

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Effects of contextual predictability on optional subject omission in Russian

Permalink

<https://escholarship.org/uc/item/1fc8s7jx>

Author

Kravtchenko, Ekaterina

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

**EFFECTS OF CONTEXTUAL PREDICTABILITY ON OPTIONAL SUBJECT
OMISSION IN RUSSIAN**

A thesis submitted in partial satisfaction
of the requirements for the degree of

MASTER OF ARTS

in

LINGUISTICS

by

Ekaterina Kravtchenko

June 2013

The Thesis of Ekaterina Kravtchenko
is approved:

Professor Pranav Anand, Chair

Professor Adrian Brasoveanu

Professor Jaye Padgett

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by
Ekaterina Kravtchenko
2013

Table of Contents

List of Figures	v
List of Tables	vi
Abstract	vii
Acknowledgments	viii
1 Introduction	1
2 Background	3
2.1 Efficient Communication	3
2.1.1 Uniform Information Density	4
2.1.2 Backward Predictability	7
2.2 Use of Attenuated Referential Expressions	8
2.2.1 Efficient Use of Referential Expressions in English	8
2.2.2 Pro-Drop	10
2.3 Omission of Subjects in Russian	14
3 General Methods	15
3.1 Materials	17
3.2 Annotation	18
4 Experiment 1	20
4.1 Methods	21
4.1.1 Design	21
4.1.2 Participants	22
4.1.3 Procedure	23
4.2 Results	23

5	Experiment 2	30
5.1	Methods	30
5.1.1	Design	30
5.1.2	Participants	32
5.2	Results	33
6	Discussion	40
6.1	Experiments 1 and 2	41
6.2	Communicative Efficiency and Audience Design	41
6.2.1	Speakers' Use of Subjective Probabilities	43
6.3	Task	44
6.4	Follow-Up	46
	Appendix A. Tables	49
A.1	Experiment 1	49
A.2	Experiment 2	49
	Appendix B. Figures	49
B.1	Experiment 1	50
B.2	Experiment 2	51
	Appendix C. Predictors	54
	Appendix D. Sources	56
	References	59

List of Figures

1	Surprisal vs. average accuracy (contextual probability)	25
2	Overt/null subject density by Surprisal (exp. 1)	26
3	Coefficient bar plot - mixed logit model (exp. 1)	28
4	Experiment 1 model fit	28
5	Overt/null subject density by Surprisal (exp. 2)	34
6	Coefficient bar plot - mixed logit model (exp. 2)	36
7	Experiment 2 model fit	36
8	Surprisal by subject form and Form of Last Mention	38
9	Coefficient bar plot - mixed logit model (exp. 1 and 2)	40
10	Overt/null subject density by Entropy (exp. 1)	50
11	Coefficient bar plot - mixed logit model (exp. 1) subset	50
12	Experiment 1 (subset) model fit	51
13	Overt/null subject density by Entropy (exp. 2)	51
14	Coefficient bar plot - mixed logit model (exp. 2) subset with interaction	52
15	Experiment 2 (subset) model fit with interaction	52
16	Coefficient bar plot - mixed logit model (exp. 2) subset	53
17	Experiment 2 (subset) model fit	53

List of Tables

1	Passage summary statistics by genre	18
2	Mixed logit model - experiment 1	27
3	Mixed logit model - experiment 2	35
4	Mixed logit model - experiments 1 and 2	39
5	Mixed logit model - experiment 1 (subset)	49
6	Mixed logit model - experiment 2 (subset) with interaction	49
7	Mixed logit model - experiment 2 (subset)	49
8	Blog sources	57
9	Play sources	57
10	Interview sources	58

Abstract

Effects of contextual predictability on optional subject omission in Russian

by

Ekaterina Kravtchenko

The role of contextual predictability in language processing and production has gained increasing attention in recent years, but in production remains relatively understudied above the phonetic/phonological levels. Further, the existence and source of predictability effects is still contested, and it remains unclear which phenomena these effects extend to, as well as what their explanatory range is (Jurafsky, 2003). I present two experiments which investigate the role of contextual predictability in the choice to omit optional subjects in Russian. The results demonstrate that when other predictors of referential expression choice are taken into account, contextual predictability, conditioned on preceding context, remains a significant predictor of whether an optional subject is pronounced or elided in colloquial text or speech. Contextual predictability conditioned on following context, in contrast, does not appear to exert significant influence after effects of preceding context are controlled for.

