Nouns slow down speech across structurally and culturally diverse languages

Frank Seifart, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nivja de Jong, Balthasar Bickel

Nouns slow down speech across structurally and culturally diverse languages

Frank Seifart[a,b,c,1], Jan Strunk[b], Swintha Danielsen[d], Iren Hartmann[d], Brigitte Pakendorf[c], Søren Wichmann[e,f], Alena Witzlack-Makarevich[g], Nivja H. de Jong[e,h], and Balthasar Bickel[i]

[a]Amsterdam Center for Language and Communication, University of Amsterdam, 1012 VT Amsterdam, The Netherlands; [b]Institut für Linguistik, University of Cologne, 50923 Cologne, Germany; [c]Laboratoire Dynamique du Langage, UMR5596, CNRS & Université de Lyon, 69007 Lyon, France; [d]Institut für Linguistik, University of Leipzig, D-04107 Leipzig, Germany; [e]Leiden University Centre for Linguistics, Leiden University, 2311 BX Leiden, The Netherlands; [f]Laboratory of Quantitative Linguistics, Kazan Federal University, 420000 Kazan, Russia; [g]Abteilung für Allgemeine Sprachwissenschaft, Institute for Scandinavian Studies, Frisian Studies, and General Linguistics, Kiel University, 24098 Kiel, Germany; [h]Leiden University Graduate School of Teaching, Leiden University, 2333 BN Leiden, The Netherlands; and [i]Department of Comparative Linguistics, University of Zurich, 8032 Zurich, Switzerland

Edited by Willem J. M. Levelt, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, and approved April 25, 2018 (received for review January 12, 2018)

By force of nature, every bit of spoken language is produced at a particular speed. However, this speed is not constant—speakers regularly 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.

speech rate | nouns | language universals | word planning | language processing

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.,

Significance

When we speak, we unconsciously pronounce some words more slowly than others and sometimes pause. Such slowdown effects provide key evidence for human cognitive processes, reflecting increased planning load in speech production. Here, we study naturalistic speech from linguistically and culturally diverse populations from around the world. We show a robust tendency for slower speech before nouns as compared with verbs. Even though verbs may be more complex than nouns, nouns thus appear to require more planning, probably due to the new information they usually represent. This finding points to strong universals in how humans process language and manage referential information when communicating linguistically.

Author contributions: F.S., J.S., and B.B. designed research; J.S. performed research; F.S., J.S., S.D., I.H., B.P., S.W., A.W.-M., and B.B. analyzed data; F.S., N.H.d.J., and B.B. wrote the paper with input from all other authors; and J.S. produced the *SI Appendix* with input from all other authors.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The complete datasets used in this study are available at https://figshare.com/s/085b09d7d82b5501df4e.

[1]To whom correspondence should be addressed. Email: f.c.seifart@uva.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1800708115/-/DCSupplemental.

Published online May 14, 2018.

