Manuscript in press at Language Documentation & Conservation, April 2020

Determinants of phonetic word duration in ten language documentation corpora: Word frequency, complexity, position, and part of speech

Jan Strunk
University of Cologne, Cologne, Germany
https://orcid.org/0000-0001-8546-1778

Frank Seifart (corresponding author)
Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany
Dynamique Du Langage (CNRS & Université de Lyon), Lyon, France
University of Cologne, Cologne, Germany
seifart@leibniz-zas.de
https://orcid.org/0000-0001-9909-2088

Swintha Danielsen
CIHA, Santa Cruz de la Sierra, Bolivia

Iren Hartmann
Leipzig University, Leipzig, Germany

Brigitte Pakendorf
Dynamique Du Langage (CNRS & Université de Lyon), Lyon, France

Søren Wichmann
Leiden University, Leiden, The Netherlands
Kazan Federal University, Kazan, Russia
Beijing Language University, Bejing, People's Republic of China

Alena Witzlack-Makarevich
Hebrew University of Jerusalem, Jerusalem, Israel
https://orcid.org/0000-0003-0138-4635

Balthasar Bickel
University of Zurich, Zurich, Switzerland

Abstract
This paper explores the application of quantitative methods to study the effect of various factors on phonetic word duration in ten languages. Data on most of these languages were collected in fieldwork aiming at documenting spontaneous speech in mostly endangered languages, to be used for multiple purposes, including the preservation of cultural heritage and community work. Here

we show the feasibility of studying processes of online acceleration and deceleration of speech across languages using such data, which have not been considered for this purpose before. Our results show that it is possible to detect a consistent effect of higher frequency of words leading to faster articulation even in the relatively small language documentation corpora used here. We also show that nouns tend to be pronounced more slowly than verbs when other factors are controlled for. Comparison of the effects of these and other factors shows that some of them are difficult to capture with the current data and methods, including potential effects of cross-linguistic differences in morphological complexity. In general, this paper argues for widening the cross-linguistic scope of phonetic and psycholinguistic research by including the wealth of language documentation data that has recently become available.

1. Introduction[1]

Speakers of all languages modulate their speech rate by pronouncing some words faster and others more slowly. Such variation can be studied by measuring the phonetic duration of words in seconds while controlling for a word's length (measured as, e.g., the number of its phonological segments). Well-known determinants of faster vs. slower pronunciation are a word's frequency, with high frequency leading to faster pronunciation, and its position in the utterance, with words in final position being pronounced more slowly. The relative strength and interaction of such factors are of theoretical interest for two reasons: First, they inform models of speech production (e.g., Bell et al. 2009; Jaeger & Buz 2017), and secondly, they shed light on processes of historical language change, since fast pronunciation over time leads to contracted and eventually to phonologically short word forms (e.g., Ernestus 2014; Sóskuthy & Hay 2017). However, most research on phonetic word duration is based on data from an exceedingly small number of well-studied languages, and systematic comparisons between different languages are even rarer. This is problematic because we know that languages vary enormously on all levels (Evans & Levinson 2009), and it is therefore unclear to what extent findings from individual languages generalize to others (Norcliffe, Harris & Jaeger 2015).

In this paper, we advance the cross-linguistic comparison of determinants of word duration by investigating the duration of content words (nouns and verbs) across a set of spoken language corpora from ten typologically, areally, and culturally diverse languages. We refer to this type of data here as "language documentation corpora" in the sense that they were collected through linguistic fieldwork in the respective speech communities with the aim of achieving a multimedia documentation of spontaneously spoken language that can be used for multiple purposes (Himmelmann 1998). Seven of the corpora used here were collected in the context of recent language documentation projects, aiming at a comprehensive documentation of language use in the respective communities (Baure, Bora, Chintang, Even, Hoocąk, N‖ng, and Texistepec) and one in the context of a large-scale study on contact-induced language change (Sakha). In order to better connect to the existing body of research on word duration in major languages, we add to this set parts of well-known and well-studied spoken corpora of Dutch and English (Figure 1), which can

[1] FS and JS wrote the paper, with input and additions from all authors; JS carried out the statistical analyses; all authors collected and annotated data (see Table 1 and Section 2.1 for details). The research of FS and JS was supported by a grant from the Volkswagen Foundation's Dokumentation Bedrohter Sprachen (DoBeS) program (89 550). FS and BP are grateful to the LABEX ASLAN (ANR-10-LABX-0081) of Université de Lyon for its financial support within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). SW's research was supported by JPICH/NWO and a subsidy of the Russian Government to support the Programme of Competitive Development of Kazan Federal University

likewise be regarded as language documentation corpora in the broad sense of the term applied here.

The set of corpora investigated here are different from those used in previous studies on word duration (e.g., Yuan, Liberman & Cieri 2006; Bell et al. 2009; Ernestus 2014; Sóskuthy & Hay 2017) primarily in two respects: First, their contents are not closely controlled for, with some containing only traditional narratives (Texistepec), some only personal narratives (Sakha), some only conversation (English and Dutch), and the others various combinations of these. Second, they differ from those used in previous studies in terms of size: They range from about 18,000 to about 56,000 words – a typical size for corpora collected in fieldwork on underdescribed and underresourced languages. In contrast, previous studies relied on corpora of hundreds of thousands, sometimes several millions or even billions of words, especially for counting word frequencies and calculating related measures as determinants of word duration.

Durational characteristics of speech are determined by a complex set of factors, some of them language-specific (Fletcher 2010). In addition to the duration of words, as studied here, previous research has also focused on phone and syllable durations and their mutual dependencies. Measurements of syllable and phone durations have been carried out in corpus phonetic studies on a range of languages with the aim of characterizing languages as a whole as belonging to distinct rhythmic types (Ramus, Nespor & Mehler 2000). Measures for this purpose include the standard deviation of the duration of vocalic (ΔV) and consonantal (ΔC) intervals, the percentage to which speech is vocalic (%V) (Ramus & Mehler 1999) and the pairwise variability index (PVI) (Grabe & Low 2002), a measure of the degree to which adjacent syllables or vocalic/consonantal intervals contrast in duration. This research revealed that languages systematically differ with respect to durational contrasts between syllables, as well as between consonantal and vocalic segments of the syllables. Results of such measurements have also been used to investigate whether languages tend to approximate rhythmic isochrony for syllable durations or inter-stress interval durations (Pike 1945; Abercrombie 1967), but this classification remains contested (Bertinetto & Bertini 2008; Arvaniti 2012; Nolan & Jeon 2014). A different approach to characterizing speech rhythm is to extract rhythmic patterns directly from the speech signal by techniques like the decomposition of the amplitude envelope (Tilsen & Arvaniti 2013; Gibbon & Li 2019). Dependencies between phone, syllable, and word durations have been found primarily in terms of shortening due to an increased number of phones in a syllable, or due to an increased number of syllables in a word (Trouvain 2004: 36), a phenomenon known as "polysyllabic shortening" (Lehiste 1972). While the current study does not address durational characteristics of units below the word level (phones or syllables), some of its results, especially the different temporal behavior of nouns and verbs, could help to fill gaps in these paradigms.

In the current paper, we pursue three descriptive goals aiming at new theoretical insights. First, we ask whether a consistent effect of higher word frequency leading to faster pronunciation of nouns and verbs can be observed in typical language-documentation data. This is by no means evident, as the previous literature stresses the need for big data to reliably detect such effects (e.g., Brysbaert & New 2009: 980; Liberman 2019: 15.2). Second, we hypothesize that word class will have an effect on word duration, specifically that nouns will be articulated more slowly than verbs, following up on earlier results on speech rate in time windows preceding nouns and verbs (Seifart et al. 2018). Third, we provide a comparison of the direction and magnitude of the effect of six factors potentially influencing word duration in each of our ten languages: In addition to (i) word frequency and (ii) word class, these are four factors that are included in our analyses as control

factors, namely (iii) word length (in phones), (iv) morphological complexity, (v) word position in an utterance, and (vi) the local speech rate in the utterance surrounding the target word.

Furthermore, this study aims to advance methodologies of studying determinants of word duration in language documentation corpora. Therefore, considerable space is devoted to discussing methodological choices and results of statistical analyses, also regarding our control factors, in particular morphological complexity and word position, in order to identify directions for future research on these factors in language documentation data. This is also one reason why, in the current study, we do not use alternative data beyond our core set of language documentation corpora, even where they would be available, e.g. for English, Dutch, and, to some extent, Sakha and Chintang. For the same reason, we focus our analyses on annotations that documentary linguists typically supply when creating multi-purpose linguistic resources (orthographic transcriptions, translations, time alignment of utterances, morphological segmentation and annotation, and part-of-speech tags) and disregard factors that would require extensive additional manual annotation, such as accentual lengthening. The relevant computational tools (R scripts) used in this study as well as data preparation and analysis logs are accessible at https://github.com/janstrunk/DeterminantsofWordDuration.

<Figure 1>

This paper proceeds as follows: In section 2 we introduce the languages and the corpora investigated here and describe the data preprocessing procedures we applied. Section 3 describes the methods used in the statistical analyses and how the factors that potentially influence word duration were defined, including discussion of issues arising from the particularities of our data when compared to previous studies. Section 4 presents and discusses our results, and section 5 concludes this study.

2. Data

2.1. Corpus characteristics

The sample of languages studied here is a convenience sample, which, however, represents broad genealogical and areal diversity. Eight of the corpora were compiled during fieldwork carried out by the authors in small speech communities, which have at most very recent writing traditions, and who speak minority, and often endangered, languages. As already mentioned, some of these corpora were compiled as part of larger language documentation collections, which also include documentation of cultural practices, ethnobotanical knowledge, etc. Others were compiled as text collections in the context of other research projects. The data selected for the current study consist mostly of monological texts, most typically traditional narratives, "indigenous texts" (Haig, Schnell & Wegener 2011), and personal narratives (a genre sometimes also called "autobiographic narrative" or "life stories").

The Baure corpus was collected in the village of Baures, in Bolivian Amazonia, first as part of a PhD project by Swintha Danielsen (2003–2007) and later in the context of the DoBeS Baure documentation project (financed by the Volkswagen foundation from 2008 until 2013), together with the linguists Femmy Admiraal and Lena Terhart and the anthropologist Franziska Riedel (see http://dobes.mpi.nl/projects/baure/project/). Baure is a Southern Arawakan language with only 10 speakers alive today, with some additional semi-speakers, who are currently very active in the revitalization of the language. Most of the longer texts in the corpus are traditional narratives,

while some are conversations. Transcription and further analysis of the data was carried out by the linguists of the project. The corpus is in ELAN format (https://tla.mpi.nl/tools/tla-tools/elan/) and offers transcriptions and translations into Spanish and English and glosses created with the Toolbox software (https://software.sil.org/toolbox/). The data used in the current study have been corrected and further annotated by Swintha Danielsen, who is a fluent speaker of the language and still continues working with the speakers, compiling data and publishing Baure materials.

The Bora corpus was collected between 2004 and 2008 by Frank Seifart in Bora communities in Northwestern Peru in the context of a documentation project that covered a total of four neighboring languages (http://dobes.mpi.nl/projects/center/). Bora is a Boran language still spoken by a few hundred adults in small communities in the Amazonian regions of Peru and Colombia, who practice small-scale horticulture and retain some traditional practices. Bora is still acquired by some children, although the community as a whole is shifting towards local Spanish. The corpus used here includes traditional narratives as well as personal narratives, such as retellings of events when speakers fell in the jungle (originally elicited to study spatial orientation of gestures). The data were transcribed and translated into Spanish mostly by native speaker assistants and morphologically analyzed using the Toolbox software by other assistants under the supervision of Frank Seifart. The original data are archived at The Language Archive (https://hdl.handle.net/1839/1ae788f0-6777-462c-bb20-b2fdb77f6491).

The Chintang corpus (http://clrp.uzh.ch) was collected between 2004 and 2009 by an interdisciplinary team led by Balthasar Bickel. Chintang is a Sino-Tibetan language from the Kiranti group spoken in the southern foothills of the Nepalese Himalayas. The data mostly consists of conversational data across all ages (Stoll et al. 2015) but also covers many other genres, including retellings of a stimulus film widely used in linguistics, the Pear Story (Chafe 1980). The recordings were transcribed by native speakers and glossed and analyzed by student assistants from the community and universities in Nepal and Germany. For present purposes, we analyzed a small subcorpus of adult speech in two variants: one including retellings of the Pear Story and one excluding these (Section 3.3.2).

The Even corpus represents the Lamunkhin dialect and was collected in the village Sebjan-Küöl in central Yakutia between 2008 and 2010. Data collection was undertaken first for a project on the role of language contact in Even dialect diversification that was initiated by Brigitte Pakendorf in the now-obsolete Max Planck Research Group on Comparative Population Linguistics and later continued in the framework of a DoBeS project (http://dobes.mpi.nl/projects/even/). Even is a dialectally diverse Northern Tungusic language spoken in numerous small communities of erstwhile nomadic hunters and reindeer breeders scattered over a vast area of northeastern Siberia. Although most dialects are by now moribund, the Lamunkhin dialect is still viable, with some children acquiring it as their home language; overall, there are probably 250-300 speakers of this dialect. However, it is under intense contact pressure from the dominant indigenous language of the region, the Turkic language Sakha (Yakut). The recordings were for the most part undertaken by Brigitte Pakendorf, with some being provided by Natalia Aralova, and they were transcribed by native speakers, with transfer to Toolbox or later to ELAN by the linguists. The morpheme analysis was done mostly by Brigitte Pakendorf, with some texts glossed by Natalia Aralova and/or student assistants in the DoBeS project, and additional annotation work by Evgeniya Zhivotova. The data included in this study largely comprise autobiographical narratives and anecdotes, with a few fairy tales, descriptions of traditions, and four Pear Stories included as well. While most recordings are monologues, with only the linguist(s) present, some were more interactive.

