	PN	Austin, 1978
Yir Yoront	Southwest Paman	PN	Alpher, 1991
Yankunytjatjara	Western Desert (Wati)	PN	Goddard, 1983

Mengerrdji/Erre/ Urningangk	Giimbiyu	NPN	Campbell 2006
Gooniyandi	Bunuban	NPN	McGregor, 1990
Gaagudju	Gaagudju	NPN	Harvey, 2011
Tiwi	Tiwi	NPN	Lee, 1987
Anindilyakwa	East Arnhem	NPN	van Egmond, 2012
Ngalakgan	Ngalakgan	NPN	Merlan, 1983
Bininj Kunwok	Gunwinygic	NPN	Evans, 2003
Wambaya	Eastern Mirndi	NPN	Nordlinger, 1993
Murrinhpatha	Southern Daly	NPN	Mansfield, 2019
Alawa	Alawa	NPN	Sharpe, 1972

4 Results

Our survey found 8 possible deprivative grammatical expressions in our sample. These were divided into two categories — deprivative constructions (discussed in Section 4.1) and deprivative morphemes (Section 4.2) depending on whether the expression was colexified on the noun phrase, or monomorphemic in nature. We also found the deprivative to have a variety of specialised applications, each of which we recommend further study on. First, we discuss the possibility of a 'deprivative construction' in two languages and suggestions for research in this direction. Next, we provide evidence of the monomorphemic deprivative expression in six languages, with extended descriptions of their specialised applications in three domains — kinship, body parts, and wh-constituents in interrogatives. Lastly, we make suggestions for future research. Our findings can be summarised in the table below:

Table 2: *Findings*

Language	Subfamily	Classification	Source	D-CON	D-MORPH
Yankunytjatjara	Western Desert (Wati)	PN	Goddard, 1983	-	+
Bilinarra	Ngumpin	PN	Meakins, Nordlinger, 2014	-	+
Warlpiri	Yapa	PN	Simpson, 1983	+	-
Wanyjirra	Ngumpin	PN	Senge, 2016	-	+
Yuwaalaraay/ Gamilaraay	Central-NSW	PN	Giacon, 2014	-	+
Mengerrdji/Erre/Urningangk	Giimbiyu	NPN	Campbell 2006	-	+
Gooniyandi	Bunuban	NPN	McGregor, 1990	-	+

Gaagudju	Gaagudju	NPN	Harvey, 2011	+	-
Abbreviations: D-CON: evidence of deprivative construction; D-MORPH: evidence of deprivative morpheme; PN: Pama-Ngyungan; NPN: Non-Pama-Nyungan; + represents positive evidence; - represents lack of evidence.					

5 Deprivative Construction

When surveying cross-linguistically, it is important to look for *functional* criteria, not formal ones (Haspelmath, 2007). In other words, our aim is to search for a mapping of form to function — and we cannot be limited in the forms we consider. Two of the languages in our sample (Gaagudju, ISO:gbu; Warlpiri, ISO:wbp) displayed the two functions — LACK and WANT — colexified on the entire noun phrase, as opposed to on a single morpheme. Due to a lack of available data, a cohesive analysis of the commonalities in the construction cannot be conducted in the present study, but below, we describe the expression of the deprivative construction in the two languages. Additionally, it does not seem that these constructions are conventionalized, which casts further doubt on its status as a true 'construction' that the users of the language employ, as opposed to the occasional collocation of meaning. Nevertheless, the examples shed light on the possible variety of grammatical expression of the deprivative, and we provide suggestions for further analysis of this construction.

5.1 Gaagudju

Gaagudju shows evidence of some deprivative constructions, present in its negative possessive constructions. These constructions are formed by using an incorporation construction of a negated noun to convey a negative possessive meaning (Harvey, 2011, p. 331) - which is typical of the privative meaning too (Ephraums, 2021). Below, example (1) illustrates a deprivative meaning from a negative possessive construction in Gaagudju.

(1) *gaayu ibärdbi arr-djee-gi Φ-m-burroo-ya ngaarndjil*
Neg Neg 1-go-EV 3I<l:FU-spear:Fish-EV
gaayu=nga=bilaarra
fish Neg=lIO=spear
'No, I am not going to go and spear fish. I do not have a spear.'
Harvey (2011, p. 331)

The morpheme typically used for expressing the negative possessive meaning is *gaayu*, and noun incorporation in Gaagudju can importantly only occur to nouns that are a) headed by an indirect object clitic, and b) part nominals (Harvey, 2011, p. 301). In the grammar's examples for negative incorporation constructions, there is clear LACK encoded via negative marking of the noun, but furthermore, most examples also display WANT. Consider example (2) below:

(2) *ibärdbi m-balaa-biri balanggid gaayu=nga*
Neg 1 :FU-cover-Aux blanket Neg=lIO
'I cannot cover up. I have no blanket.'

There are two analyses that we consider for understanding deprivative constructions. To understand both, we can distil LACK and WANT to two propositions: 1) That the participant (X) lacks the entity (Y); 2) That the lack of Y affects X negatively. The first analysis is that the deprivative represents a sort of modalized possession: one where there is the best of possible worlds where the participant has Y and is able to function normally because of it, and the world we have accessed currently due to lack of Y.

The second analysis is that the nominal with deprivative marking (Y) assigns a malefactive role to the participant X. Considering that benefactive/malefactive roles have been known to emerge from possessive constructions (Lichtenberk 2002, Margetts 2004), this is a likely grammaticalization pathway. Both analyses are tentative, though we feel more credibility lies with the second one — we recommend that future work can develop a formalisation of the semantic function of deprivative constructions, in Australia and/or across the world.

5.2 Warlpiri

In Warlpiri (Yapa, wbp), as opposed to the clausal negation of negative possessive construction such as in Gaagudju, deprivative meanings can be constructed via multiple nominal modification of the privative and source on the same nominal.

(3) *Jurru* *ka-lu* *maljarlawurlawu* *karri*
hair-ABS PRES-3pl sticks.out-ABS stand NPST
ngapa-wangu-nagka
water-PRIV.SOURCE
'Their hair gets stiff and sticks out from lack of water' (from not being washed).
Simpson (1983, p. 304)

Overall, while there is no clear evidence that these strategies for deprivative meanings are conventionalized in Gaagudju or Warlpiri, they provide interesting insights for understanding how syntactic constructions may be utilised for constructing deprivative meanings. First, we've analysed Gaagudju to show how the negative possession construction displays deprivative meanings, and highlighted Margetts (2004)'s work in how benefactives can arise from possessive construction, to suggest that a similar process may have occurred in Gaagudju. Secondly, in Warlpiri the co-occurrence of SOURCE and PRIVATIVE gives rise to deprivative meanings, and this suggests a possible grammaticalization pathway where multiple nominal modification of the sort described in Nordlinger (1998) could give rise to new nominal modifiers.

6 Deprivative Morpheme

The deprivative morpheme is here defined as a monomorphemic nominal marker. It is differentiated from the deprivative construction (discussed in the previous section) in which the deprivative function is colexified on the NP as a whole. The typical deprivative morpheme found in our sample is described in the remainder of this section. Three specialised functions of the deprivative morpheme are described following this.

Minimally the analysis suggests that the deprivative morpheme indicates that the lack of the referent of the modified nominal is in some way the cause of something else. It can be said to perform the two aforementioned semantic functions (LACK and WANT). Firstly, the negating function is identified to be similar to the typical Australian privative (Ephraums, 2021; Phillips, in press), as it denotes absence of the nominal referent (LACK). Secondly, it indicates that this absence has an effect of some kind, and is the motivation for some action (WANT). It remains to be seen whether this effect can be characterised as A) always on the participant, in essence acting to assign an EXPERIENCER or other semantic role on the participant, and whether this effect can be characterised as negative, as most examples in our data seem to reflect. For the deprivative morpheme, there is no other element in the phrase indicating either of these functions: they are colexified on the noun solely by the deprivative morpheme.

In (4) below, the marker *tjiratja* glossed by Goddard as 'DEPRIV' modifies the nominal *anku* 'sleep', indicating that the absence is the reason for the speaker being uneasy. A close English equivalent is 'for want of', as used in Goddard's translation.