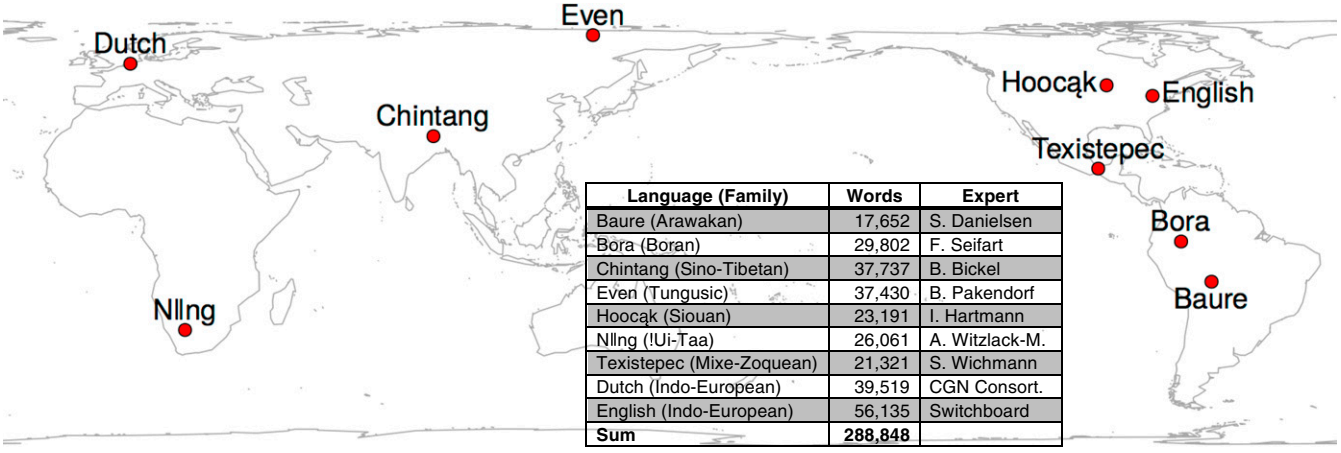
between a noun like *the teacher* and a pronoun like *she*) is subject to complex, multidimensional decision procedures which involve various internal and audience-oriented processing mechanisms (22–25) and are shaped both by general pragmatic principles (26, 27) and by language-specific and cultural 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 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 "pro-verbs" 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 (Nǁng) (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 *uhm*) (*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)



Language (Family)	Words	Expert
Baure (Arawakan)	17,652	S. Danielsen
Bora (Boran)	29,802	F. Seifart
Chintang (Sino-Tibetan)	37,737	B. Bickel
Even (Tungusic)	37,430	B. Pakendorf
Hoocąk (Siouan)	23,191	I. Hartmann
Nǁng (ǃUi-Taa)	26,061	A. Witzlack-M.
Texistepec (Mixe-Zoquean)	21,321	S. Wichmann
Dutch (Indo-European)	39,519	CGN Consort.
English (Indo-European)	56,135	Switchboard
Sum	**288,848**	

Fig. 1. Location of the nine languages and size of the corpora studied here. For detailed information, see *SI Appendix*, Table S1.