The Hoocąk corpus was collected between 2004 and 2013 in Wisconsin, USA. Hoocąk is a highly endangered Siouan language still spoken by less than 100 speakers over the age of 70. All data was transcribed with the help of native speakers and later analyzed manually in Toolbox and ELAN by Iren Hartmann and a few student assistants under her supervision. The corpus consists mostly of personal narratives, but also includes a few instructive and interactive texts.

The N‖ng corpus was collected between 2007 and 2011 in the Northern Cape province of South Africa. N‖ng belongs to the !Ui branch of the Tuu family. Once it was spoken in a wide area in South Africa's Gordonia district. As of 2020, N‖ng is a moribund language spoken by three elderly speakers. The collected corpus contains recordings from eight speakers. The data were collected primarily for the language documentation project "A text documentation of N|uu" (funded by the Endangered Language Documentation Programme (ELDP)) by Tom Güldemann, Martina Ernszt, Sven Siegmund, and Alena Witzlack-Makarevich. The corpus contains personal and traditional narratives, discussions of day-to-day issues, as well as procedural texts. For the present paper Alena Witzlack-Makarevich selected a subset of the data and extended the project's annotations.

The Sakha corpus was collected by Brigitte Pakendorf in 2002 and 2003 in the framework of a project aiming at elucidating the extent of Evenki contact influence on Sakha. Recordings focused on elderly speakers with little use of standardized Sakha and little or no knowledge of Russian living in mainly rural settlements in four different districts of the autonomous Republic Sakha (Yakutia) in Northeast Siberia (for details cf. Pakendorf 2007a: 61–64). Sakha is a divergent Turkic language spoken by some 450,000 people who still practise horse and cattle pastoralism supplemented by hunting and fishing. The language is relatively healthy, especially in rural areas, although a tendency of shift towards Russian can be detected when comparing the census data of 2002 with those of 2010. During most of the recordings at least one further native speaker of Sakha was present. The data were transcribed by native speakers and at a later stage transferred to Toolbox and glossed by Brigitte Pakendorf. The transcriptions were time-aligned with the audio files in ELAN for the purposes of this project by Evgeniya Zhivotova.

The Texistepec corpus is a set of folktales collected by Søren Wichmann in 1993 from a single speaker, Tomás López Florentino, of the language variously known as Texistepec Popoluca, Texistepequeño or, in the autodenomination, Wää Oot ('good word'). Lexical and grammatical aspects of the language were studied with another speaker, Carmen Román Telésforo, who also sometimes helped with the interpretation of the texts of Tomás López. The fieldwork was initially undertaken as part of a survey of Mixe-Zoquean languages and was not intended to be extensive, but Søren Wichmann took the opportunity of undertaking some salvage work when he met Tomás López and Carmen Román and realized their great potential as collaborators. The work, carried out over five months during 1993-1998, mainly resulted in a published text collection (Wichmann 1996), dictionary (Wichmann 2002), and a contribution to a series of filled linguistic questionnaires (Wichmann 2007). In the meantime, the language has even fewer speakers. The text collection includes the texts analyzed in the present paper, and is also publicly available as sound files in the Archive of the Indigenous Languages of the Americas (www.ailla.utexas.org).

The corpus of English we use is a subset of the Switchboard corpus of two-sided telephone conversations among English speakers from various parts of the US (Godfrey & Holiman 1993). This corpus was created in the early 1990s and has since been further annotated by many subsequent research projects and used in hundreds of studies including many on word durations. The data were collected by establishing telephone calls between volunteers who did not know each other and prompting them to discuss one out of a range of topics, such as gardening or college

education. For the purpose of the current study, we selected a set of 47 sessions, to arrive at a corpus size for English that is roughly comparable to that of the other languages. This subset includes discussions on a broad variety of 31 topics, one of which was discussed five times (public education), one three times (recycling) and the rest once or twice (e.g., auto repair, music, vacation, or sports).

The Dutch corpus we use consists of a subsection of the Corpus Gesproken Nederlands (CGN) (CGN-consortium, Language and Speech Nijmegen & ELIS Gent 2003). From this collection, we selected a corpus that is comparable in size, composition and annotation to those of the other languages. Specifically, we selected 17 sessions that document speech from the Netherlands (i.e. excluding Flanders), on the one hand, and that corresponds to spontaneous conversations (rather than read speech or similar). Furthermore, we only selected sessions that include reliable phonetic transcription and word alignments as well as syntactic annotation. Four of these sessions document face-to-face interviews of teachers and 13 sessions document spontaneous face-to-face dialogues.

<Table 1>

2.2. Phones, syllables, morphs, and words in our data

All data were transcribed, translated, morphologically analyzed (except for Dutch), and annotated with part-of-speech tags by language experts. The transcriptions apply practical orthographies, most of which are in use in the language communities and which are all close to a phonological transcription, with the exception of English and to some extent Dutch. (The Sakha and Even data are represented in Latinate orthographies that are transliterated from the Cyrillic script in use in the communities.) We use the number of orthographic characters as a proxy for the number of phones – realizations of phonological segments – as a basic unit in our analyses, as explained in detail in section 3.4.1. Our data are not annotated for syllables, and therefore we do not use syllables as a unit in our analyses. We indirectly control for polysyllabic shortening by including word length (in number of phones) as a factor in our analyses (see section 3.4.1).

The segmentation of our data into morphs – realizations of morphemes – is based on segmental morphemes throughout, i.e. it disregards suprasegmental marking, such as tonal morphemes, and cumulative exponence, i.e. a situation where different categories, e.g. number and case, are expressed by a single, indivisible marker. The languages in our sample differ widely in their degree of morphological complexity (i.e., in the number of morphs per word), ranging from an almost isolating language like N‖ng with on average 1.14 morphs per word to morphologically complex languages like Even or Bora with on average 1.91 and 2.21 morphs per word, respectively (see Table 3 for details). Section 3.3.3 discusses how such differences in morphological complexity affect frequential properties of words in the context of the current study. In the orthographic conventions used in our data, word units approximate prosodic words by typically representing clitics as affixes and writing compounds as one word. While there may be differences across corpora in the treatment of clitics and compounds as separate words, the same word segmentation was used for deriving relevant measures within each language, in particular regarding word durations, word length, and morphological complexity. For examples, see the complete lists of words extracted from the corpora and used in our study in Appendix A.

2.3. Time alignment

To study phonetic word durations, we require exact identification of word start and end times in the audio signals. The Dutch and English data we used were already time-aligned with audio

signals, providing us with such measurements. The eight documentation corpora were segmented into "annotation units" during transcription, which were time-aligned with audio, mostly for practical purposes, e.g., to display transcription and translations in the manner of subtitles. Such "annotation units" roughly correspond to utterances (and will be referred to as such in the remainder of this paper), but their size and the definition of their boundaries vary between the languages used here. While for some languages (like Texistepec), annotation units usually comprise only one clause, for others (e.g. Bora), such units are usually longer and might be better characterized as paragraphs (see Table 1 for average annotation unit lengths per language). Within these manually time-aligned annotation units, data were time-aligned at the phoneme level using the (Web)MAUS software (Kisler, Reichel & Schiel 2017), as described in Strunk et al. (2014; for similar approaches, see Sóskuthy & Hay 2017; Babinski et al. 2019). The word start and end times derived from WebMAUS' output were then manually checked and noticeable errors were corrected by listening through the sessions once or twice.

2.4. Data selection for current study

The initial data set contains a total of 318,431 words (Table 1). To optimally study the various factors that potentially affect word duration, we reduced this data set to a smaller set in a number of steps (Table 2). We first excluded all disfluent words, such as filled pauses like *uh* or *uhm*, false starts, and unclear words, i.e. words that could not be identified during transcription. We then excluded all words belonging to word classes other than the major content word classes nouns and verbs, and – among these – ambiguous words containing both a nominal and a verbal root.

<Table 2>

Only retaining nouns and verbs means that we excluded all function words, such as adpositions, auxiliaries, or pronouns. There are various reasons to exclude these from a study on word durations. First, previous research has shown that the durations of content words are affected differently than those of function words by some factors, most importantly, word frequency: While the duration of content words was significantly affected by word frequency in a study by Bell et al. (2009: 104), this was not the case for function words. Excluding function words is furthermore in line with other recent studies on word frequency and predictability in English, such as Seyfarth (2014: 143) and Sóskuthy & Hay (2017: 301), which also exclusively focus on content words. Another reason for retaining only nouns and verbs is that we can only compare word classes that can reliably be identified cross-linguistically. This is arguably the case for nouns and verbs (Haspelmath 2001; Kemmerer 2014; cf. also Seifart et al. 2018), while other word classes, including function words as well as, e.g., adverbs and adjectives are much more difficult to define cross-linguistically. Finally, note that when comparing a set of typologically very diverse languages, we cannot expect them to use comparable numbers and kinds of functional elements, especially since there are large differences in morphology between the languages in our sample (see Table 3, below). Accordingly, Himmelmann et al. (2018) report large differences between languages in the average number of words per intonation unit due to these differences in the use and nonuse of morphosyntactically independent function words. For instance, adverbial meanings might be expressed by adverbs in some languages, but by verbal inflection in others. The data set resulting from these reduction steps consists of 132,691 noun and verb tokens that were used to create frequency lists for the ten languages (further discussed in section 3.3.2).

For our study of word durations, we limit the analyses to the 100 most frequent word types (both nouns and verbs) in each language (details of this procedure will be discussed in section

3.3.2). For the purpose of identifying the 100 most frequent words (see Appendix A), word types are ranked according to their absolute frequencies in decreasing order. In order to resolve ties, we used document frequency as a second sorting level, that is, words that occurred in a higher number of different texts in a language/corpus were ranked higher than words of the same frequency that occurred in fewer different texts. Remaining ties were then arbitrarily resolved by the order of appearance of a word type in the corpus.

Limiting the analyses to the 100 most frequent word types has five advantages: First, it allows us to focus the statistical analysis on word types with (relatively) reliable frequency estimates. Second, the size of the data sets is comparable across languages. Third, it keeps the amount of analyzed data manageable, so that data can be manually checked (e.g. for peculiarities in word lists, see below). Fourth, this procedure also reduces the problem of violating the assumption of statistical independence between cases in regression modeling (cf. Bell et al. 2009: 99) because it makes it less likely to pick multiple words from one and the same utterance, whose durations might influence each other or at least be due to the same overall speech rate within this one utterance. Finally, cutting off the long tails of the frequency distributions allows fitting models with standard normality assumptions. Following the above procedure, on average, 0.74 word tokens were sampled from each utterance in our corpus, which contain an average number of 6.12 words per utterance, so that cases where multiple words were picked from one and the same utterance are relatively rare. From these 100-word lists, we further manually removed potential candidates for auxiliary-like elements (that is, functional verbs) that had been annotated as verbs in the corpora (e.g. copular or modal verbs) using blacklists. In addition, we also manually removed proper names. If a word was manually removed, then the 101st most frequent word was included to complete the 100-word list and so on. Appendix A provides the resulting final 100-word lists for each language, including information on word type, gloss, word class, different measures of frequency, as well as the mean word duration and word articulation rate for each word type. Appendix B provides tables of the word types that were manually excluded for each language, such as auxiliary-like elements or proper names. Manual inspection of the 100-word list for Chintang also led us to create a variant, reduced version of the corpus of this language, called "Chintang (no pear stories)", as explained in section 3.3.1.

The final data sets for each language that were used in our analysis (last column of Table 2) basically comprise all tokens of the 100 most frequent word types in that language, with the additional constraint, however, that all one-word utterances were excluded because we expect irregular behavior with respect to utterance-final lengthening and because we can avoid arbitrarily defining our independent variable *word position* as either zero or one for all one-word utterances (cf. section 3.4.3.).

3. Methods

3.1. Statistical modeling

Prior to modelling, i.e. carrying out multivariate statistical analyses, we plotted our data and inspected it for simple correlations between word duration and word frequency (results are reported in section 4.1). We then built linear mixed-effects models to study the effects of word frequency and word class, and to compare the effects of a total of six factors (each explained in more detail in sections 3.3-3.4), while controlling for the remaining factors, respectively, and also taking random variation between speakers and texts into account. These models were built using the statistical software R, in particular the `lme4` library (Bates et al. 2015; R Core Team 2018; cf.

also Sóskuthy & Hay 2017). Since our main goal is a comparison of word frequency effects and other effects across different languages, we need to make sure that the results from all languages are comparable. We therefore carried out parallel statistical analyses of our ten individual corpora, keeping the set of explanatory variables and the model structure as constant as possible.[2] We also refrained from model selection procedures within each language because these make the results more difficult to compare across languages (cf. also Sóskuthy & Hay 2017: 307). We also decided not to make models more complex by including possible interactions between independent variables. The structure of the statistical model applied to each of our ten languages is given in (1) using the formula notation of R (where the tilde separates the dependent variable to the left from the independent variables to the right, a plus sign separates the individual independent variables, and terms having the structure "(1|…)" are random effects). The dependent variable as well as the fixed and random factors (and the fact that we log-transformed some of them following common practice in the literature) will be explained in more detail in the following sections.

