By force of nature, every bit of spoken language is produced at a particular speed. However, this speed is not constant—speakers usually speed up and slow down. Variation in speech rate is influenced by a complex combination of factors, including the frequency and predictability of words, their information status, and their position within an utterance. Here, we use speech rate as an index of word-planning effort and focus on the time window during which speakers prepare the production of words from the two major lexical classes, nouns and verbs. We show that, when naturalistic speech is sampled from languages all over the world, there is a robust cross-linguistic tendency for slower speech before nouns compared with verbs, both in terms of slower articulation and more pauses. We attribute this slowdown effect to the increased amount of planning that nouns require compared with verbs. Unlike verbs, nouns can typically only be used when they represent new or unexpected information; otherwise, they have to be replaced by pronouns or be omitted. These conditions on noun use appear to outweigh potential advantages stemming from differences in internal complexity between nouns and verbs. Our findings suggest that, beneath the staggering diversity of grammatical structures and cultural settings, there are robust universals of language processing that are intimately tied to how speakers manage referential information when they communicate with one another.

H

human language in its most widespread form (i.e., in spontaneously spoken interactions) is locked in one-dimensional time. This was recognized by the founding father of modern linguistics, Ferdinand de Saussure, as one of the two fundamental principles of the linguistic sign, the other one being its arbitrary nature (1, 2). An unresolved question is which aspects of local variation in speech rate are universal (3, 4), which vary across languages and cultures (5), and which vary across individuals (6). For example, marking the end of utterances by slowing down speech is cross-linguistically common, but its implementation is language-specific (7). Good candidates for truly universal temporal features are the relatively fast pronunciations of frequent, and thus predictable, words (8) and second mentions of words (9). This speedup is argued to result from automated articulation (4) and has been suggested to contribute to efficient communication by spreading information more evenly across the speech signal (10, 11). Frequency effects also explain why function words, such as articles, prepositions, and pronouns, are pronounced faster than the less frequently occurring content words, such as nouns and verbs (12).

An aspect of speech rate that has received less attention is the local speech rate during the planning, rather than the actual pronunciation, of words. Speed variation before the articulatory onset of a word can provide key evidence for cognitive processes. For example, speakers have been found to slow down their speech rate before complex, infrequent, or novel words (13, 14), a finding that is consistent with the slowdown in lexical access speed that such words trigger in picture naming and related tasks (15–17). Here, we investigate speech rate in word-planning windows in naturalistic speech from nine languages to assess differences in the two major word classes usually found in languages: nouns and verbs. To our knowledge, the relative speedup or slowdown of speech preceding nouns versus verbs has never been directly studied. Related measures like response times in picture-naming experiments suggest that nouns require less planning time than verbs (18, 19). This is attributed to increased planning costs of verbs because of their relative grammatical and semantic complexity and their links with other elements in the clause, for example, subjects and objects. While it is unclear to what extent the planning demands of a word leave traces in the speed of its own articulation (20), these findings are potentially in conflict with studies suggesting slower rates for nouns than verbs in English noun/verb homophones (such as a fly vs. to fly) (21).

A factor that has been neglected in this research is how referential information is managed in connected, interactive speech. In running speech, the choice between referring expressions (e.g.,
between a noun like the teacher and a pronoun like she) is subject to complex, multidimensional decision procedures which involve various internal and audience-oriented processing mechanisms (22–25) and are shaped both by general pragmatic principles (26, 27) and by language-specific and cultural factors (22, 28, 29). What emerges as a cross-linguistically stable pattern, however, is that the use of nouns typically signals the newness of a referent (e.g., a new person or object introduced into the discourse), a new temporal or local setting, the need to disambiguate between different referents, or a shift in discourse topic or perspective (30). In all other contexts, pronouns (I saw the teacher, he [the teacher] was tired) or gaps (The teacher came in and [the teacher] sat down) are highly preferred (31, 32). Verbs are fundamentally different in this regard: Even if the same actions or states are referred to repeatedly, a verb is typically still necessary to form a complete sentence. In line with this, languages do not generally have “proverbs” to systematically replace verbs as pronouns do for nouns. While the generic nature of some verbs (e.g., to do) occasionally brings them close to such a function, this is usually confined to highly constrained syntactic contexts (as in Susan drank wine and so did Mary). Similarly, verbs can occasionally be gapped in some languages (Susan drank wine and Mary beer), but this is again subject to special syntactic constraints. In general, the use of verbs is thus the default option, regardless of the information status of the actions or states referred to, while the use of nouns is a marked option that is felicitous only in contexts of information novelty, disambiguation needs, or topic and perspective shifts. Given these additional constraints on the use of nouns, their use should correlate with a higher planning cost, slowing down speech before the noun.