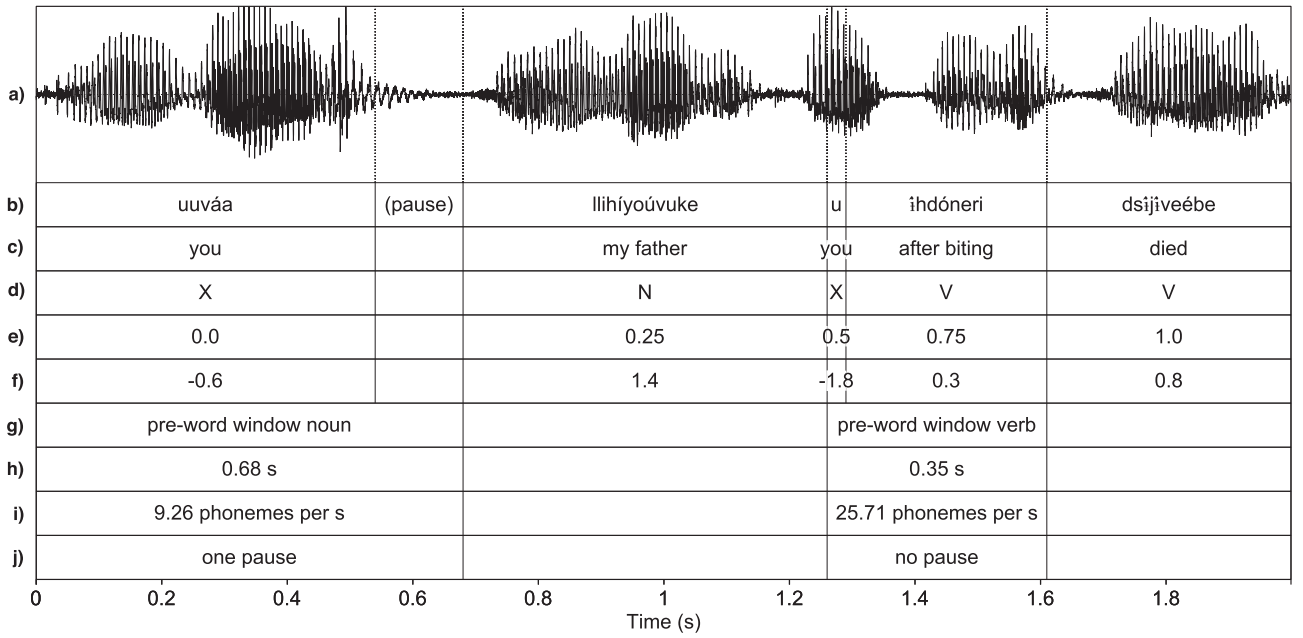


Fig. 2. Bora utterance illustrating slow articulation and presence of a pause before a noun compared with fast articulation and no pause before a verb. The example translates as "After you bit my father, he died" and is taken from a Bora mythological narrative, available online at https://hdl.handle.net/1839/00-0000-0000-000C-DFBE-1. (*a*) Waveform of audio signal; (*b*) time-aligned transcription of words; (*c*) word-by-word translation; (*d*) word class N = noun vs. V = verb vs. X = other; (*e*) position of word within utterance from 0 = start to 1 = end; (*f*) *z*-normalized word length calculated as SDs from mean word length in the language; (*g*) preword context windows for the noun *llihíyoúvuke* "my father" and the verb *dsíjíveébe* "he died," adjusted in size to word boundaries close to 500 ms before onset of target words (preword window for *ihdóneri* "after biting" not shown here); (*h*) length of preword context windows; (*i*) articulation rate of words (excluding pauses) within preword context windows; and (*j*) presence vs. absence of pauses within preword context windows. Procedures for time-aligning transcriptions and for determining position, word length, and context window size are described in *Materials and Methods*.

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, Hoocąk 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, Nllng 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). Regarding 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 than 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

