A CROSS-LANGUAGE PERSPECTIVE ON SPEECH INFORMATION RATE

FRANÇOIS PELLEGRINO
Laboratoire Dynamique du Langage, Université de Lyon and *CNRS*

CHRISTOPHE COUPÉ
Laboratoire Dynamique du Langage, Université de Lyon and *CNRS*

EGIDIO MARSICO
Laboratoire Dynamique du Langage, Université de Lyon and *CNRS*

This article is a crosslinguistic investigation of the hypothesis that the average information rate conveyed during speech communication results from a trade-off between average information density and speech rate. The study, based on seven languages, shows a negative correlation between density and rate, indicating the existence of several encoding strategies. However, these strategies do not necessarily lead to a constant information rate. These results are further investigated in relation to the notion of syllabic complexity.*

Keywords: speech communication, information theory, working memory, speech rate, cross-language study

'As soon as human beings began to make systematic observations about one another's languages, they were probably impressed by the paradox that all languages are in some fundamental sense one and the same, and yet they are also strikingly different from one another.' (Ferguson 1978:9)

1. INTRODUCTION. Charles Ferguson's quotation describes two goals of linguistic typology: searching for invariants and determining the range of variation found across languages. Invariants are supposed to be a set of compulsory characteristics, which presumably defines the core properties of the language capacity itself. Language being a system, invariants can be considered to be systemic constraints imposing a set of possible structures among which languages 'choose'. Variants can then be seen as language strategies compatible with the amount of freedom allowed by the linguistic constraints. Both goals are generally investigated simultaneously as the search for universals contrastively reveals the differences across languages. Yet linguistic typology has mostly shown that languages vary to a large extent, finding only few, if any, absolute universals, and thus unable to explain how all languages are 'one and the same' and instead reinforcing the fact that they are 'so strikingly different' (see Evans & Levinson 2009 for a recent discussion). Nevertheless, the paradox exists only if one considers both assumptions ('one and the same' and 'strikingly different') at the same level. Language is actually a communicative system whose primary function is to transmit information. The unity of all languages is probably to be found in this function, regardless of the different linguistic strategies on which they rely.

Another well-known assumption is that all human languages are overall equally complex. This statement is present in most introductory classes in linguistics or encyclopedias of language (e.g. see Crystal 1987). At the same time, linguistic typology provides extensive evidence that the complexity of each component of language grammar

* We wish to thank C.-P. Au, S. Blandin, E. Castelli, S. Makioka, G. Peng, and K. Tamaoka for their help with collecting or sharing the data. We also thank Fermín Moscoso del Prado Martín, R. Harald Baayen, Barbara Davis, Peter MacNeilage, Michael Studdert-Kennedy, and two anonymous referees for their constructive criticism and their suggestions on earlier versions of this article.

(phonology, morphology, or syntax) varies widely from one language to another, and no one claims that two languages having eleven vs. 141 phonemes (like Rotokas and !Xu respectively) are of equal complexity with respect to their phonological systems (Maddieson 1984). A balance in complexity must therefore operate within the grammar of each language: a language exhibiting a low complexity in some of its components should compensate with a high complexity in others. As exciting as this assumption looks, no definitive argument has yet been provided to support or invalidate it (see discussion in Planck 1998), even if a wide range of scattered indices of complexity have recently come into view and so far led to partial results in a typological perspective (Cysouw 2005, Dahl 2004, Fenk-Oczlon & Fenk 1999, 2005, Maddieson 2006, 2009, Marsico et al. 2004, Shosted 2006) or from an evolutionary viewpoint (see Sampson et al. 2009 for a recent discussion).

Considering that the communicative role of language has been underestimated in those debates, we suggest that the assumption of an ‘equal overall complexity’ is ill-defined. More precisely, we endorse the idea that all languages exhibit an ‘equal overall communicative capacity’ even if they have developed distinct encoding strategies partly illustrated by distinct complexities in their linguistic description. This communicative capacity is probably delimited within a range of possible variation in terms of rate of information transmission: below a lower limit, speech communication would not be efficient enough to be socially useful and acceptable; above an upper limit, it would exceed the human physiological and cognitive capacities. One can thus postulate an optimal balance between social and cognitive constraints, also taking the characteristics of transmission along the audio channel into account.

This hypothesis predicts that languages are able to convey relevant pragmatic-semantic information at similar rates and urges us to pay attention to the rate of information transmitted during speech communication. Studying the encoding strategy (as revealed by an information-based and complexity-based study; see below) is thus one necessary part of the equation, but it is not sufficient to determine the actual rate of information transmitted during speech communication.

After reviewing some historical landmarks regarding the way the notions of information and complexity have been interrelated in linguistics for almost a century, this article aims at uniting the information-based approach with a cross-language investigation. It crosslinguistically investigates the hypothesis that a trade-off is operating between a syllable-based average information density and the rate of transmission of syllables in human communication. The study, based on comparable speech data from seven languages, provides strong arguments in favor of this hypothesis. The corollary assumption predicting a constant average information rate among languages is also examined. An additional investigation of the interactions between these information-based indices and a syllable-based measure of phonological complexity is then provided to extend the discussion toward future directions, in the light of literature on the principle of least effort and cognitive processing.

2. Historical background.

The concept of information and the question of its embodiment in linguistic forms were implicitly introduced in linguistics at the beginning of the twentieth century, even before so-called INFORMATION THEORY was popularized (Shannon & Weaver 1949). They were first addressed in the light of approaches such as frequency of use (from Zipf 1935 to Bell et al. 2009) or functional load[1] (from Martinet

[1] See King 1967 for an overview of the genesis of this notion.

1933 and Twaddell 1935, to Surendran & Levow 2004). Beginning in the 1950s, they then benefited from inputs from information theory, with notions such as entropy, the communication channel, and redundancy[2] (Cherry et al. 1953, Hockett 1953, Jakobson & Halle 1956, inter alia).

Furthermore, in the quest for explanations of linguistic patterns and structures, the relationship between information and complexity has also been addressed, either synchronically or diachronically. A landmark statement was given by Zipf: 'there exists an equilibrium between the magnitude or degree of complexity of a phoneme and the relative frequency of its occurrence' (1935:49). Trubetzkoy and Joos strongly attacked this assumption: in his *Grundzüge* (1939), Trubetzkoy denied any explanatory power to the uncertain notion of complexity in favor of the notion of markedness, while Joos's criticism focused mainly on methodological shortcomings and what he considered a tautological analysis (Joos 1936; but see also Zipf's answer (1937)).

In later years, the potential role of complexity in shaping languages was discussed either with regard to its identification with markedness or by considering it in a more functional framework. The first approach is illustrated by Greenberg's answer to the self-question 'Are there any properties which distinguish favored articulations as a group from their alternatives?'. Putting forward 'the principle that of two sounds that one is favored which is the less complex', he concluded that 'the more complex, less favored alternative is called marked and the less complex, more favored alternative the unmarked' (Greenberg 1969:476–77). The second approach, initiated by Zipf's principle of least effort, has been developed by considering that complexity and information may play a role in the regulation of linguistic systems and speech communication. While Zipf mostly ignored the listener's side and suggested that least effort was almost exclusively a constraint affecting the speaker, more recent studies have demonstrated that other forces also play an important role and that economy or equilibrium principles result from a more complex pattern of conflicting pressures (e.g. Martinet 1955, 1962, Lindblom 1990). For instance, Martinet emphasized the role of the communicative need ('the need for the speaker to convey his message'; Martinet 1962:139), counterbalancing the principle of speaker's least effort. Lindblom's H&H (hypo- and hyperarticulation) theory integrates a similar postulate, leading to self-organizing approaches to language evolution (e.g. Oudeyer 2006) and to taking the listener's effort into consideration.

