To appear in P. Mondal. (Ed.), *Handbook of Linguistic Semantics - Bridging Theory, Philosophy and Cognition*. Springer.

Communicative efficiency and semantics: Language variation, universals and change

Natalia Levshina
Centre for Language Studies, Radboud University Nijmegen, The Netherlands
natalia.levshina@ru.nl

Abstract

There is substantial evidence that the pairings of form and meaning in human languages are influenced by language users' tendency to be communicatively efficient. Most importantly, we often express more accessible meanings with shorter forms, and less accessible meanings with longer forms. This helps us save time and effort. Secondly, an efficient linguistic system will contain multifunctional (ambiguous) units, which are beneficial for language users. Thirdly, the semantic relationships between meaningful units determine the relative order of these units in words and sentences, whereby more conceptually accessible elements tend to be mentioned early, and semantically associated elements tend to occur close to each other. This is efficient because it helps to save costs of language production and processing.

This chapter begins with a short introduction of the main principles of communicative efficiency, which have been discussed in the literature. Next, it provides examples of efficient form-meaning pairings and orders found across human languages, showing that the pressure for communicative efficiency can explain both language-specific phenomena and linguistic universals. The chapter also proposes possible mechanisms of language evolution that can explain the emergence of efficient form-meaning pairings. These mechanisms are based on pragmatic principles and heuristics and reflect the "invisible hand" approach to language evolution. In addition, the chapter discusses several alternative explanations of efficient form-meaning pairings, such as the iconic motivation, frequency-based explanation and the Monotonicity Hypothesis. It is argued that one of the core principles of efficiency, namely the principle of negative correlation between accessibility and costs, provides the most natural explanation for the linguistic phenomena in question.

1. Communicative efficiency: definition and main principles

Generally speaking, efficiency means minimization of the cost-to-benefit ratio. Language users communicate efficiently when they save different types of costs involved in language processing and production (most importantly, time, cognitive and articulatory effort), while making sure that they achieve their communicative goals. Under many different labels, these ideas have been around for a long time. Already Gabelentz (1891, pp. 191–195) spoke about the pressure for 'Comfort' (*Bequemlichkeit*), as opposed to the pressure for 'Clarity' (*Deutlichkeit*). Zipf (1949) formulated the Principle of Least Effort. Haiman (1983) discussed economic motivation in language, and so on. Although these and similar ideas were never away from functional approaches to language and typology, there has been a growing interest in communicative efficiency and its effects on language structure and use in recent decades and years (Hawkins, 2004; Gibson et al., 2019; Levshina & Moran, 2021; Levshina, 2022).

As it often happens with popular concepts and terms, efficiency is sometimes used in the sense which was not intended. In particular, it has been understood as only the tendency to reduce effort and time, disregarding the consequences. In this interpretation, efficiency is equated with Gabelentz' (1891) ease and opposed to clarity. However, such a definition is inadequate because a language user who fails to achieve their communicative goals cannot be considered efficient. Even a bulky complex expression can be efficient if the benefits it provides are regarded as important. For instance, legalese often involves long, wordy and complex sentences, which are hard to process. As the study by Martínez et al. (2023) suggests, the main reason for using legalese by the professionals seems the lawyers' reliance on old, pre-existing templates, which are laden with convoluted language. Recycling old templates not only helps the lawyers to save time and costs when drafting new contracts, but it also gives them confidence that the language is reliable because it has "worked" before (Martínez et al., 2023). Although there is certainly room for improvement, these important benefits may justify the processing costs of legalese. Another motivation can be to signal the authoritative nature of law (Martínez et al., 2024).

Since human communication involves diverse costs and benefits, different languages may "choose" different ways of maximizing efficiency. For example, human languages have vastly divergent systems of kinship terms, but what unites them is that their complexity (which represents the cognitive costs of learning and remembering the terms) is in a trade-off relationship with the communicative costs of the addressee picking the wrong referent. As a

result, the existing systems lie on a Pareto frontier in the space of all possible systems (Kemp & Regier, 2012). At this frontier, the cost-to-benefit ratio is minimized.

Language users who communicate efficiently, or linguistic systems that are considered efficient, usually follow several principles, which have been described in the literature. Crucially, we often observe that more predictable, expected, stereotypical, etc. meanings are usually expressed by less costly forms (normally, shorter forms with fewer articulatory details) than less predictable, expected, stereotypical ones. For example, Hawkins' principle 'Minimize Forms' says that the preference for minimization of the formal complexity of linguistic forms at the phonological, morphological, lexical and phrasal levels depends on the ease with which the intended meaning (or another property that must be transferred) can be assigned to the form (Hawkins, 2014, p. 15). In Levshina (2022), this and other manifestations of efficiency are generalized under the principle of negative correlation between accessibility and costs.

To use the notion of accessibility, we need to define it. This is not easy: accessibility has many types and levels and involves different cognitive mechanisms and processes. An important type is called conceptual accessibility, which has been defined as 'the ease with which the mental representation of some potential referent can be activated in or retrieved from memory' (Bock & Warren, 1985, p. 50), and which depends on such parameters as concreteness, imageability, animacy and discourse status, which includes givenness, recency of mention, syntactic prominence, topicality and expectancy, i.e., predictability of a referent for being mentioned in upcoming discourse (Bock & Warren, 1985; Ferreira & Yoshita, 2003; Arnold, 2010; Tanaka et al., 2011). Accessibility of referents also depends on their presence or absence in the immediately available physical context, as well as on encyclopaedic knowledge (Ariel, 2001).

One also speaks about perceptual accessibility (Gleitman et al., 2007), which determines the attention focus and is similar to perceptual salience. This is what we are likely to pay attention to or look at. Human referents are more perceptually accessible than animals and inanimate objects, but perceptual accessibility can also be increased by non-linguistic cues, such as arrows pointing to an object in a picture (Tomlin, 1997; Gleitman et al., 2007, although the effects of such cues on language production display cross-linguistic variation, e.g., Dolsheid & Penke, 2023).

Another type is lexical accessibility, or 'the ease with which the representations of word forms can be recovered from memory' (Bock & Warren, 1985, p. 52), and which is often measured in lexical decision tasks and picture naming experiments. Lexical

accessibility is influenced by different factors, which include frequency, priming, animacy, and the existence of competitors, i.e., phonologically or conceptually similar words (e.g., Gordon, 1983; Bock, 1986; Lee et al., 2022). The notion of accessibility has also been applied to grammatical distinctions and forms. For example, active verb forms are considered more accessible than passive verb forms, and it has been claimed that active constructions are on average more accessible than passives (Gleitman et al., 2007).

Moreover, accessibility is mentioned in Relevance Theory (Sperber & Wilson, 1995) as a property of the contextual assumptions needed for recovery of the intended meaning. The more frequently a certain assumption is used for inference, the greater the accessibility of this information. For example, in (1) the implicit contextual assumption necessary to interpret B's reply is that travelling with the Deutsche Bahn is normally not enjoyable. This assumption is accessible to anyone who has travelled in Germany recently or has heard the reports from international fans at the UEFA championship in 2024. To other people, it is unlikely to be accessible, so they may fail to infer the intended meaning (B doesn't expect to enjoy the trip).

(1) A: *I hope you'll enjoy your trip!*
B: *I'm travelling with the Deutsche Bahn…*

Despite their diversity, all these phenomena seem to have a common underlying cognitive and neural mechanism, namely, the degree of activation, which can be lower or higher. If it is low, we say that the item is in the long-term memory, or in the resting state; if it is high, the item is said to be in the working memory, or in the activated state (Sharwood Smith, 2017, Ch. 6). Recently mentioned and topical referents are more activated, which explains the effects of discourse status. The facilitatory frequency effects can be explained by the fact that an item with a higher resting state is easier to activate than an item with a lower resting state. Representations can be co-activated if they have been linked in previous experience. This process is known as spreading activation. It can lead to competition between multiple representations, making each of the competitors less accessible. Importantly, neural activation provides us with a general framework that accounts for diverse types of accessibility, including accessibility based on attention processes (Sharwood Smith, 2017, Ch. 7).

I propose the following definition of accessibility for a theory of communicative efficiency:

(2) Definition of Accessibility

Accessibility is the degree of activation of the mental representation of a linguistic form or meaning, which determines how easily this form or meaning can be used or retrieved in language production and comprehension at a given point of communication.

The notion of accessibility has been fruitful for explaining the choice of referential expressions in discourse (Ariel, 1990, 2001; Arnold, 2010). Referents with more accessible mental representations are usually expressed by more compact forms, such as pronouns (e.g., *she*), whereas less accessible referents tend to be expressed by longer forms, such as proper nouns and full nominal phrases (e.g., *Dr. Susan Jones* or *the professor of linguistics*). This is a perfect example of negative correlation between accessibility and costs. But we can also extend this principle to other domains, including conventionalized grammatical patterns. For example, Haspelmath's (2021) grammatical form-frequency correspondence hypothesis, which predicts shorter coding for more frequent grammatical categories in contrasts like singular vs. plural, can also be interpreted as an example of negative correlation between accessibility and costs. More examples are provided in Section 2.

