A Bathtub by Any Other Name:
the Reduction of German Compounds in Predictive Contexts

Alessandra Zarcone (alessandra.zarcone@iis.fraunhofer.de)
HumAIn Labs, Fraunhofer Institute for Integrated Circuits (IIS)
Am Wolfsmantel 33, 91058 Erlangen, Germany

Vera Demberg (vera@coli.uni-saarland.de)
Department of Language Science and Technology, Saarland University
Campus C7.2, 66123 Saarbrücken, Germany

Abstract

The Uniform Information Density hypothesis (UID) predicts that lexical choice between long and short word forms depends on the predictability of the referent in context, and recent studies have shown such an effect of predictability on lexical choice during online production. We here set out to test whether the UID predictions hold up in a related setting, but different language (German) and different phenomenon, namely the choice between compounds (e.g. *Badewanne* / bathtub) or their base forms (*Wanne* / tub). Our study is consistent with the UID: we find that participants choose the shorter base form more often in predictive contexts, showing an active tendency to be information-theoretically efficient.

Keywords: psycholinguistics, information theory, compounds, efficiency, communication

Introduction

When talking about things or living beings in the world, we can refer to them in different ways, for example choosing a proper noun (*Tom sold me some strawberries yesterday*), or a noun phrase (*The grocer sold me some strawberries*), or a pronoun (*He sold me some strawberries*). When referring to a discourse referent with a noun phrase, we often have a choice between longer or shorter nouns, or between longer or shorter versions of the same noun (*greengrocer* vs. *grocer*).

Predictability has been proposed to be one of the factors influencing the speaker's lexical choices. People make extensive use of the statistical properties of linguistic input in comprehension and production (Aylett & Turk, 2004; Fenk-Oczlon, 2001; Genzel & Charniak, 2002, 2003; Kuperberg & Jaeger, 2016; Levy, 2008; MacDonald, 2013; Manin, 2006). The Uniform Information Density Hypothesis (UID) claims that encoding mechanisms tend to optimize information transmission by distributing information over a message in a uniform way, avoiding peaks and troughs in the distribution (Levy & Jaeger, 2007). One way of keeping the distribution of information uniform is to favor shorter encodings over longer ones in more predictable contexts, to avoid spreading out less informative content over a longer stretch of the message. This tendency has been shown to shape the lexicon, as predictable words (e.g. function words) become diachronically shorter. It also influences production, as shorter encodings are chosen in more predictable contexts (Aylett & Turk, 2004; Bybee, 2007; Jaeger, 2006; Mahowald, Fedorenko, Piantadosi, & Gibson, 2013; Piantadosi, Tily, & Gibson, 2011; Kanwal, Smith, Culbertson, & Kirby, 2017; Zipf, 1949).

The information content of a discourse referent can be measured as its *Surprisal* (Hale, 2001; Levy, 2008), as follows:

$$Surprisal(r) = -\log P(r|c)$$

where, importantly, *r* is the probability of the referent, not the probability of the referring expression. The probability of the referent can be estimated by resolving all referring expressions to the referent, and then using them to estimate the referent's predictability.

Given two different encodings sharing the same meaning and syntactic category (two words which can be used interchangeably), the UID predicts that speakers would choose the shorter one in contexts where the referent is more predictable. Mahowald et al. (2013) tested this hypothesis by contrasting short / long word pairs such as *chimp / chimpanzee*. In a corpus study, they found that average Surprisal estimates for 22 word pairs were significantly higher for long forms compared to short forms. In a behavioral study with a forced-choice sentence completion task, they manipulated the predictability of the missing referent given the preceding context (supportive or neutral context). The short form was chosen significantly more often in supportive contexts (67%) than in neutral contexts (56%), supporting the hypothesis that speakers have an active preference for conveying meanings with shorter word forms in more predictive contexts. Mahowald et al. (2013), however, focused on a phenomenon that is not very pervasive in the lexicon: the target words are either acronyms (e.g. *US, UN*) or are rather long words of foreign etymology (e.g. *limousine, rhinoceros*), often referring to abstract entities (e.g. *chemotherapy, mathematics*), and shorter and longer forms (e.g. *math* and *mathematics*) may also be used in a different register or style.