Here, we aim to settle not only the question of the direction of the effect of subsequent noun versus verb use on speech rate, but also its universality. For this we use time-aligned corpora of naturalistic speech from multimedia language documentations (33). To ensure linguistic and cultural diversity, we chose a set of such corpora from languages spoken in the Amazonian rainforest (Bora and Baure), Mexico (Texistepec), the North American Midwest (Hoocąk), Siberia (Even), the Himalayas (Chintang), and the Kalahari Desert (Nilng) (Fig. 1). These seven corpora were compiled during on-site fieldwork over the past 25 y and were transcribed, translated, and annotated with word class tags by experts on the languages in collaboration with native speakers. They document naturalistic speech of various genres, including narratives, descriptive texts, and conversations, that were recorded in their original, interactive settings, such as the recording of a Bora myth illustrated in Fig. 2. While the genres covered by the corpora are diverse, all data are comparable in that they document speech which is spontaneously produced, not read out or memorized, even if texts stem from local oral traditions. We additionally used relevant sections of published corpora of spoken Dutch and English, which likewise document naturalistic spoken language annotated for word class by experts.

To assess the effects of subsequent noun versus verb use on speech rate, we used the word-class category of the lexical root contained in a word, as identified by 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). This 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 (18, 34). We investigated speedup versus slowdown effects of nouns versus verbs in time windows of ~500 ms preceding their onset (Materials and Methods and Fig. 2). This window size was set following picture- and action-naming studies that have shown that planning a single content word takes around 600 ms (35). Slowing down speech can have two independent effects (36), which we investigated in two separate studies: (i) slower articulation of words, measured as phonological segments (approximated by orthographic characters) per second (37) for all words within the time window preceding a noun or verb, and (ii) higher probability of pauses within such windows, as indicated by the presence of at least one interval ≥150 ms without articulation or with articulation of fillers only (such as English umm) (Materials and Methods). We analyzed both measures with generalized linear mixed-effects models with the word class (noun vs. verb) of the target word as the main predictor of interest. We controlled for potential slowdown at the end of utterances by including the target word’s position within the utterance, as well as the target word’s length. Our models furthermore took into account random effects caused by idiosyncrasies of individual speakers, recording sessions, and individual word forms. Inclusion of word forms takes care of the expected speedup associated with frequent and predictable items, since frequency and predictability are properties of individual word forms (38, 39) (Materials and Methods). Modeling the entire dataset revealed a significant interaction between language and the effect of word class, and we therefore fitted individual but comparable models to each language separately.