Another application of accessibility has to do with the order of meaningful elements – morphemes, words and phrases. For instance, the referents of subjects tend to be animate, given, as well as more imageable, concrete and salient than the referents of objects (see Bock & Warren, 1985). This is why subjects are produced first in most of the world's languages (Greenberg, 1963; Dryer, 2013). However, patients (which are normally encoded as direct objects) are mentioned first as subjects of passive sentences more often if the patient has been pre-activated by its lemma or picture, primed by a semantically related word or picture, or its referent has been pointed at (Gleitman et al., 2007). In Levshina's (2022) framework, these effects are explained by the more general principle of maximization of accessibility. It is efficient to produce the most accessible information at every point of discourse. When beginning a sentence, it is efficient to produce the constituent with the most accessible representation first, while conceptualizing and formulating the other, less accessible parts of the utterance at the same time. This strategy saves time in language production.

Language users also tend to maximize accessibility by placing semantically related elements closer to each other. Doing so helps to save memory costs. This idea, which is known as the principle of information locality (Futrell & Levy, 2017; Futrell 2019), also explains the principle of minimization of dependencies (Ferrer-i-Cancho, 2004; Liu, 2008; Futrell et al., 2015), as well as Hawkins' principle 'Minimize Domains', according to which

language users tend to minimize the distance separating interrelated items, and the material that needs to be processed simultaneously as the relationship between the items is processed (Hawkins, 2014, p. 11). See more examples of efficient order in Section 3.

The principle of maximization of accessibility also leads to an important but frequently overlooked preference in comprehension. When interpreting language, it is most efficient to choose the most accessible interpretation, for example, the most accessible word sense. See Levshina (2022: Ch. 4) for more information.

The mapping of meanings and forms is normally not one-to-one. In fact, it is often many-to-one (ambiguity, including homonymy, polysemy and vagueness) and one-to-many (synonymy). Chomsky (2002, p. 107) argued that the existence of ambiguity is a sign that language was not shaped for communicative needs. This is erroneous, in my view. First of all, if we consider the use of language in context, we will see that ambiguity as a threat for communication is overrated (Wasow, 2015). Context helps to boost the accessibility of the intended interpretation. Secondly, as will be explained in Section 4, semantic systems with multifunctional units have benefits both for the language producer and the addressee because such units help to save the total amount of articulatory effort and time (Piantadosi et al., 2012).

As for synonymy, it is restricted by several pragmatic principles like the Principle of No Synonymy (Goldberg, 1995, 2019), such that the number of actual alternatives in a specific instance of language use does not represent an impediment to language production. At the same time, communicative efficiency requires a certain flexibility in verbalization of the same experience as a playfield in which different formal variants may compete, so that efficient conventionalized form-meaning pairings may emerge as a result of pragmatic processes.

These most important manifestations of communicative efficiency in semantics are considered below. Section 2 provides details about the negative correlation between accessibility of meaning and the costs of the form expressing it, discussing the pragmatic mechanisms of the evolution of efficient form-meaning pairings, and several alternative explanations of their emergence. Section 3 examines information locality effects on the order of meaningful elements. After that, Section 4 focuses on the role of ambiguity and synonymy in efficient communication. Finally, Section 5 presents a summary and several open questions for future research.

2. More accessible meanings are expressed by less costly forms

2.1. Prominent illustrations

Taking an onomasiological approach (i.e., starting with a meaning and asking which forms can express it), we can say that a meaning with higher accessibility is more likely to be expressed by a less costly form than a meaning with lower accessibility. Taking a semasiological approach (i.e., from a form to possible meanings), a less costly form is more likely to invoke a high-accessibility interpretation, whereas a longer and more costly form is more likely to be associated with a low-accessibility interpretation.

The evidence for this principle is abundant both in language use and in language structure. As mentioned in Section 1, we observe negative correlation between the length of referential expressions in discourse and accessibility of their referents (Ariel, 1990, 2001; Arnold, 2010). Consider the beginning of one chapter from Agatha Christie's crime novel *The Secret Adversary*:

(3) *The Prime Minister tapped the desk in front of him with nervous fingers. His face was worn and harassed. He took up his conversation with Mr. Carter at the point it had broken off.* (A. Christie, The Secret Adversary, Chapter XXII – In Downing Street).

The first mention of the Prime Minister is a full nominal phrase. After the referent has been introduced, it is highly accessible and is therefore referred to by the shorter pronominal forms *him*, *his* and *he*. This is a typical way of using referential expressions. In many languages, one can completely drop arguments that express highly accessible referents. In that case, given and topical arguments can be omitted, whereas new focal ones, which are not accessible, are always expressed (Goldberg, 2005).

Glass (2020) also reports that verb objects are more frequently omitted in Reddit communities with special interests than in general-interest conversations. For example, home brewers often use the verb *bottle* without the object (that is, beer) because it is highly accessible for the members of that community. This strategy is efficient.

The negative correlation between accessibility and costs can also be illustrated by pairs of referents in which the referent associated with a social or professional stereotype is expressed by a shorter form. A classical example is a (female) *nurse* vs. a *male nurse*. Another illustration is the official name *Germany national football team*, which is the men's

team, vs. the *Germany women's national football team* (in German, *Deutsche Fußballnationalmannschaft* and *Deutsche Fußballnationalmannschaft der Frauen*, respectively). Also, a German family will ponder whether to buy a "normal" car (*Auto*), which drives on a fossil fuel, or to purchase an electric car (*E-Auto* or *Elektroauto*), which are still less common. When preparing for a grill party, they can choose between (meat) sausages (*Würstchen*) or vegan sausages (*vegane Würstchen*) in the supermarket. Returning to sports, field (and indoor) hockey in Germany is governed by the *Deutscher Hockey-Bund*, whereas the *Deutscher Eishockey-Bund* is responsible for ice hockey. In Russia with its strong winter sports traditions, the pattern is reversed: *xokkei* means by default 'ice hockey', whereas field hockey is expressed by the phrase *xokkei na trave* 'hockey on grass'.

Efficient pairings are restricted neither to words and phrases, nor to lexical meanings. There is a robust cross-linguistic tendency to express less frequent grammatical categories with longer forms and more frequent categories with shorter forms (Greenberg, 1966). For example, singular forms are normally more frequent than plural forms, whereas the latter are more frequent than dual forms. This means that the grammatical meaning of singular forms is usually more accessible than that of plural forms, and so on. According to Greenberg (1966), singular forms are more often formally unmarked than plural forms, and plural forms are more often unmarked than dual forms in languages of the world. Although there are languages in which both singular and plural are formally marked (e.g., Latin nouns, *lup-us* 'wolf-NOM.SG' and *lup-i* 'wolf-NOM.PL') as well as languages without any systematic number marking (e.g., Pirahã), a language with a coding asymmetry in marking will express the singular meaning with a shorter form, and the plural meaning with a longer form. An example is English: *wolf – wolv-es*. The same holds for plural and dual. This is efficient.[1]

Other examples of categories that exhibit the negative correlation between accessibility (captured by relative frequency) and costs (captured by formal length) include active and passive voice, present and future tense and positive, comparative and superlatives degrees of comparison (Greenberg, 1966; Croft, 2003; Haspelmath, 2021).

In many languages, the masculine plural form of nouns referring to human beings is considered the "default" and "unmarked" in the sense that its interpretation can include both men and women (Aikhenvald, 2016). For example, the German form *die Kollegen* 'the

[1] Some lexical categories (e.g., eyes, shoes or bees) tend to occur more often in the plural, which may contradict this theory. But interestingly, if a language also has a special number marking for certain nouns, such that the plural is the shorter form, and the singular is the longer, marked form (e.g., singulatives in Welsh, e.g., *deilen* 'leaf' vs. *dail* 'leaves'), the nouns that occur more often in the plural are also likely to have this special marking, as shown in a cross-linguistic study by Haspelmath and Karjus (2017).

colleagues' can be used to represent both female and male colleagues. At the same time, it is also the only form to be used when talking only about male colleagues. This can be due to the higher accessibility of male professionals, which existed until recently.

In order to avoid the socially undesirable ambiguity and the risk of interpreting plural forms as male-only, it is considered appropriate by many people to use these forms along with the plural feminine forms, e.g. in German, *Kolleginnen und Kollegen* 'colleagues (female) and colleagues (male)'. Since these forms are costly, other strategies have been invented, for example, the so-called *Gendernstern* 'gender-star' (an asterisk followed by the feminitive suffix, as in *Kolleg*innen* 'colleagues [male or female]'), a gap or a slash (e.g., *Kolleg_innen* and *Kolleg/innen*) or writing the suffix with a capital -I- (e.g., *KollegInnen*). In spoken language, the suffix is separated by a glottal stop to distinguish the word from the female-only form *Kolleginnen*.

These forms have provoked a heated debate. The opponents argue that they sound unnatural and spoil the German language.[2] On the other hand, there is psycholinguistic evidence from different languages that using the default masculine forms can produce a statistically significant male bias in the interpretation (e.g., Redl et al. 2021). Apparently, the male-only interpretation is still more accessible than the inclusive one. Although dictionaries and grammars mention both meanings, there is a gap between the psychological reality and what the books prescribe. When arguing for "efficient" language, the opponents of gender-inclusive language focus on saving articulation (or writing) costs, wilfully or inadvertently ignoring the social and moral costs of making half of the population less visible. Once we consider these hidden costs, the "efficiency" of using the default masculine forms becomes problematic.