We present a sentence completion study, aimed at testing if a preference for shorter encodings in more predictive contexts can also be observed for a wider and more pervasive phenomenon, that is the reduction of German compound nouns to their base form (e.g. *Badewanne*, bathtub vs. *Wanne*, tub). Similar to Mahowald et al. (2013), we manipulated the predictability of the missing referent in context, but we introduced two main changes as we (1) manipulated predictability by activating or not a relevant script knowledge scenario and (2) we adopted an open-choice task, where participants were asked to provide a missing referent. We found that also

for this different type of contrast (German compound nouns vs. base forms) predictability influences the speakers' production choices, which show a tendency for the shorter base form in more predictable contexts.

Reducing German compounds in predictive contexts

The target referring expressions in our study are German compounds and their base forms (e.g. *Badewanne* vs. *Wanne*, bathtub vs. tub). German is a morphologically-rich language with very productive compound formation, where determinative bimorphemic compounds (Olsen, 2001) are pervasive. In determinative compounds, the first constituent modifies the second one: the second constituent (head or base form) is thus a less specific or basic-level term (e.g. *Kette*, chain, could refer to a *Fahrradkette*, bike chain, or to *Halskette*, necklace, lit. "neck chain"), whereas the compound unambiguously refers to a subordinate-level concept. German compounds are typically used as naming devices (Schlücker & Hüning, 2009), whereas phrases are used to provide descriptions (*altes Papier* is used to describe paper that is old, *Altpapier* is used to refer to scrap paper). German compounds and their base forms differ with regard to their length but not with regard to their register, and thus constitute an ideal test bed for the UID hypothesis. However, compared to their base forms, German compounds are more specialized (scrap paper is a more specific type of paper, not every paper that is old qualifies as *Altpapier*, Schlücker and Hüning, 2009). Compounding provides a means of lexical specification that is exploited for ambiguity resolution: compounds, being more specialized, are preferred over their base form when naming an object in the presence a same-category distractor (e.g. *Weinglas*, wine glass / *Sektglas*, champagne glass), and the number of morphologically-complex answers increases in presence of an addressee (Bölte, Böhl, Dobel, & Zwitserlood, 2009). For this reason, we designed our stimuli such that no other referent sharing the same base with the target (e.g. *Visitenkarte*, business card vs. *Kreditkarte*, credit card) would compete with the target referent, and that the base form (e.g. *Karte*, card) could unambiguously be used to refer to the target referent.

Script-dependent predictability

The predictability of a word in context can be estimated in terms of its *Surprisal* (Hale, 2001; Levy, 2008). While Surprisal has in practice often been estimated based on its direct local context (e.g., in $n-$gram models), research has shown that the predictability of a word also depends on higher-level situational representations (Kuperberg, 2016; Kuperberg & Jaeger, 2016; Zarcone, Van Schijndel, Vogels, & Demberg, 2016) and on non-local discourse (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Xiang & Kuperberg, 2015; Asr & Demberg, 2016), which may affect the predictability of the word's referent.

Our goal is not to manipulate the predictability of a word, but rather to manipulate the predictability of a target referent via discourse context and to assess its influence on the choice

of longer vs. shorter form. In order to keep the local context as similar between conditions as possible and avoid possible confounding factors such as register, we aim to only manipulate the context outside the sentence in which the target referent occurs. One way of manipulating predictability, while at the same time keeping local context as constant as possible, is to manipulate the activation of a typical script knowledge scenario (e.g. *going to a restaurant*). Knowledge of everyday activities, including the typical events and participants involved in these scenarios (script knowledge, e.g. *reading the menu*, *ordering wine*), is stored in our memory. It is central to our everyday interaction with the environment and is frequently evoked during language comprehension and production as it allows us to communicate efficiently: typical events (such as *sitting down* after entering the restaurant) can be inferred by the comprehender without the speaker needing to mention them explicitly. Activating a situation such as a script knowledge scenario can affect a participant's accessibility in the discourse (Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Metusalem et al., 2012), which in turn can affect the way we refer to it (Vogels, Krahmer, & Maes, 2013).

Experiment

Participants

Materials

We systematically exploit script knowledge for our experimental manipulation, by introducing a script in the predictive condition (labeled *script-supportive* in Example (1) below), and avoiding to do so in the *neutral* condition. This allows us to keep the target sentence as well as the following sentence identical across conditions.