More recently, several theoretical models have been proposed to account for this regularity and to reanalyze Zipf's assumption in terms of emergent properties (e.g. Ferrer i Cancho 2005, Ferrer i Cancho & Solé 2003, Kuperman et al. 2008). These recent works have strongly contributed to a renewal of information-based approaches to human communication (along with Aylett & Turk 2004, Frank & Jaeger 2008, Genzel & Charniak 2003, Goldsmith 2000, 2002, Harris 2005, Hume 2006, Keller 2004, Maddieson 2006, Pellegrino et al. 2007, van Son & Pols 2003, inter alia), but mostly in language-specific studies (but see Kuperman et al. 2008 and Piantadosi et al. 2009).

3. Speech information rate.

3.1. Material. The goal of this study is to assess whether there exist differences in the rate of information transmitted during speech communication in several languages. The proposed procedure is based on a cross-language comparison of the speech rate and

[2] This influence is especially apparent in the fact that Hockett considered redundancy to be the first of the phonological universals: 'In every human language, redundancy, measured in phonological terms, hovers near 50%' (Hockett 1966:24), with an explicit reference to Shannon.

the information density of seven languages using comparable speech materials. The speech data used are a subset of the MULTEXT multilingual corpus (Campione & Véronis 1998, Komatsu et al. 2004). This subset consists of $K = 20$ texts composed in British English, freely translated into the following languages to convey a comparable semantic content: French (FR), German (GE), Italian (IT), Japanese (JA), Mandarin Chinese (MA), and Spanish (SP). Each text consists of five semantically connected sentences that compose either a narration or a query (to order food by phone, for example). The translation inevitably introduced some variation from one language to another, mostly in named entities (locations, etc.) and to some extent in lexical items, in order to avoid odd and unnatural sentences. For each language, a native or highly proficient speaker then counted the number of syllables in each text, as uttered in careful speech, as well as the number of words, according to language-specific rules. The appendix gives one of the twenty texts in the seven languages as an example.

Several adult speakers (from six to ten, depending on the language) recorded the twenty texts at 'normal' speech rates, without being asked to produce fast or careful speech. No sociolinguistic information on them is provided with the corpus. Fifty-nine speakers (twenty-nine males and thirty females) of the seven target languages were included in this study, for a total number of 585 recordings and an overall duration of about 150 minutes. The text durations were computed after discarding silence intervals longer than 150 ms, according to a manual labeling of speech activity.[3]

Since the texts were not explicitly designed for detailed cross-language comparison, they exhibit a rather large variation in length. For instance, the lengths of the twenty English texts range from sixty-two to 104 syllables. To deal with this variation, each text was matched with its translation in an eighth language, Vietnamese (VI), different from the seven languages of the corpus. This external point of reference was used to normalize the parameters for each text in each language and consequently to facilitate the interpretation by comparison with a mostly isolating language (see below).

The fact that this corpus was composed of read-aloud texts, which is not typical of natural speech communication, can be seen as a weakness. Though the texts mimicked different styles (ranging from very formal oral reports to more informal phone queries), this procedure most likely underestimated the natural variation encountered in social interactions. Reading probably lessens the impact of paralinguistic parameters such as attitudes and emotions and smoothes over their prosodic correlates (e.g. Johns-Lewis 1986). Another major and obvious change induced by this procedure is that the speaker has no leeway to choose his/her own words to communicate, with the consequence that a major source of individual, psychological, and social information is absent (Pennebaker et al. 2003). Recording bilinguals may provide a direction for future research on crosslinguistic differences in speech rates while controlling for individual variation. This drawback may also be seen as an advantage, however, since all fifty-nine speakers of the seven languages were recorded in similar experimental conditions, leading to comparable data.

3.2. Density of semantic information. In the present study, density of information refers to the way languages encode semantic information in the speech signal. In this view, a dense language will make use of fewer speech chunks than a sparser language for a given amount of semantic information. This section introduces a methodology to

[3] This filtering was done because in each language, pause durations vary widely from one speaker to another, probably because no explicit instructions were given to the speakers.

evaluate this density and to further assess whether information rate varies from one language to another.

Language grammars reflect conventionalized language-specific strategies for encoding semantic information. These strategies encompass more or less complex surface structures and more or less semantically transparent mappings from meanings to forms (leading to potential trade-offs in terms of complexity or efficiency; see for instance Dahl 2004 and Hawkins 2004, 2009), and they output meaningful sequences of words. The word level is at the heart of human communication, at least because of its obvious function in speaker-listener interactions and also because of its central status between meaning and signal. Thus, words are widely regarded as the relevant level for disentangling the forces involved in complexity trade-offs and for studying the linguistic coding of information. For instance, Juola applied information-theoretical metrics to quantify the crosslinguistic differences and the balance between morphology and syntax in meaning-to-form mapping (Juola 1998, 2008). At a different level, van Son and Pols, among others, have investigated form-to-signal mapping, viz. the impact of linguistic information distribution on the realized sequence of phones (van Son & Pols 2003; see also Aylett & Turk 2006). These two broad issues (the mapping from meaning to form, and from form to signal) shed light on the constraints, the degrees of freedom, and the trade-offs that shape human languages. In this study, we propose a different approach that focuses on the direct mapping from meaning to signal. More precisely, we focus on the level of the information encoded in the course of the speech flow. We hypothesize that a balance between the information carried by speech units and their rate of transmission may be observed, whatever the linguistic strategy of mapping from meaning to words (or forms) and from words to signals.

Our methodology is consequently based on evaluating the average density of information in speech chunks. The relationship between this hypothetical trade-off at the signal level and the interactions at play at the meaningful word level is an exciting topic for further investigation; it is, however, beyond the scope of this study.

The first step is to determine the chunk to use as a 'unit of speech' for the computation of the average information density per unit in each language. Units such as features or articulatory gestures are involved in complex multidimensional patterns (gestural scores or feature matrices) not appropriate for computing the average information density in the course of speech communication. By contrast, each speech sample can be described in terms of discrete sequences of segments or syllables; these units are possible candidates, though their exact status and role in communication are still questionable (e.g. see Port & Leary 2005 for a criticism of the discrete nature of those units). This study is thus based on syllables for both methodological and theoretical reasons (see also §3.3).

Assuming that for each text T_k, composed of $\sigma_k(L)$ syllables in language L, the overall semantic content S_k is equivalent from one language to another, the average quantity of information per syllable for T_k and for language L is calculated as in 1.

$$(1)\ I_L^k = \frac{S_k}{\sigma_k(L)}$$

Since S_k is language-independent, it was eliminated by computing a normalized INFORMATION DENSITY (ID) using Vietnamese (VI) as the benchmark. For each text T_k and language L, ID_L^k resulted from a pairwise comparison of the text lengths (in terms of syllables) in L and VI respectively.

$$(2)\ ID_L^k = \frac{I_L^k}{I_{\mathrm{VI}}^k} = \frac{S_k}{\sigma_k(L)} \times \frac{\sigma_k(\mathrm{VI})}{S_k} = \frac{\sigma_k(\mathrm{VI})}{\sigma_k(L)}$$

Next, the average information density ID_L (in terms of linguistic information per syllable) with reference to VI is defined as the mean of ID_L^k evaluated for the K texts.

$$(3)\quad ID_L = \frac{1}{K}\sum_{k=1}^{K} ID_L^k$$

If ID_L is superior to unity, L is 'denser' than VI since on average fewer syllables are required to convey the same semantic content. An ID_L lower than unity indicates, by contrast, that L is not as dense as VI.

Averaging over twenty texts was aimed at getting values pointing toward language-specific grammars rather than artifacts due to idiomatic or lexical biases in the constitution of the texts. Among the eight languages, each text consists of, on average, 102 syllables, for a total number of syllables per language of 2,040, which is a reasonable length for estimating central tendencies such as means or medians. Another strategy, used by Fenk-Oczlon and Fenk (1999), is to develop a comparative database made of a set of short and simple declarative sentences (twenty-two in their study) translated into each of the languages considered. This alternative approach was based on the fact that using simple syntactic structure and very common vocabulary results in a kind of baseline suitable for proceeding to the cross-language comparison without biases such as stylistic variation. However, such short sentences (ranging on average in the Fenk-Oczlon and Fenk database from five to ten syllables per sentence, depending on the language) could be more sensitive to lexical bias than longer texts, resulting in wider confidence intervals in the estimation of information density.

