

Predictability and syntactic production: Evidence from subject omission in Russian

Ekaterina Kravtchenko (eskrav@gmail.com)

Department of Linguistics; University of California, Santa Cruz; 1156 High Street
Santa Cruz, CA 95064 USA

Abstract

The quantitative study of the role of communicative efficiency in language production and comprehension has gained increasing attention recently. However, in online production most investigation has focused on the phonetic/phonological level, leaving open the question of whether the communicative pressures involved extend to the syntactic level. I present a corpus study which investigates the omission of optional clause subjects in Russian. If speakers communicate efficiently, then they are expected to preferentially omit elements that are more predictable, given preceding context. However, at the syntactic level this has only been directly demonstrated for omission of functional elements. The present study shows that even when other predictors of subject omission are taken into account, contextual predictability remains a significant predictor of whether an optional subject, a relatively complex syntactic constituent, is pronounced. This supports the hypothesis that the drive towards efficient communication is a general principle of language production.

Keywords: Psycholinguistics; language production; information theory; subject omission; Russian.

Introduction

A long-standing question in the language sciences concerns whether language is optimized for efficient communication (Givón, 1979; Hawkins, 2004; Piantadosi, Tily, & Gibson, 2011; Zipf, 1949). In the past most claims along these lines have been limited to either qualitative typological studies or appeals to intuition (Jaeger & Tily, 2011). The recent rise of information-theoretic approaches (Aylett & Turk, 2004; Genzel & Charniak, 2002; Hale, 2001; Levy & Jaeger, 2007) has led more recently to more formal and quantitative notions of efficiency. Under this approach, it is proposed that optimal communication of a message involves keeping the contextual probability of upcoming elements maximally uniform – close to, but not exceeding an arbitrary ‘channel capacity’¹. This refers to a limit on rate of information transfer beyond which the likelihood of error becomes unacceptable (Jaeger, 2010; Levy & Jaeger, 2007).

Under this framework, elements that are more contextually predictable carry less information (i.e. are redundant), with information defined in terms of probabilities: $I(Word) = -\log_2 P(Word|Context)$. This approach predicts that, subject to constraints of a speaker’s grammar, the intended meaning, and other pressures on language production, speakers will be more likely to omit those elements that are redundant. Speakers will conversely be more likely to pronounce those ele-

¹The term ‘channel capacity’ is imported from information theory, but the channel in question may be conceived of as referring to the communication channel between two interlocutors (Jaeger, 2010).

ments that are less contextually predictable. The most robust evidence for this hypothesis, in online production², currently exists at the phonetic and phonological levels; for example, involving effects of predictability on variation in syllable duration, or degree of articulatory detail (Aylett & Turk, 2004; Bell et al., 2003; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009). Jaeger (2010) demonstrates that this holds for omission of some functional elements, namely the optional complementizer *that*. This finding suggests that syntactic reduction as well is sensitive to the reduced element’s predictability in context.

However, omission of more complex constituents may present a different case. One may expect non-functional, or semantically contentful elements (i.e. those that contribute more to utterance meaning) to be less subject to omission due to functional pressures on production alone. Null subjects specifically are often argued to serve distinct syntactic or semantic functions (Carminati, 2002). There is likewise evidence that function (closed-class) and content (open-class) items may be subject to somewhat different pressures in production, with production of function words more sensitive to predictability from the preceding context (Bell et al., 2009). The question asked in this paper is whether pressure for robust (low-error) and efficient communication may partially account for whether or not speakers choose to omit optional constituents in Russian. Clause subjects can be optionally elided in Russian, although this is primarily restricted to colloquial speech and text (Zdorenko, 2009), where up to 32% of subjects are omitted. It is also restricted to contexts where the referent can be recovered in discourse (Franks, 1995). A representative example, where an embedded clause subject is omitted, is shown in (1):

(1) Maša pozvonila Pete, potomu čto ona/_ zbolela
Masha called.FEM Petia, because she/_ fell-ill.FEM
“Masha_i called Peter_j, because [she_i] was sick.”