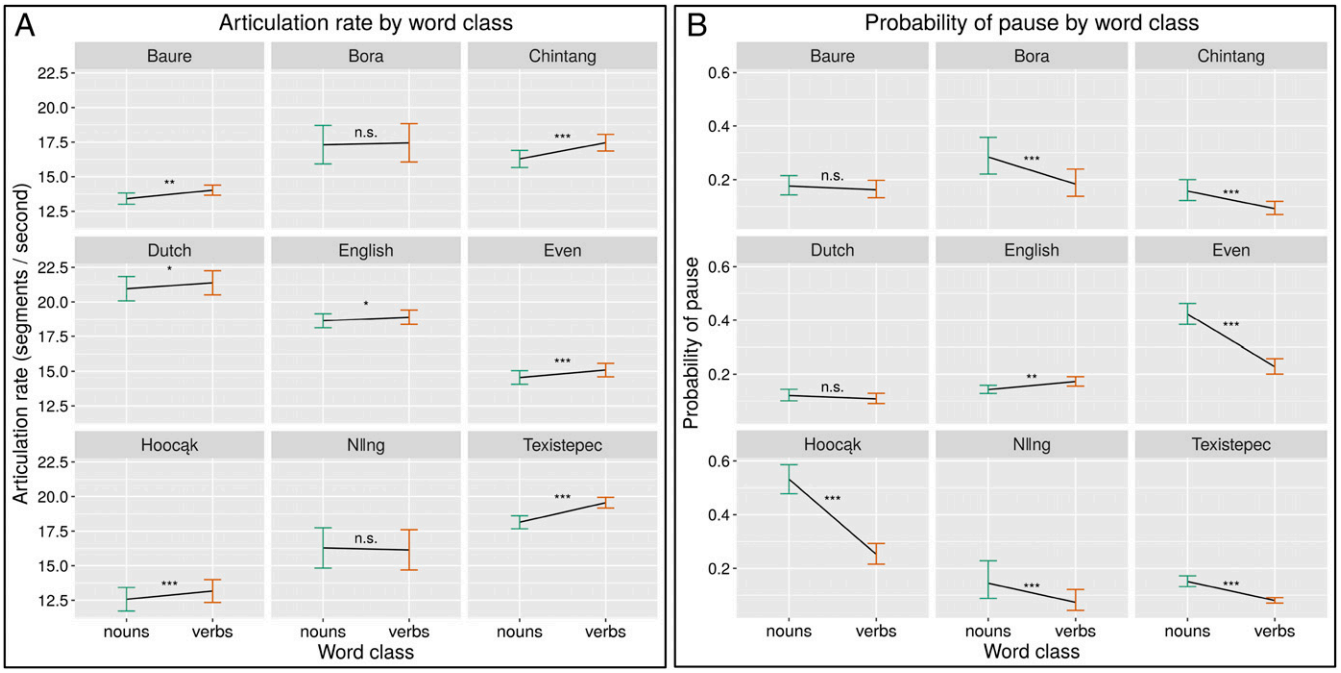


Fig. 3. Speech rate in contexts before nouns versus verbs. The effect displays show a cross-linguistic tendency for slower articulation before nouns (*A*) and a higher probability of pauses before nouns (*B*). The effect of word class (nouns vs. verbs) is plotted according to (generalized) linear mixed-effects models, with 95% confidence intervals based on these models. Both studies are based on models that are consistent across the individual languages, controlling for word position and word length as fixed factors and including random intercepts for speaker, text, and word type. The models for articulation speed included an additional interaction between word class and position, but *A* shows the overall effects of word class, averaging over positions, to simplify the visual representation (*Materials and Methods* and *SI Appendix, Supplementary Text*). Levels of statistical significance are indicated as $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; and n.s. (not significant) > 0.05.

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 finding opens 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 prefixes to nouns than as prefixes 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 furthermore 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 <sh> - /ʃ/ or Dutch <oe> - /u/ in the models below. All data were manually time-aligned during transcription at the level of annotation units (such as in Fig. 2). Annotation units correspond to turn construction units as analyzed in conversation analysis (50). They are stretches of speech, which are intonationally, grammatically, and pragmatically complete and may comprise an entire turn. The end of such units marks a point where the turn may go to another speaker or the present speaker may continue with another such unit. In our largely monological corpora, the end of such units is often characterized by minimal feedback from the listener. They represent easily recognizable major discourse boundaries (32) rather than potentially more controversial minor boundaries. To obtain accurate timing information for the beginning and end of words and pauses, we applied semiautomatic segment-level time alignment (51, 52), followed by manual corrections (the English and Dutch corpora already included word-level time alignment). For the identification of pauses, we set a minimum of 150 ms because most silent intervals in natural speech shorter than this correspond to "articulatory pauses"; 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 (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 shortened if the target word occurs near the beginning of the annotation unit 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 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 S2).

Our algorithm of defining preword windows resulted in variably 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 analysis reported here, in line with our semantically based identification of verbs (see main text). This exclusion is based on the preexisting word class annotation of auxiliaries in our subcorpora of English, Even, Hoocąk, and Texistepec. The languages Baure, Bora, Chintang, and N|ng 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 S4 and S5). In both studies of articulation rate, the dependent variable was the articulation rate in a given preword window. Articulation rate was calculated as the number of characters in the preword window divided by the length of the preword window in seconds (excluding silence between words). *SI Appendix*, Tables S6 and S7 provide detailed descriptive statistics on articulation rate. The main predictor in our models was the word class of the target word. Language-specific word class tags were converted to a common set of categories across all nine corpora: N[oun], V[erb], AUX[iliary], and OTHER. For the analyses, we only kept target words of the categories N, V, and AUX. We also excluded compound words containing both a nominal (N) and a verbal root (V or AUX) (*SI Appendix*, Tables S4 and S5). To control for potential utterance-final slowdown of the articulation rate, we included the position of the target word in the utterance as a covariate. We normalized the position by the length of the utterance so that it ranged from 0 (first word in the utterance) to 1 (last word in the utterance) (see Fig. 2 for an illustration). The normalized position of the *i*th word (counting from 1) in an utterance containing *n* words is defined as $(i - 1)/(n - 1)$. In preliminary studies, we found that longer words tended to exhibit a higher articulation rate than shorter words, consistent with earlier observations that syllable durations shrink as their number increases within a word (56). Therefore, we also included the length of the target word as a covariate in our models. We *z*-normalized word length per language by subtracting from each word's length (approximated by the number of characters in it) the mean length of word tokens in the respective corpus and by dividing the result by the SD of word token length. This procedure accounts for differences in phonological inventories (clicks, for example, exist only in N|ng), as well as for different orthographic conventions, like the use of multicharacter representations of segments, as in English <sh> - /ʃ/ or Dutch <oe> - /u/. To assess the cross-linguistic stability of effects on articulation rate, we also included the factor language (coding the nine different subcorpora: Baure, Bora, Chintang, Dutch, English, Even, Hoocąk, N|ng, and Texistepec) in our model.