Acknowledgments

I would not have been able to complete this project without the guidance of my committee members, assistance from friends within and outside of the department, and the support of my family. I would like to express particular gratitude to my advisor, Dr. Pranav Anand, for his guidance, patience, assistance with all aspects of this project, and ongoing support on so many fronts. This thesis would not have been possible without his help. I would like to thank Dr. Adrian Brasoveanu in particular for extensive assistance with statistical analysis, and many helpful discussions of the methods and theoretical issues involved. I would like to thank Dr. Florian Jaeger for initial discussion of the project, help with developing methods of testing the hypothesis, and his comments on the finished project. I would also like to thank Dr. Jaye Padgett for detailed and helpful comments on the written work, extensive discussion of the theoretical claims and implications, and support. Many thanks also go to the rest of the faculty for providing support throughout this process, and the rest of my time here – in particular, to Sandra Chung, Donka Farkas, Jim McCloskey, and Grant McGuire.

Special thanks go to Nick Deschenes, Allan Schwade, Clara Sherley-Appel and Anie Thompson for discussion of this work, editing and proofreading help, and above all for their friendship and support. Many thanks are due, as well, to participants of the Winter 2012 Research Seminar, who helped shape this work, for their time and invaluable comments. Thanks also go to Carlos Gallo, Boris Harizanov, Wendell Kimper, Roger Levy, Oliver Northrup, Maria Polinsky, Jeff Runner, Joey Sabbagh, Matt Wagers, Masaya Yoshida; and the audiences at LASC 2012, LSA 2013, and CUNY 2013, for helpful comments and feedback. Finally, I would like to thank my parents, grandparents, my sister, and my two younger brothers for encouragement, emotional and material support throughout this period, and for their enduring patience.

1 Introduction

A long-standing question in the language sciences concerns whether language is optimized for efficient communication. The basis for this claim is the idea that speakers aim to shape their utterances in a manner that increases the likelihood of efficient and low-error message transmission. Efficiency, in this context, refers to the amount of signal used to communicate a message, with less signal reflecting greater efficiency. On the other hand, error refers to the likelihood for miscommunication, including, for example, the likelihood of environmental interference during transmission of the message to the listener, or error due to processing difficulty/confusability.

Until recently, most claims along these lines have been either limited to qualitative typological studies, or appeals to intuition (Jaeger & Tily, 2011). Information-theoretic approaches (Shannon, 1948) to language analysis, in contrast, have more recently been used to formalize the notion of efficiency, allowing for quantitative investigation of some of these claims, including in online production. Under this approach, it is proposed that optimal communication of a message, within other constraints on language form or production, involves keeping the contextual probability of upcoming elements maximally constant – close to, but not exceeding an arbitrary ‘channel capacity,’ beyond which the likelihood for error becomes unacceptable (Levy & Jaeger, 2007; Jaeger, 2010).

This approach would predict that, within the constraints of a speaker’s grammar, the meaning they intend to convey, and other pressures on language production, speakers should be more likely to reduce or omit those elements that are redundant (contextually predictable), therefore avoiding sacrificing recoverability. Speakers should also be more likely to either produce, or use more signal in producing, those elements that are contextually unpredictable. The question asked here is whether pressure for robust (low-error) and efficient communication can help account for whether or not speakers choose to omit optional subjects in Russian.

Clause subjects may be optionally elided in Russian (Franks, 1995), 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 što ona/Ø zbolela
Masha called.FEM Petia, because she/Ø fell-sick.FEM
“Masha_i called Peter_j, because [she_i] was sick.”

Main clause subjects may also be omitted given appropriate context, although isolated sentences are often infelicitous (unless the subject referent is very contextually salient):

- (2) Maša pozvonila Pete včera. Ona/Ø skazala što ne pojdet na rabotu.
Masha called.FEM Petia yesterday. She/Ø said.FEM that NEG go to work
“Masha_i called Peter_j yesterday. [She_i] said that she won’t go to work.”

Gordishevsky and Avrutin (2003, 2004) note that an elided subject must typically be “pragmatically motivated,” as in response to Wh-questions, and salient situationally or temporally if lacking a linguistic antecedent. Kravtchenko (2008) also observed that intrasentential subjects may only be omitted if they co-refer with the subject (but not object) of the preceding clause¹

- (3) Maša pozvonila Pete, potomu što on/*Ø zbolel
Masha called.FEM Petia, because he/*Ø fell-sick.MASC
“Masha_i called Peter_j, because he_j was sick.”