Table 1 (second column) gives the ID_L values for each of the seven languages. The fact that Mandarin exhibits the value closest to that of Vietnamese ($ID_{MA} = 0.94 \pm 0.04$) is compatible with their proximity in terms of lexicon, morphology, and syntax. Furthermore, Vietnamese and Mandarin, which are the two tone languages in this sample, have the highest ID_L values overall. Japanese density, by contrast, is one-half that of the Vietnamese reference ($ID_{JA} = 0.49 \pm 0.02$), according to our definition of density. Consequently, even in this small sample of languages, ID_L exhibits a considerable range of variation, reflecting different grammars.

These grammars reflect language-specific strategies for encoding linguistic information, but they ignore the temporal facet of communication. For example, if the syllabic speech rate (i.e. the average number of syllables uttered per second) is twice as fast in Japanese as in Vietnamese, the linguistic information would be transmitted at the same RATE in the two languages, since their respective information densities per syllable, ID_{JA} and ID_{VI}, are inversely related. In this perspective, linguistic encoding is only one part of the equation, and we propose in the next section to take the temporal dimension into account.

LANGUAGE	INFORMATION DENSITY ID_L	SYLLABIC RATE (#syl/sec)	INFORMATION RATE
English	0.91 (± 0.04)	6.19 (± 0.16)	1.08 (± 0.08)
French	0.74 (± 0.04)	7.18 (± 0.12)	0.99 (± 0.09)
German	0.79 (± 0.03)	5.97 (± 0.19)	0.90 (± 0.07)
Italian	0.72 (± 0.04)	6.99 (± 0.23)	0.96 (± 0.10)
Japanese	0.49 (± 0.02)	7.84 (± 0.09)	0.74 (± 0.06)
Mandarin	0.94 (± 0.04)	5.18 (± 0.15)	0.94 (± 0.08)
Spanish	0.63 (± 0.02)	7.82 (± 0.16)	0.98 (± 0.07)
Vietnamese	1 (reference)	5.22 (± 0.08)	1 (reference)

TABLE 1. Cross-language comparison of information density, syllabic rate, and information rate (mean values and 95% confidence intervals). Vietnamese is used as the external reference.

3.3. Variation in speech rate. Roach (1998) claims that the existence of cross-language variations in speech rate is a language myth, due to artifacts in the communication environment or its parameters. However, he considers syllabic rate to be a matter of syllable structure and consequently to be widely variable from one language to another, leading to perceptual differences: 'So if a language with a relatively simple syllable structure like Japanese is able to fit more syllables into a second than a language with a complex syllable structure such as English or Polish, it will probably sound faster as a result' (Roach 1998:152). Consequently, Roach proposes to estimate speech rate in terms of sounds per second, to get away from this subjective dimension. He immediately identifies additional difficulties in terms of sound counting, however, due for instance to adaptation observed in fast speech: 'The faster we speak, the more sounds we leave out' (Roach 1998:152). By contrast, the syllable is well known for its relative robustness during speech communication: Greenberg (1999) reported that syllable omission was observed for only about 1% of the syllables in the Switchboard corpus, while omissions occur for 22% of the segments. Using a subset of the Buckeye corpus of conversational speech (Pitt et al. 2005), Johnson (2004) found a somewhat higher proportion of syllable omissions (5.1% on average) and a similar proportion of segment omissions (20%). The difference observed in terms of syllable deletion rate may be due to the different recording conditions: Switchboard data consist of short conversations on the telephone, while the Buckeye corpus is based on face-to-face interactions during longer interviews. The latter is more conducive to reduction for at least two reasons: multimodal communication with visual cues, and more elaborated interspeaker adaptation. In addition, syllable counting is usually a straightforward task in one's native language, even if the determination of syllable boundaries themselves may be ambiguous. In contrast, segment counting is well known to be prone to variation and inconsistency (see Port & Leary 2005:941 inter alia). In addition to the methodological advantage of the syllable for counting, numerous studies have suggested its role either as a cognitive unit or as a unit of organization in speech production or perception (e.g. Schiller 2008, Segui & Ferrand 2002; but see Ohala 2008). Hence, following Ladefoged (1975), we consider that 'a syllable is a unit in the organization of the sounds of an utterance' (Ladefoged 2007:169), and, as far as the distribution of linguistic information is concerned, it seems reasonable to investigate whether syllabic speech rate really varies from one language to another and to what extent it influences the speech information rate.

The MULTEXT corpus used in the present study was not gathered for this purpose, but it provides a useful resource for addressing this issue, because of the similar content and recording conditions across languages. We thus carried out measurements of the speech rate in terms of the number of syllables per second for each recording of each speaker (the syllabic rate, SR). Moreover, the gross mean values of SR across individuals and passages were also estimated for each language (SR_L; see Figure 1).

In parallel the 585 recordings were used to fit a model to SR using the linear mixed-model procedure[4] with Language and speaker's Sex as independent (fixed-effect) predictors and Speaker identity and Text as independent random effects. Note that in all

[4] See Baayen et al. 2008 for a detailed description of this technique and a comparison to other statistical analyses. More generally, all statistical analyses were done with R 2.11.1 (R Development Core Team 2010). The mixed-effect model was estimated using the lme4 and languageR packages. Pearson's correlations were obtained with the cor.test function (standard parameters). Because of the very small size of the language sample ($N = 7$), Spearman rho values were computed using the spearman.test function from the pspearman package, which uses the exact null distribution for the hypothesis test with small samples ($N < 22$).

regression analyses reported in the rest of this article, a *z*-score transformation was applied to the numeric data, in order to get effect estimates of comparable magnitudes.

A preliminary visual inspection of the q-q plot of the model's residuals led to the exclusion of fifteen outliers whose standardized residuals were removed from zero by more than 2.5 standard deviations. The analysis was then rerun with the 570 remaining recordings and visual inspection no longer showed deviation from normality, confirming that the procedure was suitable. We observed a main effect of Language, with highly significant differences among most of the languages: all *pMCMC* were less than 0.001 except between English and German (*pMCMC* = 0.08, n.s.), French and Italian (*pMCMC* = 0.55, n.s.), and Japanese and Spanish (*pMCMC* = 0.32, n.s.). There is also a main effect of Sex (*pMCMC* = 0.0001), with a higher SR for male speakers than for female speakers, which is consistent with previous studies (e.g. Jacewicz et al. 2009, Verhoeven et al. 2004).

Both Text ($\chi^2(1) = 269.79$, $p < 0.0001$) and Speaker ($\chi^2(1) = 684.96$, $p < 0.0001$) were confirmed as relevant random-effect factors, as supported by the likelihood ratio analysis, and were kept in subsequent analyses.

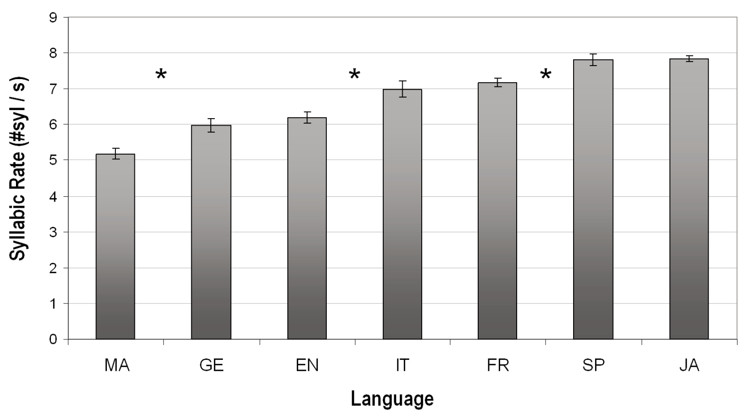

FIGURE 1. Speech rate measured in terms of the number of syllables per second (mean values and 95% confidence intervals). Stars indicate significant differences between the homogeneous subsets revealed by post-hoc analysis.