(1) *log*(*word duration*) ~ *log*(*relative frequency*) + *word class* + *word length* + *number of morphs* + *position* + *log*(*local speech rate*) + (1|*speaker*) + (1|*text*)

We use three kinds of statistical results to assess the effect of a given factor in a given language, and to compare these across languages:

1) We establish whether a given factor has a statistically significant effect on word durations in a given language or not using likelihood-ratio tests comparing a statistical model including this factor with one omitting it. The results are given in the form of a χ^2 test including the resulting *p*-value in Tables 4-5.
2) The strength of individual factors can also be compared among each other by looking at the standardized β coefficients. These express the effect of an explanatory variable on the dependent variable in terms of standard deviations (a measure that abstracts away from different units, such as the number of morphs or phones per second used to measure speech rate), that is, they answer the question by how many standard deviations the dependent variable is increased or decreased by increasing the explanatory variable in question by one standard deviation. This standardization enables the identification of the most important factors

[2] An alternative to carrying out ten separate analyses would have been to build one large model that includes language as a fixed or random factor and possible interactions between the factor language and the other explanatory variables. However, there are two reasons for using separate but parallel models, i.e. for the approach adopted here: First, the strength and direction of the effects of the various factors in such a large model involving complex interactions would be much harder to interpret. Second, the exact meaning of the explanatory variables also varies slightly from language to language. Take as an example the explanatory variable word length: Due to differences in orthographic systems, a word length of five phones may represent slightly different average word lengths in actual phonological segments in different languages if one language's orthography uses more digraphs or trigraphs than that of another language. Likewise, since our corpora for the ten individual languages differ somewhat in size, word frequencies in the different corpora are again not completely comparable. Such measures could be made more comparable, for example, by centering and scaling them (cf. also Seyfarth 2014; Sóskuthy & Hay 2017) within individual languages. However, we decided to deal with this by relying on ten parallel models and include variables without centering and scaling to allow for easier interpretation of the effects of the explanatory variables measured on their original scales (for example, in terms of the number of phones or morphs). Note, however, that we address comparability across languages by additionally providing standardized coefficients in the model summaries below and in Appendix D. These allow for a comparison between the strengths of the effects of the different variables in one model and the same variable across models.

influencing word duration within one language and also allows for a comparison of effect sizes and effect directions between languages.

3) Finally, we also estimate the strength of the effect of individual factors by calculating the contribution of an individual factor to the overall observed variation in word duration that our statistical models can explain (so-called R^2). This is done by estimating the difference (or more precisely, percentage change) in the variance explained (technically known as $\Delta R^2\%$) between a model which includes the factor in question and a model where it has been left out (following Aylett & Turk 2004: 44; Bell et al. 2009; Sóskuthy & Hay 2017: 308). We rely on proposals by Nakagawa & Schielzeth (2013) and by Johnson (2014), as implemented in the R library `MuMIn` (Bartoń 2018) in order to calculate the contribution of an explanatory variable to the variance explained by the fixed effects (marginal $\Delta R^2_{(m)}\%$) and to the variance explained by both fixed and random effects (conditional $\Delta R^2_{(c)}\%$) in our mixed-effects models. In addition, we also calculate how much of the explained variance can be uniquely attributed to the factor in question (the so-called semi-partial R^2). Our calculation here is again based on Nakagawa & Schielzeth (2013) and Johnson (2014), using an implementation from the R library `r2glmm` (Jaeger 2017). The most informative of these three measures for our purposes is marginal $\Delta R^2_{(m)}\%$ so that we only provide this measure in the summary table 5 below, while we list all three measures in the more detailed table 4 on the effect of word frequency and in Appendix D providing details on the individual mixed effects models.

3.2. Dependent variable: word duration

Word duration (measured in seconds) is the dependent variable in our statistical models. Following Bell et al. (2009) and Seyfarth (2014: 144), we use the logarithm of word duration (in our case, the natural logarithm) for two reasons: Firstly, log-transforming word duration makes its distribution slightly more normal. Secondly, we are interested in relative changes in duration rather than absolute ones (cf. Bell et al. 2009: 98): For instance, an increase in duration of 50 milliseconds is more substantial for a short word of 100 milliseconds than for a long word of 500 milliseconds. Note that after controlling for word length (see section 3.4.1), word duration will also be indicative of the relative duration of words compared to the average duration of words with the same number of phones and thus indirectly also of their articulation rate (measured in phones per second), which is also sometimes used as a variable in other studies. Recall that word durations in the current study were calculated based on automatically aligned and subsequently manually corrected word start and end times (see section 2.3).

3.3 Predictor variables

3.3.1 Word frequency

Our first independent variable of theoretical interest is word frequency. Based on studies of word frequency and word duration in English (Whalen 1991; Gahl 2008; Bell et al. 2009; Seyfarth 2014, among others), it is expected that more frequent words will have a shorter duration on average than less frequent words. However, the question that the current study asks is whether such effects will be detectable in the kind of language documentation data used here. To keep results comparable across the languages in our sample, we count word frequencies only in the corpora included in our sample for all the languages (first column of Table 1), even if alternative corpora would have been available, as is the case for English and Dutch, and to some extent Sakha and Chintang (for studies

on acoustic reduction and frequency based on large corpora of English and Dutch see, e.g., Pluymaekers, Ernestus & Baayen 2005; Seyfarth 2014).

To obtain word frequency counts, we use identical word forms as a definitional criterion for word types whose frequencies are counted, following the literature on English (Seyfarth 2014; Sóskuthy & Hay 2017: 301). That is, we count, e.g., tokens of *say* separately from tokens of *says*. We additionally require two tokens of the same word type to have identical morphological segmentations and glosses to distinguish homographs like Bora *iñe* 'this' vs. *iñe* 'moriche palm'. However, this latter requirement could not be applied to the Dutch corpus, for which no glosses or morphological segmentation are available, nor to the English corpus, for which no glosses are available, since these two corpora were not annotated with traditional morphological glossing used in language documentation projects.

Across languages with very different morphological systems, the frequential properties of words also differ. For instance, in polysynthetic languages, one will find many long word forms, of which even the most frequent have a low frequency compared to the frequencies of words in isolating languages. For example, in Chintang about 1,800 different verb forms for each verb are in regular use (Stoll et al. 2012), compared to just four in English (e.g., *play, plays, played, playing*). Because of such differences, in, e.g., the Baure corpus, four of the five most frequent word forms are forms of the verb *say* (see Appendix A). However, the differences in frequency between the most frequent words in, e.g., morphologically complex Bora compared to, e.g., morphologically simple English are not overwhelming in our data (see Appendix A). Therefore, in the current study, we apply the traditional word-form frequency measure, but take such differences into account in our interpretation of the results, and identify here as a question for future research whether, for morphologically complex languages, frequency counts should better be based on roots or (potentially derived) lemmas, rather than (inflected) word forms.

Since the corpora we used for the ten languages in our sample vary in size from about 18,000 words for Baure to about 56,000 words for English (cf. Table 1), we normalized word frequency scores by using relative frequencies (calculated by language/corpus) instead of absolute frequencies.[3] Following Bell et al. (2009), we log-transformed relative frequencies, since this leads to a somewhat more normal distribution of word frequencies; see also Seyfarth (2014) and Sóskuthy & Hay (2017), who, however, use log-transformed absolute frequencies.

As one strategy to avoid using unreliable frequency estimates especially for relatively rare words, we only analyze here the 100 most frequent word types (nouns and verbs) in each language.[4] These range in absolute frequency from 803 tokens (corresponding to a relative frequency of 0.0143) for the most frequent word in English to 9 tokens (corresponding to a relative frequency of 0.0005) for the least frequent word in the 100-word-types list for Baure. However, even among these words, we can clearly observe effects of small corpus size and the necessarily limited representativeness of the speech they document. This becomes apparent when inspecting the 100 most frequent word types in the complete corpus of Chintang (cf. Appendix A), which include such unusual items as *ebhokad* 'avocado', *saikal* 'bicycle' and *heŋga* 'bamboo basket' in 4th, 19th,

[3] Using frequency ranks instead of relative frequencies would have been another alternative (cf. also Bell et al. 2009: 100). However, frequency ranks lose potentially important information about the ratio of different frequencies, for example, about how many times more frequent the most frequent word is compared to the second most frequent word.
[4] A different approach was taken in an earlier study on a similar data set (Seifart et al. 2018), which aimed to control for word frequency. In that study, word type was included as a random factor in statistical models since this captures all aspects of word type's familiarity (and therefore its production and retrieval probabilities), including its frequency. In contrast, the approach in the current study is to explore the use of word frequency estimates on their own and to make explicit the methodological issues involved.

and 27th place, respectively. Their unexpectedly high frequency is due to the fact that 22 texts out of the 40 texts in the Chintang corpus are retellings of the Pear Story stimulus film (Chafe 1980), which involves the plucking of pears, identified as avocados by Chintang speakers, and a bicycle accident. As will be seen in section 4.2., results on word frequency effects in the Chintang corpus that includes the Pear Stories also clearly differ from the results for all the other nine languages.

We therefore created a second version of the Chintang corpus for our analyses from which we excluded all retellings of the Pear Story, called *Chintang (no pear stories)* in the following. This allows us to explore whether the fact that ordinarily less frequent words appear in the list of most frequent words due to corpus composition distorts our results on the effect of word frequency on word duration. The Chintang corpus (before the removal of the Pear Stories) was clearly the most skewed corpus in terms of contents, but - to a lesser extent - the top 100 word types in the other language documentation corpora also reflect the content of the included texts, and some of these lists appear rather different to typical lists of the top 100 word types derived from very large corpora available for English and other major languages (see Appendix A). This is particularly apparent in words that are attested in one individual text only (document frequency = 1) but that still make it onto the 100-most-frequent-word-types list because they are so frequent in that one text, and because there is an overall limited number of texts. This is why the words for 'frog', for example, are included in the Bora list (64th place) and the Hoocąk list (10th place, document frequency 2).

To evaluate our word frequency estimates, we plotted frequency ranks against absolute frequencies (in Appendix C). This confirms that word frequencies in the ten languages do indeed follow a typically Zipfian distribution (Zipf 1949), i.e. the most frequent word is about twice as frequent as the second most frequent word and about three times as frequent as the third most frequent word, etc. We take this as an indication of the validity of our word frequency measures.

Note that, in addition to unigram frequencies, that is, word frequencies proper, word durations are also known to be affected by a word's (backward or forward) predictability in context (usually estimated using bigram or trigram language models). Predictability effects capture, for instance, the fact that the word *board* in *across the board* may be pronounced faster because it is very likely to occur in the context of *across the*. The recent literature on English in fact argues that predictability in context (conditional probability) is a better predictor than (unigram) frequency for reduction effects (Aylett & Turk 2004; Bell et al. 2009; Demberg et al. 2012; Seyfarth 2014; Sóskuthy & Hay 2017) and for other processing effects (McDonald & Shillcock 2001; Baayen 2010; Piantadosi, Tily & Gibson 2011). Both Seyfarth (2014: 147–148) and Sóskuthy & Hay (2017: 305–306), for example, report mostly non-significant word frequency effects if word informativity (defined as a word's average predictability in context) is also included in the same statistical models and discuss the effect of collinearity between word frequency and word predictability/informativity. However, frequency measures, as used in the current study, do capture predictability effects at least to some extent, for two reasons. Firstly, frequency and predictability measures have been shown to be highly correlated in various languages, despite detectable differences (Piantadosi, Tily & Gibson 2011); and secondly, in the subset of morphologically complex languages, word forms are highly contextualized and unigram frequency estimates capture this dependency to some extent.

3.3.2 Word class

Word class (nouns vs. verbs) is the second independent variable of theoretical interest in our study. Word class has not been widely studied as a factor relevant to phonetic word duration. However,

a study on a data set similar to the one used here (Seifart et al. 2018) found a robust effect of word class in that the mean articulation rate of words occurring in time windows just before nouns was slower compared to words occurring just before verbs. Earlier (preliminary) analyses had also found that verbs themselves were pronounced faster than nouns themselves (Seifart & Strunk 2015). This is in line with the results reported in studies on predictability and word duration in English that did include word class/part-of-speech as a control variable, although they did not further interpret these results (like Seyfarth 2014: 145–146; Sóskuthy & Hay 2017: 305). Based on these earlier results, we predict that verbs in our data set will have a shorter duration compared to nouns when other factors are taken into account at the same time.

To study the effect of word class (nouns vs. verbs) on word duration, we use the word-class category of the lexical root contained in a word, as identified by language experts through manual annotation using language-specific criteria. Even though individual words may be nominalized or verbalized, in our data, this occurs in less than 5% of nouns and verbs, as manual inspection of about 10% of the corpora of each of the languages studied here revealed. Using the category of the lexical root also captures more closely the distinction between "object words" and "action words", which is known to be more relevant to language processing than the syntactic surface categories of words (Vigliocco et al. 2011; Kemmerer 2014).

3.4 Control variables

3.4.1 Word length

A word's length in terms of number of phones obviously strongly influences its phonetic duration, and must therefore be controlled for when studying the effect of frequency and word class on word duration. We did this by including word length (the number of phones) as a baseline control variable in our models of word duration. The number (and kind) of phones contained in the lexical representation of word forms are expected to explain most of the variation in the duration of different words, since words containing more phones will naturally have a higher duration on average than words containing fewer phones.