1. [script-supportive] Sandra kam nach einem langen Arbeitstag nach Hause und ließ sich ein Schaumbad ein.
   Sandra came home after a long workday and drew herself a bubble bath.

   [neutral] Sandra kam nach einem langen Arbeitstag nach Hause.
   Sandra came home after a long work day.

   [target] Glücklich legte sie sich in die _____ und schloss die Augen. Erst als das Wasser kalt geworden war, stand sie auf und trocknete sich ab.

She happily got into the _____ and closed her eyes. Only when the water had become cold did she get up and dry herself off.

The script-supportive context introduces a scenario (e.g. *taking a bath*) by mentioning at least one script-relevant event or participant, whereas in neutral contexts, no script-specific event or participant is mentioned before the gap. The target sentence is missing a participant of the scenario (e.g. *Badewanne*, bathtub). The target word is always a concrete referent, which is part of the script. This also allows us to avoid possible confounds related to whether the target is abstract or concrete, or related to the conceptual complexity of a word (Lewis & Frank, 2016). After the target sentence, another short context sentence follows, completing the story and disambiguating the scenario also in the neutral condition.

We constructed 36 items as in Example (1), each in a neutral and script-supportive version, and used them to create two lists of 36 items each (18 neutral, 18 script-supportive)[1]. In order to rule out specificity as a confounding factor, we excluded the possibility that the base form of the target concept (*Wanne* / tub) could be misinterpreted as another concept sharing the same base form, so that both the compound and the base form would unequivocally refer to the same discourse referent.

Norming study

Our study targets the effect of the referent predictability on the form of the referring expression. In order to confirm (1) that there was no other concept than the target that could be referred to with the base form and (2) that our conditions did indeed differ with regard to the incremental predictability of the target, we estimated the predictability of the target referents from a comprehension perspective with an incremental (left-context only) norming study.

The sentence completion study was also conducted on Prolific. Forty native German speakers took part in the norming study and were excluded from participating in the main study. Also the norming study was approved by the Ethical Review Board. The same sentences were used in the norming study as in the main study. The norming study was a completion study where the sentences were presented up to but not including the critical referent (e.g. *Sandra kam nach einem langen Arbeitstag nach Hause. Glücklich legte sie sich in die _____* // Sandra came home after a long work day. She happily got into the _____). Participants were randomly assigned to one of the two lists and each of them saw each of the 36 items in one version only. They were asked to provide one or more words to complete each sentence.

Our first analysis is concerned with confirming that the target short form in our items was not ambiguous, i.e., that it could not be used to refer to other concepts than the target

---

[1] We tried to avoid lexical overlap between the content words in the context and the modifier of the compound, but 3 items did contain a small overlap (*Schaumbad - Badewanne*, *Radtour - Fahrradkette* and *Fernsehsendung - Fernseher*).

referent. Therefore we looked for completions that shared the same base form with our target long form but referred to a different object. We only found very few cases where this happened ($< 1\%$), affecting only three items (*Schreibwarenladen*, stationery shop, for *Friseurladen*, hair salon; and *Untertopf*, flower pot saucer, for *Blumentopf*, flower pot; *Handtasche*, purse, and *Brieftasche*, wallet, for *Einkauftasche*, shopping bag). We therefore conclude that the items were well-constructed in terms of eliciting the target referent.

The second goal of the norming study was to test the predictability of the target referent. The target referent (either in its compound or base form) was used to complete the sentence 74% of the time after script-supportive contexts but just 22% of the time after neutral contexts. This large difference between conditions confirms that the manipulation of our experimental items in terms of predictability of the target referent is effective. We note that, compared to previous work, the target word was slightly more predictable in neutral contexts (22% of completions matched the target referent in the neutral contexts, whereas only 1.6% did in Mahowald et al. 2013).