These observations address the various contexts in which subject omission is permitted. However, they do not otherwise address why speakers might choose or prefer to elide subjects, given the option of doing so, or whether subject omission might confer a communicative benefit in some contexts.

The remainder of the paper describes background literature on these topics, and two experiments that were designed to determine the extent to which contextual predictability influences speakers’ choice to omit clause subjects. The results of the first experiment demonstrate that speakers’ decisions to drop optional subjects are influenced by contextual predictability conditioned on preceding context, with more predictable subjects more likely to be dropped. In contrast, the results of the second (exploratory) experiment suggest that predictability conditioned

¹The possible significant and relevance of the subject-antecedent bias is discussed further in section 2.2.2, as well as in section 6.4.

on following context may not play a role in subject omission, outside of, possibly, environments where subject omission is particularly likely to occur.

2 Background

This section briefly discusses theories of communicative efficiency, as well as how they may bear on a speaker's choice to pronounce or reduce/omit an optional linguistic element. Cross-linguistic accounts of *pro*-drop (omission of referential verb arguments) are briefly reviewed; as well as research on both use of attenuated referring expressions, and argument omission, within the framework adopted here.

2.1 Efficient Communication

A number of theories address how speakers might structure their utterances in order to communicate their intended message more efficiently or accurately. Examples include availability-based accounts, at least in some formulations (e.g. Ferreira & Dell, 2000); ambiguity avoidance accounts (e.g. Clark & Fox Tree, 2002); dependency processing accounts (e.g. Hawkins, 2004); and redundancy-avoidance accounts (e.g. Aylett & Turk, 2004; Jaeger, 2010). In order to demonstrate that a drive towards greater communicative efficiency plays a role in language production, these accounts have often focused on investigating whether speakers preferentially use more 'efficient' utterance variants, when given otherwise (more-or-less) meaning-equivalent alternatives, in expressing their intended message.

Several theories of language production have specifically addressed the question of whether speakers aim to make their utterances more efficient by omitting material less necessary for accurate transmission of the message (e.g., Aylett & Turk, 2004; Genzel & Charniak, 2002; Zipf, 1949). The most robust evidence for this has been at the phonetic and phonological levels – e.g. syllable duration, or articulatory detail (Aylett & Turk, 2004; Bell et al., 2003; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009). One theoretical approach, Uniform Information Density (Levy & Jaeger, 2007; Jaeger, 2010), posits that speakers attempt to minimize both

communicative effort, and transmission error at all levels of production, by preferentially reducing or omitting more redundant or predictable elements in their utterances, in effect arriving at an optimal ‘trade-off.’

2.1.1 Uniform Information Density

Uniform Information Density is a theory of efficient communication which proposes that speakers organize their utterances so as to optimally avoid both excessively high and low concentration of information, per time segment of speech. Under the assumptions of this framework, speakers face two competing pressures:

1. Communicate the maximum amount of information possible per time segment of speech, for maximally efficient (i.e. least costly in terms of effort) communication (Zipf, 1949). To this end, speakers would be expected to omit redundant element of their speech, in order to minimize processing and production (e.g. articulatory) effort.

2. Communicate the given information in a manner maximally likely to accurately communicate the intended meaning across to the recipient. Since redundancy may aid recoverability (Nadig & Sedivy, 2002), and communication occurs through a (sometimes literally) noisy channel, speakers would be expected to retain some redundant material to increase likelihood of accurate message transmission.

The communicatively optimal outcome is proposed to be an [arbitrary²] ‘trade-off’ of the above pressures. In order to evaluate whether a particular production choice is more efficient than another, it is first necessary to formalize or quantify efficiency in a manner that would make accounts such as UID empirically testable. Taking UID as an example, and the framework primarily considered in this paper, it would be efficient to omit those utterance elements that are more contextually predictable, or redundant. This minimizes production effort, but without excessively increasing the likelihood of miscommunication.

²The trade-off that is judged optimal – proposed to be an aim to communicate close to, but below channel capacity (see below) – may differ between situations and individuals.

Information-theoretic approaches (Shannon, 1948, 1951) can be used to quantify the role of information, or contextual predictability in language production. Information theory demonstrates that information is transmitted most efficiently through a noisy channel, if it is done so at a constant rate below channel capacity (defined probabilistically). In this framework, the probability of occurrence of a linguistic event, such as a word or phoneme, given preceding material, is used as a measure of its contextual probability, or redundancy. Communication is therefore more efficient, all else being equal, if the probability of any given linguistic event neither exceeds nor falls too far below channel capacity (see figures below). It should be noted that this is not possible, or expected, to be achieved in all cases, given that language (and its production) is subject to multiple other pressures or constraints. Contextual probability is typically quantified in *bits*, or the negative log-probability of an event given preceding context.