As mentioned above (section 2.2), we use the number of orthographic characters as a proxy for the number of phones (cf. also Seyfarth 2014: 144–145). This is relatively unproblematic for the fieldwork corpora because these use orthographies that are close to phonological transcription. But even for languages with deep, historically grown orthographies, such as English and Dutch, it has been shown that the correlations between word length in orthographic characters and word length in phonological segments are extremely high (Piantadosi, Tily & Gibson 2011), justifying our approach also for these languages. However, since the languages in our sample do differ in the complexity of their orthographic systems and their propensity to use digraphs and trigraphs, the exact meaning of the variable *word length* in terms of the actual number of phones does differ slightly between languages, which is one more reason why we chose to build separate statistical models for the individual languages rather than one complex model covering all of them at the same time (see section 3.1.).

Table 3 gives an overview of the distribution of word lengths in the ten languages in our sample (in addition to their morphological complexity, cf. section 3.4.2). Since a log-transformation does not lead to a more normal distribution of the variable *word length*, we follow Sóskuthy & Hay (2017) in using the raw number of phones rather than log-transforming it (cf. also Seyfarth 2014: 144–145).

<Table 3>

Note that the factor word length correlates with word frequency, due to the well-known fact that more frequent words tend to be shorter than less frequent words (Zipf 1935; Piantadosi, Tily & Gibson 2011). Note also that, in some recent studies on English (such as Seyfarth 2014; Sóskuthy & Hay 2017), word length in terms of number of phones (or orthographic length) is used as only one measurement among a number of related measures of word length. Among the commonly used additional measures are the number of syllables in a word and average syllable duration. Other recent studies attempt to predict a word's expected duration based on its phonemic content (Tang & Bennett 2018) or to predict the expected duration of a word in the actual context it appears in by using a text-to-speech system (TTS) (Seyfarth 2014). As there are no trained text-to-speech systems for the languages in our corpus, except for Dutch and English, we cannot use this more sophisticated baseline control variable.[5]

Note that by including word length as a control factor, we also control, to some extent, for polysyllabic shortening, i.e. for the phenomenon that longer words are expected to be pronounced faster in terms of phones per second.

3.4.2 Morphological complexity

The second control variable in our model is morphological complexity, i.e. the number of morphs that a word form is composed of. The reason for including this control variable is that the additional effort of assembling a complex word might be reflected in slower pronunciation. Consider, for example, the perfectly normal and frequent Chintang verb form *ma-u-tup-yokt-a-ŋ-ni-hẽ* (NEGATIVE-3.SUBJECT-meet-NEGATIVE-PAST-1SG.OBJECT-PLURAL.SUBJECT-INDICATIVE.PAST) 'They did not meet me'. It is important to control for this given that the ten languages in our sample differ widely in their degree of overall morphological complexity (see Table 3). Recall that the Dutch corpus we use does not provide morphological segmentations of words and that therefore we omitted this variable from the analysis of the Dutch data. Since English is not particularly morphologically complex, this variable has typically not been included in studies on English word duration (for example, Bell et al. 2009; Seyfarth 2014; Sóskuthy & Hay 2017). Relevant studies so far report contradicting results on the effect of morphological complexity on word duration: Warner et al. (2006) report a null effect in an experimental study on Dutch which compared homophonous verb forms with and without a morphological boundary. But Plag et al. (2017) show that segments such as English word-final *s* are longer if they represent separate morphemes than non-morpheme counterparts.

One might think that because we focus here on the 100 most frequent word types in each language, morphological complexity might be less relevant for the following reason: These more frequent words might have a lower average morphological complexity and less variation in morphological complexity than less frequent words, which we exclude. However, note that we do include parts of speech that tend to be morphologically complex (nouns and, especially, verbs), while we exclude those that tend to be morphologically simple (function words, particles, etc.). At any rate, inspection of these sets of 100 word types (see Appendix A) clearly shows that they in

[5] We also worry that using predictions of word duration derived from a statistically trained text-to-speech system as a control variable (Demberg et al. 2012) may actually model variation in word duration as part of this baseline that is due to other, theoretically more interesting variables and thus lead to underestimating the importance of these theoretically more interesting variables because the TTS model has learned relevant regularities from its training corpus.

fact contain many morphologically complex words: They consists of about 1.5-2.0 morphs on average for most languages, except N‖ng with 1.16 morphs and English with 1.24 morphs.

3.4.3 Word position

Word duration is not only determined by properties of word forms, as captured by the factors discussed so far, but also by the position of a word (token) in an utterance, in particular with respect to the boundaries of prosodic units, such as intonation phrases and utterances. Utterance-final lengthening effects, and to a lesser extent utterance-initial lengthening effects, are well-documented for English and other languages (Oller 1973; Klatt 1976; Yuan, Liberman & Cieri 2006; Fletcher 2010). To control for word position effects across languages, we include the position of the target word in the utterance as another control factor (cf. also Aylett & Turk 2004: 41; Sóskuthy & Hay 2017; a different approach is taken by Seyfarth 2014 who excludes all words adjacent to potential utterance boundaries). We normalize position by the length of the utterance so that it ranges from 0 (first word in the utterance) to 1 (last word in the utterance). The normalized position of the ith word (counting from 1) in an utterance containing n words is defined as $(i - 1)/(n - 1)$.

Utterances in the current study are defined as the annotation units that were set by language experts during the transcription of the data. Recall that their size and exact definition varies between the languages studied here (see section 2.3). In particular, longer annotation units may include more than one prosodic phrase, each of which potentially exhibits lengthening effects at its boundaries. Moreover, it is well known that final words (and within final words, final segments) are disproportionately strongly affected by such boundary-adjacent lengthening, while penultimate and antepenultimate words are not or hardly affected (Yuan, Liberman & Cieri 2006; Turk & Shattuck-Hufnagel 2007). Therefore, a word's position in an utterance as coded in the current study is a rather coarse measure to capture boundary-adjacent lengthening. We chose this measure here in order to have a measure of position suitable for all data available for each language in the current study. A related study comparing the same languages specifically with respect to utterance-final lengthening (Seifart et al. In press) chose to define word position more carefully by only considering comparisons among the final four words of utterances, at the expense of discarding more than 75% of the available data.

3.4.4 Local speech rate

A word may also be pronounced relatively quickly or slowly simply because it is embedded in an utterance with high or low overall speech rate. In order to control for this effect, we include local speech rate as a variable in our models of word duration, following Bell et al. (2009) and Seyfarth (2014), among others. We calculate local speech rate as the number of phones per second in the complete utterance that contains the target word, where, again, utterances are defined as annotation units set by language experts. In this calculation, we include (silent and filled) pauses when measuring the length of the utterance but exclude the target word itself by subtracting its length in phones from the overall number of phones in the utterance and its duration from the overall duration of the utterance. Unlike in the current study, in previous studies, (local) speech rate has usually been calculated based on syllables. For example, Seyfarth (2014: 144) used the number of syllables per second, while Sóskuthy & Hay (2017: 305) used the average syllable duration, that is, the inverse speech rate. However, the fieldwork corpora used here are not annotated for syllable boundaries, and therefore, we calculate speech rate in phones per second instead.

3.5. Random effects

Speech rate modulations, including articulation speed, are well known to vary strongly between individual speakers (e.g., Johnson, Ladefoged & Lindau 1993; Jacewicz et al. 2009). To account for such variation, we include what is called a random effect or random intercept (as opposed to the fixed factors described in section 3.4) for speaker in our mixed-effects models. We deal with this variation in the form of a random effect, rather than as fixed effect (predictor or control) variable, because we are not interested here in sociolinguistic differences between speakers, and, accordingly, these have not been strategically sampled. Another source of variation are differences between individual texts and recordings included in the corpora used here and between the genres they represent (e.g., traditional narratives vs. conversation). Therefore we include another random effect for text in our ten mixed-effects models.[6] The inclusion of both of these is a standard procedure in recent studies on word duration (e.g., Bell et al. 2009; Seyfarth 2014; Sóskuthy & Hay 2017).

As mentioned above (footnote 4), a previous study on a similar data set also included random effects for word type (Seifart et al. 2018), as an implicit way of controlling for word familiarity (including frequency), while in the current study, we attempt to study frequency effects more directly and therefore chose not to include a random effect of word type. Note also that some studies include both: for instance, Seyfarth (2014: 147) included both word frequency as a fixed effect and word type as a random effect in his mixed-effects models. However, this is potentially problematic since word frequencies can be considered as properties of word types, and thus including random effects for word type may actually model some of the variation in word duration that we would otherwise attribute to word frequency (as Seyfarth 2014: 147 himself notes).

4. Results and discussion

4.1. Simple correlations between word frequency and word duration

To get a first impression of the effects of word frequency on word duration, we plotted frequency ranks for the 100 most frequent word types (nouns and verbs) on the x-axis against ranks of average word durations (as given in the tables in Appendix A) on the y-axis (Figure 2). We then calculated simple bivariate correlations (i.e. not controlling for word length or other factors) between word frequency ranks and mean word duration ranks using the Spearman rank correlation coefficient. As can be seen in Figure 2, the correlations between word frequency and word duration are negative in all languages except in Chintang. That is, a higher frequency is associated with a lower mean word duration, as expected, in nine out of ten languages. In the complete Chintang corpus, there is also a negative correlation, but it is very weak. In the reduced Chintang corpus (without the Pear Stories), there appears to be an unexpected, but even weaker positive correlation (more frequent words tend to have longer mean durations).

<Figure 2>

4.2. Frequency effects in linear mixed-effects models

[6] Since we do not further discuss random effects below, we also note here that in some of our corpora, one speaker usually only occurs in one text and one text usually contains utterances from only one or two speakers. In these cases, the random effects for speaker and text are so strongly correlated that only one of them turns out significant in the multivariate model (cf. Appendix D).

Our statistical analysis simultaneously assesses the effects of word frequency and the other factors introduced above using linear mixed-effects models for each of the ten languages (plus the reduced Chintang corpus without the Pear Stories) based on the model structure given in Formula (1) above. This procedure allows us to study the effect of frequency and word class, which we are most interested in, while controlling for the effects of the other variables (word length, morphological complexity, position, and local speech rate). Detailed results for fixed and random effects for these eleven models are provided in Appendix D. Before we turn to the discussion of word frequency, word class, and the other individual factors, we note here that the results summarized in Table 4 show that the multivariate models for the ten languages (as a whole) are each able to explain a fairly large proportion of the variance in word duration, as indicated by the R^2 values (in columns 3 and 4 of Table 4): These range from 0.2977 to 0.5682 for conditional $R^2_{(c)}$ (taking both fixed and random effects into account) and from 0.2328 to 0.5145 for marginal $R^2_{(m)}$ (considering fixed effects only) (a model able to predict variation in word duration perfectly could attain a theoretical maximal R^2 value of 1). The model for Hoocąk shows the best performance overall, while the two alternative models for Chintang explain the least amount of variance in word duration and, consistent with the lack of clear correlations for Chintang in Figure 2, they have the worst performance of all eleven models evaluated here.

Table 4 also summarizes the effect of (log) word frequency on (log) word duration, which is additionally visualized in the form of effects plots in Figure 3. As Table 4 and Figure 3 show, there is a statistically significant negative effect of word frequency on word duration, i.e. the effect we expected, in seven out of our ten languages (Baure, Dutch, English, Even, N‖ng, Sakha, and Texistepec), confirming results from the simple bivariate correlations for these languages. According to the standardized β coefficients (which measure the influence of a predictor on the dependent variable in terms of standard deviations and can therefore be used to compare different predictors to each other and to compare one and the same predictor across different models), the effect of word frequency is strongest (in the negative direction) in English (β = -0.2858) followed closely by Sakha (β = -0.1927) and Dutch (β = -0.1866). Baure, N‖ng, Even, and Texistepec exhibit a slightly weaker but still comparable negative effect of word frequency on word duration, with β values of -0.1597 (Baure), -0.1414 (N‖ng), -0.1151 (Even), and -0.0959 (Texistepec), respectively. In one language, namely, Hoocąk, word frequency also has a small negative effect on word duration (β = -0.0118) but does not make a significant contribution to the model according to the likelihood ratio test.

In Bora, the effect of word frequency is positive against the prediction but very small (β = 0.0038) and word frequency is again not a significant predictor in the model according to the likelihood ratio test. It is surprising that there is no significant word frequency effect in the multivariate model for Bora given the strong correlation between frequency and duration (as plotted in Figure 2). However, additional testing showed that the correlation between the two predictors word length and word frequency is particularly strong for our Bora data set (including the 100 most frequent words, nouns and verbs only), compared to the other languages. This fact may perhaps make word frequency somewhat redundant in the multivariate model of word duration for Bora.

An even more surprising result is that word frequency has a significant positive effect on word duration for the entire corpus of Chintang (with a relatively small β coefficient of 0.0660) (cf. Figure 3). This result would suggest that in Chintang, more frequent words have longer durations, i.e. are pronounced more slowly than less frequent words, which goes against our expectation and is also at odds with the fact that the bivariate correlation between word frequency

and word duration was negative also for Chintang. However, recall that the frequency counts sampled from this Chintang corpus were heavily biased towards the idiosyncratic content of the Pear Stories and this results in high relative frequencies for otherwise uncommon words like *avocado* and *bicycle*. However, in the reduced Chintang corpus (without the Pear Stories), the effect is still in the unexpected positive direction ($\beta = 0.0257$). The likelihood ratio test indicates that word frequency does not make a significant contribution to predicting word durations, but we cannot exclude the possibility that this is an artefact of the reduced sample size. As noted above, Chintang is also exceptional with regard to the overall R^2 achieved by our models: The marginal R^2 value of 0.2328 for the entire Chintang corpus is the lowest of all ten languages in our sample, and the marginal R^2 value of 0.2892 for the reduced Chintang corpus without the Pear Stories is only slightly better than the one for the complete Chintang corpus.