The presence of different oral styles in the corpus design (narrative texts and queries) is likely to influence SR (Kowal et al. 1983) and thus explains the main random effect of Text. In addition, the main fixed effect of Language supports the idea that languages make different use of the temporal dimension during speech communication. Consequently, SR can be seen as resulting from several factors: the particular language, the nature of the production task, and variation due to the speaker (including the physiological or sociolinguistic effect of sex).

3.4. A REGULATION OF THE SPEECH INFORMATION RATE. We investigated here the possible correlation between ID_L and SR_L, with a regression analysis on SR, again using the linear mixed-model technique. ID is now considered in the model as a numerical covariate, in addition to the factors taken into account in the previous section (Language, Sex, Speaker, and Text).

We observed a highly significant effect of ID (*pMCMC* = 0.0001) corresponding to a negative slope in the regression. The estimated β value for the effect of ID is

$\beta = -0.137$ with a 95% confidence interval in the range [–0.194, –0.084]. This significant regression demonstrates that the languages of our sample exhibit regulation, or at least a relationship, between their linguistic encoding and their speech rate.

Consequently, it is worth examining the overall quantity of information conveyed by each language per unit of time (and not per syllable). This so-called INFORMATION RATE (IR) encompasses both the strategy of linguistic encoding and the speech settings for each language L. Again IR is calculated using VI as an external point of reference, where $D_k(spkr)$ is the duration of the Text number k uttered by speaker.

$$(4)\quad IR_k(spkr) = \frac{S_k}{D_k(spkr)} \times \frac{\overline{D_k}(\mathrm{VI})}{S_k} = \frac{\overline{D_k}(\mathrm{VI})}{D_k(spkr)}$$

Since there is no a priori motivation to match one specific speaker of language L to a given speaker of VI, we used the mean duration for text k in Vietnamese $\overline{D_k}(\mathrm{VI})$.[5] It follows that IR_k is higher than 1 when the speaker has a higher syllabic rate than the average syllabic rate in Vietnamese for the same text k. Next, IR_L corresponds to the average amount of information conveyed per unit of time in language L, and it is defined as the mean value of $IR_k(spkr)$ among the speakers of language L.

Each speaker can definitely modulate his/her own speech rate to deviate from the average value as a consequence of the sociolinguistic and interactional context. Our hypothesis, however, is that language identity would not be an efficient predictor of IR due to functional equivalence across languages. To test this hypothesis, we fitted a model to IR using the linear mixed-model procedure with Language and speaker's Sex as independent (fixed-effect) predictors and Speaker identity and Text as independent random effects. Again, both Text ($\chi^2(1) = 414.75$, $p < 0.0001$) and Speaker ($\chi^2(1) = 176.55$, $p < 0.0001$) were confirmed as relevant random-effect factors. Sex was also identified as a significant fixed-effect predictor ($pMCMC = 0.002$) and, contrary to our prediction, the contribution of Language was also significant for pairs involving Japanese and English. More precisely, Japanese contrasts with the six other languages ($pMCMC < 0.001$ for all pairs), and English significantly contrasts with Japanese, German, Mandarin, and Spanish ($pMCMC < 0.01$) and with Italian and French ($pMCMC < 0.05$). Our hypothesis of equal IR among languages is thus invalidated, even if five of the seven languages cluster together (GE, MA, IT, SP, and FR).

Figure 2 displays ID_L, SR_L, and IR_L on the same graph. For convenience, ID_L, which is unitless, has been multiplied by 10 to be represented on the same scale as SR_L (left axis). The interaction between information density (gray bars) and speech rate (striped bars) is visible since the first one increases while the second one decreases. The black dotted line connects the information rate values for each language (triangles, right axis). English ($IR_{\mathrm{EN}} = 1.08$) shows a higher information rate than Vietnamese ($IR_{\mathrm{VI}} = 1$). In contrast, Japanese exhibits the lowest IR_L value of the sample. Moreover, one can observe that several languages may reach very similar IR_L values with different encoding strategies: Spanish is characterized by a fast rate of low-density syllables, while Mandarin exhibits a 34% slower syllabic rate with syllables 'denser' by a factor of 49%. In the end, their information rates differ only by 4%.

3.5. SYLLABIC COMPLEXITY AND INFORMATION TRADE-OFF. Among possible explanations of the density/rate trade-off, it may be put forward that if the amount of informa-

[5] The speaker's sex was actually taken into consideration: the normalization was based on the mean duration among either Vietnamese female speakers or Vietnamese male speakers, depending on *spkr* sex.

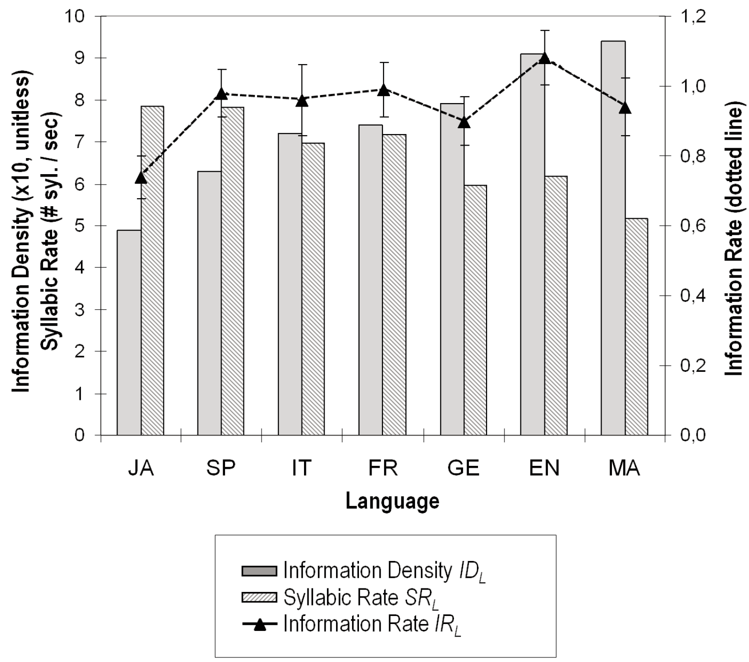

FIGURE 2. Comparison of the encoding strategies of linguistic information. Gray and striped bars respectively display information density (ID_L) and syllabic rate (SR_L) (left axis). For convenience, ID_L has been multiplied by a factor of 10. Black triangles give information rate (IR_L) (right axis, 95% confidence intervals displayed). Languages are ranked by increasing ID_L from left to right (see text).

tion carried by syllables is proportional to their complexity (defined as their number of constituents), then information density is positively related to the average syllable duration and, consequently, negatively correlated to syllabic rate. We consider this line of explanation is this section.

TOWARD A MEASURE OF SYLLABIC COMPLEXITY. A traditional way of estimating phonological complexity is to evaluate the size of the repertoire of phonemes for each language (Nettle 1995), but this index is unable to cope with the sequential language-specific constraints existing between adjacent segments or at the syllable level. It has therefore been proposed to consider this syllabic level directly, either by estimating the size of the syllable inventory (Shosted 2006) or by counting the number of phonemes constituting the syllables (Maddieson 2006). This latter index has also been used extensively in psycholinguistic studies investigating the potential relationship between linguistic complexity and working memory (e.g. Mueller et al. 2003, Service 1998), but it ignores all nonsegmental phonological information such as tone, though it carries a very significant part of the information (Surendran & Levow 2004).

Here, we define syllable complexity as a syllable's number of CONSTITUENTS (both segmental and tonal). In the seven-language dataset, only Mandarin requires taking the tonal constituent into account, and the complexity of each of its tone-bearing syllables is thus computed by adding 1 to its number of phonemes, while for neutral tone syllables, the complexity is equal to the number of phonemes. This is also the case for the syllables of the six other languages.[6] An average syllabic complexity for a language may therefore be computed given an inventory of the syllables occurring in a large corpus of this language. Since numerous studies in linguistics (Bybee 2006, Hay et al. 2001, inter alia) and psycholinguistics (see e.g. Cholin et al. 2006, Levelt 2001) point toward the relevance of the notion of frequency of use, we computed this syllabic complexity index in terms of both type (each distinct syllable counting once in the calculation of the average complexity) and token (the complexity of each distinct syllable is weighted by its frequency of occurrence in the corpus).