(4) *Kurun-na* *kuya* *ngara-nyi,*
spirit(NOM)-1sg(NOM) bad(NOM) stand-PRES
anku-tjiratja
sleep-DEPRIV(NOM)
'I'm uneasy, for want of sleep'
Yankunytjatjara (kdd), Goddard, 1983 (p. 132)

Further examples of the deprivative morpheme are seen in the following three examples (5-7) from Wayjirra, Gooniyandi and Yuwaalaraay/Gamilaraay.

(5) *Jiya* *ngu=lu* *gun.ga* *garri-nya*
kangaroo.ABS REAL=3AUG.SBJ dead BE-PAST
ngawa-wunyja
water-LACK
'The kangaroos were dead, being thirsty.'
Wanyjirra (ddj), Senge (2016, p. 324)

(6) *Manka* *wangmarrawinti* ***kampawinyja***
ear he went mad **water-[DEPRIV]**
'He went mad from thirst'
Gooniyandi, McGregor (1990, p. 282)

In (7), the deprivative-marked nominal *dhamiyaa-nginda* falls into the scope of the negative marker *waal*, indicating that the speaker does not have want of the tomahawk.

(7) *Waal* *ngaya* ***dhamiyaa-nginda***
Not 1SG **tomahawk-WANT**
'I don't want the tomahawk' (i.e., the axe is not mine')
Yuwaalaraay/Gamilaraay, Giacon (2014, p. 82)

6.1 Kinship Application of Deprivative Morpheme

The deprivative morpheme in Yankunytjatjara, as identified by Goddard (1983:132), has a specialised semantic function only when it modifies a kinship term. Here it serves to indicate that "the actor behaved as he or she did on account of a lack of care or proper treatment of his or her relative". It does not indicate an absence of the referent of the nominal itself. This is shown by (8): the deprivative morpheme denotes absence of proper treatment of his son, not absence of the son himself. A further examples of this kinship application in Yankunytjatjara is (9) in which the cause of speaking "like this" is the lack of proper treatment of the son. This function of the deprivative was not found in any other languages in our sample.

(8) *Pika-ri-ngu* *paluru* ***katja-tjiratja***
angry-INCHO-PAST DEF(NOM) **son-DEPRIV(NOM)**

'He got angry, over lack of proper treatment of his son'
Yankunytjatjara, Goddard (1983, p. 132)

(9)	*Kuta-ngku*	*malany-tju*	*alatji*	*wangka-nyi*
	senior brother-ERG	junior brother-ERG	like this	talk-PRES
	Katja-tjiratja-ngku			
	son-DEPRIV-ERG			

‘(A man’s) older brother(s), and younger brother(s) speak like this, over lack of proper treatment of (his) son’ (i.e., would defend son from unfair criticism).
Yankunytjatjara, Goddard (1983, p. 132)

6.2 Body Part Application of Deprivative Morpheme

The deprivative morpheme identified in Wanyjirra has a specialised function when attached to body part nominals. It indicates not the absence of the body part itself, but absence of the normal function of the body part. In (10) the suffix *wuja* attached to the nominal *barndawurru* ‘back’ indicates that there is a lack of normal function of a back. Likewise *milba-wunyja* (11) denotes lack of sight but (presumably) not lack of eyes.

(10) ***milba-wunyja***
eye-LACK
‘blind’
Wanyjirra, Senge (2015, p. 158)

(11) ***barndawurru-wuja***
back-LACK
‘something wrong with [his] back’
Wanyjirra, Senge (2015, p. 158)

Note that the Wanyjirra nominal privative marker can also perform this function: [body part]-[PRIV] can indicate either absence of the body part OR absence of its function (Senge, 2015, p.158, 207-208). Senge therefore states that the semantic range of the privative is broader than that of the deprivative. These two possible functions are not uncommon of privatives/nominal negators across Australian languages (Ephraums, 2021).

6.3 Interrogative Application of Deprivative Morpheme

The deprivative morphemes identified in Wanyjirra and Yuwaalaraay/Gamilaraay sometimes appear in interrogative constructions. Here, the morpheme modifies a wh-constituent, as shown in (12-16). Based on the data available to us, the function could be broadly translated as asking ‘for want of what?’, or ‘what is the lack that causes this event?’. No evidence of this was found in any other languages in our sample containing deprivative morphemes.

(12)	***Nyamba-wuja***	*nyila*	*gula*	*gun.ga*	*wandi-nya*
	what-LACK	DIST1	NEG	dead	FALL-PAST

‘Why didn’t it die?’
Wanyjirra, Senge (2015, p. 247)

(13) ***Nyamba-wuja*** *burja-burja* *yan-i*
what-LACK RDP-run GO-PAST
'Why was he running?'
Wanyjirra, Senge (2015, p. 431)

(14) ***nyamba-wuja=wayi=lu***
what-LACK=IGNOR=3AUG.SBJ
CS#: 'I don't know what's wrong with them.'
Wanyjirra, Senge (2015, p. 494)

(15) I don't know what did the girl hit his mother for.
minya-nginda=waa *nhama* / *birralii-djuul-u, ngambaadhi*
what-WANT=IGNOR 3.DEF / child-DIM-ERG, mother.PD
nguungu *buma-y*
3SG.DAT hit-PST
'I don't know what she wanted, that kid, that she hit her mother.'
Yuwaalaraay/Gamilaraay, Giacon (2014, p. 83)

(16) Why (you) are those two men hitting each other
Minya-nginda *nhama* *dhayn* *buma-la-waa-nha*
what-WANT 3.DEF man hit-RECP-MOV-PRS
'What are those men fighting for?'
Yuwaalaraay/Gamilaraay, Giacon (2014, p. 83)

7 Conclusion

The deprivative category as we have identified it in our sample represents an under-researched form of negation. In our sample of 20 languages, it was found in 8 languages, with monomorphemic expressions far outnumbering that of in a construction. Our survey and discussion here represent a preliminary description of this category: further study is needed to establish more robust syntactic similarities, especially for the deprivative construction. Evidence of predictable, conventionalized syntactic slots would be key evidence towards that goal. The study also makes suggestions for possible grammatical pathways for the deprivative, arising from assignment of malefactive role to participants, similar to that in Oceanic languages, or perhaps from multiple nominal marking eventually converging. Further work can also be done towards understanding the deprivative as an expression of modalized possession, in which case it represents a particularly interesting case as an exhibit of nominal modality. Lastly, more insight into the specialised kinship, body part and interrogative applications of deprivative morphemes can further our understanding of how nominal modification interfaces with those particular domains. All of these recommendations can be best enacted with the availability of more language data, and description of endangered languages which are often disregarded in wider theory.

8 References

Blake, B. (1987). *Australian Aboriginal grammar*. London: Routledge.
Campbell, L. (2006). *A sketch grammar of Urningangk, Erre and Mengerrdji: The Giimbiyu languages of Western Arnhem Land* [Unpublished honours dissertation]. University of Melbourne.
Dench, A., & Evans, N. (1988). Multiple case marking in Australian languages. *Australian Journal of Linguistics*, 8(1), 1–47.

Dixon, R. M. W. (1976). Topic A: The derivational affix “having.” In R. M. W. Dixon (Ed.) *Grammatical categories in Australian languages*. Canberra: Australian Institute of Aboriginal Studies ; Humanities Press.

Ephraums, G. (2021). *“Lacking” in Australian languages: A survey of the nominal privative* [Unpublished dissertation]. Monash University.

Giacon, J. (2014). *A grammar of Yuwaalaraay and Gamilaraay* [Unpublished doctoral dissertation]. The Australian National University.

Goddard, C. (1983). *A semantically-oriented grammar of the Yankunytjatjara dialect of the Western Desert language* [Unpublished doctoral dissertation]. The Australian National University.

Harvey, M. (2011). *A grammar of Gaagudju*. New York: De Gruyter Mouton.

Haspelmath, M. 2007. Pre-established categories don’t exist: Consequences for language description and typology. *Linguistic Typology,* 11(1), 119–132.