But why is it efficient when more frequent meanings are expressed by shorter forms? The total length of a text is shorter when the units that occur often in that text are short, whereas rare units can be long. In Information Theory, this is known as efficient compression (Cover & Thomas, 2006). It has been argued that Zipf's law of abbreviation, according to which frequent words tend to be shorter than infrequent words, is a manifestation of compression as a universal principle of human and animal communication (Ferrer-i-Cancho, Hernández-Fernández et al., 2013; Bentz & Ferrer-i-Cancho, 2016). Zipf himself illustrated his principle of Least Effort in language using the metaphor of a language user as an artisan

[2] The German state of Bavaria has recently forbidden gender-inclusive forms with asterisks, colons and other symbols in official documents and in public places like schools and universities.

who wants to do their work (communication) as efficiently as possible with their tools (words). The artisan can reduce the amount of work by reducing the mass of the tools that are used frequently (Zipf, 1949, pp. 60–61). Therefore, by reducing frequently used words or grammatical forms, language users save effort and time in the long run.

More recently, there has been work arguing that word lengths (in characters of phonemes) can be best predicted by informativity, or the average surprisal (negative logarithimically transformed conditional probability) of a word given its neighbours (Piantadosi et al., 2011). It is claimed that the correlation of word lengths with the words' contextual predictability helps to maintain unform information density, which ensures efficient information transfer (Fenk & Fenk, 1980; Jaeger, 2006; Levy & Jaeger, 2007; but see Levshina, 2022a, p. 21). However, later examinations using cleaned-up data and more diverse samples of languages have shown that only frequency displayed a systematic negative correlation with word length across languages (Meylan & Griffiths, 2021; Koplenig et al., 2022; Levshina, 2022b; Pimentel et al., 2023). At the same time, there is sufficient evidence that deletion of phonemes, phonological duration and amount of articulatory detail depend on conditional probability of the target units (phonemes, syllables and words) given their context – computed on average and in a specific context (e.g., Aylett & Turk, 2004, 2006; Cohen Priva, 2008; Seyfarth, 2014).

It is reasonable to expect that efficient language users reduce their time and effort whenever they can without jeopardizing the successful transfer of information. This explains the shortening of forms that represent highly accessible meanings. Which statistical measures (frequency, conditional probability given context, average predictability, etc.) are the best for representing accessibility in different situations is a question that requires more investigation.

2.2.Pragmatic mechanisms and evolution of efficient form-meaning pairings

How do efficient form-meaning pairings emerge? Obviously, language users do not aim to optimize language for communication. Only very rarely do we consciously decide on the use of efficient form-meaning mappings (e.g., when we negotiate the use of acronyms and abbreviations). Instead, efficient form-meaning pairings emerge as a result of an "invisible hand" process (Keller, 1994), similar to a path in a field, because of the "shortcuts" taken by language users when they pursue their communicative goals. Effective and easy "shortcuts" have a chance of becoming widely spread and even conventionalized in the language. In this

regard, language change is similar to biological evolution. However, unlike stochastic mutations in DNA, the emergence of efficient forms is not random. It is driven by general principles of human communication, which have been formulated in several pragmatic theories.

One of the most important contributions is Grice's (1975) cooperative principle, which assumes that humans communicate rationally and expect the same from their interlocutors. The tendency to use less costly expressions for transfer of more accessible information can be covered by the maxim of Quantity (more specifically, the submaxim 'Do not make your contribution more informative than required') and one of the submaxims of Manner, which says 'Be brief (avoid unnecessary prolixity)' (Grice, 1975). As for the use of more costly expressions for transfer of less accessible information, it can be explained by the second submaxim of Quantity, 'Make your contribution as informative as is required', and one of the submaxims of Manner, 'Avoid ambiguity'. For example, if the speaker says, 'She is smart' in the situation where several competing referents have a similar level of activation, the referent intended by the speaker is less accessible, which results in ambiguity.

Grice's maxims distinguish between the quantity of information (how much is communicated) and the formal properties of linguistic expressions (how it is communicated). However, in practice they are often closely related. Marked grammatical categories, for example, have both more coding material and are more informative than their unmarked counterparts, which can leave the grammatical meaning unspecified. Similarly, 'male nurse' is both more verbose and more informative than simply 'nurse'. In addition, frequent words, which are more compact, tend to be more polysemous because they more readily acquire new meanings (Zipf, 1949; Harmon & Kapatsinski, 2017). This is why it is possible to present more parsimonious accounts. A prominent example is Horn (1984), who argued that most of Grice's maxims can be replaced by two principles: the Q principle 'say as much as you can', and the R principle 'say no more than you must', which reflect the tension between the speaker's and the addressee's effort. When the addressee hears a relatively complex and verbose expression and knows that there exists a simpler, less effortful way of saying a similar thing, they will infer that the message is also somehow semantically "marked", for example, representing a non-typical, less salient member of the category (Horn, 1984, p. 22). In other words, the more costly form signals a less accessible meaning. Following the R principle, the addressee infers that this extra effort was necessary, otherwise the R principle would have been violated by the speaker. As a result of language use, the less costly alternative can then become associated with the more accessible meaning, representing a

stereotypical or particularly salient member of the category (Horn, 1984; see also Bybee, 1994 on grammaticalization of zero), whereas the marked alternative can become associated with the marked meaning based on the Q principle.

Horn's theory was questioned by Levinson (2000), for whom the conflation of formal prolixity and amount of information was too radical. In particular, the Q principle in the discussion of the division of pragmatic labour is primarily focused on linguistic form (prolixity), whereas in other contexts, such as scalar implicatures, its application involves only differences in informativity. For instance, the word 'all' is semantically stronger, or more informative than 'some' (which is why 'I've eaten some chocolates' implicates 'I didn't eat all chocolates'), but it is not more formally complex and does not require more effort from the speaker to produce. To address such issues, Levinson formulated his own I-, M- and Q-heuristics. Discussing them in full detail is beyond the scope of this chapter, but it is useful to give examples of how these heuristics can explain efficient language patterns.

The I-heuristic says, in a short version, 'What is expressed simply is stereotypically exemplified' (Levinson, 2000, p. 37). An example is *I saw a nurse*, which implicates that the referent of *a nurse* is female. Another illustration is English nominal compounds, such as *a bread knife* 'a knife for cutting bread' or *a kitchen knife* 'a knife used in the kitchen'. These stereotypical (and accessible) interpretations are expressed by simple, low-cost forms. Similarly, *an ice knife* is normally a knife for cutting and carving ice. But if someone wants to speak about the material a knife is made of, they will say, for example, *a knife made of ice*, following Levinson's M-heuristic: 'What's said in an abnormal way isn't normal' (Levinson, 2000, p. 38). In this case, the less accessible meaning is expressed by a more costly form. Finally, the Q-heuristic says, in a concise formulation, 'What isn't said, isn't' (Levinson, 2000, p. 35). It has to do with informational strength, as in scalar implicatures (cf. *some* vs. *all*).[3] However, as mentioned above, more information usually means longer forms. Thus, for the purposes of explaining communicatively efficient form-meaning pairings all these theories display more similarities than differences.

The pragmatic theories also make one important point. A linguistic form does not only convey a conceptual meaning about some state of affairs in the world, but it also has a procedural meaning, guiding the addressee towards the intended interpretation. By using a low-cost form, the language producer signals that the meaning is highly accessible. By using

[3] Many examples of the Q-heuristic reflect another core principle of efficient communication, namely, the positive correlation between benefits and costs (see Levshina, 2022, Section 1.4.1), so this heuristic is also relevant from the perspective of communicative efficiency.

a high-cost form, they signal that the meaning is not very accessible. This idea is central in Ariel's (1990, 2001) Accessibility Theory, according to which the speaker guides the addressee's retrieval of a referent by using different coding strategies that serve as cues about the degree of the referent's accessibility: for example, pronouns signal higher accessibility than full nominal phrases.

The general principles of rational communicative behaviour have been formalized and tested empirically in Rational Speech Act Theory (Frank & Goodman, 2012; Degen, 2023), including a penalty for using formally complex expressions in the presence of less costly alternatives (Bergen et al., 2016). At the same time, it is a matter of ongoing debate to what extent these models capture real communication. Some empirical evidence shows that the rational models, which involve the hearer performing recursive "reverse-engineering" of the speaker's communicative intentions (Degen, 2023), do not predict addressees' communicative behaviour better than a baseline model driven solely by literal word meaning and priors reflecting the contextual salience of referents (Sikos et al., 2021). It seems that the rational behaviour with the aim of being communicatively efficient is only "good-enough", similar to "good-enough" language comprehension and production (Ferreira, 2003; Goldberg & Ferreira, 2022).

Even though we do not have a fully spelled out mathematical model explaining efficient communication yet, there is plenty of evidence that linguistic forms adjust to changes in the accessibility of meanings. For example, when a new gadget or technology become popular, the name is often shortened by clipping, e.g., *telephone > phone*, *application > app,* German *Automobil > Auto* 'car', or with the help of an acronym, e.g., *short message service > SMS*, *artificial intelligence > AI*.

Consider the change in the frequencies of the German words *Auto* and *Automobil* in Google Books Ngrams in Figure 1. Although *Automobil* was on the rise in the beginning of the 20th century, it was taken over by *Auto*, which has become the overwhelmingly dominant variant. At the same time, the longer variant *Automobil* did not disappear completely. It is used now in pragmatically marked elevated or ironic contexts. This semantic change is efficient because it helps to preserve the negative correlation between communicative costs and accessibility of meaning.