The case of optional *that*-omission, discussed in (Jaeger, 2010), illustrates how a speaker's production choices might reflect UID principles. The examples below, taken from the above paper, demonstrate two cases where the probability of a complement clause onset, based on matrix verb CC-subcategorization frequencies, determines the overall likelihood of omission of the optional complementizer *that*:

(4) My boss confirmed {that/...} we were absolutely crazy.

(5) My boss thinks {that/...} I am absolutely crazy.

In example (4), the relatively low probability of a complement clause following verbs such as *confirm* leads to speakers preferentially pronouncing the complementizer (Jaeger, 2010; Jaeger, Levy, & Ferreira, 2010). Pronouncing the optional *that* spreads the information carried by the complement clause onset over more words (*that* + 1st word), increasing its redundancy – and therefore (hypothetically) staying below channel capacity, thus minimizing transmission error. In contrast, the high probability of a complement clause following verbs with high CC-subcategorization frequency, such as *think* in example (5), leads speakers to prefer omission of the complementizer, reducing the redundancy of the clause onset, and (hypothetically) communicating closer to channel capacity, thereby conserving effort.

These effects are hypothesized to arise at all levels of production (Jaeger, 2010), reflecting a more general drive towards greater communicative efficiency above the lexical level. Although it has been most conclusively demonstrated at the phonetic and phonological levels, there is increasing evidence that preferential omission of more contextually predictable elements occurs at other levels of production. This includes optional adjunct (instrument) omission in English (Brown & Dell, 1987), optional object omission in English (Resnik, 1996), referential expression choice in English (Tily & Piantadosi, 2009), the choice to distribute information over one or two clauses in English (Gómez Gallo, Jaeger, & Smyth, 2008), use of optional contractions in English (Frank & Jaeger, 2008), the (arguably non-meaning-equivalent) alternation between *some* and *some of* (Degen & Jaeger, 2011), and optional case-marker omission in Japanese (Kurumada & Jaeger, 2012). As might be noted, most investigations to date have been restricted to English, with exception of Kurumada and Jaeger (2012); Brown and Dell (1987) and Resnik (1996), in addition, only provide suggestive evidence that more contextually predictable arguments or adjuncts are elided, as the actual predictability in context of the omitted elements is not measured.

UID predicts that effects should be evident at all ‘choice points’ in production, i.e. when multiple forms are available to express the same message. In these instances, speakers would be expected to choose the variant that transmits the message most accurately and efficiently - or close to, but just below the channel capacity. The most basic prediction is that speakers are expected to pronounce those elements that are less predictable given preceding context, and omit those elements that are more predictable (or redundant) given preceding context. At word level, this is expected in cases of optional argument omission, *pro*-drop, and ellipsis (Jaeger, 2006, 2010). However, it has only been empirically demonstrated for omission of functional elements, e.g. optional *that*-omission (Jaeger, 2010), with most investigations, as mentioned, limited to English.

The effect of redundancy avoidance on reduction or omission of more semantically or syntactically complex elements, such as verb arguments, is more difficult to test experimentally within this framework. In this respect, it may be more feasible to investigate optional omission

of elements such as pronouns. For example, sentence recall experiments have been used to investigate whether speakers preferentially recall and reproduce sentences (whether correctly or incorrectly) that more closely conform to those that UID would predict to be optimal (Jaeger et al., 2010; Kurumada & Jaeger, 2012). However, in this paradigm, it seems unlikely that participants might misremember the presence or omission of constituents any more complex than pronouns, or optional functional elements:

(6) ✓ Masha told me {**that**/Ø} she was too sick to go to work.

(7) ? Masha told me that Russia won {**the World Cup**/Ø}.

Omission of more complex items, such as verb arguments or adjuncts, may also be more difficult to investigate in corpora, given, for example, the difficulty of determining the site, or form of elided elements. In some contrast, if the question is reframed to ask, for example, whether more predictable referents (not specific words or forms) are more likely to be omitted or referred to using attenuated forms – where intended reference is unambiguous – one does not need to determine the precise form or structure of the omitted material. In this spirit, Tily and Piantadosi (2009) provide evidence that the choice of referring expression in English is sensitive to the predictability of referent mention in context, with shorter and less informative expressions such as pronouns preferred for more predictable referents. Building up on this finding, the optionality and frequency of subject omission in Russian provides an optimal test case of the hypothesis that *omission* of non-functional elements is sensitive to these pressures, as well.