<Table 4> <Figure 3>

To further evaluate the importance and size of the effect of word frequency on word duration in our ten languages, we calculated several alternative measures of the amount of variance in word duration that the factor word frequency can explain (cf. columns 6 to 9 in Table 4), that is, the portion of the variance that is *uniquely* explained by the predictor word frequency in our models. While languages differ somewhat with regard to the amount of this variance, the values for marginal $\Delta R^2_{(m)}$ as well as for semi-partial R^2 are generally in the same ballpark for those languages where word frequency has a significant negative effect: Percentage change in marginal R^2 ($\Delta R^2_{(m)}$) ranges from almost 20% in English (19.55%), 7.90% in Dutch, 5.84% in Baure, 5.72% in Sakha, 3.94% in N‖ng, 3.20% in Even, to 1.64% in Texistepec, while semi-partial R^2 values range from 0.10 for English, followed by 0.05 for Dutch and Sakha, 0.03 for N‖ng, 0.02 for Baure and Even, to 0.01 for Texistepec (that is, values range from almost 20% of variation in word duration uniquely explained by word frequency in our English data to about 1% in our corpus of Texistepec). Values for $\Delta R^2_{(m)}$ and semi-partial R^2 are negligible for Bora ($\Delta R^2_{(m)} = -0.02\%$, semi-partial $R^2 = 0$, that is, the model actually improves after omitting word frequency) and Hoocąk ($\Delta R^2_{(m)} = 0.08$, semi-partial $R^2 = 0$), respectively, where the effect of word frequency is also not statistically significant according to likelihood ratio tests. The exceptional language Chintang, where the effect of word frequency is reversed, has intermediate values of 2.02% for $\Delta R^2_{(m)}$ and 0.0049 for the semi-partial R^2. While percentage changes of maximally around 20% (for English) in the variance explained by all fixed factors when the variable word frequency is added or removed (marginal $\Delta R^2_{(m)}$) may not seem like a lot, it has to be kept in mind that we expect that most of the variation in word durations will already be explained by word length (that is, the number of phones in a word type). We take this up in section 4.4.

4.3. Effects of word class on word duration

The most important results regarding the effects of word class on word duration in our linear mixed-effects models are reported in Table 5 (β coefficients, $\Delta R^2_{(m)}$, and statistical significance based on likelihood ratio tests) and graphically represented in Figure 4 (detailed results are given in Appendix D). These show that in seven out of our ten languages, verbs have significantly shorter durations than nouns (Baure, both versions of Chintang, Dutch, English, Even, Sakha, and Texistepec). Word class is not a statistically significant predictor for word durations in Bora and N‖ng. Results on Hoocąk are exceptional in our sample in that verbs appear to have longer durations than nouns. This could possibly be due to word order: Additional testing showed that in

Hoocąk, even more so than in the other verb-final languages (like Even and Sakha), verbs are very often utterance-final, whereas nouns almost never occur in this position. Thus final lengthening would disproportionally affect verbs in this language. Interestingly, the effect of position in Hoocąk is also negative (that is, the later in an utterance a word occurs, the shorter it is, see Section 4.4). It may thus be that the part-of-speech effect and the final lengthening effect cannot really be distinguished in the data for this language and therefore the effects of part-of-speech and position appear to be different than expected.

Note that results from our models on the factor word class do not mean that the absolute duration of verbs is shorter on average than the duration of nouns. Instead, the results show that word duration for verbs is shorter than that of nouns if the effects of word length and of morphological complexity (which are both usually higher for verbs than for nouns in the languages studied here) are accounted for. This procedure also accounts, in principle at least, for differences in word order, e.g. the fact that in verb-final languages, like Even, and Sakha, verbs would also be often affected by final lengthening. However, the potentially strong word position effects in Hoocąk still need to be appropriately dealt with in future research.

4.4. Comparison of all six fixed factors in the linear mixed-effects models

Results from our multivariate models on the comparison of the effects of all six fixed factors across the ten languages (frequency, word class, and the four control factors word length, morphological complexity, word position, and local speech rate) are also given in Table 5. Figure 4 shows the direction and strength of each of the six fixed effects in each language based on the standardized β coefficients. Recall that standardized β coefficients measure the effect a particular predictor has on the dependent variable in terms of standard deviations. This unitless measure allows for a comparison of effect sizes between different predictors (possibly originally measured using different scales and units of measurement) within and across models. Recall also that marginal $\Delta R^2_{(m)}$ values, as given in Table 5, measure the proportion of variance in word duration that the model is additionally able to predict when the fixed effect in question is added to the model. Both in Table 5 and Figure 4, we additionally provide significance stars based on likelihood ratio tests (cf. Appendix D for more details on the models for individual languages).

<Table 5>

As can be seen in Table 5 and Figure 4, the effects on word duration of the six factors studied here are similar across the ten languages in some respects, but also different in others. We compare their effects by discussing them one by one, starting with those whose effects are the most uniform across the ten languages, and then moving on to the ones where languages diverge.

The baseline control variable word length (in phones) uniformly leads to longer durations, as is of course expected. Its standardized β coefficients are all positive and also of roughly the same magnitude in all languages, with a range from 0.35 for English to 0.67 for Baure. It is also by far the most important predictor for word duration in all languages, with $\Delta R^2_{(m)}$ values ranging from 21.78% for English to 67.07% for Even. The β and $\Delta R^2_{(m)}$ values are always the highest for word length compared to the values for all other factors in a particular language. There seems to be a natural (though not perfect) correlation between the average word length of a language (cf. Table 3) and the strength of the effect of word length on word duration. For instance, N‖ng and English have relatively short words (3.45 phones per word on average for N‖ng and 3.70 for English) and correspondingly display relatively small effects of word length on word duration (β

= 0.45 for N‖ng and β = 0.35 for English) compared to the other languages, whereas Bora and Hoocąk have longer words of 7.13 and 6.64 phones on average, respectively, and accordingly also exhibit a stronger effect of word length on word duration (β = 0.59 for Bora and β = 0.60 for Hoocąk).

The results for local speech rate are also consistent across languages. As expected, this control factor has a negative effect in all ten languages (including the reduced Chintang corpus without the Pear Stories) in that a higher local speech rate is associated with a shorter duration of the target word. Its effect size ranges from a standardized β coefficient of -0.08 and a $\Delta R^2_{(m)}$ value of 1.81% in Baure to a β value of -0.28 and a $\Delta R^2_{(m)}$ value of 26.61% for the reduced Chintang corpus.

Compared to other factors, the two predictor variables, word frequency and word class, also have relatively uniform effects on word durations in our data. As already discussed in sections 4.2-4.3, seven languages display a significant effect in the expected direction for word frequency, with a statistically significant reversal in one, probably biased, version of the Chintang corpus. Likewise, seven language exhibit a significant word class effect in the expected direction, while results are reversed with regard to word class in the Hoocąk corpus, and statistically not significant in other corpora.

For the variable position of a word in an utterance, aimed primarily at capturing final lengthening, the results differ strongly between languages. Positions later in the utterance are associated with longer word durations – as expected – in five languages (Baure, Dutch, English, N‖ng, and Texistepec), as well as in the complete corpus of Chintang. But, unexpectedly, position later in a clause leads to statistically shorter durations in two languages (Even and Hoocąk). The effect of position is not statistically significant in Bora, the reduced Chintang corpus without the Pear Stories, and Sakha. Note that – unlike for frequency – the biased contents of the complete Chintang corpus should not distort results on final lengthening, so the lack of significance in the reduced version of the Chintang corpus is here likely due only to the reduced sample size.

Regarding the unexpected results for Even and Hoocąk, additional testing showed that these are actually the two languages in which verbs, as against nouns, occur most often in final position. As mentioned above, this may explain the unexpectedly long durations of words occurring toward the end of utterances in these corpora.

Note that a study on a similar data set that focused on the durations of the final four words of utterances only, (Seifart et al. In press), we found that in all ten languages studied here final words were significantly lengthened, with mixed results for prefinal positions. The mixed results of the current study thus appear to be at least partially due to the fact that, in the current study, we assigned a position value to every word in an utterance, while final lengthening primarily affects the last word in an utterance only. In addition, many of the annotation units in the language documentation corpora most probably include various (smaller) intonation phrases, which would only be detectable by additional prosodic annotation. These may each trigger final lengthening on words we here coded as medial. In line with this consideration, in the current study there is a tendency for effects of position to be strongest in languages in which the annotation units, which are used as proxies for utterances, are shortest (see Table 1) and that thus more closely approximate intonation phrases, e.g., N‖ng and Texistepec, while they are less clear in languages in which annotation units are longer, e.g., Bora.

The effect of the number of morphs in a word type on word duration differs the most across the ten languages in the current study. While there are statistically significant positive effects (more morphs result in longer word duration) in five languages (Chintang [both in the full and in the

reduced corpus], Even, Hoocąk, N‖ng, and Sakha), there are significant effects in the opposite direction (more morphs result in a shorter word duration) in three other languages (Bora, English, and Texistepec). There is no significant effect of morphological complexity in Baure. Recall that morphological complexity could not be tested for one language, Dutch, due to a lack of relevant morphological annotations. The effect of morphological complexity also generally seems to be weaker than the effect of other factors such as word frequency (except for Chintang and Hoocąk). There could be collinearity between word class and morphological complexity if verbs are much more complex morphologically than nouns in a language or vice versa. In such a case, the variables word class and number of morphs may perhaps compete for the same variance in word duration. The fact that three languages display an unexpected negative effect may thus indicate that the effect of morphological complexity on word duration is either strongly dependent on other, yet unknown factors (such as the distinction between inflection, derivation, and perhaps cliticization), or is not captured properly in our analysis.

<Figure 4>

5. Conclusion

The first question we asked in this study was whether frequency effects on word durations are observable in the corpora extracted from language documentations studied here. We conclude that this is generally the case, as long as the frequency counts are not based on data sets that are heavily biased in terms of content, for example, by stemming in their majority from one experiment. The prospects are thus good for studying word frequency in necessarily relatively small language documentation corpora, which may include a variety of narrative texts of different genres, collected for a variety of purposes. In addition, our results provide evidence for reduction of frequent words in five languages (Baure, Even, N‖ng, Sakha, and Texistepec) in which word durations have never been studied before.

The second question we asked was whether nouns are pronounced more slowly than verbs across the languages studied here. We conclude that this is indeed overwhelmingly the case, both in terms of uniformity across languages and in terms of the strength of the effect. These results suggest that word class as a factor for the deceleration of articulation has not only been underestimated in previous studies on phonetic word duration, but also that in spontaneous speech, nouns appear to be articulated more slowly than verbs, while previous experimental studies have found the opposite (Szekely et al. 2005; Vigliocco et al. 2011). Our results on word class are in line with the general, but not exceptionless, cross-linguistic trend observed in another cross-linguistic corpus study (Seifart et al. 2018) regarding articulation speed and pauses in time windows just preceding nouns and verbs. Both of these results can be attributed to the overall higher information load of nouns in discourse. This appears to outweigh increased processing costs of verbs because of their relative grammatical and semantic complexity, and their intrinsic dependencies with other elements in the clause, e.g., subjects and objects, which have been used to explain slower articulation of verbs in experimental studies (Szekely et al. 2005; Vigliocco et al. 2011).

The third aim of this study was a comparison of the effects of six factors on word duration, including four factors that were included as control factors in the analyses. Here, we conclude that across languages, word durations are uniformly most strongly affected by word length and by the speech rate in the surrounding context. The expected effects could thus be reliably replicated using the methods applied here and the data set used here. Compared to other factors, results for word frequency and word class (nouns vs. verbs) were also relatively clear, as just discussed. The results

for the remaining two factors, morphological complexity and word position, were more mixed across languages. It is reasonably clear that lack of prosodic annotation is the reason why word position did not yield clearer results in the current study. Regarding morphological complexity, more research, involving also more annotation work, is needed to understand how it affects word durations across diverse languages. Note that morphological complexity also plays an important role for the frequency estimates across languages in that the most frequent word forms in morphologically complex languages include more individual inflected word forms belonging to the same lemma, adding to the importance of a deeper understanding of this factor.

An additional aim of this study has been to identify key methodological issues for the exploitation of multi-purpose language documentation corpora for cross-linguistic research on the phonetic realization of words. The first thing to note here is that such research is nowadays greatly facilitated by automatic phoneme-level time alignment procedures, from which accurate word start and end times can be obtained. We showed that differences in composition across corpora in terms of, e.g., narratives vs. conversation, do not constitute an obstacle in principle for such studies. Regarding the relatively small size of the corpora, which is due to fieldwork projects having limited funding given the time and expertise needed to transcribe and annotate such data, we showed that, again, this does not in principle preclude such corpora for being used for such studies, even though non-significance in such datasets may in some cases be simply due to the reduced number of data points. We propose a number of techniques to deal with data sparsity, among them focusing on the 100 most frequent word types for each language, which also allows for manual inspection of data.