Results and Discussion

Results are summarized in the effect displays in Fig. 3, showing that all nine languages exhibit a significant slowdown before nouns compared with verbs with respect to at least one of our two ways of measuring slowdown. Only one language (English)
exhibits a significant slowdown before verbs, and only when measured in terms of pause probability (see SI Appendix, Supplementary Text for details). The overall tendency for slowdown before nouns is striking because the culturally and linguistically vastly diverse populations in our sample display remarkable differences in many respects; for example, in overall speed and the range of variation (Fig. 3 and SI Appendix, Table S6). For instance, Hoocaka speakers articulate more slowly and pause more often in the context of both nouns and verbs than Dutch speakers do. Language or culture-specific facts may also mask the observed effect in individual studies for individual languages. For instance, Nling words are so short (on average 4.61 segments per word) that there is little room for differences in articulation rate within words. We have presently no explanation for the exceptional behavior of English regarding pauses, except for speculating that English noun planning might be “easier” because the gap option (as opposed to the pronoun option) is far less common than in the other languages, reducing choice efforts. Another possibility is that our English corpus is based on telephone rather than face-to-face interactions, but evidence so far suggests that speakers are not strongly influenced by the visual presence of listeners in reference production (9, 23). Whatever the reason, this result highlights the need for a diverse sample, such as that represented here, including languages other than English, which has been found to be exceptional in other studies also (40).

The overall results, based on models with data from all nine languages taken together, show that, across our diverse sample, the slowdown effect before nouns prevails: Regarding articulation, the effect is small but robust, causing around 3.5% slower articulation rate before nouns than before verbs, despite strong variation overall and a few exceptions found in specific utterance positions in individual languages (see SI Appendix, Supplementary Text for details). Depending on pauses, across all nine languages, the probability of pauses before nouns is about 60% greater than before verbs, and, in the majority of languages, the odds of pauses before nouns are about twice as high as before verbs (see SI Appendix, Supplementary Text for details). Compared with other factors, the effect of word class is also surprisingly strong: In statistical models of all our data taken together, this effect is about two times stronger than the effect of a target word’s length and more than eight times stronger than the effect of its position within the utterance (SI Appendix, Tables S9, S20, S33, and S44).

**Conclusion**

Our results from naturalistic speech contradict experimental studies showing faster planning of nouns (18, 19) and thus suggest that the effect of referential information management overrides potential effects of higher processing costs of verbs. As such, these results resonate with earlier findings of cross-linguistic parallels in the timing of turn taking (5, 41) and point to strong universals of language processing that are grounded in how humans manage information. But our present findings indicate that speech rate variation is universally constrained also at a fine-grained level, within turns and depending on which kinds of content words are used: Pragmatic principles of noun use and the slowdown associated with new information converge to create a uniform pattern of speech rate variation across diverse languages and cultures. Our finding has several implications. First, models of language processing need to more systematically incorporate aspects of information management in interactive speech (41–43). Second, while speech rate in corpora is mostly studied in terms of the articulation of a word, speech rate variation before words of different types is a measure with great potential to gain insights into the mechanisms of language production. Third, naturalistic corpus studies on widely diverse languages allow detection of signals that
do not suffer from the sampling bias in much of current theorizing about language and speech (33, 44). Most such work is still largely based on educated speakers of a small number of mostly Western European languages, and it remains unclear whether findings generalize beyond this (40, 45, 46). Finally, by revealing patterns linked to specific word classes, our findings open avenues for explaining how grammars are shaped through the long-term effects of fast pronunciation, such as phonological reduction (47) and the emergence of grammatical markers (4). In particular, slower speech and more pauses before nouns entail a lower likelihood of contraction of independent words. This explains the fact that cross-linguistically fewer function words become fused as preprefixes to nouns than as preprefixes to verbs, a fact so far little understood (48).

Materials and Methods

Corpus Characteristics. All data were collected, transcribed, and annotated by experts during on-site fieldwork on the languages (see Fig. 1 and SI Appendix, Table S1 for details). The transcriptions in the language documentation corpora use established orthographies or orthographies developed during fieldwork and in consultation with native speakers. All of these orthographies are fairly “shallow” by generally applying one-to-one mappings of phonological segments to orthographic characters. In our analyses, we used these orthographic characters as proxies for phonological segments. This is further justified by the fact that correlations between word length in orthographic characters and word length in phonological segments are extremely high, even for languages with relatively deep orthographies, such as English and Dutch (49). Nevertheless, we control for multicharacter representations, such as English /ʃp/, /ʃ/, or Dutch /œ/, /œ/ in the models below. All data were manually time-aligned during transcription at the level of annotation units, as in Fig. 2) and exclude windows at the beginning of utterances; for example, before stop consonants, as previous studies have shown (53–55).