Main clause subjects are also frequently omitted in casual speech, given appropriate context, although isolated sentences may be infelicitous, unless the subject referent is contextually salient. The use of null subjects in Russian appears to reflect constraints found in other languages on use of null, or reduced, referential expressions (Franks, 1995);

²A number of studies have demonstrated that more stable or conventionalized linguistic forms can reflect similar pressures (e.g., Piantadosi et al., 2011); it is not however clear that these findings can be straightforwardly generalized to online production, as they may result from distinct learning or acquisition biases.

Table 1: Passage summary by genre. This table shows averages for each document type, along with standard deviation values.

	Plays [8]	Interviews [8]	Blogs [8]	Total [24]
Length (Words)	374.6 (121.6)	426.1 (43.0)	438.0 (84.6)	412.5 (89.8)
Length (Sentences)	55.4 (22.4)	40.6 (9.2)	37.0 (12.9)	44.3 (17.2)
# Data Points	44.9 (14.0)	42.3 (9.1)	48.5 (16.0)	45.3 (13.0)
Subject Omission	36.6% (14.5)	20.4% (15.4)	28.3% (19.8)	28.5% (17.3)

Gordishevsky & Avrutin, 2003). Previous quantitative investigations suggest that null elements are typically reserved for more accessible or topical referents (Givón, 1983; Gundel, Hedberg, & Zacharski, 1993; Kameyama, 1985); similar proposals have also been made by e.g. Ariel (1990).

The optional omission of syntactic constituents has previously been investigated using an information-theoretic framework (Resnik, 1996). Although not directly addressing the role of contextual predictability, Resnik’s work suggests that speakers are more likely to leave out optional objects in English (as in ‘I ate [dinner]’) if their meaning is more predictable, or inferable given the preceding verb. Resnik shows that verbs which have a more restricted set of (probable) direct objects, or that are more informative as to the identity of their objects, such as ‘eat’ or ‘drink’ (as opposed to ‘take,’ or ‘see’), are also more likely to be produced without those objects. Resnik’s study does not however directly look at the information carried by the objects (i.e. how predictable the objects themselves are in context). In this study, a verb’s entropy with respect to the semantic properties of its objects (i.e. how much uncertainty there is, upon encountering the verb, regarding the identity of the object that follows it) only correlates with how likely those objects are to be explicitly mentioned. Similarly, Brown and Dell (1987) show that more inferable instruments are more likely to be omitted, but do not directly measure the predictability of the omitted material. The work described here, therefore, presents for the first time a direct investigation of the role of contextual predictability in the omission of complex syntactic constituents, and the first quantitative investigation of its role in the omission of subjects.

Finally, work investigating communicative pressures on production cross-linguistically has to date has been sparse. In general, most work in linguistic production has been limited to English (Jaeger & Norcliffe, 2009), and work on comprehension to a few well-investigated and typically closely related languages (Anand, Chung, & Wagers, 2011), leaving open the question of whether postulated production pressures or preferences, as well as their observed interactions, are universal, or language-specific. This study helps address this question by looking at pressures on omission of clause subjects in Russian.

Materials

A corpus of 24 brief passages was created using material from 8 personal blogs, 8 interviews, and 8 plays³, by 21 different authors. Each passage was selected from the beginning of the text, to avoid discontinuity effects. Only passages with subject omission present were considered (register effects on omission, and corresponding rates, are discussed in Zdorenko, 2009). Although source selection was limited and non-randomized, the overall rate of subject omission is comparable to that typically found in personal speech (~20-35%), as shown in Table 1.

All animate subjects, and concrete inanimate subjects mentioned in the text – i.e., those that might be plausible candidates for omission – were included in the experiment, for a total of 1085 data points. The subsequent analysis was restricted crucially to cases where dropping or pronouncing the subject did not categorically alter or substantially bias meaning, as alternation in those cases would not be optional. Subjects with contrastive focus, where contrast with the preceding subject is intended, were excluded, as omission instead strongly implies co-reference with it. It was additionally restricted to the canonical subject cases (Nominative or Dative), as other ‘subjects’ (9 data points; 0.8%) are typically analyzed as impersonal constructions, or are difficult to distinguish from topicalization.