These indices were computed from large, syllabified, written corpora gathered for psycholinguistic purposes. Data for French are derived from the LEXIQUE 3 database (New et al. 2004); data for English and German are extracted from the WebCelex database (Baayen et al. 1993); data for Spanish and Italian come from Pone 2005; data for Mandarin are extracted from Peng 2005 and converted into pinyin using the Chinese Word Processor (© NJStar Software Corp.); and data for Japanese are computed from Tamaoka & Makioka 2004.

Data are displayed in Table 2. For each language, the second column gives the size of the syllable inventories, and the third and fourth columns give the average syllabic complexity for types and tokens, respectively. On average, the number of constituents estimated from tokens is 0.95 smaller than the number of constituents estimated from types. Leaving the tonal constituent of Mandarin syllables aside, this confirms the well-known fact that shorter syllables are more frequent than longer ones in language use. Japanese exhibits the lowest syllabic complexity—whether per types or per tokens—while Mandarin reaches the highest values.

LANGUAGE	SYL INVENTORY SIZE N_L	SYL COMPL (TYPE) (avg #const/syl)	SYL COMPL (TOKEN) (avg #const/syl)
English	7,931	3.70	2.48
French	5,646	3.50	2.21
German	4,207	3.70	2.68
Italian	2,719	3.50	2.30
Japanese	416	2.65	1.93
Mandarin	1,191	3.87	3.58
Spanish	1,593	3.30	2.40

TABLE 2. Cross-language comparison of syllabic inventory and syllable complexities (in terms of number of constituents).

SYLLABIC COMPLEXITY AND INFORMATION. The paradigmatic measures of syllabic complexity, for types and tokens, echo the syntagmatic measure of information density previously estimated on the same units—syllables. It is thus especially relevant to evaluate whether the trade-off revealed in §3.4 is due to a direct relation between the syllabic information density and syllable complexity (in terms of number of constituents).

In order to assess whether the indices of syllabic complexity are related to the density/rate trade-off, their correlations with information density, speech rate, and informa-

[6] Further investigations would actually be necessary to exactly quantify the average weight of the tone in syllable complexity. The value '1' simply assumes an equal weight for tonal and segmental constituents in Mandarin. Furthermore, lexical stress (such as in English) could also be considered as a constituent, but to the best of the authors' knowledge, its functional load (independently from the segmental content of the syllables) is not precisely known.

tion rate were computed. The very small size of the sample ($N = 7$ languages) strongly limits the reliability of the results, but nevertheless gives insight into future research directions. For the same reason, we estimated the correlation according to both Pearson's correlation analysis (r) and Spearman's rank correlation analysis (ρ), in order to potentially detect incongruent results. Eventually, both measures of syllabic complexity are correlated to ID_L. The highest correlation is reached with the complexity estimated on the syllable types ($\rho = 0.98$, $p < 0.01$; $r = 0.94$, $p < 0.01$) and is further illustrated in Figure 3. Speech rate (SR_L) is negatively correlated with syllable complexity estimated from both types ($\rho = –0.98$, $p < 0.001$; $r = –0.83$, $p < 0.05$) and tokens ($\rho = –0.89$, $p < 0.05$; $r = –0.87$, $p < 0.05$). These results suggest that syllable complexity is engaged in a twofold relationship, with the two terms of the trade-off highlighted in information density rate (ID_L and SR_L), without enabling one to disentangle the causes and consequences of this scheme. Furthermore, an important result is that no significant correlation is evidenced between the syllabic information rate and indices of syllabic complexity, both in terms of type ($\rho = 0.16$, $p = 0.73$; $r = 0.72$, $p = 0.07$) and token ($\rho = 0.03$, $p = 0.96$; $r = 0.24$, $p = 0.59$).

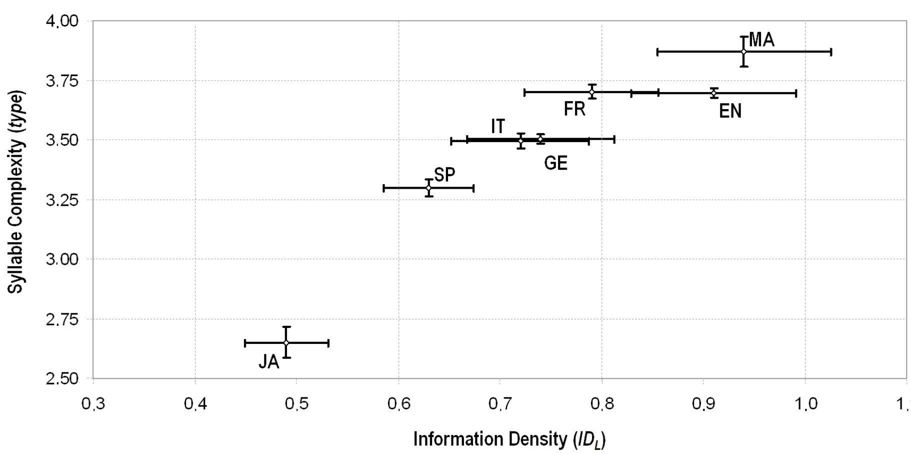

FIGURE 3. Relation between information density and syllable complexity (average number of constituents per syllable type). 95% confidence intervals are displayed.

4. DISCUSSION.

4.1. TRADE-OFF BETWEEN INFORMATION DENSITY AND SPEECH RATE. Two hypotheses motivate the approach taken in this article. The first one states that, for functional reasons, the rate of linguistic information transmitted during speech communication is, to some extent, similar across languages. The second hypothesis is that this regulation results in a density/rate trade-off between the average information density carried by speech chunks and the number of chunks transmitted per second.

Regarding information density, results show that the seven languages of the sample exhibit a large variability, highlighted in particular by the ratio of one-half between Japanese and Vietnamese ID_L. Such variation is related to language-specific strategies, not only in terms of pragmatics and grammars (what is explicitly coded and how) but also in terms of word-formation rules, which, in turn, may be related to syllable complexity (see Fenk-Oczlon & Fenk 1999, Plotkin & Novack 2000). One may object that, in Japanese, morae are probably more salient units than syllables in accounting for linguistic encoding. Nevertheless, syllables give a basis for a methodology that can be strictly transposed from one language to another, as far as average information density is concerned.

Syllabic speech rate varies significantly as well, both among speakers of the same language and crosslinguistically. Since the corpus was not explicitly recorded to investigate speech rate, the variation observed is an underestimation of the potential range from slow to fast speech rates, but it provides a first approximation of the 'normal' speech rate by simulating social interactions. Sociolinguistic arguments pointing to systematic differences in speech rates among populations speaking different languages are not frequent, one exception being the hypothesis that links a higher incidence of fast speech to small, isolated communities (Trudgill 2004). Additionally, sociolinguists often consider that, within a population, speech rate is a factor connected to speech style and is involved in a complex pattern of status and context of communication (Brown et al. 1985, Wells 1982).

Information rate is shown to result from a density/rate trade-off illustrated by a very strong negative correlation between ID_L and SR_L. This result confirms the hypothesis suggested fifty years ago by Karlgren (1961) and revived more recently (Greenberg & Fosler-Lussier 2000, Locke 2008):

> It is a challenging thought that general optimalization rules could be formulated for the relation between speech rate variation and the statistical structure of a language. Judging from my experiments, there are reasons to believe that there is an equilibrium between information value on the one hand and duration and similar qualities of the realization on the other. (Karlgren 1961:676)

However, IR_L exhibits a greater than 30% degree of variation between Japanese (0.74) and English (1.08), invalidating the first hypothesis of a strict cross-language equality of rates of information. The linear mixed-effect model nevertheless reveals that no significant contrast exists among five of the seven languages (German, Mandarin, Italian, Spanish, and French) and highlights the fact that texts themselves and speakers are very significant sources of variation. Consequently, one has to consider the alternative loose hypothesis that IR_L varies within a range of values that guarantee efficient communication, fast enough to convey useful information and slow enough to limit the communication cost (in its articulatory, perceptual, and cognitive dimensions). A deviation from the optimal range of variation defined by these constraints is still possible, however, because of additional factors such as social aspects. A large-scale study involving many languages would eventually confirm or invalidate the robustness of this hypothesis, answering questions such as: What is the possible range of variation for 'normal' speech rate and information density? To what extent can a language depart from the density/rate trade-off? Can we find a language with both a high speech rate and a high information density?