Algorithm for Determining Preword Windows. We use the term “preword window” for the immediate context before individual (target) words in which we measured speech rate. Based on the time frame known to be relevant for word planning (31), we chose a preword window size of 500 ms; that is, relatively local windows containing at most a few (context) words. We only consider preword windows occurring within utterances (i.e., annotation units, as in Fig. 2) and exclude windows at the beginning of utterances. This ensures that changes in articulation rate and pauses can be attributed to the target word and not to phenomena occurring outside of utterances, for example, to turn-taking constraints, parallel nonlinguistic activities, conceptual planning of an utterance as a whole, etc.

If there is silence within an annotation unit at 500 ms before the onset of the target word, the preword window is exactly 500 ms long. The size of preword windows is adjusted when there are context words that are only partially included in the 500-ms window because we consider articulation rate information from whole words, not parts of words. To determine which of these context words should be included in the preword window, we define a word’s midpoint as the word start time + word end time)/2. If a context word’s midpoint occurs within a 500-ms preword window before the target word, the whole word is included in this preword window. If this word’s start time is outside the 500-ms preword window, the preword window size is enlarged to include the whole context word. In such cases, the preword window is slightly larger than 500 ms. If the midpoint of a context word preceding the target word is outside the 500-ms window but its endpoint is within the 500-ms window, this context word is not counted as part of the preword window. Instead, the start time of the preword window is set to the end time of this excluded context word, and the window is shortened. The preword window in such cases may still contain pauses as well as words.
of which the midpoints fall inside the 500-ms interval. Fig. 2 illustrates both shortened and enlarged context windows. A preword window can also be shorter than the word class it occurred within (SI Appendix, Table S3), since we do not consider pauses between annotation units. If a target word has only one or two words before it, it can be the case that the 500-ms preword window extends to before the first word. In such cases, the preword window start time is set to the start time of the first word, and the length of the preword window is shortened accordingly. The mean length of preword windows is 456 ms (SD 164 ms) and thus slightly shorter than 500 ms, but roughly comparable for all languages (SI Appendix, Table S3). Our algorithm of defining preword windows resulted in variously sized windows. However, window length does not systematically covary with parts of speech (SI Appendix, Table S3), and this justifies averaging the length per window when computing articulation rate. (We also considered explicitly modeling the window length variation, but the truncations induced by our algorithm would have necessitated overly complex and nonstandard models, which were, moreover, not equally applicable to the analysis of articulation rate and the analysis of pause probability.)

Verbs and Auxiliaries. Some languages, like English, distinguish a category of functional verbal elements (auxiliaries, AUX) from ordinary content verbs (V). We excluded all known auxiliaries from the analyses reported here, in line with our semantically based identification of verbs (see main text). This exclusion is based on the different word class annotations in our corpora of English, Even, Hoocąk, and Texistepec. The languages Baure, Bora, Chintang, and Nling do not have any auxiliaries. However, auxiliaries are not annotated differently from content verbs in the corpus we used for Dutch, despite the strong similarity with English. To make sure that excluding auxiliaries in some languages but not others did not lead to spurious differences between languages, we also carried out alternative analyses in which all verbal target words, including auxiliaries, were included in the category of verbs. The results of these alternative analyses are summarized in SI Appendix, Supplementary Text and fully converge with the results in Fig. 3.

Analyses of Articulation Rate. For the analyses of articulation rate, we discarded all preword windows that contained disfluencies (filled pauses such as uh or um or false starts) or only consisted of a silent pause (SI Appendix, Tables S6 and S7) or filled pause. We therefore also included preword windows that contain only pauses as well as preword windows that contain a disfluency, such as a filled pause (hesitation) or a false start (SI Appendix, Tables S26 and S27). Like in the articulation rate study, our main analysis excluded auxiliaries from verbs while an additional alternative analysis included auxiliaries as verbs in corpora where auxiliaries were identified as such (SI Appendix, Supplementary Text and Tables S43–S48).