In summary, the current study has shown the feasibility of applying advanced statistical methods to study word durations in language documentation corpora in order to obtain theoretically interesting results, and has identified methodological issues for the further development of such studies. This sets the stage for further explorations at the interface of documentary linguistics and quantitative approaches to widen the cross-linguistic scope of corpus phonetics (Liberman 2019) and related fields of inquiry.

References

Abercrombie, David. 1967. *Elements of General Phonetics*. Chicago: Aldine.

Arvaniti, Amalia. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40(3). 351–373. doi:10.1016/j.wocn.2012.02.003.

Aylett, Matthew & Alice Turk. 2004. The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech* 47(1). 31–56. doi:10.1177/00238309040470010201.

Baayen, R. Harald. 2010. Demythologizing the word frequency effect: A discriminative learning perspective. In Gonia Jarema, Gary Libben & Chris Westbury (eds.), *Methodological and Analytic Frontiers in Lexical Research (Part I)* (The Mental Lexicon 5), vol. 3, 436–461. Amsterdam: John Benjamins. doi:10.1075/ml.5.3.10baa.

Babinski, Sarah, Rikker Dockum, J. Hunter Craft, Anelisa Fergus, Dolly Goldenberg & Claire Bowern. 2019. A Robin Hood approach to forced alignment: English-trained algorithms and their use on Australian languages. *Proceedings of the Linguistic Society of America* 4(1). 3-1–12. doi:10.3765/plsa.v4i1.4468.

Bartoń, Kamil. 2018. *MuMIn: Multi-Model Inference*. https://CRAN.R-project.org/package=MuMIn (5 April, 2019).

Bates, Douglas, Martin Mächler, Benjamin M. Bolker & Steven C. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. doi:10.18637/jss.v067.i01.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111. doi:10.1016/j.jml.2008.06.003.

Bertinetto, Pier Marco & Chiara Bertini. 2008. On modeling the rhythm of natural languages. *Proceedings of the 4th International Conference on Speech Prosody, SP 2008*, 427–430.

Bickel, Balthasar, Sabine Stoll, Martin Gaenszle, Novel Kishore Rai, Elena Lieven, Goma Banjade, Toya Nath Bhatta, et al. 2011. *Audiovisual corpus of the Chintang language, including a longitudinal corpus of language acquisition by six children: ca. 650,000 words transcribed and translated, of which ca. 450,000 glossed, plus paradigm sets and grammar sketches, ethnographic descriptions, photographs*. Nijmegen: The Language Archive. https://hdl.handle.net/1839/00-0000-0000-0005-6F41-C@view.

Brysbaert, Marc & Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4). 977–990. doi:10.3758/BRM.41.4.977.

CGN-consortium, Language and Speech Nijmegen & ELIS Gent. 2003. *Corpus Gesproken Nederlands*. Nijmegen: Nederlandse Taalunie.

Chafe, Wallace L. 1980. *The Pear Stories. Cognitive, Cultural, and Linguistic Aspects of Narrative Production* (Advances in Discourse Processes 3). Norwood, NJ: Ablex Publishing Company.

Danielsen, Swintha, Franziska Riedel, Femmy Admiraal & Lena Terhart. 2009. *Baure Documentation*. Nijmegen: The Language Archive. https://hdl.handle.net/1839/00-0000-0000-000D-8382-B@view.

Demberg, Vera, Asad B. Sayeed, Philip J. Gorinski & Nikolaos Engonopoulos. 2012. Syntactic surprisal affects spoken word duration in conversational contexts. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 356–367. Stroudsburg, PA: Association for Computational Linguistics.

Ernestus, Mirjam. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* (SI: Usage-Based and Rule-Based Approaches to Phonological Variation) 142. 27–41. doi:10.1016/j.lingua.2012.12.006.

Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5). 429–492. doi:10.1017/S0140525X0999094X.

Fletcher, Janet. 2010. The Prosody of Speech: Timing and Rhythm. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The Handbook of Phonetic Sciences, Second Edition*, 521–602. Chichester: Blackwell. doi:10.1002/9781444317251.ch15.

Fox, John. 2003. Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software* 8(15). 1–27. doi:10.18637/jss.v008.i15.

Gahl, Susanne. 2008. Time and Thyme Are not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech. *Language* 84(3). 474–496. doi:10.1353/lan.0.0035.

Gibbon, Dafydd & Peng Li. 2019. Quantifying and Correlating Rhythm Formants in Speech. The 3rd International Symposium on Linguistic Patterns in Spontaneous Speech (LPSS 2019) Speech communication: Technology, learning, and pathology November 21-22, 2019, Academia Sinica. https://export.arxiv.org/pdf/1909.05639.

Godfrey, John & Edward Holiman. 1993. *Switchboard-1 Release 2 LDC97S62*. Philadelphia: Linguistic Data Consortium.

Grabe, Esther & Ee Ling Low. 2002. Durational variability in speech and the Rhythm Class Hypothesis. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7* (Phonology and Phonetics 4–1), 515–546. Berlin, Boston: De Gruyter Mouton. doi:10.1515/9783110197105.515. doi:10.1515/9783110197105.515.

Güldemann, Tom, Martina Ernszt, Sven Siegmund & Alena Witzlack-Makarevich. 2011. *A Text documentation of N|uu*. London: ELAR. https://elar.soas.ac.uk/Collection/MPI194591.

Haig, Geoffrey, Stefan Schnell & Claudia Wegener. 2011. Comparing corpora from endangered language projects: Explorations in language typology based on original texts. In Geoffrey Haig, Nicole Nau, Stefan Schnell & Claudia Wegener (eds.), *Documenting Endangered Languages. Achievements and Perspectives*, 55–86. Berlin, Boston: De Gruyter Mouton. doi:10.1515/9783110260021.55.

Hammarström, Harald, Robert Forkel & Martin Haspelmath (eds.). 2018. *Glottolog 3.3*. Jena: Max Planck Institute for the Science of Human History. doi:10.5281/zenodo.1321024. https://glottolog.org/ (8 December, 2018).

Hartmann, Iren. 2013. *Hoocąk Corpus*. Leipzig: MPI-EVA.

Haspelmath, Martin. 2001. Word Classes and Parts of Speech. In Neil J. Smelser & Paul B. Baltes (eds.), *International Encyclopedia of the Social & Behavioral Sciences*, 16538–16545. Oxford: Pergamon. doi:10.1016/B0-08-043076-7/02959-4.

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36(1). 161–195. doi:10.1515/ling.1998.36.1.161.

Himmelmann, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases: a cross-linguistic interrater study. *Phonology* 35(2). 207–245. doi:10.1017/S0952675718000039.

Jacewicz, Ewa, Robert A. Fox, Caitlin O'Neill & Joseph Salmons. 2009. Articulation rate across dialect, age, and gender. *Language variation and change* 21(2). 233–256. doi:10.1017/S0954394509990093.

Jaeger, Byron. 2017. *r2glmm: Computes R Squared for Mixed (Multilevel) Models*. https://CRAN.R-project.org/package=r2glmm (30 June, 2019).

Jaeger, T. Florian & Esteban Buz. 2017. Signal Reduction and Linguistic Encoding. In Eva M. Fernández & Helen Smith Cairns (eds.), *The Handbook of Psycholinguistics*, 38–81. Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118829516.ch3.

Johnson, Keith, Peter Ladefoged & Mona Lindau. 1993. Individual differences in vowel production. *The Journal of the Acoustical Society of America* 94(2). 701–714. doi:10.1121/1.406887.

Johnson, Paul C. D. 2014. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution* 5(9). 944–946. doi:10.1111/2041-210X.12225.

Kemmerer, David. 2014. Word classes in the brain: implications of linguistic typology for cognitive neuroscience. *Cortex* 58. 27–51. doi:10.1016/j.cortex.2014.05.004.

Kisler, Thomas, Uwe Reichel & Florian Schiel. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45. 326–347. doi:10.1016/j.csl.2017.01.005.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59(5). 1208–1221. doi:10.1121/1.380986.

Lehiste, Ilse. 1972. The Timing of Utterances and Linguistic Boundaries. *The Journal of the Acoustical Society of America* 51(6B). 2018–2024. doi:10.1121/1.1913062.

Liberman, Mark Y. 2019. Corpus Phonetics. *Annual Review of Linguistics* 5(1). 91–107. doi:10.1146/annurev-linguistics-011516-033830.

McDonald, Scott A. & Richard C. Shillcock. 2001. Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing. *Language and Speech* 44(3). 295–322. doi:10.1177/00238309010440030101.

Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods In Ecology and Evolution* 4(2). 133–142. doi:10.1111/j.2041-210x.2012.00261.x.

Nolan, Francis & Hae-Sung Jeon. 2014. Speech rhythm: a metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1658). 20130396. doi:10.1098/rstb.2013.0396.

Norcliffe, Elisabeth, Alice C. Harris & T. Florian Jaeger. 2015. Cross-linguistic psycholinguistics and its critical role in theory development: early beginnings and recent advances. *Language, Cognition and Neuroscience* 30(9). 1009–1032. doi:10.1080/23273798.2015.1080373.

Oller, D. Kimbrough. 1973. The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America* 54(5). 1235–1247. doi:10.1121/1.1914393.

Pakendorf, Brigitte. 2007a. *Contact in the prehistory of the Sakha (Yakuts): Linguistic and genetic perspectives* (LOT Dissertation Series 170). Utrecht: LOT.

Pakendorf, Brigitte (ed.). 2007b. *Documentation of Sakha (Yakut)*. Leipzig: MPI-EVA.

Pakendorf, Brigitte, Dejan Matić, Natalia Aralova & Alexandra Lavrillier. 2010. *Documentation of the dialectal and cultural diversity among Ėvens in Siberia*. Nijmegen, Leipzig: DOBES, MPIP, MPI-EVA.

Piantadosi, Steven T., Harry Tily & Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America* 108(9). 3526–3529. doi:10.1073/pnas.1012551108.

Pike, Kenneth L. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53(1). 181–216. doi:10.1017/S0022226715000183.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118(4). 2561–2569. doi:10.1121/1.2011150.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org.

Ramus, F. & J. Mehler. 1999. Language identification with suprasegmental cues: a study based on speech resynthesis. *Journal of the Acoustical Society of America* 105(1). 512–521. doi:10.1121/1.424522.

Ramus, Franck, Marina Nespor & Jacques Mehler. 2000. Correlates of linguistic rhythm in the speech signal. *Cognition* 75(1). 265–292. doi:10.1016/S0010-0277(00)00101-3.

Šavrič, Bojan, Tom Patterson & Bernhard Jenny. 2019. The Equal Earth map projection. *International Journal of Geographical Information Science* 33(3). 454–465. doi:10.1080/13658816.2018.1504949.

Seifart, Frank. 2009. Bora documentation. In Frank Seifart, Doris Fagua, Jürg Gasché & Juan Alvaro Echeverri (eds.), *A multimedia documentation of the languages of the People of the Center. Online publication of transcribed and translated Bora, Ocaina, Nonuya, Resígaro, and Witoto audio and video recordings with linguistic and ethnographic annotations and descriptions*. Nijmegen: The Language Archive. https://hdl.handle.net/1839/00-0000-0000-0008-38E5-2.

Seifart, Frank & Jan Strunk. 2015. Local variation in speech rate cross-linguistically: Noun-to-verb ratio and other factors. Functional Cognitive Linguistics (FunC) Friday lecture, KU Leuven, 30 January 2015.

Seifart, Frank, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nikolaus P. Himmelmann & Balthasar Bickel. In press. The extent and degree of utterance-final lengthening in spontaneous speech from 10 languages. *Linguistic Vanguard*.

Seifart, Frank, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nivja H. de Jong & Balthasar Bickel. 2018. Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences of the United States of America* 115(22). 5720–5725. doi:10.1073/pnas.1800708115.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1). 140–155. doi:10.1016/j.cognition.2014.06.013.

Sóskuthy, Márton & Jennifer Hay. 2017. Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166. 298–313. doi:10.1016/j.cognition.2017.05.032.

Stoll, Sabine, Balthasar Bickel, Elena Lieven, Netra P. Paudyal, Goma Banjade, Toya N. Bhatta, Martin Gaenszle, et al. 2012. Nouns and verbs in Chintang: children's usage and surrounding adult speech. *Journal of Child Language* 39(2). 284–321. doi:10.1017/S0305000911000080.

Stoll, Sabine, Taras Zakharko, Steven Moran, Robert Schikowski & Balthasar Bickel. 2015. Syntactic mixing across generations in an environment of community-wide bilingualism. *Frontiers in Psychology* 6. 82. doi:10.3389/fpsyg.2015.00082.

Strunk, Jan, Florian Schiel & Frank Seifart. 2014. Untrained Forced Alignment of Transcriptions and Audio for Language Documentation Corpora using WebMAUS. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, 3940–3947. Reykjavik: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/1176_Paper.pdf.

Szekely, Anna, Simonetta D'Amico, Antonella Devescovi, Kara Federmeier, Dan Herron, Gowri Iyer, Thomas Jacobsen, Anal'a L. Arévalo, Andras Vargha & Elizabeth Bates. 2005. Timed Action and Object Naming. *Cortex* 41(1). 7–25. doi:10.1016/S0010-9452(08)70174-6.