Koch, H. (2014). 2. Historical relations among the Australian languages: genetic classification and contact-based diffusion. In *Historical relations among the Australian languages: genetic classification and contact-based diffusion* (pp. 23–90). New York: De Gruyter Mouton.

Lichtenberk, F. (2002). The Possessive-Benefactive Connection. *Oceanic Linguistics,* 41(2), 439–474.

Margetts, A. 2004. From implicature to construction: Emergence of a benefactive construction in Oceanic. *Oceanic Linguistics,* 43(2), 445–468.

McGregor, W. (1990). *A functional grammar of Gooniyandi* (Vol. 22). Amsterdam: John Benjamins Publishing Company.

Miestamo, M., Bakker, D. & Arppe, A. (2016). Sampling for variety. *Linguistic Typology*, 20(2). 233–296.

Meakins, F., & Nordlinger, R. (2013). *A grammar of Bilinarra: An Australian aboriginal language of the Northern Territory*. New York: Mouton de Gruyter.

Nordlinger, R. (1998). *Constructive case: Evidence from Australian languages*. CSLI Publications.

Saulwick, A. (1996). *To have and to hold: The semantics of the proprietive case in Australian languages* [Unpublished doctoral dissertation]. University of Melbourne.

Senge, C. (2015). *A Grammar of Wanyjirra, a language of Northern Australia* [Unpublished doctoral dissertation]. The Australian National University.

Simpson, J. H. (1983). *Aspects of Warlpiri Morphology and Syntax* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.

"Looks like we've got a problem": A Corpus-based Analysis of Discourse Markers

Suzy Park
The University of Edinburgh

Abstract. This paper focuses on the phrase 'it looks like' which is distinguished from inflected forms of the phrasal verb "look like" and aims to establish it as a multiword discourse marker (MDM), by analyzing its properties as a discourse marker and multiword expression (MWE). Based on the sequences extracted from 140 million token-sized social media texts, I observed and classified the syntactic and semantic characteristics of the sequences. Through this process, 3 significant characteristics of the sequence as a discourse marker were revealed: non-propositionality, multifunctionality, and initiality. Furthermore, its properties as a MWE, most notably its idiomaticity, will also be discussed. In this study, I discussed how this MDM, frequent in colloquial contexts, may be explicitly determined in order to be properly interpreted in natural language understanding tasks, and thus, to be registered in multiword language resources.

Keywords: corpus linguistics; discourse analysis; multiword discourse markers

---

1 Introduction

This paper aims to examine certain sequences that contain multiword discourse markers (MDM) such as 'it looks like', which has not received much attention in this field. The sequence 'it looks like' analyzed as a frozen multi-word expression in this study, is distinguished from the standard usages of the verb-particle construction 'to look like'.

The multiword sequence 'it looks like', at first glance, may be considered as an inflected form of a verb-particle construction 'to look like', which is defined as 'to have an appearance that is very similar to something' (Merriam-Webster). The following examples demonstrate such usage of this phrasal verb:

(1) They *look like* their father.
(2) It *looks like* a small, intricate computer.

The sentences below, in contrast, do not seem to carry the same meaning:

(3) It looks like we will be arriving soon.
(4) Looks like it is going to snow tomorrow.

Initially, (3) and (4) may appear to be ungrammatical: a clause is preceded by the preposition *like*, and (4) is even further complicated with subject omission. However, it can be commonly found in various online forums and social media, where informal speech patterns are dominant. Here are some examples observed from online blog posts, Twitter, and Reddit:

(5) (a) [Blog post] Looks like I'll be working tonight!
(b) [Blog post] It looks like I have no choice.
(c) [Twitter] It looks like lots of accounts are being wiped out.

(d) [Twitter] It looks like they've made a few changes from the teaser.
(e) [Reddit] …at first glance it looks like our living standard is increasing…
(f) [Reddit] It looks like 2021 was unpleasant for many people…

The sequences 'It looks like' and 'Looks like' appear as a multi-word expression (MWE) in the above examples. Their syntactic and semantic properties are distinguished from the ordinary usages of the verbal phrase 'to look like': in (5), they mostly appear in the front of the sentence, whereas the verbal construction 'to look like' may appear in any syntactically adequate positions; they do not follow inflectional changes such as 'looked like' or 'looking like', differently from the ordinary verbal phrases 'to look like'; moreover, as mentioned above, in (5), after the preposition 'like', a full sentence occurs, which is grammatically inacceptable; and finally, the sequences may simply be omitted or replaced by a discoursal adverbial such as 'reportedly' or 'allegedly', named 'Hearsay Markers' in Fraser (2009) in:

(6) A: Is the game still on?
B: Reportedly, the game was postponed because of rain
(7) I won't live in Boston. Allegedly, all the politicians are corrupt.

As this phenomenon is particularly frequent in colloquial texts, in this study, online user-generated texts such as tweets or blogs that contain people's informal utterances were examined. Based on the examination of the syntactic and semantic properties of the multi-word expression *it looks like*, I suggest that it should be distinguished from the standard verb-particle construction 'to look like' and should be classified as an MDM that shares numerous characteristics with other discourse markers. In this way, the sequence should be registered in MWE language resources so that natural language processing tasks can perform properly.

2 Previous Studies

In the definition sections of the entry 'look like' in current dictionaries, it is hard to find a description of the colloquial usages of *it looks like.* Table 1 summarizes the definitions of the Cambridge, Merriam-Webster, Collins, Oxford Learner's and Oxford English Dictionaries:

Table 1: *Definitions of current dictionaries.*

Dictionary	**Meaning of *look like***	**Colloquial usage of *It looks like***
Cambridge	to be similar in appearance to someone or something	N/A
Merriam-Webster	to have an appearance that is very similar to (someone or something): to resemble (someone or something)	N/A
Collins	N/A (only *look exactly like* can be found)	(In American English) i. It seems that there will be; (e.g., *It looks like rain*) ii. It seems as if
Oxford Learner's	To have a particular appearance	No formal definition, but states that *like* can replace *as if* or *as though* in informal American English
OED	To have the appearance of being, to seem to be; to resemble in appearance.	Chiefly followed by it: it appears to be the case; it seems likely.
	With gerund, verbal noun, or noun: to give promise or indication of, show a likelihood of.	Originally U.S. With following clause: it appears as though, seems likely that (something is the case).

As seen in Table 1, the only dictionary that provides an account (although far from sufficient) of the properties of 'it looks like' is the OED and Collins dictionary.

Previous works on discourse markers tended to focus on the description and classification of the different functions of a single-word marker (Ferrara, 1997; Jucker, 1993; Lee-Goldman, 2010). According to Fraser (1999), discourse markers do not constitute a separate syntactic category, and there are three sources of discourse markers: conjunction, adverb, and prepositional phrases. A few idioms like 'still and all' or 'all things considered' are also considered. Nonetheless, multiword discourse markers consisting of a verbal construction such as 'it looks like' were hardly mentioned in the literature.

Der (2010) described semantic-pragmatic features and formal-syntactic features of discourse markers, which may serve as certain criteria in order to define various types of discourse markers. According to Siepmann (2005), MWEs are collocations composed of two or more words acting as a single unit, and real MDMs usually have their own syntactic structure.

Considering the difficulties MWEs pose on their identification (Der, 2010), in this paper I attempt to clarify 'it looks like' as an MDM especially notable in informal speech by analysing its semantic and syntactic properties present in blog corpora.

3 Data and Methodology

This study used the Blog Authorship Corpus, which was created by Schler et al. (2006). It is a compilation of 681,288 blog posts from 19,320 bloggers at blogger.com. The blog posts were written in or before 2004, and there were approximately 35 posts per blogger. Token size was over 140 million. A blog corpus seemed like the data of best fit for two reasons: first, it is a form of social media where colloquial uses of language can be easily found. Second, there is diversity in terms of gender and age of speaker. Concordance analysis was conducted using Antconc, and theories on hedge devices were employed to classify the different functions of 'It looks like'.

4 'It looks like' as a discourse marker

4.1 Non-propositionality

Table 2: *Examples of "It looks like" as a discourse marker.*

It looks like	the team has flourished through the years.
It looks like	there are many options available to me.
It looks like	she might have to have surgery.
It looks like	it's about to rain.
Looks like	I'll have to find something else to do.