These results support the idea that, despite the large variation observed in phonological complexity among languages, a trend toward regulation of the information rate is at work, as illustrated here by Mandarin and Spanish reaching almost the same average information rate with two opposite strategies: slower, denser, and more complex for Mandarin vs. faster, less dense, and less complex for Spanish. The existence of this density/rate trade-off may thus illustrate a twofold least-effort equilibrium in terms of ease of information encoding and decoding on the one hand, vs. efficiency of information transfer through the speech channel on the other.

In order to provide a first insight on the potential relationship between the syntagmatic constraints on information rate and the paradigmatic constraints on syllable formation, we introduced type-based and token-based indices of syllabic complexity. Both are positively correlated with information density and negatively correlated with syllabic rate. But one has to be cautious with these results for at least two reasons. The first is that the language sample is very small, which leads to results that have no typological range (this caveat is obviously valid for all of the results presented in this article). The second shortcoming is due to the necessary counting of the constituents for each syllable, leading to questionable methodological choices mentioned earlier for phonemic segmentation and for weighting tone and stress dimensions.

4.2. Information-driven regulations at the individual level. Another noticeable result is that the syllabic complexity indices do not correlate with the observed information rate (IR_L). Thus, these linguistic factors of phonological complexity are bad—or at least insufficient—predictors of the rate of information transmission during speech communication. These results provide new crosslinguistic arguments in favor of a regulation of the information flow.

This study echoes recent work investigating informational constraints on human communication (Aylett & Turk 2004, Frank & Jaeger 2008, Genzel & Charniak 2003, Keller 2004, Levy & Jaeger 2007), with the difference that it provides a cross-language perspective on the average amount of information rather than a detailed, language-specific study of the distribution of information (see also van Son & Pols 2003). All of these studies have in common the assumption that human communication may be analyzed through the prism of information theory, and that humans try to optimally use the channel of transmission through a principle of uniform information density (UID). This principle postulates 'that speakers would optimize the chance of successfully transmitting their message by transmitting a uniform amount of information per transmission (or per time, assuming continuous transmission) close to the Shannon capacity of the channel' (Frank & Jaeger 2008:939). Frank and Jaeger proposed that speakers would try to avoid spikes in the rate of information transmission in order to avoid 'wasting' some channel capacity. This hypothesis is controversial, however, especially because what is optimal from the point of view of Shannon's theory (transmission) is not necessarily optimal for human cognitive processing (coding and decoding). It is thus probable that transmission constraints also interact with other dimensions such as probabilistic paradigmatic relations, as suggested in Kuperman et al. 2007, or attentional mechanisms, for instance.

More generally, information-driven trade-offs could reflect general characteristics of information processing by human beings. Along these lines, frequency matching and phase locking between the speech rate and activations in the auditory cortex during a task of speech comprehension (Ahissar et al. 2001) would be worth investigating in a cross-language perspective to elucidate whether these synchronizations are also sensitive to the information rate of the languages considered. Working memory refers to the structures and processes involved in the temporary storage and processing of information in the human brain. One of the most influential models includes a system called the phonological loop (Baddeley 2000, 2003, Baddeley & Hitch 1974) that would enable one to keep a limited amount of verbal information available to the working memory. The existence of time decay in memory span seems plausible (see Schweickert & Boruff 1986 for a mathematical model), and several factors may influence the capacity of the phonological loop. Among others, articulation duration, phonological similarity, phonological length (viz. the number of syllables per item), and phonological complexity (viz. the number of phonological segments per item) are often mentioned (Baddeley

2000, Mueller et al. 2003, Service 1998). Surprisingly, mother tongues of the subjects and languages of the stimuli have not been thoroughly investigated as relevant factors per se. Tasks performed in spoken English vs. American Sign Language have indeed revealed differences (Bavelier et al. 2006, Boutla et al. 2004), but without determining for certain whether they were due to the distinct modalities, to the distinct linguistic structures, or to neurocognitive differences in phonological processing across populations. Since there are, however, significant variations in speech rate among languages, cross-language experiments would probably provide major clues for disentangling the relative influence of universal general processes vs. linguistic parameters. If one assumes constant time decay for the memory span, it would include a different number of syllables according to different language-specific speech rates. By contrast, if linguistic factors matter, one could imagine that differences across languages in terms of syllabic complexity or speech rate would influence memory spans. As a conclusion, we would like to point out that cross-language studies may be very fruitful for revealing whether memory span is a matter of syllables, words, quantity of information, or simply duration. More generally, such cross-language studies are crucial both for linguistic typology and for language cognition (see also Evans & Levinson 2009).

Appendix: Translations of text p8

English: Last night I opened the front door to let the cat out. It was such a beautiful evening that I wandered down the garden for a breath of fresh air. Then I heard a click as the door closed behind me. I realised I'd locked myself out. To cap it all, I was arrested while I was trying to force the door open!

French: Hier soir, j'ai ouvert la porte d'entrée pour laisser sortir le chat. La nuit était si belle que je suis descendu dans la rue prendre le frais. J'avais à peine fait quelque pas que j'ai entendu la porte claquer derrière moi. J'ai réalisé, tout d'un coup, que j'étais fermé dehors. Le comble c'est que je me suis fait arrêter alors que j'essayais de forcer ma propre porte!

Italian: Ieri sera ho aperto la porta per far uscire il gatto. Era una serata bellissima e mi veniva voglia di starmene sdraiato fuori al fresco. All'improvviso ho sentito un clic dietro di me e ho realizzato che la porta si era chiusa lasciandomi fuori. Per concludere, mi hanno arrestato mentre cercavo di forzare la porta!

Japanese: 昨夜、私は猫Weに出してやるために玄関を開けてみると、あまりに気持のいい夜だつたので、新鮮な空気をす吸おうと、ついふらっと庭へ降りたのです。すると後ろでドアが閉まつて、カチッと言う音が聞こえ、自分自身を締め出してしまったことに気が付いたのです。挙げ句の果てに、私は無理矢理ドアをこじ開けようとしているところを逮捕されてしまったのです。

German: Letzte nacht habe ich die haustür geöffnet um die katze nach draußen zu lassen. Es war ein so schöner abend daß ich in den garten ging, um etwas frische luft zu schöpfen. Plötzlich hörte ich wie tür hinter mir zufiel. Ich hatte mich selbst ausgesperrt und dann wurde ich auch noch verhaftet als ich versuchte die tür aufzubrechen.

Mandarin Chinese: 昨晚我打开前门放猫出去的时候，看到夜色很美，就走下台阶，想到花园里呼吸呼吸新鲜空气。当时只听到身后咔哒一声，发现自己被锁在门外了。更糟的是，当我试图撬开门的时候被警察逮捕了。

Spanish: Anoche, abrí la puerta del jardín para sacar al gato. Hacía una noche tan buena que pensé en dar un paseo y respirar el aire fresco. De repente, se me cerró la puerta. Me quedé en la calle, sin llaves. Para rematarlo, me arrestaron cuando trataba de forzar la puerta para entrar.

REFERENCES

Ahissar, Ehud; Srikantan Nagarajan; Merav Ahissar; Athanassios Protopapas; Henry Mahncke; and Michael M. Merzenich. 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences* 96.23.13367–72.

Aylett, Matthew P., and 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.