Tang, Kevin & Ryan Bennett. 2018. Contextual predictability influences word and morpheme duration in a morphologically complex language (Kaqchikel Mayan). *Journal of the Acoustical Society of America* 144(2). 997–1017. doi:10.1121/1.5046095.

Tilsen, Sam & Amalia Arvaniti. 2013. Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America* 134(1). 628–639. doi:10.1121/1.4807565.

Trouvain, Jürgen. 2004. *Tempo variation in speech production: implications for speech synthesis* (Phonus 8). Saarbrücken: Institut fur Phonetik, Universitat des Saarlandes.

Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35(4). 445–472. doi:10.1016/j.wocn.2006.12.001.

Vigliocco, Gabriella, David P. Vinson, Judit Druks, Horacio Barber & Stefano F. Cappa. 2011. Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience and biobehavioral reviews* 35(3). 407–426. doi:10.1016/j.neubiorev.2010.04.007.

Warner, Natasha, Erin Good, Allard Jongman & Joan Sereno. 2006. Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics* 34(2). 285–293. doi:10.1016/j.wocn.2004.11.003.

Whalen, D. H. 1991. Infrequent words are longer in duration than frequent words. *The Journal of the Acoustical Society of America* 90(4). 2311–2311. doi:10.1121/1.401072.

Wichmann, Søren. 1996. *Cuentos y colorados en popoluca de Texistepec*. Copenhagen: C.A. Reitzel.

Wichmann, Søren. 2002. *Diccionario analítico del popoluca de Texistepec*. México, D.F.: Universidad Nacional Autónoma de México.

Wichmann, Søren. 2007. *Popoluca de Texistepec*. México, D.F.: Colegio de México.

Yuan, Jiahong, Mark Liberman & Christopher Cieri. 2006. Towards an Integrated Understanding of Speaking Rate in Conversation. *Interspeech 2006*, 541–544.

Zipf, George Kingsley. 1935. *The psycho-biology of language: an introduction to dynamic philology*. Boston: Houghton Mifflin.

Zipf, George Kingsley. 1949. *Human behavior and the principle of least effort; an introduction to human ecology.* Cambridge, Mass.: Addison-Wesley Press.

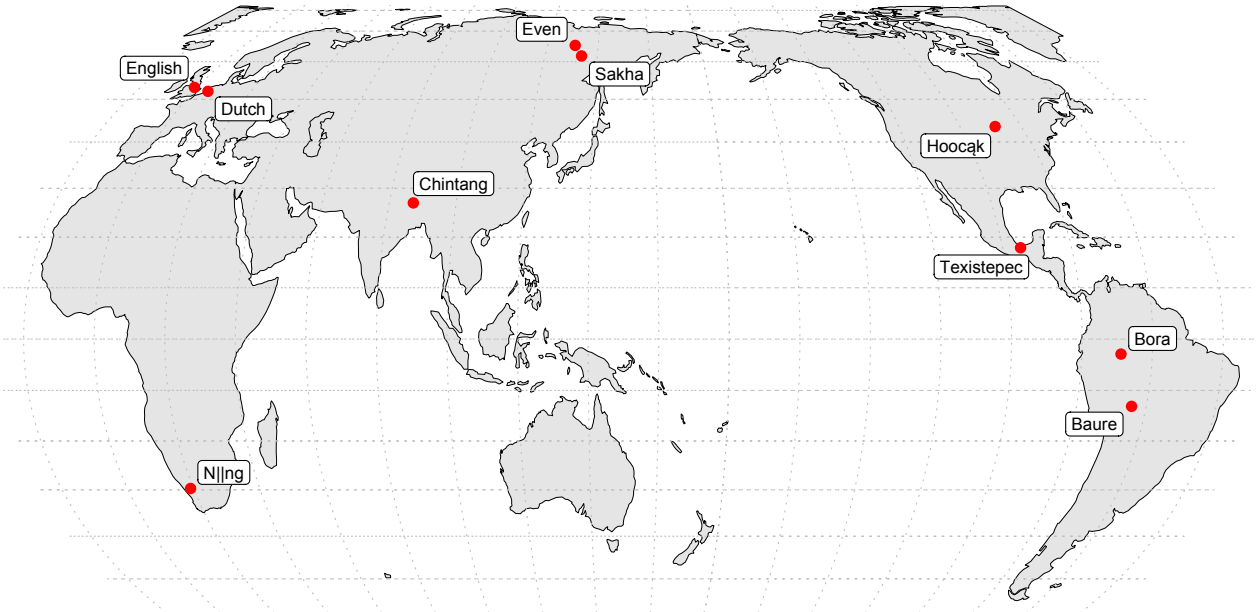

Figure 1. Location of languages included in the study on an Equal Earth map projection, retaining the relative size of areas (Šavrič, Patterson & Jenny 2019)

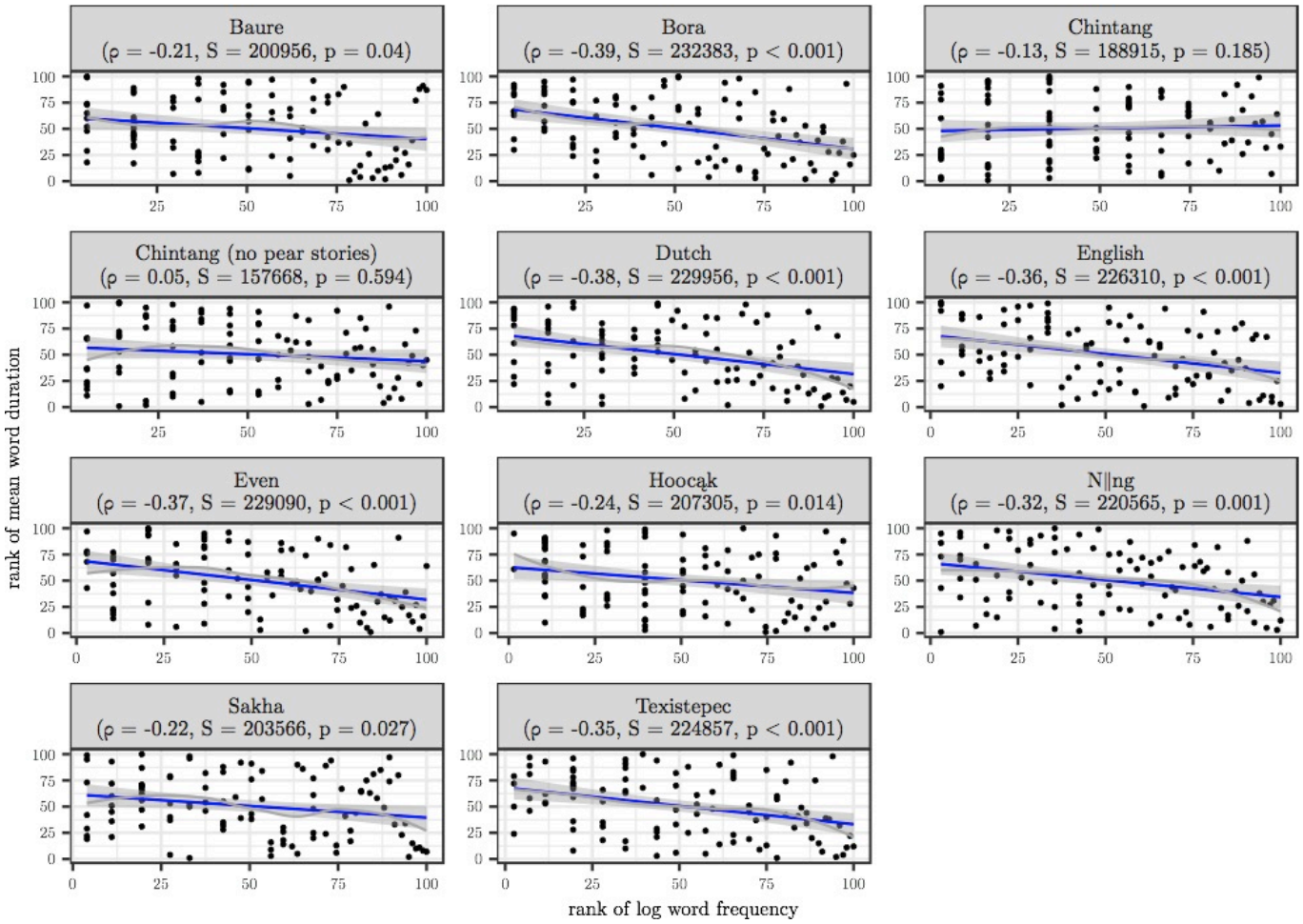

Figure 2. Bivariate correlation between word frequency and average word duration based on Spearman's nonparametric rank correlation coefficient. Plots are based on ranks of word frequency and duration. Correlation coefficients and test results are given beneath the strip titles where ρ indicates the direction and strength of the correlation, ranging from -1 to 1. Negative correlation coefficients indicate that higher word frequencies are associated with shorter average word durations. The p-values indicate whether the correlation coefficient is statistically significantly different from zero (i.e. no correlation) or not. Note that we here use ranks, and not relative frequency and duration as such, since rank correlation coefficients do not require that the data follow a normal distribution, which certainly is not the case for word frequencies (cf. Appendix C). Note that for sake of completeness, we report significance tests and p-values here, but we have strong reservations against interpreting these because they are based on simple correlations which cannot replace our actual statistical analyses (see Section 4.2).

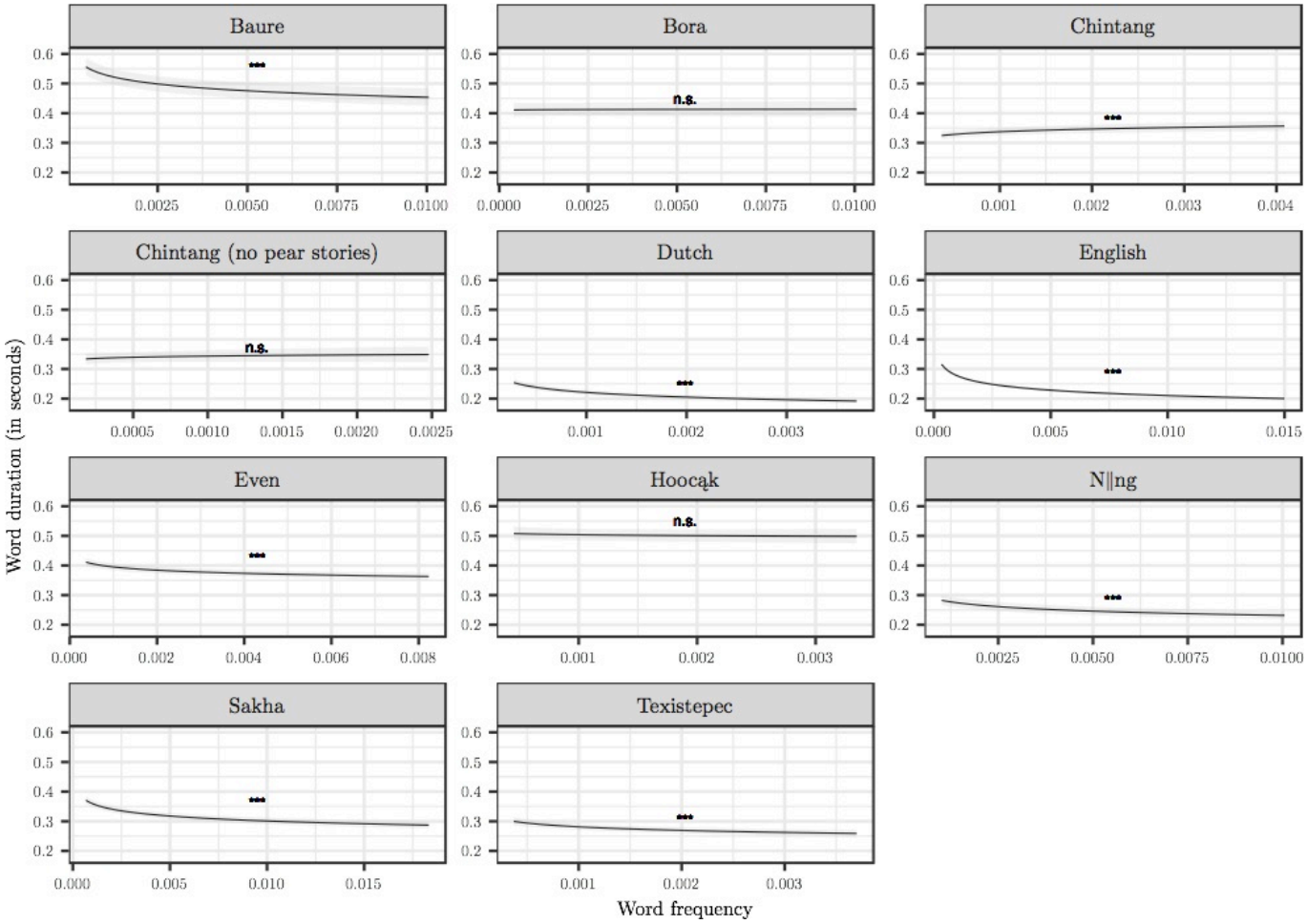

Figure 3. Display of the effect of word frequency on word duration in the individual languages, derived from the eleven mixed-effects models using the R library effects by Fox (2003). Note: n.s. $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