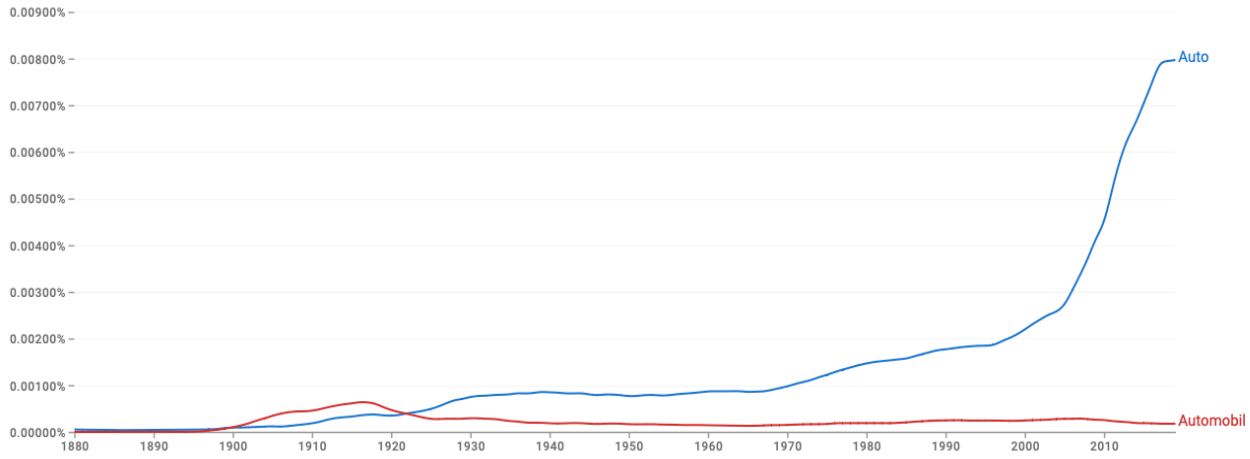

Figure 1. Probabilities of *Auto* and *Automobil* in the German (2019) Google Books corpus (Google Books Ngram Viewer, https://books.google.com/ngrams/).

Another important strategy is lexical replacement. For example, in English the word *automobile* was replaced by the shorter word *car* (Zipf, 1965[1935], p. 33). An additional illustration is from Tenepaja Tzeltal, a Mayan language spoken in Mexico (Witkowski & Brown, 1983). The conquest led to import of sheep to this mountainous region. At first, a sheep was referred to by a composite expression, *tunim čih* 'cotton deer'. With time, sheep became widely spread and played an increasingly important role in the life of the community. As a result of the increase in the meaning's accessibility, the word *čih* began to designate sheep on its own, as the adjective 'cotton' was dropped. The name for deer, who apparently became less frequently mentioned than sheep, became *te?tikil čih* 'wild sheep'. Notably, in Bachajón Tzeltal, a closely related language spoken in the lowlands, where sheep are less common, the name for sheep still stands for 'cotton deer'.

Overall, we can say that meanings "compete" for shorter and more easily accessible forms. The more accessible meanings usually win. The longer forms are often not discarded and begin to specialize in less accessible meanings (including pragmatically marked ones).

These ideas find support in artificial language learning experiments. Kanwal et al. (2017) used two unfamiliar plant-like objects as stimuli. Each of the objects was introduced under two names. One name was short and ambiguous because it was used to name both objects, whereas the other was long and referred to the object in a unique way. For example, Plants A and B were called each 50% of the time *zop*, which was short and ambiguous. In addition, Plant A was called 50% of the time *zopekil*, whereas Plant B was called *zopudon* (both are long and unique names). The participants saw one of the plants more frequently than the other one. After the learning phase, the participants performed a director-matcher task, in which

one of the participants was the director, who had to describe an object, and the other was the matcher, who had to guess which object was meant. The experiment showed that the directors used the shorter names like *zop* more often to name the more frequent objects. Crucially, two conditions had to be met for the efficient form-meaning mapping to emerge in the experiment. The first condition was a pressure for accuracy: the participants were interested in maximizing the number of correct matches. The second condition was a pressure for saving time, when the longer names took visibly longer to transmit than the shorter names.

An experimental proof for the causal link between frequency (accessibility) of grammatical meanings and formal marking has been provided by Kurumada and Grimm (2019), whose miniature language learning experiments revealed that learners preferred using plural marking on nouns that were less probable to occur with plural meaning, although the linguistic input did not contain that bias. Levshina (2019), likewise, used an artificial language learning experiment to show that participants had a small but statistically significant preference for efficient form-meaning pairings in expression of causation events. When taught pairs of a shorter and a longer allomorph for expressing two causing events that occurred with different frequencies, language learners chose the shorter allomorph to describe the more frequent causing event more often than to express the less frequent one.

2.3. Competing explanations: the case of causative constructions

Although the form-meaning pairings discussed above are efficient, other accounts of their emergence have been proposed in the literature. This subsection discusses some of them. A phenomenon that allows us to compare different accounts is expression of causation. Consider the examples in (4).

(4) a. Harry Potter raised the cup.
b. Harry Potter caused the cup to rise.

Both sentences express the same caused event: the cup was raised. Likewise, the causer is also the same: Harry Potter. However, the causing events are different. In (4a), in which the lexical causative verb *raise* is used, the most likely interpretation is that Harry raised the cup in the most expected way, i.e., using his hands. In (4b), which contains a periphrastic causative *cause to rise*, the choice of the form suggests that the cup rose because Harry had done something special – for example, used magic (cf. McCawley, 1976).

The "division of labour" between the lexical causative in (4a) and the periphrastic causative in (4b) is efficient. The causing event expressed in (4a) is more accessible than the interpretation generated by (4b), based on the frequency of the events in the world.[4] The pragmatic principles outlined in Section 2.2 allow us to explain how such division of labour can emerge. When the addressee sees or hears a sentence in (4b), they will assume that the costly expression has been chosen by the speaker or writer for a good reason, and that a less accessible interpretation is meant. With time, this division of labour can become conventionalized.

However, one also finds other accounts in the literature. First of all, the formal differences between the causative constructions may be motivated by the Iconicity Principle (Haiman, 1983): language users tend to choose forms that resemble (or isomorphic with) the conceptualizations that they want to express. Lexical causatives like *raise, kill, break*, etc. have no distance between the causing and caused events, which are merged in one lexeme. In contrast, periphrastic (or analytic) causatives like *cause* + *to*-Infinitive express the very abstract cause and its effect by different words, which can be separated by other elements (in our example, the noun phrase *the cup*). The greater distance and lower formal integration in the more complex causatives perfectly mirror the meaning these forms express – that is, less direct causation.

Note that iconicity and communicative efficiency are not mutually contradictory. One complication in this debate is the existence of many different "flavours" of iconicity (cf. Haspelmath, 2008). Some of them may facilitate language processing, which makes them potential contributors to communicative efficiency, for example, iconicity of sequence, when the sequence of forms matches the sequence of experiences. Another type of iconicity, which is called by Haspelmath (2008) iconicity of contiguity, means that forms that belong together semantically occur next to each other. The reader may recognize in this definition the principle of information locality, which in its turn is a manifestation of the principle of maximization of accessibility. Yet another illustration is iconic correspondences between referents and signs in sign languages, which can help to save time and reduce memory costs, especially when encoding complex events (Slonimska et al., 2020). Unlike these examples of "efficient" iconicity, iconicity of cohesion (more cohesive meanings are represented by more cohesive forms) is not necessary for explaining the typical form-meaning pairings in

[4] Even in the magical world, I would assume that it is easier to use your hands than magic to raise a normal-sized cup.

causative constructions. Instead, the greater cohesion of direct causatives is explained by their more advanced degree of grammaticalization (or lexicalization in lexical causatives like *raise* or *fell*), which has to do with their greater frequency in comparison with indirect causatives (Haspelmath, 2008, p. 20).

There are several theoretical issues with the iconicity account of causative constructions. First of all, directness and indirectness of causation are very difficult to define. They are not semantic primitives. In fact, many different semantic distinctions have been interpreted as criteria for (in)directness of causation. For instance, Haiman (1983) mentions the presence or absence of physical contact between the Causer and the Causee, whereas Verhagen and Kemmer (1997) focus on who is the main source of energy for the caused event – the Causer or another entity. Some researchers discuss clusters of conceptual features related to (in)directness. For example, Bohnemeyer et al. (2010) list such features as mediation, contact and force dynamics. Levshina (2022, p. 158) identifies six parameters: spatiotemporal integration, physical contact, the source of energy, affectedness of the Causee, length of the causation chain and the presence of impingements. These distinctions are often correlated (Levshina, 2016), but they are not equivalent. Martin and Schäfer (2014) demonstrate that in order to explain the semantic differences between lexical and analytic causatives in different languages, the direct vs. indirect distinction has to be defined in different ways. It is unclear how to reconcile the multifaceted character of (in)directness with the iconicity account without being circular.