One of the most notable properties of 'it looks like' as a discourse marker is its non-propositionality, which leads to its optionality. As can be seen from the table above, separating "it looks like" from the rest of the sentence does not affect the truth condition of the original sentence. In addition, since "it looks like" is attached to a fully formed sentence, it becomes an optional component, which is another important property of a discourse marker.

4.2 Multifunctionality

4.2.1 Non-expletive it

(8) It looks like it's going to be wet.
(9) It looks like it was hit by a tornado.
(10) It looks like it is a story about a child of Noah and Allie.

Before discussing the roles of 'it looks like', it seems adequate to discern ambiguous cases where it is hard to ascertain whether or not it is used as a discourse marker. Although (8), (9), and (10) have identical forms, (8) is unambiguous as 'it looks like' only adds the speaker's uncertainty to the prediction 'it's going to be wet'. (9) and (10) are not so clear. In order to qualify as a discourse marker, the subject it must be expletive; that is, it should not refer to any objects. (9) and (10) could be interpreted as 'it looks like' acting as a discourse marker, but also a situation where the two 'its' are coindexed and like is used as a colloquial substitute for 'as if'.

4.2.2 Speculation

Hedge expressions were defined by Lakoff (1973) as fuzzy expressions, which cause the propositions to be 'more' or 'less' true. Representative examples of hedges include 'sort of', 'really', 'technically', and 'very' (Itani, 1995). Its role as blurring and expanding the boundaries of a statement has allowed hedges to signal the uncertainty of the speaker in an utterance. In this perspective, 'it looks like' constructions can be used as a hedge device that transforms definitive statements into mere speculations. Let us consider:

(11) It looks like it's finished.
(12) Looks like she's angry with me.

(11) and (12) illustrate this usage, as it would be natural to add, for instance, 'but I am not sure' at the end of the statement.

4.2.3 Mitigation

Hedges can also contribute to lessening the impact of negative sentiments and act as a means of mitigation. In the following examples:

(13) It looks like they did a crappy job.
(14) It looks like I will never talk to him again.

(13) and (14) are examples of 'it looks like' playing such a role, when associated with expressions that deliver a negative opinion. Identifying words or utterances that carry negative emotions would help recognize this effect of 'it looks like' constructions in NLP.

4.2.4 Politeness

Hedging expressions are often linked to politeness, and politeness can be achieved through distancing the speaker from the utterance (Rounds, 1982; Skelton, 1988). For instance:

(15) Looks like I figured it out.
(16) It looks like I'm road tripping to Laughlin this weekend!
(17) Looks like you need a little fun.

(15) denotes an instance where the speaker tries not to give the impression that they are 'boasting' by using 'looks like' as a hedge. (16) can be understood as a fairly objective statement, but the speaker adds distance to their own decision or judgment, which may be construed as an attempt to appear less assertive. (17) is an example of a suggestion, and the effect of politeness is added through the hedging discourse marker.

4.3 Initiality

Conjunctions or conjunctive adverbs, often named as connectives, are examples of discourse markers that can appear not only in utterance-medial, but even sentence-medial positions.

However, this does not indicate that a discourse marker can appear in any position freely; rather, for 'it looks like', it seems to appear in the initial position of a proposition. Therefore, it can be preceded by adverbs (e.g., 'so', 'now'), or appear after a conjunction (e.g., 'and', 'but'). This characteristic helps differentiate its use as a discourse marker with inflected forms of the phrasal verb 'look like', as they can appear in sentence-medial positions, especially in relative clauses (e.g., ‘This is not what it looks like.’)

5 ‘It looks like’ as an Informal MWE

5.1 Idiomaticity

5.1.1 Syntactic Idiomaticity

This relates to the properties of ‘it looks like’ as a discourse marker, namely non-propositionality and optionality. In the original constructions of the phrasal verb “look like,” the verb is followed by a noun phrase (NP), not a clause. However, the informal MWE “it looks like” precedes a complete sentence, and this derivation implies that it has different syntactic properties from its components. Thus, it is also possible to say that “it looks like” is syntactically idiomatic, and syntactically independent from its components and the clause it modifies.

5.1.2 Semantic Idiomaticity

At first glance, the expression seems to retain the meaning of the phrasal verb ‘look like’, as it can be paraphrased as ‘seem’ or ‘resemble’. At the same time, the MWE is not fully compositional, as the expression ‘it looks like’ acts as a single unit and can be replaced by a single-word adverb such as ‘apparently’ or ‘seemingly’. This semantic idiomaticity, or proverbiality supports the idea that it should be recognized as a distinct unit with specific meanings and functions.

5.1.3 Pragmatic Idiomaticity

As noted above, ‘it looks like’ has multiple functions as a discourse marker, including politeness and uncertainty. This multifunctionality is an example of pragmatic idiomaticity, where an expression is reserved to convey emotions or attitudes in a restricted context.

5.2 Fixed Expressions

Baldwin & Kim (2010) classified lexicalized MWEs into 3 categories, depending on how much syntactic variation is possible. Highly idiomatic expressions that act as a single unit are less flexible than collocations that happen to co-occur frequently, but do not have lexical, semantic, syntactic, or pragmatic idiomaticity.

Among the 3 categories, which are fixed expressions, semi-fixed expressions, and syntactically flexible expressions, ‘it looks like’ constructions should be labeled as a fixed expression: inflections or morphosyntactic variations are not allowed, and no other element (e.g., adverbs) can come between any of the components. This rigidity confirms the notion that this is a pragmatically and lexically idiomatic construction that should be studied as a unit of its own, not as an altered form of ‘to look like’.

Another peculiarity that is visible is the fact that it can be omitted. Subject omission is often found in informal speech (Haegeman, 1997; Weir, 2012), and MWEs often cause utterances to become more or less informal (Baldwin & Kim, 2010).

6 Further Studies

As cross-language variation is an important feature of MWEs, further comparison with other MWEs that display similar properties would help deepen the understanding of this phenomenon. Extracting the rules and patterns of part of speech (POS) can contribute to the field of NLP as well, and an example work in progress is shown below:

Table 3: *An Example of Part of Speech Classification of 'It looks like'.*

(It) looks like + subjective pronoun (SP)	Discourse Marker
(It) looks like + objective pronoun (OP)	Used as an inflected form of 'look like'
(It) looks like + it	May be used as a discourse marker, or a colloquial form of 'look as if'; disambiguation may require manual classification depending on context, or intricate syntactic rules
(It) looks like + NP (excluding it, SP, OP)	If ends with NP, inflected form of 'look like'; if NP is followed by VP, discourse marker
(It) looks like + indefinite pronoun (e.g., someone, everyone, nothing)	Mostly used as a discourse marker, with a VP following

Also, this study did not include analysis of phonetic/prosodic features because it did not deal with spoken corpora, but further research with speech corpora might uncover properties that support its role as a discourse marker. Lastly, since dictionaries state that this construction is commonly used in American English, sociolinguistic studies on how its distribution varies depending on different varieties of English could also be of interest.

7 Conclusion

In this paper, I argued that 'it looks like' should be considered as a multiword discourse marker, based on its properties both as a discourse marker and a multiword expression. Its function is best described as a hedge device, which often signals uncertainty or politeness. It also distances the speaker from the utterance. Although there are still issues that are unresolved-from solving ambiguity to figuring out the frequency and distribution of different uses of the discourse marker depending on the context-establishing 'it looks like' as a separate marker from its usual usage will be the starting point for its recognition in natural language processing.

8 References

Baldwin, T., & Kim, S. N. (2010). Multiword expressions. *Handbook of natural language processing*, 2, 267-292.

Constant, M. et al. (2017). Multiword Expression Processing: A Survey. *Association for Computational Linguistics*, 43(4), 837-892.

Dér, C. (2010). On the status of discourse markers. *Acta Linguistica Hungarica*, 57(1), 3-28.

Ferrara, K. (1997). Form and function of the discourse marker anyway: implications for discourse analysis. *Linguistics,* 35, 343-378.

Jucker, A. (1993). The discourse marker well: A relevance-theoretical account. *Journal of Pragmatics*, 19, 435-452.