Measuring Probability Experimentally

In order to obtain a measure of subject predictability, given preceding material, an experiment was conducted, based on the design of a similar study measuring the effects of predictability on use of English referring expressions (Tily & Piantadosi, 2009)⁴. 70 native speakers of Russian were recruited online to participate in a ‘guessing game,’ where they were presented with 2 corpus passages, each from a different genre and in counterbalanced order. Passage text was revealed to participants incrementally, as they were asked to guess the identity of upcoming concealed subjects, by clicking on any previously mentioned referent, or on ‘Someone or Something New,’ if they believed that no previously

³Although dialogue in plays is arguably artificial, one would expect roughly the same pressures on production to play out – this is in fact supported by Manin (2006), which looked at the relation between contextual predictability and word length in a variety of literary texts.

⁴The code for running the experiment was obtained from Harry Tily upon request, and used with minimal modification.

Table 2: This table shows the beta coefficients associated with each main effect in the model, as well as corresponding standard errors, z -values, significance levels, and contribution to log-likelihood. (z) indicates the predictor was centered and standardized.

	Coef β	SE(β)	z	p	p_{χ^2}
Intercept	-1.20	0.27	-4.4	<.0001	
Information	-0.30	0.12	-2.4	<.05	0.015
Last Mention Form: Null	0.45	0.19	2.4	<.05	0.021
Coreferent with Preceding Subj.: Yes	1.33	0.22	6.0	<.0001	1.6e-09
Last Mention Function: Subj.	0.60	0.28	2.1	<.05	0.031
Distance in Words to Last Mention (z)	-0.34	0.11	-3.1	<.01	0.002

mentioned referent was a plausible continuation. After each guess, participants were given immediate accuracy feedback, as well as ‘points’ for accurate guesses⁵, and passage text up to the next guessing point was revealed. Given free word order in Russian, and the consequent impossibility of determining the ‘position’ of an omitted subject, the guessing point in each case was the clause onset.

Following Manin (2006) and Tily and Piantadosi (2009), average guessing accuracy for each data point, across participants, was used as a proxy for predictability. A measure of the information carried by each subject was obtained using the following equation:

$$I(Subject_i) = -\log_2 P(Response = \text{correct} + 1)$$

Here, $Subject_i$ is the subject of the i^{th} clause in the data set, and $P(Response = \text{correct})$ reflects the probability of participants accurately guessing this subject’s identity, given preceding context. 1 was added to the count of participants that accurately guessed the subject’s identity, given the relatively high number of subjects that were never accurately guessed – as the log transformation otherwise produced an infinitely high measure of information for these cases⁶.

Corpus Study

To test whether more predictable subjects are preferentially omitted, the dataset was first additionally annotated for several control predictors:

- *Form of last mention*: Whether the most recent coreferent mention is null (1) or overt (0).
- *Grammatical function of last mention*: Whether the grammatical function of the most recent coreferent mention is subject (1) or other (0).

⁵Points were given in order to encourage participants to make accurate guesses, rather than guessing randomly, or repeatedly guessing ‘something new.’ Some form of accuracy feedback was unavoidable without purposefully omitting the previously concealed and referentially unambiguous subjects, which would substantially disrupt the flow and interpretability of the passage.

⁶It should be noted that using either the raw predictability measure (average accuracy), or setting an arbitrary upper bound for the information measure (4 bits) instead, produced the same results in the subsequent analysis.

- *Coreferent with preceding subject*: Whether the clause subject in question is coreferent with the preceding subject (1) or not (0).
- *Distance to last mention in words*: The distance to the most recent coreferent mention in words.
- *Number of previous mentions*: The number of times the referent has previously been mentioned in the discourse.
- *Number of preceding referents*: The number of individual referents introduced in the preceding discourse.

These predictors were chosen, following Tily and Piantadosi (2009), based on the expectation that the same factors which facilitate the use of pronouns also facilitate omission (cf. Gundel et al., 1993; Kameyama, 1985). Thus more salient subjects (Ariel, 1990) – e.g., those mentioned most recently, those having fewest competing referents, and those mentioned more times in the discourse – were expected to be preferentially omitted. Similarly, subjects whose most recent mention was null, or in parallel syntactic position (Arnold, 1998; Gordon, Grosz, & Gilliom, 1993), were expected to be preferentially omitted. As genre was collinear with the other predictors, it was not included in the analysis.