In addition to these fixed factors, we included several random factors in our models as controls: (*i*) the speaker of the utterance, (*ii*) the text/recording in which the utterance occurred, and (*iii*) the specific word type of the target word (defined as a specific word form with a specific morphemic structure). We included word type to model differences between individual target words, such as their meaning associations, polarity, emotional values, their complexity, etc. The word type factor also captures the effect of a word's frequency on preceding pauses and articulation rate (12) since a crucial property of a word form is its frequency and thereby the extent to which a speaker is familiar with it. The reason for dealing with frequency and familiarity in this manner, rather than using frequency counts for each word form, lies in the nature of the language documentation corpora used here. Except for Chintang, Dutch, English, and Even, our corpora effectively represent the entirety of text material available for a given language in the sample. This implies that frequency counts can only be obtained from the relatively small corpora under investigation themselves, and such counts would not reflect the accumulated experience of a speaker, thus invalidating estimates.

For the statistical models of articulation rate, we used linear mixed-effects regression models (57), as implemented in the lmer function provided by R package lme4 (58, 59). Model comparison with the Akaike information criterion (AIC) and (two-sided) likelihood ratio tests revealed a significant interaction between word class and language (*SI Appendix*, Tables S8 and S9), and so we created models for the individual languages based on the structure of the best-fitting cross-linguistic model (*SI Appendix*, Tables S10–S18). This choice ensures the comparability of the language-specific models in terms of the magnitude and direction of the observed word class effects in the different languages. (We additionally confirmed that results did not change substantially when models for individual languages were further reduced in complexity by additional model comparisons within each language.) The *P* values in *SI Appendix*, Table S9 are based on likelihood ratio tests that 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. 3*A* 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 as the main study. 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 (silent 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 (62). We used the same predictor variables and random factors as in the study on articulation rate but now with a *logit* link function and assuming a binomial distribution (as implemented in the glmer function in the lme4 package) (58, 59). Model comparison with likelihood ratio tests and AIC revealed again a significant interaction between language and word class (*SI Appendix*, Tables S32 and S33), and so we fit models separately for each language based on the structure of the best-fitting cross-linguistic model (*SI Appendix*, Tables S34–S42). *P* values and effect plots (Fig. 3*B*) were based on the same procedure as in the articulation rate study, to ensure comparability. Effect sizes were derived as probability ratios (relative risks) and odds ratios, both when including and excluding auxiliaries (*SI Appendix*, Table S49).

Data Availability. The complete datasets used in this study are available at https://figshare.com/s/085b09d7d82b5501df4e.

ACKNOWLEDGMENTS. We thank all native speakers that provided data and all assistants that helped annotate the data. We acknowledge the comments of Damián 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 Dokumentation Bedrohter Sprachen (DoBeS) 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 [MesAndLin(g)k, 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