Fraser, B. (2009). An account of discourse markers. *International review of Pragmatics*, 1(2), 293-320.

Haegeman, L. (1997). Register variation, truncation, and subject omission in English and in French1. *English Language & Linguistics*, 1(2), 233-270.

Itani, R. (1995). A relevance-based analysis of Lakoffian hedges: sort of, a typical and technically. *UCL Working Papers in Linguistics*, 7, 87-105.

Jucker, A. H. (1993). The discourse marker well: A relevance-theoretical account. *Journal of pragmatics*, 19(5), 435-452.

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts, *UC Berkeley Previously Published Works*, 458-508.

Lee-Goldman, R. (2011). No as a discourse marker. *Journal of Pragmatics*, 1-46.

Sag, I. et al. (2002). Multiword expressions: a pain in the neck for NLP. *Proceedings of Computational Linguistics and Intelligent Text Processing*, 1-16.

Rounds, P. (1982). Hedging in written academic discourse: Precision and flexibility. *University of Michigan (Mimeo)*.

Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. W. (2006, March). Effects of age and gender on blogging. In *AAAI spring symposium: Computational approaches to analyzing weblogs* (Vol. 6, pp. 199-205).

Skelton, J. (1988). The care and maintenance of hedges. *ELT Journal*, 42(1), 37-43.

Weir, A. (2012). Left-edge deletion in English and subject omission in diaries1. *English Language & Linguistics*, 16(1), 105-129.

Hidden Behind a Smile: Acquisition and Societal Perception of Speech Sounds in Children with Cleft Lip and Palate

Emily Shepherdson
University of York

Abstract. Cleft lip and palate (CLP) is a birth defect causing the upper lip, hard palate and/or the soft palate (the velum) not to fuse together correctly in the womb. There is very little research on the effects it has on a child's speech sound acquisition. Particularly, how speech and language therapy (SLT) supports a child in reaching a similar intelligibility as their non-CLP peers. It is suggested that a child with CLP that has undergone SLT intervention should have "good quality, intelligible (understandable) speech by age 5-6" (CLAPA, 2021). I explore the validity of this claim utilising pre-existing knowledge on phoneme acquisition, as well as gathering the general public's view on CLP through a perception study. I then compared this to a study by Gordon-Brannan & Hodson (2000) which suggested a percentage intelligibility which, if a child fell below this, indicated the presence of a speech development disorder. Overall, I discovered that the idea that a child will have developed good quality and intelligible speech by the age of 5-6 years after SLT intervention is not entirely accurate as the CLP children's average intelligibility still indicated a lag in speech development. I also discovered that much of the general public has a relatively poor understanding of CLP. These findings also suggest that there is little understanding of speech development disorders as a whole. If this is the case, adults may not be able to identify when a child requires SLT intervention in cases of less prominent speech issues.

Keywords: L1 acquisition; speech and language therapy

---

1 Introduction

Cleft lip and palate (CLP) is a genetic mutation which causes the upper lip, and the palates, to not fuse together correctly in the womb. CLP is the most common facial birth defect in the UK, affecting 1 in 700 babies (NHS, 2019). It does not only affect eating and drinking, but also speech sound production. In this perception study, I aim to detail how people perceive the quality of speech in children with CLP, as well as concluding the accuracy of the claims that speech and language therapy will offer a child with CLP 'good quality, intelligible (understandable) speech by age 5-6' (CLAPA, 2021). This research aims to provide a basis for future research into the intelligibility of plosives and fricatives produced by children with CLP.

1.1 The Effects on Speech Sound Acquisition

Cleft lip and palate (CLP) is, as explained previously, a genetic mutation causing a lack of fusing in the lip and palate in the womb. Despite surgeons aiming to have these spaces closed in surgery by around 12 months (CLAPA, 2021), speech sound acquisition has already begun (Kuhl, 2004). This means that there will likely be delays to acquisition of speech, as well as compensatory articulations produced subconsciously by the child to compensate for their inability to produce certain sounds.

The sounds predominantly affected are plosives (/p/, /d/, /k/) and fricatives (/f/, /θ/, /s/, /ʃ/) and by extension affricates (/ʧ/, /ʤ/). They are affected as they require a degree of closure in the oral cavity to be produced which is often lost through the air escaping through the palate and out from the nasal cavity. This denies the plosive of its required 'pop' on the release of air, and the fricative of the air being released orally through a narrow channel to produce the necessary vibrations. CLP is also noted to cause hypernasality in speech due to the air being allowed to exit through the nasal cavity, however this area will not be discussed to any length in the report as measuring nasality would require this data to have been collected in a lab using a Nasometer which was not possible for this project.

2 Methodology

2.1 Selecting Participants

As this is a perception study, it was vital that I create a list of key demographics whom I wanted to collect my data from. From here, it was concluded that in order to judge where children may have errors in their speech sound production, participants would need to be native or fluent speakers of English. The research is also audio based and so it was imperative that participants had a high quality of hearing with no known hearing loss.

2.2 Research Method

To collect this data, I used a survey which asks participants to listen to an audio clip of a child saying 1 or 2 words and answer a range of questions. These questions asked participants to decide whether the child has cleft lip and palate (CLP) or not, how they reached their conclusion, to transcribe the speech, and a scale from 1-5 stating how intelligible they believe the speech to be – 1 being completely unintelligible and 5 being adultlike. I selected the audios from a CLP corpus (Cleland et al., 2020) and a control group from a typically developing corpus (Eshky et al., 2018). As only 50% of children with CLP have access to speech and language therapies (NHS, 2016), it was also vital that I select a corpus in which the CLP children have undergone, or are currently undergoing, speech and language therapy (SLT) which is confirmed by the fact that the CLP corpus is collected throughout regular SLT sessions.

To select the audios, I chose the manners of articulation which are most affected by CLP - plosive, fricative and affricate (Nagarajan et al., 2009) - and found words in the CLP corpus which include these sounds for each place of articulation in the English Phoneme Inventory. I have selected one sound from either voiced or voiceless to represent each sound due to the constraints of the corpora. Then, I matched a child with CLP to a typically developing child by age and gender to ensure speech has been allowed to develop in a similar environment over the same period of time. This allowed the intelligibility of speech sounds in children with CLP and typically developing children to be directly compared to one another with accuracy.

3 Research Findings

In this section, I will be exploring the findings of my research through the analysis of the responses to my survey. This section will address the intelligibility of the sounds produced by children with cleft lip and palate (CLP) in comparison to the intelligibility of their typically developing (TD) peers, as well as looking into the overall quality of speech from both groups. I also aim to conduct targeted analysis of the individual speech sounds to investigate possible issue points and possible instances of compensatory articulations. After this initial exploration, I also aim to determine whether the general public view children with CLP as beginning to assimilate to the speech of their TD peers through their perception of what determines the presence of a phonological development disorder such as CLP.

3.1 Intelligibility of Speech Sounds

To begin my exploration into the intelligibility of this speech, I first looked further into literature discussing the level of intelligibility that a typically developing child should have at various ages. Flipsen (2006) conducted a study into child intelligibility and found that 4-year-olds have an average intelligibility of 96.82%. Another study conducted by Gordon-Brannan & Hodson (2000), also researching child intelligibility proposed that children with an intelligibility rating of 66% or less by the age of four likely have a phonological development disorder requiring intervention. When reporting on the intelligibility of child speech in this report, I will be using these percentages as a basis to judge the quality of speech. All speakers in this research are between the ages of five and nine, therefore these ratings should provide a good point of comparison for where the child should be in their intelligibility. During this analysis, it is vital to remember that the quality of speech is likely to vary between each child so these percentages are simply a point of reference rather than a fixed rating.