In the neutral condition, we obtained a total of 161 target completions, whereas in the script-supported conditions, there were a total of 530 target completions. The relatively large number of target completions also in the neutral condition (albeit fewer than those in the script-supportive condition) allows us to estimate whether an effect of predictability on referring expression form can already be found in the norming study. This analysis was not planned at the outset of the study (as we aimed to assess this point using the main study reported in the next section), but nevertheless provides a valuable data point. We found that the rate of short forms in the neutral condition was lower (58%) than in the script-supported condition (69%), which is in line with the UID hypothesis. However, a linear mixed effects model with form as a response variable and context type as the predictor variable did not support this difference as significant ($\beta = 0.4180, SE = 0.2406, z = 1.737, p = 0.082$).

Procedure

The norming study allowed us to test the referent's predictability in an incremental comprehension setting, that is how expected a referent would be when one incrementally processes the sentence. In the main study, we aim at testing our hypothesis that the choice of referring expression during production is influenced by the referent predictability. We employ a left- and right-context cloze task, following the rationale that (1) language producers normally already know what they will be talking about (Clark & Wasow, 1998; Ford, 1982; Griffin & Bock, 2000; Griffin, 2003), and are not guessing what might come up, as in the incremental cloze setting; and (2) it can be difficult to reliably estimate the usage of long vs. short forms, when participants enter completions

which do not refer to the target concept[2].

We presented the complete contexts and only omitted the target referent as in 1. The right-hand context is now present, because in a production perspective speakers typically know in advance what they are going to say. As the whole context is now available, participants will likely guess the correct discourse referent. However, we predict that their perception of the incremental predictability of the discourse referent (its predictability from the comprehender's point of view) will lead them to choose a shorter encoding (base form) more often during production in more predictable contexts.

Participants were asked to read the contexts and to type a word to complete each sentence. We used the same lists as those in the norming study, but in this study we used the complete contexts and only omitted the target referent as in 1. Participants were randomly assigned to one of the two lists and each of them saw each of the 36 items in one version only.

Results

As the main study was also a sentence completion study, it was possible for participants to complete the sentences with a referent other than the target referent, as in the norming study. When both left and right context were provided, the target referent (either in its compound or base version) was used to complete the sentence 89% of the time (93% in script-supportive contexts and 85% in neutral ones). We excluded data points with nontarget referents from our analysis. None of the excluded referents was a same-category distractor or a non-target long form sharing the same base form as the target referent. This confirms that in our materials the full compound (*Badewanne*) and the shorter head (*Wanne*) were meaning-equivalent and that any effect of long vs. short form can not be justified based on disambiguation of the target concept.

Participants showed a baseline preference for the base form independent of context (it was used 66% of the time, whereas the compound version was used 34% of the time) and chose the base form more often in script-supportive contexts (72%) than in neutral contexts (61%). We analyzed the completions with a mixed-effect logistic regression with full random effects structure and with the type of completion (compound vs. base) as a binary dependent variable. The effect of context (neutral vs. script-supportive) on what form was chosen was significant ($\beta = 0.854, SE = 0.182, z = 4.69, p < 0.001$)[3].

We also tested a model including a predictor that accounts for the potential difference in ambiguity of the long form compared to the short form: for instance, the word *Nagel* could be a fingernail but also a nail to be hammered, while *Fingernagel* only has one meaning. This could favor longer forms in cases where the base form is ambiguous.
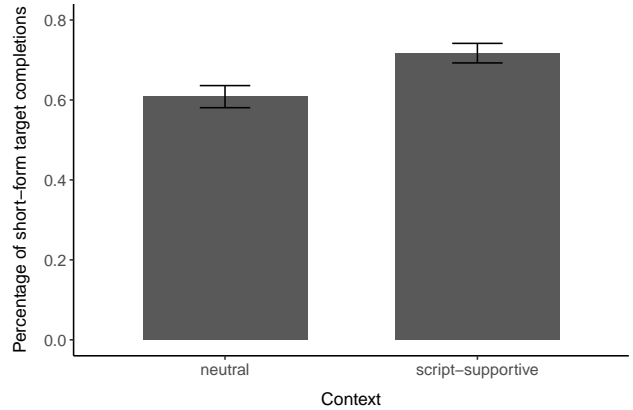

Figure 1: Effect of referent predictability on length of referring expression.