Second, if we take a broad cross-linguistic perspective, we discover that not all semantic differences between more or less compact or formally integrated causative constructions boil down to (in)directness of causation. Already Dixon (2000) presented a multivariate account of cross-linguistic variation of causative constructions, which was empirically supported by a quantitative study in Levshina (2022, Ch. 7). Other parameters, such as intentional vs. accidental causation, also tend to be associated with the formal differences. Wolff (2003) found that English participants of an experimental study avoided using lexical causatives to describe situations with mediated (indirect) causation. However, the participants were more willing to use lexical causatives in such situations if the Causer was sentient or acting intentionally.

Crucially, the cross-linguistically observed associations between forms and meanings are easily explained by the differences in accessibility between different types of causation. Unmediated causation is more accessible than mediated causation, intentional causation is more accessible than unintentional causation, and so on, because we speak of unmediated,

intentional etc. causation more often (see the corpus counts in Levshina, 2022, Ch. 7). Thus, the principle of negative correlation between accessibility and costs provides a more comprehensive account of the cross-linguistic variation than the iconicity explanation based on the isomorphism of cohesive forms and meanings.

The functional approach to semantics and typology developed here is closely related to the usage-based approach, whose famous proponent has been Joan Bybee (1985, 2010). The central role in usage-based linguistics is played by frequency. An important effect of frequent repetition of a linguistic expression in language use is automatization and routinization of articulation. As a result, the expression becomes shortened and its components become one chunk, e.g. *going to* becomes *gonna*. It is very likely that these processes can be responsible for the formal reduction of frequently used causatives, e.g., when morphological causatives become no longer analysable. However, not all processes can be explained by chunking and optimization of articulation routines. For instance, the English causative *make* + bare infinitive was originally used with the *to*-infinitive, but lost it, as it became the most frequent causative. It is unlikely that this was due to routinization of articulation and chunking because there can be other phonological material between the two verbs, e.g., *She made her little daughter laugh*.

Yet another explanation can be found in formal semantics – at least, for word pairs differing in their morphological complexity. The Monotonicity Hypothesis formulated by Koontz-Garboden (2007) says that operations involved in word formation do not remove semantic operators from lexical semantic representations. For example, the adjective *red* names a simple state, the verb *redden* names a change into that state, and the participle *reddened* names a state brought about by a change of state event named by the verb *redden*.

Looking more closely at causatives, one can see why this view may be problematic. Many languages have anticausative forms that are derived from causative (or semantically, causal) ones. For example, *romper* in Spanish means 'break$_{tr}$'. This is a causal verb, which can be classified as a lexical causative. But there is also a reflexive verb *romperse* 'break$_{intr}$', or 'become broken', which is formed from the causal verb with the help of the reflexive morpheme and whose meaning is called inchoative. A standard formal semantic representation of this causative–inchoative pair is presented below (Koontz-Garboden, 2009, p. 81):

(5) a. Inchoative *break* [e.g., *romperse*]
λ*x*[*BECOME not-whole(x)*]

b. Causative *break* [e.g., *romper*]
λ*x*λ*y*[*y CAUSE BECOME not-whole(x)*]

Since the semantic representation of the inchoative in (5a) is less complex than that of the causative in (5b), some semanticists believe that the inchoative form is derived from the causative form by deleting the causal operator CAUSE. This goes against the Monotonicity Hypothesis. Koontz-Garboden argues, however, that this is not a problem, and claims that reflexive inchoatives like *romperse* still contain the CAUSE operator as part of their lexical semantic representation. In such change of state events, "the individual undergoing the change is also the EFFECTOR participant in the event that causes the change of state event" (Koontz-Garboden, 2009, p. 86). That is, in an event represented by an inchoative verb like *romperse*, the broken object is also "responsible" for its own change of state. For other types of derivation there are analogous explanations.

Although this theory is originally aimed at capturing derivational processes and does not make claims about semantic complexity, one could say that we observe a correlation between the amount of information expressed by a word and its formal complexity, which can be interpreted as an example of form-meaning isomorphism, or as evidence of the principle of positive correlation between benefits and costs in the theory of efficiency presented here. These would be competing explanations against the principle of negative correlation between accessibility and costs.

There are some problems with such explanations. There is also no consensus among formal semanticists, as far as anticausatives are concerned (e.g., Horvath & Siloni, 2011), and the amount of information expressed by a word is difficult to measure independently of theories. How do we tell, for example, whether the broken object is in some way "responsible" or not responsible for its change of state? In contrast to such theories, the accessibility account is based on independent criteria, such as frequency, predictability or surprisal (cf. Haspelmath et al., 2014 for frequency data). The meanings that are more accessible, such as 'break$_{tr}$' tend to be expressed cross-linguistically by simpler and shorter forms than the meanings that are less accessible – such as 'break$_{intr}$', or 'become broken'. Crucially, the asymmetries can go in either direction, which leads to emergence of pairs like *romper* > *romperse,* in which the causal verb is more basic and the non-causal verb is more

complex, and pairs like *dance* > *make dance*, in which the non-causal verb is more basic and the causal expression is more complex.

But why do these meanings tend to be more or less accessible across human languages? This has to do with semantics, in particular, with the degree of spontaneity of events, as they are construed and described by language users (Haspelmath, 2016). For situations like breaking, the causal 'break$_{tr}$' meaning is more accessible than the non-causal, inchoative meaning 'break$_{intr}$' because we do not often speak of things as getting broken on their own. Instead, we normally speak of an external effector (a person, instrument or force) breaking something, for reasons of perceptual salience, social importance, legal responsibility, etc. This is why the causal verb in this pair is often shorter and less complex than the inchoative verb. In contrast, for situations like dancing, the non-causal meaning is more accessible than the causal one. Verbs like *dance* rarely occur in causal contexts (e.g., we do not often make someone dance). This event is normally construed as a spontaneous one, requiring no external agent. Most people dance because they feel like dancing. This is why languages tend to have a longer causative form for such pairs of events.

To summarize, the efficiency explanation of causatives based on the principle of negative correlation between accessibility and costs should be preferred to the others because of several reasons. First, it includes the semantic distinctions beyond the problematic (in)directness of causation. Second, it is applicable to different grammatical domains and phenomena: not only morphological derivations, but also all kinds of analytic and periphrastic causatives. Moreover, it is based on falsifiable criteria (such as frequency) which can be collected from corpora or tested in experiments.

3. Maximization of accessibility and the order of meaningful elements

Efficiency manifests itself not only in the length and complexity of linguistic forms, but also in their order. As discussed in Section 1, language users prefer to use orders that maximize accessibility. For instance, language producers tend to present more accessible constituents before less accessible ones. In many languages, sentences typically begin with more accessible information.[5] Although in many cases a prominent role is played by discourse accessibility

[5] Note, however, that this tendency is not universal: in some languages it is more conventional to present newsworthy and therefore less accessible information first (Mithun, 1992; Chafe, 1994), cf. also languages with dominant OSV and OVS orders (Dryer, 2013).

(e.g., givenness and topicality), semantic properties are also important. Participants of numerous experiments preferred to begin sentences with constituents (subjects or objects) with animate (Tanaka et al., 2011), concrete and imaginable (Bock & Warren, 1985), or semantically primed referents (Bock, 1986).

In other positions in a sentence, semantic effects on word order have also been observed. An example is variation between the double-object dative (e.g., *Give me the money*) and the *to*-dative (e.g., *Give the money to me*) in English. In the double-object dative, the Recipient comes before the Theme, whereas in the *to*-dative, the Theme is followed by the Recipient. Numerous multifactorial studies reveal that the double-object dative is preferred if the Recipient is animate, and the *to*-dative is preferred if the Recipient is inanimate (e.g., Bresnan et al., 2007), other factors controlled for. At the same time, the effect size of animacy can vary across languages and even varieties of the same language (Bresnan & Hay, 2008).

Another manifestation of maximization of accessibility is the principle of information locality (Futrell & Levy, 2017), which says that words strongly associated semantically and/or grammatically are placed close to each other. This helps to save processing costs of keeping the words in the memory and intergrating the information they express (cf. Yngve, 1960; Gibson, 1998). It also reduces the costs associated with high surprisal, which serves as a good predictor of online processing difficulty.

This principle can explain many phenomena, such as the tendency to minimize dependency distances (Ferrer-i-Cancho, 2004; Liu, 2008; Gildea & Temperley, 2010; Futrell et al., 2015) and the domains necessary for the recognition of constituents (Hawkins, 2004). Another illustration is the order of multiple adjectives in a nominal phrase. The semantic class of the adjective plays an important role in determining its position. The order of adjectives in English is OPINION + SIZE + AGE + SHAPE + COLOUR + ORIGIN + MATERIAL + PURPOSE + NOUN, e.g., *a large [SIZE] wooden [MATERIAL] dining [PURPOSE] table*, or *a beautiful [OPINION] old [AGE] German [ORIGIN] castle*. Notably, these rules seem to be explained by the differences in the mutual information between head nouns and adjectives representing different semantic classes. The lower the (pointwise) mutual information, the further from the noun the adjective is expected to be, as demonstrated in Hahn et al. (2018) and Futrell (2019). Consider evaluative (OPINION) adjectives, which are placed the farthest from the noun, e.g., *a beautiful red dress*, but not *a red beautiful dress*. Evaluative adjectives do not restrict the set of referents, but communicate the speaker's attitude, which depends on their subjective state. What is beautiful for one person, can be ugly for another. The mutual information of such an adjective and its head noun is then relatively low. Notably, as a linguistic

unit develops more subjective meanings, its position also moves towards the periphery of a constituent or clause (Traugott, 2010), which supports the principle. Yet, the order of adjectives has been explained by numerous other semantic and information-theoretic variables (see an overview in Scontras, 2023), the relative contribution of which may depend on the predominant word order of adjectives and nouns in a language (Dyer et al., 2023). More cross-linguistic research is needed to disentangle all these factors.