We used a Boolean variable to code the existence of a (silently or filled) pause in a given preword context window. We defined silent pauses as periods of silence between two words (uttered by the same speaker as part of one utterance) that were at least 150 ms long. SI Appendix, Tables S28–S31 give descriptive statistics on pause probability, as well as confidence intervals estimated with an exact binomial method. We used the same predictor variables and random factors as in the study on articulation rate but now with a pause on likelihood ratio tests to compare the final model to alternative models where the relevant factor was dropped. To control the false discovery rate (FDR), we also adjusted P values based on the Benjamini and Hochberg (BH) method (60), using R's p.adjust function. The effect plots in Fig. 3A were produced using the effect function from the R library effects (61). They show significance based on adjusted P values (BH). Models based on an alternative dataset that included the known auxiliaries of English, Even, Hoocąk and Texistepec in the category of verbs (SI Appendix, Tables S19–S24) followed exactly the same procedure and are not shown here. To better assess effect sizes, we furthermore calculated the predicted articulation rate difference between nouns and verbs, distinguishing between positions at the beginning and at the end of utterances (SI Appendix, Table S25).

Analysis of Pause Probability. We investigated the probability that a preword window before a noun versus before a verb contained at least one silent and/or filled pause. We therefore also included preword windows that contain only pauses as well as preword windows that contain a disfluency, such as a filled pause (hesitation) or a false start (SI Appendix, Tables S26 and S27). Like in the articulation rate study, our main analysis excluded auxiliaries from verbs while an additional alternative analysis included auxiliaries as verbs in corpora where auxiliaries were identified as such (SI Appendix, Supplementary Text and Tables S43–S48).

We used a Boolean variable to code the existence of a (silently or filled) pause in a given preword context window. We defined silent pauses as periods of silence between two words (uttered by the same speaker as part of one utterance) that were at least 150 ms long. SI Appendix, Tables S28–S31 give descriptive statistics on pause probability, as well as confidence intervals estimated with an exact binomial method. We used the same predictor variables and random factors as in the study on articulation rate but now with a pause on likelihood ratio tests to compare the final model to alternative models where the relevant factor was dropped. To control the false discovery rate (FDR), we also adjusted P values based on the Benjamini and Hochberg (BH) method (60), using R's p.adjust function. The effect plots in Fig. 3A were produced using the effect function from the R library effects (61). They show significance based on adjusted P values (BH). Models based on an alternative dataset that included the known auxiliaries of English, Even, Hoocąk and Texistepec in the category of verbs (SI Appendix, Tables S19–S24) followed exactly the same procedure and are not shown here. To better assess effect sizes, we furthermore calculated the predicted articulation rate difference between nouns and verbs, distinguishing between positions at the beginning and at the end of utterances (SI Appendix, Table S25).

Data Availability. The complete datasets used in this study are available at https://figshare.com/s/0B5E090da7d2b5051df4e.

ACKNOWLEDGMENTS. We thank all native speakers that provided data and all assistants that helped annotate the data. We acknowledge the comments of Damian E. Blasi, Sebastian Sauppe, Volker Dellwo, and Sabine Stoll. The research of F.S. and J.S. was supported by grants from the Volkswagen Foundation's Dokumentationsprojekt Bedrohter Sprachen (DoBoS) program (80 110, 83 522, 86 292, and 89 550) and the Max Planck Institute for Evolutionary Anthropology; the research of S.W. was supported by a European Research Council (ERC) Advanced Grant [MesAndLimGk, Grant 295918, Principal Investigator: W. Adelaar] and by a subsidy of the Russian Government to support the Programme of Competitive Development of Kazan Federal University. B.P. is grateful to Laboratoires d’Excellence (LABEX) Advanced...
Seifert et al.