To find a percentage of intelligibility from my data, I based instances of 'correct transcriptions' on the inclusion of the highlighted phoneme in one or more appropriate positions in the word. I have also made allowances for minor errors in the understanding of each phoneme – for example, 'happy' was allowed where the correct transcription was 'puppy'. I also allowed homophones as correct transcriptions. From here, I was able to find my percentage intelligibility through the sum of:

$$100\ x\ \left(\frac{Number\ of\ participants\ able\ to\ correctly\ identify\ the\ phoneme}{Number\ of\ overall\ participants\ (40)}\right)$$

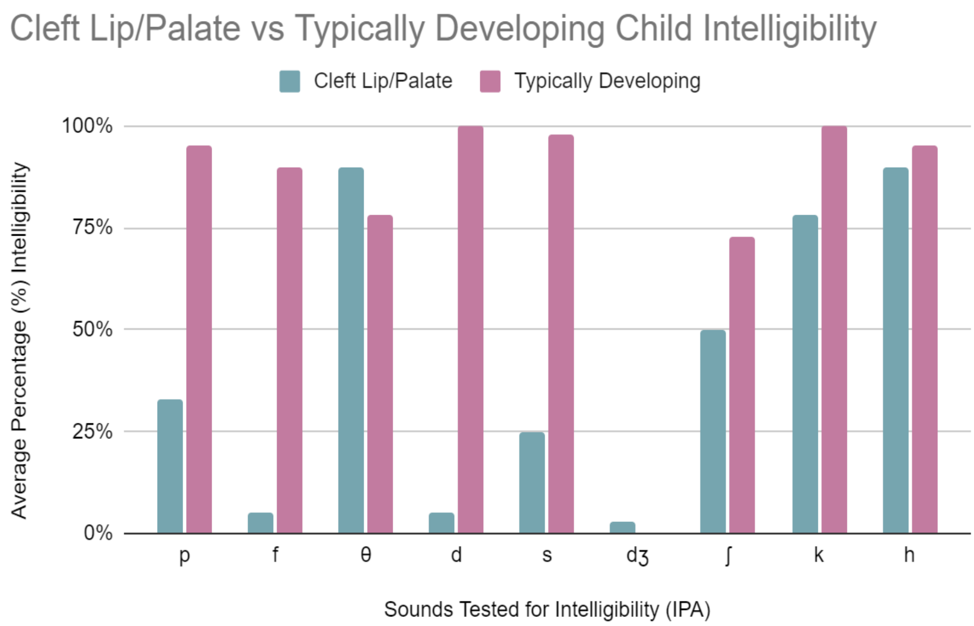


Figure 1: *Average Percentage of Phoneme Intelligibility in CLP vs TD.*

To look first at the case of the typically developing children, we can see that there is a significantly high proportion of these children able to be identified as producing these sounds correctly by unfamiliar adults. Although there seems to be this high level of correct identification, it is interesting to note that only 1/3 of phonemes were produced to a standard that allowed the children to reach to or go beyond the proposed 96.82% intelligibility that children of this age have previously been seen to achieve. These phonemes are /d/ (100%), /s/ (98%), and /k/ (100%) - although, /p/ and /h/ came close to reaching this bracket of intelligibility as they both achieved a 95% intelligibility rating. In fact, these TD children were only able to achieve an average intelligibility of 81% in their production of plosives and fricatives. Perhaps this is due to the children being Scottish and therefore producing phonemes in a way that was unfamiliar to a largely English participation group.

This percentage is also heavily influenced by the 0% intelligibility rating of the voiced postalveolar affricate /dʒ/ in which 90% of participants transcribed 'bridge' as 'bread'. However, this is possibly not an issue of the intelligibility of the affricate, but instead due to an assumption of the transcription through the use of vowels. The Scottish child in this audio produces the word 'bridge' as /bɹɛd͡ʒ/ whereas an English participant may anticipate the vowel in this word to be the near-close front unrounded vowel /ɪ/. Therefore, they may have unintentionally opted to transcribe a word which is familiar in English pronunciation with an almost identical set of phonemes which, in this case, is /bɹɛd/ (bread). This perhaps invites an exploration into the perception and production of vowels in a future study but does not offer an accurate intelligibility rating for /dʒ/ due to multiple outside sociophonetic factors. If this was removed from the data as an anomaly, the average intelligibility would be 91% which is still lower than expected but much closer to 96.82%.

To look now at the intelligibility of the sounds produced by children with cleft lip and palate, it is noticeable that very few of the children producing speech sounds have been able to reach or even come close to an intelligibility of 66%. We again only see 1/3 of the phonemes reaching the 66% that is seen as the upper boundary for identification of intervention being required for a possible phonological developmental disorder. The sounds that reached and exceeded this rating are /θ/ (90%), /k/ (78%), and /h/ (90%). In fact, the average intelligibility of the plosives, fricatives and affricates produced by children with CLP is just 42%. This is considerably lower than the 66% boundary that suggests issues with phonological development and suggests a lack of assimilation to the level of speech of their TD peers. It also suggests that the impact that CLP has had is still very much noticeable in their speech and this contradicts claims that speech and language therapy should allow speech to be intelligible by the age of 5-6 years (CLAPA, 2021).

First to address the voiceless glottal fricative /h/, it is interesting to see such a high level of intelligibility, but it is not unexpected. This is a sound that is produced in the glottis with no other constrictions of the tongue or lips meaning that its production is not particularly affiliated with any issues that CLP may cause. It is, however, produced orally which has previously been explored as an issue of CLP as the gaps in the palate can cause air to leave through the nasal cavity, causing hypernasality in some articulations. As previously addressed, these are not the appropriate circumstances in which to properly explore nasality of sounds but some of these findings could be used as a basis for further study.

The 78% intelligibility of /k/ is also a noteworthy feature as it is not only a plosive meaning that air being released nasally hinders the 'pop' of air required, and therefore the production of this sound, it is also a velar plosive - the velum, where this sound is produced, is also known as the soft palate. As CLP often affects mainly the soft palate or both the hard and soft palates (CLAPA, 2021), I had anticipated a considerably lower intelligibility rating. However, as I will explore further later in this paper (pg.11-12) there is evidence suggesting that dorsal sounds are produced to a higher quality with fewer compensatory articulations. Along with this, the high level of intelligibility also insinuates a high standard of speech and language therapy in allowing this child to have clarity in their speech sound articulation.

Again, to refer to outside sociophonetic influences on responses, I believe that the incredibly high 90% intelligibility of the voiceless dental fricative /θ/ is likely an anomaly. This is because the dental sounds can be heavily impacted by CLP as there can be an impact on the development of the teeth (Haque & Alam, 2015). Here, I believe that a majority of participants have understood the word 'teeth' as /ti:f/ as 7.5% of participants stated they had, but through context allowed themselves to orthographically transcribe the word simply as 'teeth' – creating the assumption that they must have understood the word as /ti:θ/. Though this is speculation, I believe that context is evidence enough to assume that this rating may not be the most accurate representation of a child's intelligibility of the /θ/ phoneme.

As well as the higher ratings of intelligibility, it is also important to mention the lowest ratings of intelligibility. These come from /f/ (5%), /d/ (5%), /s/ (25%) and /dʒ/ (3%). In this section, I will not be addressing /dʒ/ as the lower rating is likely due to similar circumstances as previously discussed for the sound in the typically developing dataset. To look first at the voiced alveolar plosive /d/ and the voiceless alveolar fricative /s/, the intelligibility is expected to be lower due to the impact of CLP on the alveolar. At the ages of 5 and 8 years respectively, these children will likely not have ungone often necessary alveolar bone grafting surgery which aims to build up the alveolar bone that may not have been developed in the womb due to the cleft (Guo et al., 2011). Though the lower intelligibility is expected, 25% intelligibility is an interesting feature as it is fairly high for the circumstances.

The voiceless labiodental fricative is another of the sounds that was anticipated to be largely affected by CLP as it is both a fricative and also involves labial articulation. An interesting feature is the fact that unlike many of the other sounds where participants have identified articulatory compensations in phonemes such as /h/, this sound is most frequently completely omitted from transcriptions. 42.5% of participants identified the word 'phone' as some variation of the word 'own' which shows the child has either made little attempt to produce this sound and suggests an inability to do so. I would agree with this verdict as I would transcribe the audio as /əʊn/.