Another well-known example is related to the order of verbal derivational and inflectional morphemes, which is usually as follows (Bybee, 1985):

(6) Valence > Voice > Aspect > Tense > Mood > Agreement (Person and Number)

Bybee argues that the position of a morpheme is determined by the semantic effect of that morpheme on the root meaning. The categories on the left of the scale in (6) have a greater effect than the ones on the right. For example, valence changes the number and role of participants involved in the event, which can be seen as a dramatic impact on the semantics of the verb. An additional piece of evidence is that valence-related differences are often lexicalized, as in the causative–inchoative pair *kill* and *die*. In contrast, the scope of mood is the whole proposition, so it is less relevant for the semantics of the verb itself. As for agreement markers, such as person and number inflections, they refer to the participants of the event and have little impact on the semantics of the verb.

These tendencies can be explained by maximization of accessibility and, more specifically, information locality. If we take aspect markers, which have a strong impact on the meaning of the verb, they would also be less freely applicable to different verbs than more peripheral markers. For example, an imperfective marker is more compatible with durative verbs than with punctual ones. Therefore, mutual information of the root and the affix will be relatively high. In contrast, a person marker has fewer restrictions on the root. Their mutual information would thus be lower. In a corpus study of the order of morphemes in Japanese and Sesotho (a Southern Bantu language spoken in Lesotho and South Africa), Hahn, Degen and Futrell (2021) found that mutual information, which represents the strength of association between morphemes, goes down as the distance between the morphemes increases. This decay is interpreted as evidence of an efficient trade-off between memory and surprisal costs. We can thus conclude that the cross-linguistically preferred order of morphemes is efficient because it helps to save processing costs.

The semantic phenomena discussed here are only a fraction of a large number of examples demonstrating the efficient use of word and morpheme order in the world's languages. For a bigger picture, see Levshina (2022, Ch. 3).

Switching the perspective from typical distances to possible distances, we find indications that stronger semantic attraction between elements can allow them to be separated by a larger distance. A case in point is island constructions. Although there are different factors at play, such as the syntactic role of the extracted constitutent and its discourse status (Abeillé et al., 2020; Cuneo & Goldberg, 2022), which determine the acceptability of island constructions, it is also enhanced if the extracted constituent and the gap are more strongly semantically and/or syntactically related. According to Chaves and Putnam (2020, p. 207), (7a) is more felicitous than (7b) because the extracted referent is part of the frame evoked by the verb *read*, but not *drop*.

(7) a. *What did you read a book about?*
b. *?What did you drop a book about?*

The content of a book is relevant in a book-reading scene evoked in (7a), and irrelevant in the book-dropping scene in (7b). For an illustration, compare the sentence *I read a book about linguistics* with *I dropped a book about linguistics*. Mentioning the topic of a book in the former sentence is more felicious than in the latter. The reason why (7a) is considered acceptable by language users is that the level of mutual accessibility of the extracted referent and the frame, which is high enough to enable successful processing of such sentences. Based on this finding, we can formulate a more general testable hypothesis: if required, semantically and/or syntactically related elements with higher mutual accessibility can be separated by a greater distance than less strongly associated elements.

When speaking about the principle of negative correlation between accessibility and costs in Section 2, I mentioned pragmatic principles that may be responsible for the emergence and use of efficient form-meaning pairings. Are such principles also at work here? We cannot exclude that language users have certain mutual expectations about the order of presenting information in discourse. For instance, one of the submaxims of Manner in Grice's (1975) theory says, 'Be orderly'. However, it seems more likely that the order of constituents and smaller units is in most cases a result of the language producer's automatic decisions, rather than rational behaviour that involves pragmatic reasoning of the Gricean type. According to MacDonald's (2013) Production-Distribution-Comprehension account, the form of utterances

is strongly affected by language users' tendency to minimize difficulty in incremental language production, which requires parallel planning and execution of utterances. When a language user produces highly accessible words and other units, they save time and processing costs. For the comprehender such orders are also easy to process because the addressee is exposed to the statistical tendencies in language, which are determined by the language producer's preferences in language production (MacDonald, 2013).

4. Ambiguity and synonymy are necessary for efficiency

4.1. Ambiguity and efficiency

If we look at linguistic units out of context, we will see that many, if not most of them display ambiguity of diverse types, from homonymy (e.g., *bank*, 'a financial institution' and 'the land alongside a river or a lake') to polysemy (e.g., *tree*, 'a tall woody perennial plant' and 'a diagram that shows branches coming from a central node') and vagueness (e.g., *aunt*, 'the mother's sister' and 'the father's sister'). Such one-to-many form-meaning pairings are very common. As mentioned in Section 1, the presence of ambiguity in human languages has been interpreted by Chomsky (2002) as evidence that they were not shaped for communicative needs.

This claim is unfounded. First of all, when we look at language in context, in which disambiguating information is abundant, we see no clear indications that ambiguity is costly. In fact, what is truly surprising, is not that people sometimes have difficulties with ambiguity, but that they experience them very rarely (MacKay, 1987, p. 133). The reason is that more probable interpretations of potentially ambiguous units become highly accessible when they are activated by the context.

Note that not everything needs to be disambiguated. We will not waste our effort on disambiguation if it does not lead to useful contextual effects (Sperber & Wilson, 1995). For example, if someone owes you money and wants to pay it back saying, *I must go to a bank*, for the creditor it does not matter if the debtor is going to a bank office or to an ATM. The underspecified information does not produce any cognitive effects, and we do not spend our processing effort trying to retrieve it.

In fact, language users save time and effort by using shorter and potentially ambiguous forms. Since short forms are limited, one-to-many pairings are unavoidable. However, this potential ambiguity is usually not a problem thanks to contextual cues. Therefore, ambiguity is efficient (Piantadosi et al., 2012).

The extent of potential ambiguity in language is argued to be determined by the trade-off between the language producer's and comprehender's effort. According to Zipf (1949), there are two competing forces: the Speaker's Economy (the Force of Unification), and the Auditor's Economy (the Force of Diversification). If the Speaker's Economy operates unchecked, it will result in a vocabulary of one very short word referring to all *N* possible meanings. If the Auditor's Economy operates on full scale, it will lead to *N* words, each expressing a unique meaning. The existing systems of 'manageable' ambiguity are a result of the compromise between these forces operating in context.

Why do shorter forms tend to be more ambiguous, rather than longer ones? It seems that the accessibility of a unit (not only its meaning, but perhaps also its form) plays a role. Shorter units are more frequent (cf. Zipf's law of abbreviation), and more frequent units are more easily available for new meanings, as shown experimentally by Harmon and Kapatsinki (2017). Another possible explanation is that more frequent units may have stronger statistical associations between diverse contextual cues and the intended interpretations. This means that language users can rely on the disambiguating information more.

4.2. Synonymy and efficiency

As mentioned earlier, Zipf (1949) argued that it is in the interest of the speaker to have as few forms as possible to choose from. He called it the Force of Unification. But how costly is the presence of synonyms for language production in reality? Opinions differ. According to Goldberg (2019, p. 26), unbiased decisions are more difficult to make, therefore fully equivalent forms would represent a challenge for language production. At the same time, Gardner et al. (2021) argue that the availability of different morphosyntactic options the speaker can choose from does not represent difficulties for the speaker.

Regardless of the production costs of true synonymy, speakers do not seem to face this challenge very often. First of all, the Principle of Contrast (Clark, 1987), which is also known as the Principle of No Synonymy (Goldberg, 1995) in Construction Grammar, says that if two forms are distinct, they should also differ functionally – for example, in register (e.g., *ask* vs. *inquire*), dialect (e.g., *lift* in British English vs. *elevator* in American English),

connotation (e.g., *frugal* vs. *stingy*), and so forth (Goldberg, 2019: 25–26). The pragmatic origins of this principle are principles similar to the ones discussed in Section 2.2, for example, Levinson's (2000) Q-heuristic 'What isn't said, isn't': in the presence of a salient alternative to some expression, a language user will not over-extend the meaning of this expression to include the meaning of that alternative. Although the Principle of No Synonymy has been contested in Construction Grammar (for example, Uhrig [2015] argues that different constructions may be very similar semantically), it is important to highlight the fact that this principle operates at the level of a specific usage event. In other words, the distinct forms should not be fully interchangeable in a given context. The speaker rarely needs to choose between fully equivalent forms (Goldberg: 2019, p. 26). Similar to what has been said about ambiguity, synonymy is rare if one considers context.

The Principle of No Synonymy has important repercussions for efficient language use, in particular, for the use of reduced and full forms. For example, Levshina and Lorenz (2022) investigated the factors responsible for the choice between the reduced form *wanna* and the full form *want to* in the spoken part of the British National Corpus. If language users used these forms efficiently, they would use *wanna* in contexts with higher contextual predictability and *want to* in contexts with lower predictability. However, hardly any predictability-related effects were detected. A possible explanation is that language users do not see *want to* and *wanna* sufficiently often as real alternatives because the forms have very different sociolinguistic and stylistic profiles. In particular, *wanna* is markedly informal, being used by younger speakers, for whom it also serves as an identity marker. Levshina and Lorenz hypothesize that if the difference between the full and reduced form is highly salient (at least, in a given context), efficiency-related effects involving predictability will not be observed.