We estimated the number of senses as the number of synsets for the base forms and the full compounds in GermaNet (Hamp & Feldweg, 1997; Henrich & Hinrichs, 2010)[4] and included the difference between number of senses of the base form and number of senses for the full compounds as a predictor. The effect of sense difference did however turn out to not be a significant predictor of word form choice ($\beta = -0.3865, SE = 0.2311, z = -1.672, p = 0.09$) and including this predictor did not significantly improve model fit ($\chi = 0.8904, p = 0.34$).

Discussion

The UID hypothesis states that speakers choose among alternative linguistic encodings the one that optimizes information transmission by distributing information in a uniform way, for example by favoring shorter encodings in contexts where a discourse referent is more predictable. In contrast to previous work (Mahowald et al., 2013), we tested this prediction on a class of referring expressions (German compounds and their base forms) that is more pervasive than abbreviations. Our experimental design additionally contributes to the literature on UID effects by tightly controlling the stimuli to avoid confounds in terms of register and using a more naturalistic sentence completion task. We ran an incremental norming study with left context only in order to both exclude the possibility that the referent form could be explained in terms of specificity, and to estimate the incremental predictability of the target referent from the point of view of a comprehender.

Our main sentence completion study included both left and right context, in order to mimic the information available during production, when speakers plan ahead what they are going to say. The norming study showed that our discourse referents are less unpredictable in neutral contexts than those in Mahowald et al. (2013) - 22% of target completions in the

---

[2]Note that this is the reason why Mahowald et al. (2013) resorted to a forced-choice design.

[3]The R^2 of the correlation between the fitted and observed values was 0.42 for this model, whereas it was 0.38 without the context predictor.

[4]We accessed Release 15.0 (May 2020) of GermaNet using the GermaNet Rover web tool
https://weblicht.sfs.uni-tuebingen.de/rover/

neutral contexts in our study vs. 1.6% in Mahowald et al. (2013). Despite that, our manipulation of script-driven predictability had a significant effect on the choice of encoding: participants used the base form more often in script-supportive contexts (72%) than in neutral contexts (61%). Predictability on referring expression may also have influenced the completions from the norming study, where short forms were preferred more often in the script-supported condition (69%) than in the neutral condition (58%), but the effect did not reach significance ($p = 0.082$). Our results successfully replicate findings for lexical choice during on-line processing, and is consistent with the results reported in Mahowald et al. (2013). Our results furthermore extend previous findings to the choice between base and compound forms in German (e.g. *Wanne* vs. *Badewanne*).

Would our results also be compatible with alternative rational accounts of language processing other than the UID? The Rational Speech Act framework (RSA) proposes that speaker choice in language production and comprehender interpretation can be accounted for as a rational communicative process (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013). While a lot of the work in the RSA framework has addressed pragmatic phenomena such as the usage and interpretation of quantifiers, the theory does make specific predictions also for our setting: rational speakers reason about the utility of alternative utterances with respect to their communicative success and their production effort (or cost). If two variants are identical in terms of the communicative success of the interaction, the speaker is predicted to choose the more cost efficient (i.e., shorter or more common) word. On the other hand, if one of the forms is at a risk of being misunderstood by the listener, the speaker would be more likely to choose the longer one (Bergen & Goodman, 2015). In many settings, the UID and the RSA would therefore make identical predictions: high predictability often corresponds to a low risk of misunderstanding, and highly surprising contents on the other hand would usually come with an increased risk of misunderstanding.

The RSA study that seems most closely related to the present manipulation is a study by Graf, Degen, Hawkins, and Goodman (2016), who show that speakers, when not constrained by contextual considerations, tend to refer to objects using basic-level labels, which are typically shorter (e.g. *dog* vs. *German shepherd*). The use of a more specific label is motivated by the contextual informativeness of a description, for example if a distractor of the same basic level is available (e.g. *dalmatian* vs. *German shepherd*). An important difference to our study is that the presence of a same-category distractor in their manipulation (e.g. another type of dog), while we made sure in our study that the use of the more specific version (the compound) was not necessary, as shown by the lack of completions referring to other concepts compatible with the base form of the target referent in the norming study and by an overall preference for the base form in the main study. In the case of our stimuli, materials are carefully con-trolled such that there is little risk of misunderstanding at the point of the target referent, and the following material (available to the participants of the main experiment) fully disambiguates the stories (and the referents). This is also supported by the almost complete absence of non-target completions in the norming study. Therefore, one could argue that the present experiment makes a contribution also in showing that the choice between base form and the full compound is also determined by predictability, and not only by specificity.