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. Vaissière 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, UK).
5. Stivers T, et al. (2009) Universals and cultural variation in turn-taking in conversation. *Proc Natl Acad Sci USA* 106:10587–10592.
6. Dellwo V, Leemann 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 Cognit* 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* 84:474–496.
9. Bard EG, et al. (2000) Controlling the intelligibility of referring expressions in dialogue. *J Mem Lang* 42:1–22.
10. Aylett M, Turk 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. *Cognit Psychol* 61:23–62.
12. Bell A, Brenier JM, Gregory M, Girand C, Jurafsky D (2009) Predictability effects on durations of content and function words in conversational English. *J Mem Lang* 60: 92–111.
13. Fox Tree JE, Clark HH (1997) Pronouncing "the" as "thee" to signal problems in speaking. *Cognition* 62:151–167.
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: 344–380.
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, Druks J, 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. *The Handbook of Psycholinguistics*, eds Fernández EM, Cairns HS (John Wiley & Sons, Hoboken, NJ), pp 38–81.
21. Conwell E (2017) Prosodic disambiguation of noun/verb homophones in child-directed speech. *J Child Lang* 44:734–751.
22. Kibrik AA, Khudyakova MV, Dobrov GB, Linnik A, Zalmanov DA (2016) Referential choice: Predictability and its limits. *Front Psychol* 7:1429.
23. Arnold JE (2008) Reference production: Production-internal and addressee-oriented processes. *Lang Cogn Process* 23:495–527.
24. Ariel M (2014) *Accessing Noun-Phrase Antecedents* (Routledge, London).
25. Gatt A, Krahmer E, van Deemter K, van Gompel RPG (2014) Models and empirical data for the production of referring expressions. *Lang Cogn Neurosci* 29:899–911.
26. Levinson SC (2000) *Presumptive Meanings. The Theory of Generalized Conversational Implicature* (MIT Press, Cambridge, MA).
27. Grosz 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. Stoll 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 Slobin*, 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. Givón T (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, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing* (Univ of Chicago Press, Chicago).
33. Seifart F (2012) The threefold potential of language documentation. *Potentials of Language Documentation: Methods, Analyses, and Utilization, Language Documentation & Conservation Special Publication*, eds Seifart F, et al. (Univ of Hawai'i Press, Manoa, HI), pp 1–6.
34. Kemmerer D (2014) Word classes in the 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, Pinget A-F, Quené 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. Sóskuthy M, Hay J (2017) Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166:298–313.
40. Henrich 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. Hagoort P, Levinson SC (2014) Neuropragmatics. *The Cognitive Neurosciences*, eds Gazzaniga MS, Mangun GR (MIT Press, Cambridge, MA), pp 667–674.
44. Moran S, et al. (2018) A universal cue for grammatical categories in the input to children: Frequent frames. *Cognition* 175:131–140.
45. Anand P, Chung S, Wagers M (2015) Widening the net: Challenges for gathering linguistic data in the digital age. Available at https://www.nsf.gov/sbe/sbe_2020/2020_pdfs/Wagers_Matthew_121.pdf. Accessed January 25, 2018.
46. Norcliffe E, Harris AC, Jaeger TF (2015) Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. *Lang Cogn Neurosci* 30:1009–1032.
47. Zipf GK (1949) *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley Press, Cambridge, MA).
48. Himmelmann NP (2014) Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90:927–960.
49. Piantadosi ST, Tily H, Gibson E (2011) Word lengths are optimized for efficient communication. *Proc Natl Acad Sci USA* 108:3526–3529.
50. Sacks H, Schegloff EA, Jefferson G (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 50:696–735.
51. Strunk J, Schiel F, Seifart F (2014) Untrained forced alignment 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.
52. Kisler 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.
53. Goldman-Eisler F (1968) *Psycholinguistics: Experiments in Spontaneous Speech* (Academic, London).
54. Hieke AE, Kowal S, O'Connell DC (1983) The trouble with "articulatory" pauses. *Lang Speech* 26:203–214.
55. Campione E, Véronis J (2002) A large-scale multilingual study of silent pause duration. *Speech Prosody* 2002:199–202.
56. Lehiste I (1970) *Suprasegmentals* (MIT Press, Cambridge, MA).
57. Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59:390–412.
58. 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.
59. Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67.
60. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300.
61. Fox J (2003) Effect displays in R for generalised linear models. *J Stat Softw* 8:1–27.
62. Subbiah M, Rajeswaran V (2017) Proportion: A comprehensive R package for inference on single Binomial proportion and Bayesian computations. *SoftwareX* 6: 36–41.