AYLETT, MATTHEW P., and ALICE TURK. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* 119.3048–58.

BAAYEN, R. HARALD; DOUGLAS J. DAVIDSON; and D. M. BATES. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59.390–412.

BAAYEN, R. HARALD; RICHARD PIEPENBROCK; and H. VAN RIJN. 1993. The CELEX lexical database. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

BADDELEY, ALAN D. 2000. The episodic buffer: A new component of working memory? *Trends in Cognitive Science* 4.417–23.

BADDELEY, ALAN D. 2003. Working memory: Looking back and looking forward. *Nature Reviews: Neuroscience* 4.829–39.

BADDELEY, ALAN D., and GRAHAM J. HITCH. 1974. Working memory. *The psychology of learning and motivation: Advances in research and theory, vol. 8*, ed. by Gordon H. Bower, 47–89. New York: Academic Press.

BAVELIER, DAPHNE; ELISSA L. NEWPORT; MATTHEW L. HALL; TED SUPALLA; and MRIM BOUTLA. 2006. Persistent difference in short-term memory span between sign and speech. *Psychological Science* 17.12.1090–92.

BELL, ALAN; JASON M. BRENIER; MICHELLE GREGORY; CYNTHIA GIRAND; and DAN JURAFSKY. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60.1.92–111.

BOUTLA, MRIM; TED SUPALLA; ELISSA L. NEWPORT; and DAPHNE BAVELIER. 2004. Short-term memory span: Insights from sign language. *Nature Neuroscience* 7.9.997–1002.

BROWN, BRUCE L.; H. GILES; and J. N. THAKERAR. 1985. Speaker evaluation as a function of speech rate, accent and context. *Language and Communication* 5.3.207–20.

BYBEE, JOAN L. 2006. *Frequency of use and the organization of language*. New York: Oxford University Press.

CAMPIONE, ESTELLE, and JEAN VÉRONIS. 1998. A multilingual prosodic database. Paper presented at the 5th International Conference on Spoken Language Processing, Sydney, Australia.

CHERRY, E. COLIN; MORRIS HALLE; and ROMAN JAKOBSON. 1953. Toward the logical description of languages in their phonemic aspect. *Language* 29.1.34–46.

CHOLIN, JOANA; WILLEM J. M. LEVELT; and NIELS O. SCHILLER. 2006. Effects of syllable frequency in speech production. *Cognition* 99.205–35.

CRYSTAL, DAVID. 1987. *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.

CYSOUW, MICHAEL. 2005. Quantitative method in typology. *Quantitative linguistics: An international handbook*, ed. by Gabriel Altmann, Reinhard Köhler, and Rajmund G. Piotrowski, 554–78. Berlin: Mouton de Gruyter.

DAHL, ÖSTEN. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.

EVANS, NICHOLAS, and STEPHEN C. LEVINSON. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32.429–48.

FENK-OCZLON, GERTRAUD, and AUGUST FENK. 1999. Cognition, quantitative linguistics, and systemic typology. *Linguistic Typology* 3.2.151–77.

FENK-OCZLON, GERTRAUD, and AUGUST FENK. 2005. Crosslinguistic correlations between size of syllables, number of cases, and adposition order. *Sprache und Natürlichkeit: Gedenkband für Willi Mayerthaler*, ed. by Gertraud Fenk-Oczlon and Christian Winkler, 75–86. Tübingen: Gunther Narr.

FERGUSON, CHARLES A. 1978. Historical backgrounds of universals research. *Universals of human language, vol. 1: Method and theory*, ed. by Joseph H. Greenberg, 7–33. Stanford, CA: Stanford University Press.

FERRER I CANCHO, RAMON. 2005. Decoding least effort and scaling in signal frequency distributions. *Physica A: Statistical Mechanics and its Applications* 345.275–84.

FERRER I CANCHO, RAMON, and RICHARD V. SOLÉ. 2003. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences* 100.3.788–91.

Frank, Austin F., and T. Florian Jaeger. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the 30th annual conference of the Cognitive Science Society (CogSci08)*, 939–44. Austin, TX: Cognitive Science Society.

Genzel, Dmitriy, and Eugene Charniak. 2003. Variation of entropy and parse trees of sentences as a function of the sentence number. *Proceedings of the 2003 conference on Empirical Methods in Natural Language Processing*, ed. by Michael Collins and Mark Steedman, 65–72. East Stroudsburg, PA: Association for Computational Linguistics. Online: http://acl.ldc.upenn.edu/acl2003/emnlp/index.htm.

Goldsmith, John A. 2000. On information theory, entropy, and phonology in the 20th century. *Folia Linguistica* 34.1–2.85–100.

Goldsmith, John A. 2002. Probabilistic models of grammar: Phonology as information minimization. *Phonological Studies* 5.21–46.

Greenberg, Joseph H. 1969. Language universals: A research frontier. *Science* 166.473–78.

Greenberg, Steven. 1999. Speaking in a shorthand—A syllable-centric perspective for understanding pronunciation variation. *Speech Communication* 29.159–76.

Greenberg, Steven, and Eric Fosler-Lussier. 2000. The uninvited guest: Information's role in guiding the production of spontaneous speech. *Proceedings of the CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modeling*, Kloster Seeon, Germany.

Harris, John. 2005. Vowel reduction as information loss. *Headhood, elements, specification and contrastivity*, ed. by Philip Carr, Jacques Durand, and Colin J. Ewen, 119–32. Amsterdam: John Benjamins.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hawkins, John A. 2009. An efficiency theory of complexity and related phenomena. In Sampson et al., 252–68.

Hay, Jennifer; Janet Pierrehumbert; and Mary E. Beckman. 2001. Speech perception, well-formedness, and the statistics of the lexicon. *Phonetic interpretation: Papers in laboratory phonology 6*, ed. by John Local, Richard Ogden, and Rosalind Temple, 58–74. Cambridge: Cambridge University Press.

Hockett, Charles F. 1953. Review of Shannon & Weaver 1949. *Language* 29.1.69–93.

Hockett, Charles F. 1966. The problem of universals in language. *Universals of language*, 2nd edn., ed. by Joseph H. Greenberg, 1–29. Cambridge, MA: MIT Press.

Hume, Elizabeth. 2006. Language specific and universal markedness: An information-theoretic approach. Paper presented at the annual meeting of the Linguistic Society of America, Colloquium on Information Theory and Phonology, Albuquerque, NM.

Jacewicz, Ewa; Robert A. Fox; Caitlin O'Neill; and Joseph Salmons. 2009. Articulation rate across dialect, age, and gender. *Language Variation and Change* 21.2.233–56.

Jakobson, Roman, and Morris Halle. 1956. *Fundamentals of language*. The Hague: Mouton.

Johns-Lewis, Catherine. 1986. Prosodic differentiation of discourse modes. *Intonation in discourse*, ed. by Catherine Johns-Lewis, 199–220. San Diego: College-Hill Press.

Johnson, Keith. 2004. Massive reduction in conversational American English. *Spontaneous speech: Data and analysis (Proceedings of the 1st Session of the 10th International Symposium)*, ed. by Kiyoko Yoneyama and K. Maekawa, 29–54. Tokyo: The National International Institute for Japanese Language.

Joos, Martin. 1936. Review of Zipf 1935. *Language* 12.3.196–210.

Juola, Patrick. 1998. Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics* 5.3.206–13.

Juola, Patrick. 2008. Assessing linguistic complexity. *Language complexity: Typology, contact, chang*e, ed. by Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson, 89–108. Amsterdam: John Benjamins.

Karlgren, Hans. 1961. Speech rate and information theory. *Proceedings of the 4th International Congress of Phonetic Sciences (ICPhS)*, Helsinki, 671–77.

Keller, Franck. 2004. The entropy rate principle as a predictor of processing effort: An evaluation against eye-tracking data. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Barcelona, 317–24.

King, Robert D. 1967. Functional load and sound change. *Language* 43.4.831–52.