Studies on Language Complexity (ASLAN) (ANR-10-LABX-0081) of the 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). The research of B.B. was supported by Swiss National Science Foundation Grant CRSII1_160739 (“Linguistic Morphology in Time and Space”).

1. de Saussure F (1916) Cours de linguistique générale (Payot, Lausanne, Switzerland).
2. Blasi DE, Wichmann S, Hammarström H, Stadler PF, Christiansen MH (2016) Sound-meaning association biases evidenced across thousands of languages. Proc Natl Acad Sci USA 113:10818–10823.
3. Vaisièrè J (1983) Language-independent prosodic features. Prosody: Models and Measurements, Springer Series in Language and Communication, eds Cutler A, Ladd DR (Springer, Heidelberg), pp 53–66.
4. Bybee JL (2010) Language, Usage and Cognition (Cambridge Univ Press, Cambridge, MA).
5. Sivert S, et al. (2009) Universals and cultural variation in turn-taking in conversation. Proc Natl Acad Sci USA 106:10587–10592.
6. Dello V, Leeman A, Kolly M-J (2015) Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. J Acoust Soc Am 137:1513–1528.
7. Ordin M, Polyanskaya L, Laka I, Nespor M (2017) Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. Mem Cogn 45: 863–876.
8. Gahl S (2008) Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. Language 84:874–946.
9. Bard EG, et al. (2000) Controlling the intelligibility of referring expressions in dialogue. J Mem Lang 42:1–22.
10. Ayliet M, Tunk A (2004) The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. Lang Speech 47:31–56.
11. Jaeger TF (2010) Redundancy and reduction: Speakers manage syntactic information density. Cogn Psychol 61:23–60.
12. Bell A, Brenier JM, Gregory M, Girand C, Jurafsky D (2009) Probadability effects on durations of content and function words in conversational English. J Mem Lang 60: 92–111.
13. Fox Tree JL, Clark HH (1997) Pronouncing the “th” as “thee” to signal problems in understanding. Cognition 62:135–157.
14. de Jong NH (2016) Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. Int Rev Appl Linguist Lang Teach 54:113–132.
15. Bates E, et al. (2003) Timed picture naming in seven languages. Psychon Bull Rev 10: 171–178.
16. Jescheniak JD, Levelt WJM (1994) Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. J Exp Psychol Learn Mem Cogn 20:824–843.
17. Levelt WJ, Roelofs A, Meyer AS (1999) A theory of lexical access in speech production. Behav Brain Sci 22:1–38, discussion 38–75.
18. Vigliocco G, Vinson DP, Druss K, Barber H, Cappa SF (2011) Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. Neurosci Biobehav Rev 35:407–426.
19. Szekely A, et al. (2005) Timed action and object naming. Cortex 41:7–25.
20. Jaeger TF, Buz E (2017) Signal reduction and linguistic encoding.
21. Indo-Aryan Languages and Linguistic Typology: Implications for Cognitive Neuroscience. Cortex 58:27–51.
22. Kibrik AA, Khudyakova MV, Dobrov GB, Linnik A, Zalmanov DA (2016) Referential processes.
23. Ariel M (2014) Accessing Noun-Phrase Antecedents
24. Ariel M (2014) The Handbook of Discourse: A Quantitative Cross-Language Study
25. Gatt A, Krahmer E, van Deemter K, van Gompel RPG (2014) Models and empirical data for the production of referring expressions.
26. Levinson SC (2000) Presumptive Meanings. The Theory of Generalized Conversational Implicature (MIT Press, Cambridge, MA).
27. Gross BJ, Joshi AK, Weinstein S (1995) Centering: A framework for modeling the local coherence of discourse. Comput Linguist 21:202–225.
28. Bickel B (2003) Referential density in discourse and syntactic typology. Language 79: 708–736.
29. Stolz S, Bickel B (2009) How deep are differences in referential density? Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac któin, eds Guo J, et al. (Psychology Press, New York), pp 543–555.
30. Fox B (1987) Discourse Structure and Anaphora: Written and Conversational English (Cambridge Univ Press, Cambridge, UK).