3.2 Quality of Speech

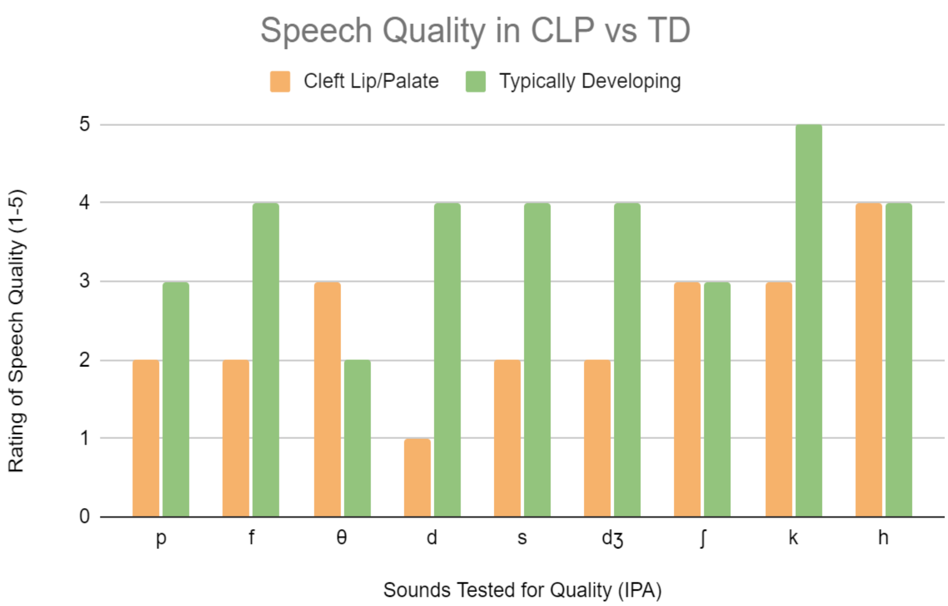


Figure 2: *Average Quality of Speech in CLP vs TD Children from a Rating of 1-5.*

As previously stated, one of the defining aims of the speech and language therapy that is provided to children with CLP is to ensure children have a good quality of speech by 5-6 years in order to be understood when entering into school (CLAPA, 2021). To test this, I have asked participants to give each audio a rating of 1-5. 1 being completely incomprehensible and 5 being almost adultlike. Of course, it must be considered that the children are only speaking single words and so this may not be an indication of their speech quality on a whole, but it does allow participants to consider how individual phonemes contribute to the overall quality of speech. From this, we can determine the quality of the child's ability to produce individual phonemes to affect speech quality. The above graph shows an average rating for the quality of speech for each of the phonemes.

One noticeable point is the considerably higher quality of speech in typically developing children than in their CLP peers with an average quality of speech at 4 whereas the children with CLP only reach an average of 2. However, it is perhaps unreasonable to make a direct comparison between the two as TD children will of course have a higher quality of speech due to there being little to no delay in their speech sound acquisition. As such, this data set should simply remain a demonstration of what is possible for children of a certain age to achieve, rather than a point to make criticisms towards delayed speech sound acquisition.

To return to the point of the average rating of the quality of speech for CLP being just 2, this suggests that the quality of speech is, overall, relatively low as it indicates that the speech is, for the most part, unintelligible. This dispute claims that speech by these ages should be to a good standard after SLT intervention.

To begin a discussion of the individual qualities of speech, I would first like to address the phoneme with the lowest quality of speech – /d/ - which was the only of the CLP phonemes to be scored an average of just 1. Personally, I would transcribe the speech of 'daddy mended' in the audio as /gæ.gi mɛn.gɛg/ as the voiced velar plosive/g/ is very prominent in the articulation. The participants often agreed with this verdict with 32.5% producing a transcription containing the /g/ phoneme. This offers an interesting insight into how speech quality can be greatly affected by the compensatory articulations produced by children with CLP. In this case, we see difficulty in producing an alveolar sound - which as previously mentioned is impacted by CLP and is often not improved until surgery at a later age (Coots, 2012) - being compensated for with a velar sound while maintaining the voicing and manner.

Though in this instance, it can be somewhat explained as a necessary compensation due to this being prior to alveolar bone grafting surgery, this is a common and yet often unexplained phenomenon in people with repaired cleft and alveolar ridge. Gibbon and Crampin (2001) explored the "retracted or 'backed' tongue placement" in CLP patients with repaired palates who were seemingly still attempting to "compensate for previous structural deficiencies" (K Govathi & Hari, 2017), particularly in producing alveolar stops. In this data, we begin to see the startings of such a pattern with the children often having a higher quality of speech in their production of the dorsal sounds. This is because there are fewer misarticulations that require compensations due to the predisposition of the child retracting the tongue as a remnant from producing so many of their other sounds this way out of necessity. The quality of speech ratings for dorsal sounds are also often higher as some compensations for the coronal and dorsal sounds can make the child seem as though they are speaking gibberish, which is of course not intelligible or good quality speech to an unfamiliar adult.

Other examples of the dorsal sounds being of a higher quality are that the voiceless palato-alveolar fricative /ʃ/ and the voiceless velar fricative /k/ are both rating a 3. While still relatively low overall as we must remember that CLP is still present, the quality exceeds the average rating.

3.3 The Public's Perception of CLP Speech

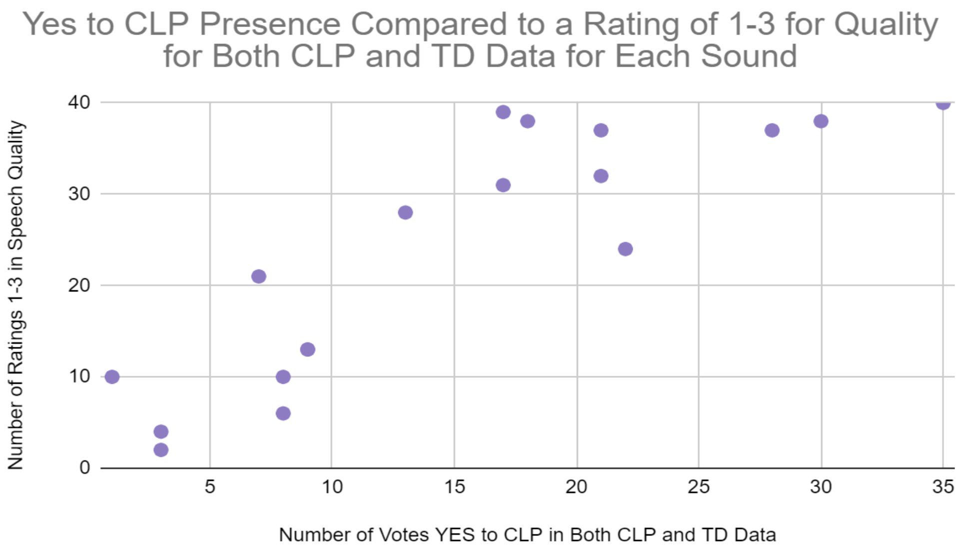


Figure 3: *Votes for YES to Speaker Having CLP Compared to Number of 1-3 Quality Ratings.*

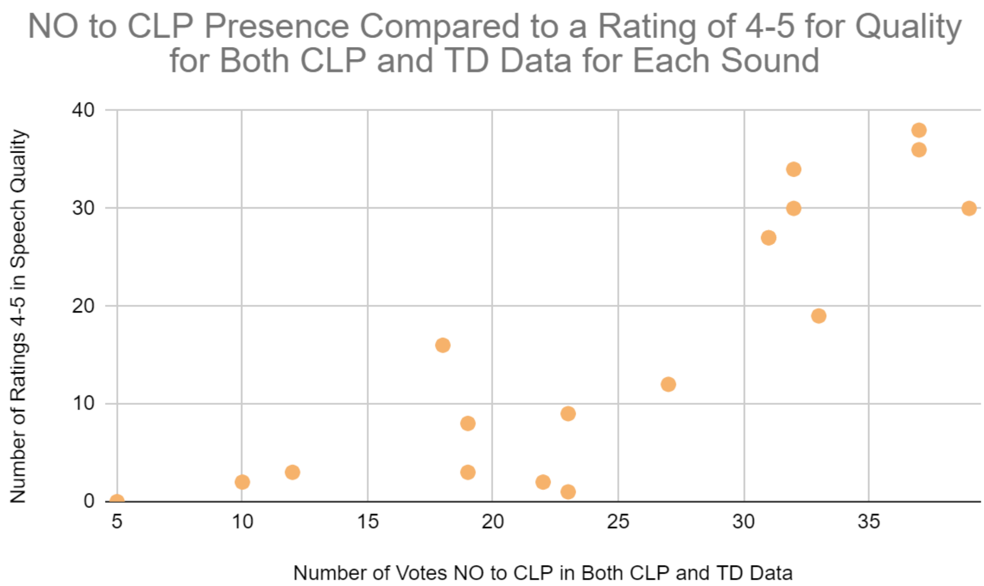


Figure 4: *Votes for NO to Speaker Having CLP Compared to Number of 4-5 Quality Ratings.*