At the same time, the availability of functionally similar alternatives is pragmatically important for the efficient division of labour to emerge. For example, we derive the meaning of an untypical causing event of the sentence with periphrastic causative in *Harry Potter caused the cup to rise* because there is a more obvious alternative with the lexical causative *raise* (see Section 2.3). Thus, synonymy both enables and constrains efficient language use.

5. Summary and open questions

To summarize, the pressure for communicative efficiency influences the structure and order of linguistic signs in languages of the world. First, more accessible meanings tend to be expressed by shorter and more accessible forms, as supported by numerous illustrations from grammar and lexicon in this chapter. Second, more semantically accessible (animate and concrete) referents tend to occur first in discourse. In addition, semantically associated linguistic units (morphemes, words and constituents) tend to be placed closer to each other, as we see from the cross-linguistic tendencies for dependency length minimization and related phenomena. Finally, ambiguity and synonymy are not threats for efficient communication. On the contrary, they serve as its necessary conditions.

Although these and similar ideas have been expressed by many scholars in the past, like Gabelentz (1891) and Zipf (1935, 1949), the systematic research on communicative efficiency is still relatively young. It is not surprising then that there remain numerous open questions. One of such issues has already been mentioned. The problem is that there is no universal measure of accessibility that would be applicable across different contexts and languages. As shown in Section 1, there exist many different levels and sources of accessibility (semantic, discourse-related, perceptual, lexical, constructional, etc.) that determine the degree of activation of competing forms and meanings in complex and language-specific ways. More research on accessibility is needed for improving our understanding of communicative efficiency and avoiding the risk of circularity.

In addition, we need to understand better how efficient form-meaning mappings emerge in the process of language evolution. For example, how much rational Gricean behaviour is involved in the creation and propagation of shorter forms for accessible meanings, and the shifts of meanings of longer forms? How are frequency, length, polysemy and other factors interrelated?

To answer these difficult questions, we need interdisciplinary research, to which all available types of data and methods should make their contribution, including analysis of typological and diachronic data, fine-grained usage-based analysis of language-specific alternations, and experiments in real and artificial language learning and communication. We will also profit from approaches involving information theory and animal communication research. The study of communicative efficiency will also benefit from causal models (e.g., Baayen et al. [2016] and Levshina [2021]), which capture how some linguistic variables may cause changes in the others.

References

Abeillé, A., Hemforth, B., Winckel, E., & Gibson, E. (2020). Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition*, 204, 104293.

Aikhenvald, A.Y. (2016). *How gender shapes the world*. Oxford University Press.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. Routledge.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), Text representation: Linguistic and psycholinguistic aspects (pp. 29–87). John Benjamins Publishing Company. https://doi.org/10.1075/hcp.8.04ari.

Arnold, J.E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, 4(4), 187–203. https://doi.org/10.1111/j.1749-818x.2010.00193.x

Aylett, M., & Turk, A. (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, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of Acoustical Society of America*, 119(5), 3048–3058.

Baayen, R.H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, 30(11), 1174–1220. https://doi.org/10.1080/02687038.2016.1147767

Bentz, Ch., & Ferrer-i-Cancho, R. (2016). Zipf's law of abbreviation as a language universal. Capturing Phylogenetic Algorithms for Linguistics. In Ch. Bentz, G. Jäger & I. Yanovich (Eds.), *Proceedings of the Leiden workshop on capturing phylogenetic algorithms for linguistics*. University of Tübingen. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/68558.

Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9(20). http://dx.doi.org/10.3765/sp.9.20.

Bock, J.K. (1986). Meaning, sound, and syntax: lexical primingin sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 575–586.

Bock, J.K., & Warren, R.K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21, 47–67.

Bohnemeyer, J., Enfield, N.J., Essegbey, J., & Kita, S. (2010). The macro-event property: The segmentation of causal chains. In J. Bohnemeyer & E. Pederson (Eds.), *Event representation in language: Encoding events at the language-cognition interface* (pp. 43–67). Cambridge University Press.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R.H. (2007). Predicting the dative alternation. In G. Bouma, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Royal Netherlands Academy of Science.

Bresnan, J., & Hay, J. (2008). Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua*, 118(2), 245–259.

Bybee, J.L. (1985). *Morphology: A study of the relation between meaning and form*. John Benjamins.

Bybee, J.L. (1994). The grammaticization of zero. Asymmetries in tense and aspect systems. In W. Pagliuca (Ed.), *Perspectives on grammaticalization* (pp. 235-254). John Benjamins.

Bybee, J.L. (2010). *Language, usage, and cognition*. Cambridge University Press.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Chaves, R.P., & Putnam, M.T. (2020). *Unbounded dependency constructions: Theoretical and experimental perspectives*. Oxford University Press.

Chomsky, N. (2002). An interview on minimalism. In N. Chomsky, *On Nature and Language* (pp. 92–161). Cambridge University Press.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Lawrence Erlbaum Associates.

Cohen Priva, U. (2008). Using information content to predict phone deletion. In N. Abner & J. Bishop (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics* (pp. 90–98). Cascadilla Proceedings Project.

Cover, Th.M., & Thomas, J.A. (2006). *Elements of Information Theory* (2nd ed.). Wiley.

Croft, W. (2000). *Explaining language change: an evolutionary approach*. Longman.

Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge University Press.

Cuneo, N., & Goldberg, A. (2022). Islands effects without extraction: The discourse function of constructions predicts island status. In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society* (pp. 2303 – 2309). URL https://escholarship.org/uc/item/1w83v9mm.

Degen, J. (2023). The Rational Speech Act framework. *Annual Review of Linguistics,* 9, 519-540. https://doi.org/10.1146/annurev-linguistics-031220-010811.

Dixon, R. M. W. (2000). A typology of causatives: Form, syntax and meaning. In R. M. W. Dixon & A.Y. Aikhenvald (Eds.), *Changing valency: Case studies in transitivity* (pp. 30–83). Cambridge University Press.

Dolscheid, S., & Penke, M. (2023). Attention vs. accessibility: the role of different cue types for non-canonical sentence production in German. *Frontiers in Languange Sciences*, 2, 1256471. https://doi.org/10.3389/flang.2023.1256471.

Dryer, M.S. (2013). Order of Subject, Object and Verb. In M.S. Dryer & M. Haspelmath (Eds.), WALS Online (v2020.3) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7385533.

Fenk, A., & Fenk, G. (1980). Konstanz im Kurzzeitgedächtnis - Konstanz im sprachlichen Informationsfluß. *Zeitschrift für experimentelle und angewandte Psychologie* XXVII, 3, 400–414.

Ferreira, F. (2003). The misinterpretation of canonical sentences. *Cognitive Psychology*, 47(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7.

Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, 32(6), 669–692. https://doi.org/10.1023/A:1026146332132.

Ferrer-i-Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E*, 70(5), 056135.

Ferrer-i-Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M. J., & Semple, S. (2013). Compression as a universal principle of animal behavior. *Cognitive Science*, 37(8), 1565–1578. https://doi.org/10.1111/cogs.12061.

Frank, M.C., & Goodman, N.D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998.

Futrell, R. (2019). Information-theoretic locality properties of natural language. In *Proceedings of the First Workshop on Quantitative Syntax (Quasy, SyntaxFest 2019, Paris)* (pp. 2–15). ACL. https://doi.org/10.18653/v1/W19-7902

Futrell, R., & Levy, R. (2017). Noisy-context surprisal as a human sentence processing cost model. In M. Lapata, Ph. Blunsom & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long papers* (pp. 688–698). EACL. https://www.aclweb.org/anthology/E17-1065

Futrell, R., Mahowald K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33), 10336–10341.

Gabelentz, G. von der. (1891). *Die Sprachwissenschaft: ihre Aufgaben, Methoden und bisherigen Ergebnisse*. T.O. Weigel.

Gardner, M., Uffing, E., Van Vaeck, N., & Szmrecsanyi, B. (2021). Variation isn't that hard: Morphosyntactic choice does not predict production difficulty. *PLOS One*, 16(6), 1–39. https://doi.org/10.1371/journal.pone.0252602

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1–76.

Gibson, E., Futrell, R., Piantadosi, S., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Science*, 23(5), 389-407. https://doi.org/10.1016/j.tics.2019.02.003

Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science*, 34(2), 286–310.

Glass, L. (2020). Verbs describing routines facilitate object omission in English. *Proceedings of the Linguistic Society of America*, 5(1), 44–58. https://doi.org/10.3765/plsa.v5i1.4663.

Gleitman, L.R., January, D., Nappa, R., & Trueswell, J.C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57, 544–569.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. University of Chicago Press.

Goldberg, A. E. (2005). Argument realization: the role of constructions, lexical semantics and discourse factors. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive grounding and theoretical extensions* (pp. 17– 44). John Benjamins.

Goldberg, A. E. (2019). *Explain me this: creativity, competition, and the partial productivity of constructions*. Princeton University Press.