Komatsu, Masahiko; Takayuki Arai; and Tsutomu Sugarawa. 2004. Perceptual discrimination of prosodic types. Paper presented at Speech Prosody 2004, Nara, Japan.

Kowal, Sabine; Richard Wiese; and Daniel C. O'Connell. 1983. The use of time in storytelling. *Language and Speech* 26.4.377–92.

Kuperman, Victor; Mirjam Ernestus; and R. Harald Baayen. 2008. Frequency distributions of uniphones, diphones, and triphones in spontaneous speech. *Journal of the Acoustical Society of America* 124.6.3897–908.

Kuperman, Victor; Mark Pluymaekers; Mirjam Ernestus; and R. Harald Baayen. 2007. Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America* 121.4.2261–71.

Ladefoged, Peter. 1975. *A course in phonetics*. Chicago: University of Chicago Press.

Ladefoged, Peter. 2007. Articulatory features for describing lexical distinctions. *Language* 83.1.161–80.

Levelt, Willem J. M. 2001. Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences* 98.23.13464–71.

Levy, Roger, and T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems 19*, ed. by Bernhard Schölkopf, John Platt, and Thomas Hofmann, 849–56. Cambridge, MA: MIT Press.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modelling*, ed. by William J. Hardcastle and Alain Marchal, 403–39. Dordrecht: Kluwer.

Locke, John L. 2008. Cost and complexity: Selection for speech and language. *Journal of Theoretical Biology* 251.4.640–52.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian. 2006. Correlating phonological complexity: Data and validation. *Linguistic Typology* 10.1.106–23.

Maddieson, Ian. 2009. Calculating phonological complexity. *Approaches to phonological complexity*, ed. by François Pellegrino, Egidio Marsico, Ioana Chitoran, and Christophe Coupé, 85–110. Berlin: Mouton de Gruyter.

Marsico, Egidio; Ian Maddieson; Christophe Coupé; and François Pellegrino. 2004. Investigating the 'hidden' structure of phonological systems. *Berkeley Linguistics Society* 30.256–67.

Martinet, André. 1933. Remarques sur le système phonologique du français. *Bulletin de la Société de linguistique de Paris* 33.191–202.

Martinet, André. 1955. *Économie des changements phonétiques: Traité de phonologie diachronique*. Berne: Francke.

Martinet, André. 1962. *A functional view of language*. Oxford: Clarendon.

Mueller, Shane T.; Travis L. Seymour; David E. Kieras; and David E. Meyer. 2003. Theoretical implications of articulatory duration, phonological similarity and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29.6.1353–80.

Nettle, Daniel. 1995. Segmental inventory size, word length, and communicative efficiency. *Linguistics* 33.359–67.

New, Boris; Christophe Pallier; Marc Brysbaert; and Ludovic Ferrand. 2004. *Lexique 2*: A new French lexical database. *Behavior Research Methods, Instruments, & Computers* 36.3.516–24.

Ohala, John J. 2008. The emergent syllable. *The syllable in speech production*, ed. by Barbara L. Davis and Krisztina Zadjó, 179–86. New York: Lawrence Erlbaum.

Oudeyer, Pierre-Yves. 2006. *Self-organization in the evolution of speech*, trans. by James R. Hurford. (Studies in the evolution of language.) Oxford: Oxford University Press.

Pellegrino, François; Christophe Coupé; and Egidio Marsico. 2007. An information theory-based approach to the balance of complexity between phonetics, phonology and morphosyntax. Paper presented at the annual meeting of the Linguistic Society of America, Anaheim, CA.

Peng, Gang. 2005. Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics* 34.1.134–54.

PENNEBAKER, JAMES W.; MATTHIAS R. MEHL; and KATE G. NIEDERHOFFER. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology* 54.547–77.

PIANTADOSI, STEVEN T.; HARRY J. TILY; and EDWARD GIBSON. 2009. The communicative lexicon hypothesis. *Proceedings of the 31st annual meeting of the Cognitive Science Society (CogSci09)*, 2582–87. Austin, TX: Cognitive Science Society.

PITT, MARK A.; KEITH JOHNSON; ELIZABETH HUME; SCOTT KIESLING; and WILLIAM RAYMOND. 2005. The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication* 45.90–95.

PLANK, FRANZ. 1998. The co-variation of phonology with morphology and syntax: A hopeful history. *Linguistic Typology* 2.2.195–230.

PLOTKIN, JOSHUA B., and MARTIN A. NOWAK. 2000. Language evolution and information theory. *Journal of Theoretical Biolog*y 205.1.147–60.

PONE, MASSIMILIANO. 2005. Studio e realizzazione di una tastiera software pseudo-sillabica per utenti disabili. Genova: Università degli Studi di Genova thesis.

PORT, ROBERT F., and ADAM P. LEARY. 2005. Against formal phonology. *Language* 81.4.927–64.

R DEVELOPMENT CORE TEAM. 2010. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: http://www.R-project.org.

ROACH, PETER. 1998. Some languages are spoken more quickly than others. *Language myths*, ed. by Laurie Bauer and Peter Trudgill, 150–58. London: Penguin.

SAMPSON, GEOFFREY; DAVID GIL; and PETER TRUDGILL (eds.) 2009. *Language complexity as an evolving variable*. (Studies in the evolution of language 13.) Oxford: Oxford University Press.

SCHILLER, NIELS O. 2008. Syllables in psycholinguistic theory: Now you see them, now you don't. *The syllable in speech production*, ed. by Barbara L. Davis and Krisztina Zadjó, 155–76. New York: Lawrence Erlbaum.

SCHWEICKERT, RICHARD, and BRIAN BORUFF. 1986. Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12.419–25.

SEGUI, JUAN, and LUDOVIC FERRAND. 2002. The role of the syllable in speech perception and production. *Phonetics, phonology and cognition*, ed. by Jacques Durand and Bernard Laks, 151–67. Oxford: Oxford University Press.

SERVICE, ELISABET. 1998. The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology* 51A.283–304.

SHANNON, CLAUDE E., and WARREN WEAVER. 1949. *The mathematical theory of communication*. Urbana: University of Illinois Press.

SHOSTED, RYAN K. 2006. Correlating complexity: A typological approach. *Linguistic Typology* 10.1.1–40.

SURENDRAN, DINOJ, and GINA-ANNE LEVOW. 2004. The functional load of tone in Mandarin is as high as that of vowels. *Proceedings of Speech Prosody 2004*, Nara, Japan, 99–102.

TAMAOKA, KATSUO, and SHOGO MAKIOKA. 2004. Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. *Behavior Research Methods, Instruments, & Computers* 36.3.531–47.

TRUBETZKOY, NICHOLAI S. 1939. *Grundzüge der Phonologie*. Travaux du cercle linguistique de Prague 7. [Translated into French as *Principes de phonologie*, by J. Cantineau. Paris: Klincksieck, 1949.]

TRUDGILL, PETER. 2004. Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8.3.305–20.

TWADDELL, W. FREEMAN. 1935. On defining the phoneme. *Language* 11.1.5–62.

VAN SON, ROB J. J. H., and LOUIS C. W. POLS. 2003. Information structure and efficiency in speech production. *Proceedings of Eurospeech 2003*, Geneva, Switzerland, 769–72.

VERHOEVEN, JO; GUY DE PAUW; and HANNE KLOOTS. 2004. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech* 47.3.297–308.

WELLS, JOHN C. 1982. *Accents of English, vol. 1*. Cambridge: Cambridge University Press.

ZIPF, GEORGE K. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press.
ZIPF, GEORGE K. 1937. Statistical methods and dynamic philology. *Language* 13.1.60–70.

DDL - ISH
14 Avenue Berthelot
69363 Lyon Cedex 7
France
[Francois.Pellegrino@univ-lyon2.fr]
[Christophe.Coupe@ish-lyon.cnrs.fr]
[Egidio.Marsico@ish-lyon.cnrs.fr]

[Received 8 October 2009;
revision invited 15 April 2010;
revision received 30 July 2010;
revision invited 31 March 2011;
revision received 18 April 2011;
accepted 18 May 2011]