Different overt subject forms (pronouns, names, descriptions) were collapsed in the model⁷. Subjects of conjoined clauses were included in the analysis – it should be noted, however, that the effect of information content on subject omission remains if this data is removed from the analysis ($\beta = -0.29$, $z = -2.3$, $p < .05$). Given the difficulty of determining ‘correct’ guesses for plural referents, only singular referents were considered. Similarly, only those that were previously mentioned in the discourse were included, given that a correct ‘Someone or Something New’ guess for a previously unmentioned referent does not indicate that the participant has accurately guessed the concealed subject’s identity. A total of 720 utterances (362 overt and 358 null subjects), from an initial 1085 (66.4%), were included in the model.

⁷Crucially, if the analysis is restricted to null subjects and pronouns only (# utterances: 656), a marginally significant effect of Information, in the same direction, remains ($\beta = -0.24$, $z = -1.9$, $p = .06$) – suggesting that the results below generalize to the more specific pronoun/zero alternation.

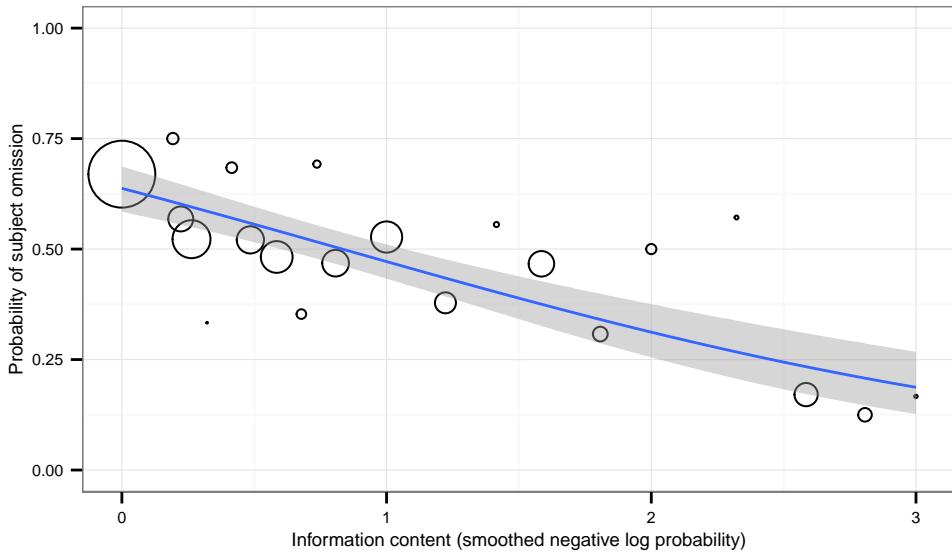


Figure 1: This plot shows the probability of subject omission as a function of information content. The points represent average omission likelihood for each ‘information’ value, with size proportional to the number of cases. The line represents a logistic regression curve, with the shaded region corresponding to the 95% confidence interval.

A generalized mixed-effects logistic regression analysis, arrived at using χ^2 -test model comparison, showed that speakers preferentially omit more predictable subjects ($\beta = -0.30$, $z = -2.4$, $p < .05$). Passage was included as a random intercept, to adjust for passage and author-specific subject omission rates. A summary of the model can be found in Table 2. This effect holds after controlling for other expected predictors of subject omission (*number of previous mentions* and *number of preceding referents* did not contribute significantly to model fit, and were therefore excluded). All control predictors behave as expected: subjects are more likely to be omitted when the form of last mention is ‘null,’ when they are coreferent with the preceding subject, and when the grammatical function of the last mention is subject. They are less likely to be omitted as distance in words to last mention increases. The probability of subject omission, as a function of information content, is plotted in Figure 1.

In summary, speakers are significantly more likely to omit subjects when they are more predictable in context. Further, the effect of information carried by the referent, on likelihood of subject omission, continues to hold after other likely predictors of subject omission are controlled for.

Discussion

The corpus study described in this paper tested the hypothesis that the choice to omit optional subject constituents in Russian is subject to communicative pressures – specifically, the pressure to avoid overly redundant elements in speech and text. The analysis demonstrates that this is indeed the case: speakers are more likely to omit subjects whose identity is easier for comprehenders to guess, based on preced-

ing material. Crucially, this effect holds after controlling for other likely predictors of subject omission. This suggests that online production of syntactic constituents is subject to the same pressure for communicative efficiency as is production of more functional, non-constituent elements at other levels of processing.