31. Kim G (1983) Topic continuity in discourse: An introduction. Topic Continuity in Discourse: A Quantitative Cross-Language Study, ed Givón T (John Benjamins, Amsterdam, The Netherlands), pp 1–41.
32. Chafe WL (1994) Discourse, Conscience, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing (Univ of Chicago Press, Chicago, IL).
33. Seifert F (2012) The threefold potential of language documentation. Potentials of Language Documentation: Methods, Analyzes, and Utilization, Language Documentation & Conservation Special Publication, eds Seifert F, et al. (Univ of Hawai’i Press, Manoa, HI), pp 1–6.
34. Kemmerer D (2014) Word classes in the human brain: Implications of linguistic typology for cognitive neuroscience. Cortex 58:27–51.
35. Indefrey P, Levelt WJM (2004) The spatial and temporal signatures of word production components. Cognition 92:101–144.
36. Bosker HR, Plat F, Vente H, Sanders T, de Jong NH (2013) What makes speech sound fluent? The contributions of pauses, speed, and repairs. Lang Test 30:159–175.
37. Koreman J (2006) Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. J Acoust Soc Am 119:582–596.
38. Seyfarth S (2014) Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. Cognition 133:140–155.
39. Sökuthy M, Hay J (2017) Changing word usage predicts changing word durations in New Zealand English. Cognition 166:298–313.
40. Heinrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? Behav Brain Sci 33:61–83, discussion 83–135.
41. Levinson SC (2016) Turn-taking in human communication-Origins and implications for language processing. Trends Cogn Sci 20:6–14.
42. Levelt WJM (1989) Speaking: From Intention to Articulation (MIT Press, Cambridge, MA).
43. Hagop S, et al. (2015) Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. Lang Cogn Neurosci 30:1099–1032.
44. Zeff G (1949) Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology (Addison-Wesley Press, Cambridge, MA).
45. Himmelmann NP (2014) Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. Language 90:927–960.
46. Plantadosi ST, Tily H, Gibson E (2011) Word lengths are optimized for efficient communication. Proc Natl Acad Sci USA 108:3526–3529.
47. Sacks H, Schegloff EA, Jefferon G (1974) A simplest systems for the organization of turn-taking for conversation. Language 50:696–735.
48. Sturk J, Schiel F, Seifert F (2014) Untrained forced aligned of transcriptions and audio for language documentation corpora using WebMAUS. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), eds Calzolari N, et al. (European Language Resources Association, Reykjavik, Iceland), pp 3940–3947.
49. Kehler T, et al. (2016) BAS speech science web services—An update of current developments. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), eds Calzolari N, et al. (European Language Resources Association, Paris, France), pp 3880–3885.
50. Goldman-Eisler F (1968) Psycholinguistics: Experiments in Spontaneous Speech (Academic, London).
51. Hekje AE, Kowal S, O’Connell DC (1983) The trouble with “articulatory” pauses. Lang Speech 26:203–214.
52. Campione E, Vérinon J (2002) A large-scale multilingual study of silent pause duration. Speech Prosody 2002:193–202.
53. Lehiste I (1970) Suprasegmentals (MIT Press, Cambridge, MA).
54. Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects models modeling with crossed random effects for subjects and items. J Mem Lang 59:390–412.
55. R Core Team (2018) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna), Version 3.4.4. Available at https://www.r-project.org/. Accessed March 15, 2018.
56. Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67.
57. Benjamin V, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300.
58. Fox J (2003) Effect displays in R for generalised linear models. J Stat Softw 8:1–7.
59. Subbiah M, Rajeswaran V (2017) Proportion: A comprehensive R package for inference on binomial proportion and Bayesian computations. SoftwareX 6: 36–41.