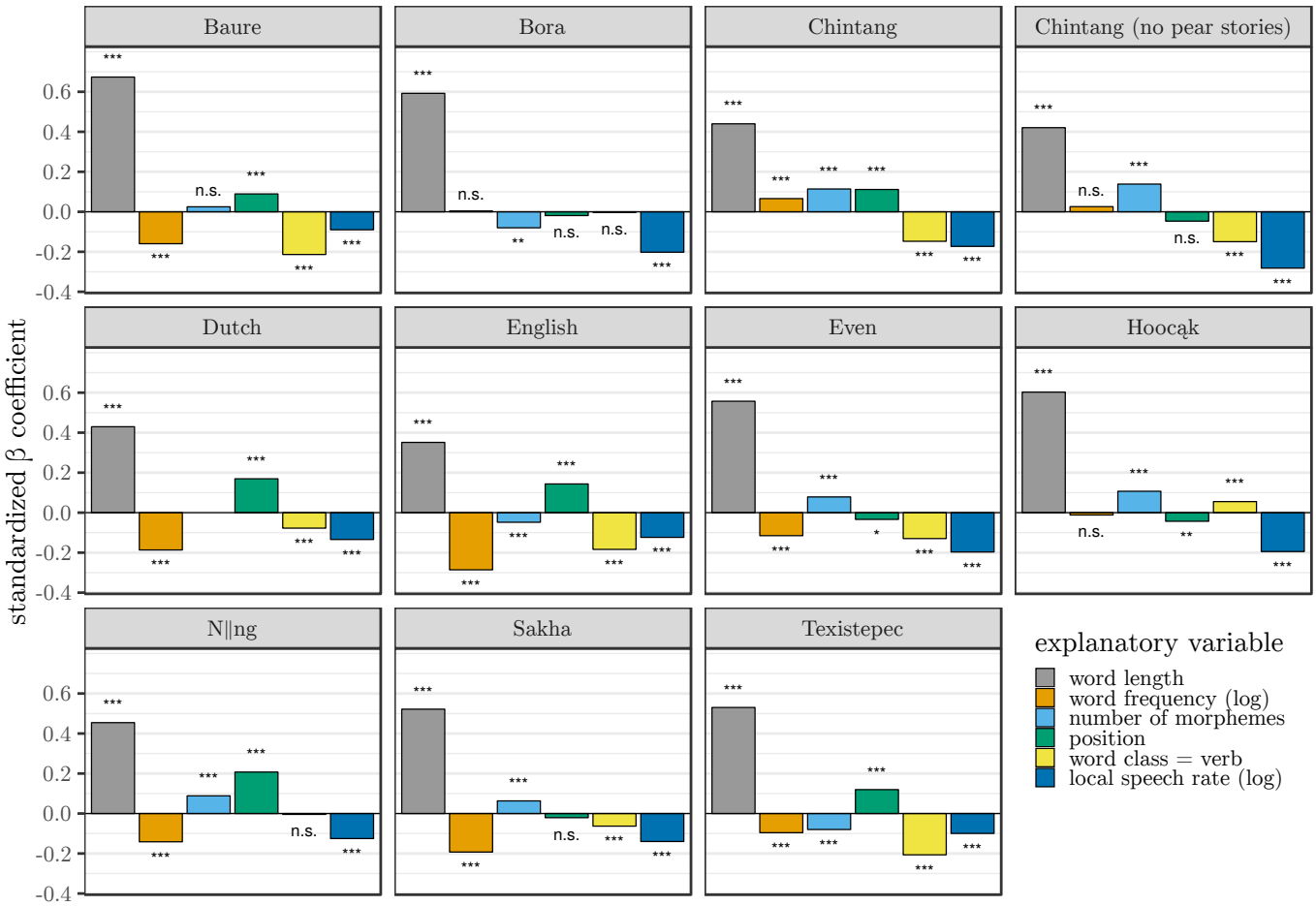

Figure 4. Comparison of standardized β coefficients for all six fixed effects in the mixed-effects models for the ten languages with significance stars based on likelihood ratio tests: n.s. $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1. Languages represented in the current study, with reference to the code assigned by Glottolog (Hammarström, Forkel & Haspelmath 2018), where further information on the languages can be found, and some descriptive parameters for the corpora used here. Note: "Unit length" stands for average length of annotation units in words (see section 2.2.)

Language				**Corpus**				
Language	**Glottocode**	**Family**	**Word order**	**Spea-kers**	**Texts**	**Words**	**Unit length**	**Reference**
Baure	baur1253	Arawakan	VSO	12	37	17 563	4.06	Danielsen et al. (2009)
Bora	bora1263	Boran	SOV	46	37	28 679	7.21	Seifart (2009)
Chintang	chhi1245	Sino-Tibetan	SOV	74	40	37 731	4.18	Bickel et al. (2011)
Dutch	dutc1256	Indo-	SOV	42	17	39 448	6.87	CGN-consortium
English	stan1293	Indo-	SVO	80	47	56 136	8.09	Godfrey & Holiman
Even	lamu1253	Tungusic	SOV	32	67	37 394	7.73	Pakendorf et al. (2010)
Hoocąk	hoch1243	Siouan	SOV	28	62	23 176	7.89	Hartmann (2013)
N‖ng	nuuu1241	!Ui-Taa	SVO	7	33	25 850	3.32	Güldemann et al.
Sakha	yaku1245	Turkic	SOV	25	16	31 139	7.93	Pakendorf (2007b)
Texistepec	texi1237	Mixe-	VSO	1	6	21 315	3.90	Wichmann (1996)
total/mean				***347***	***362***	***318 431***	***6.12***	

Table 2. Data selection and final data sets per language (number of word tokens)

language	all words	no disfluencies	nouns and verbs only	no ambiguous words (data set for word lists)	100 most frequent word types	word types manually excluded	no one-word utterances (final data set)
Baure	17 563	17 046	7 917	7 843	2 189	1 993	1 924
Bora	28 679	28 086	14 176	14 175	3 128	2 869	2 863
Chintang	37 731	36 916	15 242	15 239	2 959	2 816	2 734
Dutch	39 448	37 650	10 724	10 724	5 278	2 696	2 694
English	56 136	54 678	18 654	18 651	10 433	6 139	6 136
Even	37 394	34 340	18 299	18 299	4 393	3 602	3 582
Hoocąk	23 176	22 573	10 454	10 321	2 129	2 021	2 017
N‖ng	25 850	24 795	12 143	12 143	6 010	6 010	5 706
Sakha	31 139	29 972	16 581	16 581	4 422	3 354	3 347
Texistepec	21 315	21 202	8 858	8 715	3 160	2 788	2 738
Total	**318 431**	**307 258**	**133 048**	**132 691**	**44 101**	**34 288**	**33 741**
Chintang (no pear stories)	20 097	20 097	8 172	8 169	1 441	1 303	1 292

Table 3. Mean word length and morphological complexity per language (based on the complete corpora and words of all categories, corresponding to the word counts in the column *all words* in Table 2). Note that the Dutch corpus does not provide information about the morphological segmentation of words.

	word length in phones			word length in morphs		
Language	mean	std. dev.	median	mean	std. dev.	median
Baure	5.73	3.20	5	1.86	1.25	1
Bora	7.13	3.48	7	2.21	1.10	2
Chintang	5.14	2.65	5	1.81	1.27	1
Dutch	3.85	2.32	3	n/a	n/a	n/a
English	3.70	2.03	3	1.09	0.32	1
Even	5.79	2.78	5	1.91	1.05	2
Hoocąk	6.64	3.20	6	1.71	1.04	1
Nǁng	3.45	1.83	3	1.14	0.36	1
Texistepec	5.77	2.84	5	1.68	0.82	1
Sakha	5.14	2.56	4	1.81	0.98	2

Table 4. Summary of the word frequency effect on word duration based on the multivariate models. Statistical significance is calculated according to likelihood ratio tests (cf. the last four columns). Columns two to four give the number of observations per language as well as the overall portion of variation in word duration that the complete multivariate model is able to explain based on the fixed effects only (so-called marginal $R^2_{(m)}$) and fixed and random effects combined (so-called conditional $R^2_{(c)}$). Columns five and six provide information about the effect of word frequency in terms of this factor's coefficient (in the original scale), its standardized β coefficient. Columns seven to nine provide the percentage change in marginal and conditional variation explained (R^2) if the word frequency factor is omitted from the model ($\Delta R^2_{(m)}$% and $\Delta R^2_{(c)}$%) and, finally, the semi-partial correlation between word frequency and word duration (*s.-p.* R^2), that is, the amount of variation in word duration that can be uniquely attributed to the factor word frequency (also considering the other independent variables in the model at the same time).

Language	n	$R^2_{(m)}$	$R^2_{(c)}$	coefficient	β	$\Delta R^2_{(m)}$%	$\Delta R^2_{(c)}$%	s.-p. R^2	χ^2	*df*	*p*	
Baure	1 924	0.3239	0.3552	-0.0681	-0.1597	5.8441	1.3303	0.0244	46.96	1	<0.001	***
Bora	2 863	0.3556	0.4108	0.0020	0.0038	-0.0213	-0.0273	0.0000	0.05	1	0.826	n.s.
Chintang	2 734	0.2328	0.2977	0.0390	0.0660	2.0194	1.1119	0.0049	13.88	1	<0.001	***
Chintang (no pear stories)	1 292	0.2892	0.3297	0.0167	0.0257	0.2284	-0.0190	0.0008	1.09	1	0.296	n.s.
Dutch	2 694	0.4100	0.4650	-0.1079	-0.1866	7.9001	6.6115	0.0505	149.92	1	<0.001	***
English	6 136	0.3801	0.4133	-0.1197	-0.2858	19.5514	16.3843	0.1045	688.26	1	<0.001	***
Even	3 582	0.3711	0.4225	-0.0404	-0.1151	3.2049	2.7168	0.0189	70.85	1	<0.001	***
Hoocąk	2 017	0.5145	0.5682	-0.0094	-0.0118	0.0851	0.0040	0.0003	0.62	1	0.430	n.s.
N‖ng	5 706	0.3745	0.4333	-0.0857	-0.1414	3.9417	3.8114	0.0268	164.99	1	<0.001	***
Sakha	3 347	0.4773	0.5037	-0.0774	-0.1927	5.7201	5.7505	0.0503	178.64	1	<0.001	***
Texistepec	2 738	0.4313	0.4394	-0.0635	-0.0959	1.6365	1.3517	0.0124	34.36	1	<0.001	***

Table 5. Comparison of the direction and strength of the fixed effects in the models for the ten individual languages based on standardized β coefficients and percentage change in marginal R^2 when leaving out the fixed effect in question. Levels of statistical significance based on likelihood ratio tests comparing models with and without the fixed effect in question: n.s. $p \geq 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	word frequency (log)			word class = verb (vs. noun)			word length		
	β	$\Delta R^2_{(m)}$%		β	$\Delta R^2_{(m)}$%		β	$\Delta R^2_{(m)}$%	
Baure	-0.1597	5.8441	***	-0.2136	5.9740	***	0.6737	34.4366	***
Bora	0.0038	-0.0213	n.s.	-0.0026	-0.0151	n.s.	0.5923	38.0563	***
Chintang	0.0660	2.0194	***	-0.1475	5.1095	***	0.4399	56.2432	***
Chintang (no pear stories)	0.0257	0.2284	n.s.	-0.1495	4.1954	***	0.4206	43.3012	***
Dutch	-0.1866	7.9001	***	-0.0775	1.9876	***	0.4297	41.8685	***
English	-0.2858	19.5514	***	-0.1837	7.4276	***	0.3514	21.7815	***
Even	-0.1151	3.2049	***	-0.1296	2.8687	***	0.5572	67.0652	***
Hoocąk	-0.0118	0.0851	n.s.	0.0553	0.7170	***	0.6026	44.8522	***
N‖ng	-0.1414	3.9417	***	-0.0045	-0.0080	n.s.	0.4544	37.8865	***
Sakha	-0.1927	5.7201	***	-0.0631	0.5802	***	0.5218	29.5126	***
Texistepec	-0.0959	1.6365	***	-0.2066	4.9379	***	0.5303	36.9072	***

	number of morphs			position			local speech rate (log)		
	β	$\Delta R^2_{(m)}$%		β	$\Delta R^2_{(m)}$%		β	$\Delta R^2_{(m)}$%	
Baure	0.0250	-0.1241	n.s.	0.0893	1.8783	***	-0.0898	1.8094	***
Bora	-0.0801	0.4307	**	-0.0190	0.0519	n.s.	-0.2021	12.3645	***
Chintang	0.1140	2.7733	***	0.1116	4.7834	***	-0.1729	12.7666	***
Chintang (no pear stories)	0.1385	2.9264	***	-0.0469	0.8022	n.s.	-0.2817	26.6149	***
Dutch				0.1693	7.5910	***	-0.1342	7.6702	***
English	-0.0481	0.4548	***	0.1436	4.9897	***	-0.1236	5.2302	***
Even	0.0786	0.8280	***	-0.0337	0.1680	*	-0.1967	13.2355	***
Hoocąk	0.1070	1.5440	***	-0.0431	0.4938	**	-0.1945	8.0100	***
N‖ng	0.0885	1.5692	***	0.2073	11.7731	***	-0.1252	6.5637	***
Sakha	0.0631	0.4411	***	-0.0208	0.0944	n.s.	-0.1400	5.5948	***
Texistepec	-0.0799	0.5960	***	0.1195	3.0774	***	-0.0995	2.2384	***

Author e-mail address:
frank.seifart@berlin.de