These findings are most straightforwardly compatible with theories of efficient communication that make reference to omission or reduction of highly predictable, or redundant, elements in online language production. The theory of Uniform Information Density (UID; Jaeger, 2010), for example, makes the prediction that these effects should be present at all levels of production, while most similar theories (e.g., Aylett & Turk, 2004; Jurafsky, Bell, Gregory, & Raymond, 2001) do not make reference to production beyond the phonetic/phonological levels. This study therefore provides further direct evidence that the effects of predictability on production extend to the syntactic level, including to the omission of full syntactic constituents. The findings are also compatible with theories of efficient reference production specifically (Arnold, 2008), which propose that speakers use less informative or reduced expressions for those referents that their addressees are more likely to expect in context, and that using such expressions is in general more efficient, in terms of speaker effort.

In that light, it is worth noting that the control predictors used in this study are themselves taken by e.g. Arnold to be cues for expected reference. The Expectancy hypothesis for reference production is outlined further in Arnold (2008), but would predict that any properties which increase the likeli-

hood of subsequent mention of a particular referent – such as parallelism of grammatical function and recency of mention (Arnold, 1998, 2008) – should correlate with higher rates of subject omission. With respect to the present study, then, including these control predictors in the model provides a stronger test of whether contextual predictability (information content) has a unique influence on production. In contrast to the reference-specific account outlined above, Uniform Information Density more generally goes beyond the claim that specific cues in discourse play a role in shaping production, but rather makes a broader claim that production is driven by predictability-based pressures at all levels.

From a methodological standpoint, following Manin (2006) and Tily and Piantadosi (2009), the task in this experiment might be seen as an alternative method for estimating the probabilities associated with linguistic events in context. Potentially supplementing corpus and elicited production methods for providing probability estimates, it may in principle provide a more accurate estimate of the probabilities that comprehenders subjectively assign to various linguistic events. For example, there appear to be systematic differences between probability estimates obtained using corpus studies and traditional cloze tasks (Smith & Levy, 2011); it remains unclear which provides a better representation of speakers' subjective probabilities, in either production or comprehension. These issues notwithstanding, the methodology used in this study may also be a more feasible option for obtaining similar probability estimates for linguistic events in languages where no suitably large corpora are available.

Finally, currently there is poor consensus on what precisely causes any given language to display a conventionalized or systematic pattern of subject omission (such as that found in some Romance languages). Although richness of subject-verb morphological agreement appears to play a role (Chomsky, 1982), it is neither a sufficient nor necessary criterion (Jaeggli & Safir, 1989), with many languages that demonstrate no morphological agreement nevertheless demonstrating extensive *pro-drop*⁸ – including, for example, Chinese and Japanese. I propose that the average contextual predictability, or information content, of pronouns may play a role in the licensing and frequency of *pro-drop* cross-linguistically. Pronouns may in general be considered good candidates for elision, as they are relatively uninformative, and in some languages effectively replicate grammatical information provided by the verb. Conventionalized omission (or reduction) of relatively uninformative linguistic elements would, on average, increase efficiency of communication. There is, for example, evidence that average predictability is a unique predictor of phoneme deletion in English (Cohen Priva, 2008), and word length cross-linguistically (Piantadosi et al., 2011).

A language in which the intended reference of a pronoun

⁸Here I will use the term *pro-drop* to refer only to referential subject omission.

is, on average, easier to predict from preceding context (due, for example, to more predictable patterns of co-reference), might then be more likely to license systematic subject omission. Cross-linguistically, null pronouns often preferentially or obligatorily co-refer with preceding subjects, to a greater extent than overt pronouns (Carminati, 2002)⁹. This is in line with the account proposed here, as one would expect any given language to preferentially use the less informative (i.e. null) form for the more predictable pattern of co-reference. For subjects, this is typically to co-refer with preceding subjects (Chambers & Smyth, 1998; Smyth, 1994). In short, effects of average contextual predictability could potentially both help predict which languages are more likely to license subject omission, and account for why null subjects, to a greater extent than their overt counterparts, preferentially co-refer with preceding subjects. This hypothesis is in line with recent resurged interest in functional explanations for typological patterns (Jaeger & Tily, 2011), and may be reserved for future work.