Now to move away from how the children produce the sounds, I have also chosen to look further into how cleft lip and palate as a whole is perceived by the general public. First, I must note that in order to analyse these ratings, I have chosen to categorise a speech quality rating of 1-3 as the lower quality band, and 4-5 as the higher quality band. To represent this data, I have produced scatter graphs displaying the correlation first between the number of people who voted "Yes" when asked if they believed the child in the audio has CLP compared to those who gave a rating of 1-3 (the lower band of speech quality rating), and then another representing those who voted "No" when asked the same question compared to those who gave a rating of 4-5 (the higher band of speech quality rating). This is to determine how people perceive CLP, as well as offering an interpretation of how well speech and language therapy has allowed the children with CLP to assimilate with their TD peers. This data has been included for each of the sounds, both from the CLP and TD corpora. Though the typically developing children of course do not have CLP and can therefore not be directly associated with the speech development issue that come with CLP, it is still interesting to assess if there is a general consensus of the public assuming that a lower quality of speech equals the presence of CLP.

Though the correlation in both graphs is relatively weak, it is very apparent that there is a positive correlation between saying that a child does have cleft lip and palate and rating the speech as a lower quality, the same can also be said for the opposite. Although it is possible that the children who do have CLP generally tend to have a lower quality of speech, it is interesting that this correlation continues with the typically developing children who were incorrectly assumed to have CLP presumably somewhat based on the rating of a lower quality of speech. This suggests that there is a general assumption among the public that a lower quality of speech directly corresponds to the presence of CLP.

The clear consensus that CLP is associated with a poorer, less intelligible quality of speech suggests that there are perhaps issue points to be addressed in claims of speech being intelligible and of a good quality by the ages of the children involved in this study.

Although we can suggest that there are some points to explore with the quality and intelligibility of speech being less than expected by the ages of these participants, we must also look into how CLP is perceived by society as a whole. As part of this study, participants were also asked to explain how they had reached their conclusions with regard to the child having CLP or not. I feel that some of the responses raise issue points behind the education sounding CLP, especially given it is the most common facial birth defect in the UK. Conclusions included multiple mentions of the participant saying that the child has a cleft lip and palate because there seems to be the presence of a lisp. Though CLP certainly can cause similar problems in producing alveolar fricatives /s/ and /z/, they are certainly not the same articulatory issue. There were also multiple mentions of a “Yes” vote for CLP being based on a child’s issues seemingly with intonation. Again, CLP can certainly have an impact on this but they are not synonymous speech issues. This suggests that generally, the public appear to have a poor understanding both of what a typically developing child may sound like at each age, and generally of phonological development disorders as a whole. This could present a wider issue of guardians of children not being able to properly identify deficiencies in the quality of speech and possibly unintentionally allow a child to go without SLT intervention. This invites a further exploration into education around this subject.

4 Conclusion

Through this perception study, it is apparent that there are perhaps issues that are not allowing children to reach their full potential of speech. These data suggest that children with CLP overall have not been able to achieve the assumed position of having good quality and relatively high intelligibility of speech by the ages of 5-6 and beyond. This highlights areas to focus on in the continued support of a child’s developing speech and possibility invites a reassessment of the constraints on improving a child’s speech.

To note, this work is not intended to be a criticism of speech and language therapists who do vital work in improving the speech of children, and in turn improving their confidence and wellbeing. This research does, however, perhaps offer insight into how we, as a society, view the speech of people with atypical speech patterns. It also provides a basis for further research in speech acquisition and perception, and contributes to the knowledge which will inevitably aid in children having access to their fundamental right of communication.

5 References

CLAPA. (2021). Speech: How is it Treated? Retrieved: 21/07/2021, from *CLAPA*: <https://www.clapa.com/what-is-cleft-lip-palate/how-is-it-treated/>.

CLAPA. (2021). Surgery: How is it Treated? Retrieved: 22/07/2021, from *CLAPA*: <https://www.clapa.com/what-is-cleft-lip-palate/how-is-it-treated/>.

CLAPA. (2021). What is Cleft Lip & Palate? Cleft Palate. Retrieved: 22/07/2022, from *CLAPA*: <https://www.clapa.com/what-is-cleft-lip-palate/>.

Cleland, J., Lloyd, S., Campbell, L., Crampin, L., Palo, J. P., Sugden, E., Wrench, A., & Zharkova, N. (2020). The impact of real-time articulatory information on phonetic transcription: ultrasound-aided transcription in cleft lip and palate speech. *Folia Phoniatrica et Logopaedica*, 72, 120-130.

Coots, B. K. (2012). Alveolar Bone Grafting: Past, Present, and New Horizons. *Seminars in Plastic Surgery*, 26(4), 178-83.

Eshky, A., Ribeiro, M. S., Cleland, J., Richmond, K., Roxburgh, Z., Scobbie, J., & Wrench, A. (2018). Ultrasuite: A repository of ultrasound and acoustic data from child speech therapy sessions. *Proceedings of the 19th Annual Conference of the International Speech Communication Association*, *India,* 1888-1892.

Flipsen, P. (2006). Measuring the intelligibility of conversational speech in children. *Clinical Linguistics & Phonetics*, 20(4), 303-12.

Gibbon, F. E., & Crampin, L. (2001). An electropalatographic investigation of middorsum palatal stops in an adult with repaired cleft palate. *Cleft Palate Craniofacial Journal*, 38, 96-105.

Gordon-Brannan, M., & Hodson, B. (2000). Intelligibility/Severity Measurements of Prekindergarten Children's Speech. *American Journal of Speech-Language Pathology*, 9, 141-150.

Guo, J., Li, C., Zhang, Q., Wu, G., Deacon, S. A., Chen, J., Hu, H., Zou, S., & Ye, Q. (2011). Secondary bone grafting for alveolar cleft in children with cleft lip or cleft lip and palate. *Cochrane Database of Systematic Reviews*, 6, No.: CD008050.

Haque, S., & Alam, M. K. (2015). Common Dental Anomalies in Cleft Lip and Palate Patients. *Malaysian Journal of Medical Sciences*, 22(2), 55-60.

K Govathi, N., & Hari, P. (2017). A Study on Patterns of Compensatory Articulation Errors with Reference to Age of Surgery in Children with Repaired Cleft Lip and Palate. *Global Journal of Otolaryngology*, 7(2), 19.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–43.

Nagarajan, R., Savitha, V. H., & Subramaniyan, B. (2009). Communication disorders in individuals with cleft lip and palate: An overview. Retrieved: 16/08/2021, from *US National Library of Medicine National Institutes of Health Search database*: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2825064/>.

NHS. (2016). *Cleft Speech & Language Study*. Retrieved 22/07/2022, from *Bristol Speech & Language Therapy Research Unit*: <https://www.nbt.nhs.uk/bristol-speech-language-therapy-research-unit/bsltru-research/cleft-speech-language-study/>.

NHS. (2019). Cleft lip and palate. Retrieved 22/07/2021, from *NHS*: <https://www.nhs.uk/conditions/cleft-lip-and-palate/>.

Acknowledgements

I would like to thank my supervisors – Dr Tamar Keren-Portnoy and Dr George Bailey – for offering their expert knowledge of Child Language Acquisition and Phonetics respectively, and without whom this project would not have been possible. I would also like to thank the Laidlaw Foundation for offering me the opportunity to conduct this research!

contents