Goldberg, A.E., & Ferreira, F. (2022). Good-enough language production. *Trends in Cognitive Sciences*, 26(4), 300–311. https://doi.org/10.1016/j.tics.2022.01.005.

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior*, 22(1), 24–44.

Greenberg, J.H. (1963). Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In J.H. Greenberg (Ed.), *Universals of Human Language* (pp. 73–113). MIT Press.

Greenberg, J.H. (1966). *Language Universals, with Special Reference to Feature Hierarchies*. Mouton.

Grice, H.P. (1975). Logic and Conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics*, Vol. 3, Speech Acts (pp., 41 – 58). Academic Press.

Hahn, M., Degen, J., Goodman, N., Jurafsky, D., and Futrell, R. (2018). An information-theoretic explanation of adjective ordering preferences. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society (CogSci)* (pp. 1766–1771).

Hahn, M., Degen, J., & Futrell, R. (2021). Modeling word and morpheme order in natural language as an efficient trade-off of memory and surprisal. *Psychological Review*, 128(4), 726–756. https://doi.org/10.1037/rev0000269

Haiman, J. (1983). Iconic and economic motivation. *Language*, 59(4), 781–819.

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, 98, 22–24.

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics*, 19(1), 1–33. https://doi.org/10.1515/COG.2008.001

Haspelmath, M. (2016). Universals of causative and anticausative verb formation and the spontaneity scale. *Lingua Posnaniensis*, 58(2), 33–63.

Haspelmath, M. (2021). Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics*, 57(3), 605–633. https://doi.org/10.1017/S0022226720000535.

Haspelmath, M., & Karjus, A. (2017). Explaining asymmetries in number marking: Singulatives, pluratives and usage frequency. *Linguistics*, 55(6), 1213–1235.

Haspelmath, M., Calude, A., Spagnol, M., Narrog, H. & Bamyacı, E. (2014). Coding causal–noncausal verb alternations: A form–frequency correspondence explanation. *Journal of Linguistics*, 50(3), 587–625.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars.* Oxford University Press.

Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford University Press.

Horn, L.R. (1984). Towards a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Georgetown University Round Table on Languages and Linguistics* (pp. 11–42). Georgetown University Press.

Horvath, J., Siloni, T. (2011). Anticausatives: against reflexivization. *Lingua*, 121, 2176–2186.

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. [Ph.D. thesis, Stanford University].

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165, 45–52. https://doi.org/10.1016/j.cognition.2017.05.001

Keller, R. (1994). *On language change: The invisible hand in language.* Routledge.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science, 336*(6084), 1049–1054. https://doi.org/10.1126/science.1218811

Koontz-Garboden, A. (2007). States, changes of state, and the monotonicity hypothesis. PhD Thesis, University of Stanford.

Koontz-Garboden, A. (2009). Anticausativization. *Natural Language and Linguistic Theory* 27, 77–138. https://doi.org/10.1007/s11049-008-9058-9

Koplenig, A., Kupietz, M., & Wolfer, S. (2022). Testing the relationship between word length, frequency, and predictability based on the German Reference Corpus. *Cognitive Science*, 46(6), e13090. https://doi.org/10.1111/cogs.13090.

Kurumada, Ch., & Grimm, S. (2019). Predictability of meaning in grammatical encoding: Optional plural marking. *Cognition*, 191, 103953. https://doi.org/10.1016/j.cognition.2019.04.022.

Lee, C., Lew-Williams, C., & Goldberg, A.E. (2022). Accessibility factors that lead to good-enough language production. In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society* (pp. 2303 – 2309). URL https://escholarship.org/uc/item/828625bm.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

Levshina, N. (2016). Why we need a token-based typology: A case study of analytic and lexical causatives in fifteen European languages. *Folia Linguistica*, 50(2), 507–542.

Levshina, N. (2019). Linguistic Frankenstein, or How to test universal constraints without real languages. In K. Schmidtke-Bode, N. Levshina, S. M. Michaelis & I. Seržants (Eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence* (pp. 203–223). Language Science Press. https://langsci-press.org/catalog/book/220.

Levshina, N. (2021). Cross-linguistic trade-offs and causal relationships between cues to grammatical Subject and Object, and the problem of efficiency-related explanations. *Frontiers in Psychology*, 12, 648200. https://doi.org/10.3389/fpsyg.2021.648200

Levshina, N. (2022). *Communicative efficiency: Language structure and use*. Cambridge University Press.

Levshina N., & Lorenz, D. (2022). Communicative efficiency and the Principle of No Synonymy: Predictability effects and the variation of *want to* and *wanna*. *Language and Cognition*, 14(2), 249-274. doi:10.1017/langcog.2022.7

Levshina, N., & Moran, S. (2021). Efficiency in human languages: corpus evidence for universal principles. *Linguistics Vanguard*, 7(s3), 20200081. https://doi.org/10.1515/lingvan-2020-0081.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt & Th. Hoffman (Eds.), *Advances in Neural Information Processing Systems (NIPS),* vol. 19 (pp. 849–856). MIT Press.

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2), 159–191.

MacDonald, M.C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 226. https://doi.org/10.3389/fpsyg.2013.00226.

MacKay, D. G. (1987). *The Organization of perception and action: A theory for language and other cognitive skills*. Springer.

Martin, F, & Schäfer, F. (2014). Causation at the syntax-semantics interface. In B. Copley & F. Martin (Eds.), *Causation in Grammatical Structures* (pp. 209-244). Oxford University Press.

Martínez, E., Mollica, F., & Gibson, E. (2023). Even lawyers do not like legalese. *Proceedings of the National Academy of Sciences*, 120(23), e2302672120. https://doi.org/10.1073/pnas.2302672120 .

Martínez, E., Mollica, F., & Gibson, E. (2024). Even laypeople use legalese. *Proceedings of the National Academy of Sciences*, 121(35), e2405564121. https://doi.org/10.1073/pnas.240556412.

McCawley, J. (1976). Remarks on what can cause what. In M. Shibatani (Ed.), *Syntax and Semantics VI: The grammar of causative constructions* (pp. 117–129). Academic Press.

Meylan, S., & Griffiths, T. L. (2021). The challenges of large-scale, web-based language datasets: word length and predictability revisited. *Cognitive Science*, 45(6), e12983. https://doi.org/10.1111/cogs.12983.

Mithun, M. (1992). Is basic word order universal? In: D. L. Payne (Ed.), *Pragmatics of word order flexibility* (pp. 15–61). John Benjamins.

Piantadosi, S., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526.

Piantadosi, Steven, Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122, 280–291. https://doi.org/10.1016/j.cognition.2011.10.004

Pimentel, T., Meister, C., Wilcox, E., Mahowald, K., & Cotterell, R. (2023). Revisiting the optimality of word lengths. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 2240–2255), Singapore. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.137.

Redl, T., Frank, S. L., de Swart, P., & de Hoop, H. (2021). The male bias of a generically-intended masculine pronoun: Evidence from eyetracking and sentence evaluation. *PLoS ONE*, 16(4), e0249309. https://doi.org/10.1371/journal.pone.0249309

Scontras, G. (2023). Adjective ordering across languages. *Annual Review of Linguistics*, 9, 537–376.

Sikos, L., Venhuizen, N.J., Drenhaus, H., & Crocker, M.W. (2021). Reevaluating pragmatic reasoning in language games. *PLoS ONE*, 16(3), e0248388. https://doi.org/10.1371/journal.pone.0248388.

Sharwood Smith, M. (2017). *Introducing language and cognition: A map of the mind*. Cambridge University Press.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1), 140–155.

Slonimska, A., Özyürek, A., & Capirci, O. (2020). The role of iconicity and simultaneity for efficient communication: The case of Italian Sign Language (LIS). *Cognition*, 200, 104246. https://doi.org/10.1016/j.cognition.2020.104246.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Blackwell Publishers.

Tanaka, M.N., Branigan, H.P., McLean, J.F., & Pickering, M.J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, 65(3), 318–330. https://doi.org/10.1016/j.jml.2011.04.009.

Tomlin, R. (1997). Mapping conceptual representations into linguistic representations: the role of attention in grammar. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 162–189). Cambridge University Press.

Traugott, E. C. (2010). Revisiting subjectification and intersubjectification. In K. Davidse, L. Vandelanotte & H. Cuyckens (Eds.), *Subjectification, intersubjectification and grammaticalization* (pp. 29–71). De Gruyter Mouton.

Uhrig, P. (2015). Why the Principle of No Synonymy is overrated. *Zeitschrift für Anglistik und Amerikanistik*, 63(3), 323–337. https://doi.org/10.1515/zaa-2015-0030

Verhagen, A., & Kemmer, S. (1997). Interaction and causation: Causative constructions in modern standard Dutch. *Journal of Pragmatics*, 27, 61–82.

Wasow, Th. (2015). Ambiguity avoidance is overrated. In S. Winkler (Ed.), *Ambiguity: Language and communication* (pp. 29–47). De Gruyter Mouton.

Witkowski, S. R., & Brown, C. H. (1983). Marking-reversals and cultural importance. *Language*, 59(3), 569–582.

Wolff, Ph. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition,* 88, 1–48.

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104(5), 444–466.

Zipf, G. K. (1965[1935]). *The psychobiology of language: An introduction to dynamic philology*. MIT Press.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.