Summarizing, theories of efficient communication such as Uniform Information Density (UID) predict that the probability of omission of an optional linguistic element is inversely correlated with the information that it carries in context. Further, UID specifically predicts that these pressures should play an active role at all points in discourse where speakers have a choice between different ways of formulating a message, and that their influence is not limited to any specific level of processing, or only to functional linguistic elements. The findings described in this paper support this hypothesis, and demonstrate that more predictable syntactic constituents are preferentially omitted, when they are more predictable given preceding context. This provides direct evidence that communicative pressures on production extend to the syntactic level, and, crucially, that UID is indeed a general principle of production.

Acknowledgments

I am very grateful to Pranav Anand for extensive assistance with all aspects of this project, as well as to Adrian Brasoveanu and Jaye Padgett for their assistance with methods, and feedback on earlier versions of this manuscript. I also thank four anonymous CogSci reviewers.

References

Anand, P., Chung, S., & Wagers, M. (2011). Widening the net: Challenges for gathering linguistic data in the digital age. White paper 121, SBE 2020: Future Research in the Social, Behavioral & Economic Sciences. NSF.

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

⁹Although it has been proposed that null and overt pronouns demonstrate a division of labor, with overt pronouns in *pro-drop* languages preferentially co-referring with objects or non-topical elements (Carminati, 2002), the evidence on this has been mixed, and suggests rather that null subject pronouns may have a *stronger* preference for co-reference with subjects (e.g., Ueno & Kehler, 2010).

Arnold, J. E. (1998). *Reference form and discourse patterns*. Doctoral Dissertation, Stanford University.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4), 495-527.

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, 31-56.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92-111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113(2), 1001-1024.

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, 19(4), 441-472.

Carminati, M. (2002). *The processing of Italian subject pronouns*. Doctoral Dissertation, UMass, Amherst.

Chambers, C. C., & Smyth, R. (1998). Structural paraparallel and discourse coherence: a test of centering theory. *Journal of Memory and Language*, 39, 593-608.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.

Cohen Priva, U. (2008). Using information content to predict phone deletion. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*.

Franks, S. (1995). *Parameters of Slavic morphosyntax*. Oxford: Oxford University Press.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of ACL-2002* (p. 199-206).

Givón, T. (1979). *On understanding grammar*. New York, NY: Academic Press.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-linguistic study* (p. 1-41). Amsterdam: John Benjamins.

Gordishevsky, G., & Avrutin, S. (2003). Subject and object omissions in child Russian. In *Proceedings of IATL 19*.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17(3), 311-347.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274-307.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Association for computational linguistics* (p. 1-8).

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

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23-62.

Jaeger, T. F., & Norcliffe, E. (2009). The cross-linguistic study of sentence production: State of the art and a call for action. *Language and Linguistic Compass*, 3(4), 866-887.

Jaeger, T. F., & Tily, H. (2011). Language processing complexity and communicative efficiency. *WIRE: Cognitive Science*, 2(3), 323-335.

Jaeggli, O., & Safir, K. J. (1989). The null subject parameter and parametric theory. In O. Jaeggli & K. J. Safir (Eds.), *The null subject parameter* (p. 1-44). Dordrecht: Kluwer Academic Publishers.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (p. 229-254). Amsterdam: John Benjamins.

Kameyama, M. (1985). *Zero anaphora: The case of Japanese*. Doctoral Dissertation, Stanford University.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems 19* (p. 849-856). Cambridge, MA: MIT Press.

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

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

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, 61, 127-159.

Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In *Proceedings of the 33th Annual Meeting of the Cognitive Science Society*.

Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, 23, 197-229.

Tily, H., & Piantadosi, S. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In *Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference*.

Ueno, M., & Kehler, A. (2010). The interpretation of null and overt pronouns in Japanese: Grammatical and pragmatic factors. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (p. 2057-2062).

Zdorenko, T. (2009). Subject omission in Russian: a study of the Russian National Corpus. *Language and Computers*, 71(1), 119-133.

Zipf, G. (1949). *Human behaviour and the principle of least effort*. Cambridge, MA: Addison-Wesley.