Based on the original RSA formulation, our results may therefore seem to be incompatible with the RSA, where the longer form should only be chosen if at the end of the utterance, successful communication is at risk. This means that speakers should happily choose the shorter form in both conditions in our experiment (especially in the non-incremental main experiment where they are aware of the end of the sentence). Nevertheless, we find a high rate of long forms in the neutral context condition. As Levy (2018) points out however, the cost term in RSA can be instantiated in terms of the UID (peaks and troughs in information density are costly). Under this formulation, the RSA is also consistent with the findings of our experiment.

Conclusions

Our study provides additional evidence for information density control during language production and efficient communicative design not only as a diachronic process that shapes the lexicon but also during active use of language. The high-level predictability of the discourse referent, influenced by the activation of the relevant script knowledge scenario, has an influence on the lower-level surface form that is chosen to refer to the target. The choice of a shorter form when the referent is more predictable, that is when it is less loaded with information, and of a longer form when the referent is less predictable (thus more informative) is an efficient strategy to optimize information transmission by distributing information over a message in a uniform way.

Acknowledgements

References

Asr, F. T., & Demberg, V. (2016). But vs. although under the microscope. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 38).

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.

Bergen, L., & Goodman, N. D. (2015). The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, *7*(2), 336–350.

Bölte, J., Böhl, A., Dobel, C., & Zwitserlood, P. (2009). Effects of referential ambiguity, time constraints and addressee orientation on the production of morphologically complex words. *European Journal of Cognitive Psychology*, *21*(8), 1166–1199.

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford University Press.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive psychology*, *37*(3), 201–242.

Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. *Frequency and the Emergence of Linguistic Structure*, *45*, 431.

Ford, M. (1982). Sentence planning units: implications for the speaker's representation of meaningful relations underlying sentences. MIT Press.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 199–206).

Genzel, D., & Charniak, E. (2003). Variation of entropy and parse trees of sentences as a function of the sentence number. In *Proceedings of the 2003 conference on Empirical methods in natural language processing* (pp. 65–72).

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–184.

Graf, C., Degen, J., Hawkins, R. X., & Goodman, N. D. (2016). Animal, dog, or dalmatian? Level of abstraction in nominal referring expressions. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review*, *10*(3), 603–609.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*(4), 274–279.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* (pp. 1–8).

Hamp, B., & Feldweg, H. (1997). GermaNet - a Lexical-Semantic Net for German. In *Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*.

Henrich, V., & Hinrichs, E. (2010, May). GernEdiT - the GermaNet editing tool. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*.

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. Unpublished doctoral dissertation, Stanford University.

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Language-users choose short words in predictive contexts in an artificial language task. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? what the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, *31*(5), 602–616.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117–129.

Levy, R. P. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R. P. (2018). Communicative Efficiency, Uniform Information Density, and the Rational Speech Act Theory. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (p. 684–689).

Levy, R. P., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In *Proceedings of the 20th Conference on Neural Information Processing Systems (NIPS)*.

Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition*, *153*, 182–195.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318.

Manin, D. (2006). Experiments on predictability of word in context and information rate in natural language. *Journal of Information Processes*, *6*(3), 229–236.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*(4), 545–567.

Olsen, S. (2001). Copulative compounds: a closer look at the interface between syntax and morphology. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 2000* (pp. 279–320). Dordrecht: Kluwer.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.

Schlücker, B., & Hüning, M. (2009). Compounds and phrases. A functional comparison between German A+N compounds and corresponding phrases. *Italian Journal of Linguistics/Rivista di Linguistica*, *21*(1), 209–234.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Jour-

nal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443.

Vogels, J., Krahmer, E., & Maes, A. (2013). Who is where referred to how, and why? the influence of visual saliency on referent accessibility in spoken language production. *Language and cognitive processes*, *28*(9), 1323–1349.

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*(6), 648–672.

Zarcone, A., Van Schijndel, M., Vogels, J., & Demberg, V. (2016). Salience and attention in surprisal-based accounts of language processing. *Frontiers in Psychology*, *7*, 844.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. New York: Hafner.