2.1.2 Backward Predictability

In contrast to predictability conditioned on preceding context, very little is known about the role of ‘backward predictability’ (i.e. predictability conditioned on following material) in production. There is limited evidence from Frank and Jaeger (2008), which looked at optional verb+auxiliary contractions in English, that the conditional probability of BE and HAVE (but not NOT), given the following context, is predictive of the likelihood of contraction. Bell et al. (2003) and Bell et al. (2009) more straightforwardly demonstrate that backward predictability is

a significant predictor of word duration, with both content and function words pronounced with shorter duration when they are more predictable given following context. Pluymaekers, Ernestus, and Baayen (2005), similarly, demonstrate that it is a significant predictor of articulatory reduction. There is, however, little evidence that it plays a role at other levels of production.

Summarizing, although it would appear plausible that predictability conditioned on following context influences production, in a manner similar to that of preceding context, there is currently no theoretical account of this, nor substantial empirical evidence for it at most levels of production. In Russian, it may be somewhat difficult to distinguish the influence of backward predictability and more straightforward null subject ‘recoverability,’ given morphological agreement on the verb, although it could be argued that the latter is subsumed by the former. It may be possible to investigate whether backward predictability effects are in fact subsumed by effects of morphological recoverability. The second experiment described in this paper addresses the more general question of whether speakers preferentially omit subjects that are more predictable given following context.

2.2 Use of Attenuated Referential Expressions

This section provides a brief overview of previous research on omission of verb arguments in an information-theoretic framework (Resnik, 1996); and the experiments conducted by Tily and Piantadosi (2009) concerning use of attenuated referring expressions in English, which formed the basis of the work described in this paper. It also briefly discusses omission of referential arguments cross-linguistically.

2.2.1 Efficient Use of Referential Expressions in English

Resnik (1996): This work demonstrates that speakers are more likely to leave out optional objects in English (as in ‘I ate [dinner]’), if their meaning is more predictable given the preceding verb. The prediction UID makes in the case of optional verb arguments is that the more likely their mention is given preceding context, the more likely speakers are to omit them. Verbs that 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 e.g. ‘take,’ or ‘see’), are also more likely to be produced without those objects. In other words, this is indirect evidence that more contextually predictable, or inferable objects are also more likely to be omitted. The study does not, however, directly look at the information carried by, or Surprisal of the objects (i.e. how predictable the objects themselves are in context). Formally, a verb’s entropy with respect to the semantic properties of its objects (how much uncertainty there is, upon encountering the verb, regarding what [type of] object follows it) correlates with how likely those objects are to be explicitly mentioned.

Tily & Piantadosi (2009): The experiments described in this paper are modeled after those in Tily and Piantadosi (2009). As in cases of omission or phonetic/morphosyntactic reduction, UID makes the prediction that shorter and less informative referring expressions should be preferentially used for more contextually probable referents. In order to test the hypothesis that choice of referring expression is made with considerations of communicative efficiency, Tily and Piantadosi (2009) looked at referring expression choice in English: definite descriptions, such as ‘the President of the United States of America;’ proper names, such as ‘Obama;’ and pronouns, such as ‘he’. The hypothesis was that shorter and less informative referring expressions³ should be used for more predictable referents. The experimental task presented participants with a ‘game,’ where they were asked to guess the identity of upcoming NPs (in any syntactic position).

Participants read text passages selected from a corpus of WSJ text annotated for coreference (Weischedel, Pradhan, Ramshaw, & Micciulla, 2008). The passage text was shown up to the mention of the first (concealed) NP, at which point participants were asked to guess its identity: whether it was one of any referents previously mentioned in the discourse, or “Something new.” After they made a guess, they received immediate feedback with respect to their accuracy, as well as points for accurate guesses. The rest of the passage text, including the referent in question, were then revealed up to the next guessing point. The average guessing accuracy

³Definite descriptions are on average longest, and names are longer than pronouns; both definite descriptions and names are also more informative than pronouns.

across participants, for each NP, was used as a the primary measure of whether that referent was predictable given preceding context, with greater guessing accuracy reflecting greater referent predictability or redundancy. Tily and Piantadosi (2009) used an additional measure of entropy, or uncertainty regarding the identity of the upcoming referent: whether guesses as to an NP's identity, across participants, are 'spread out' over many referents, or concentrated on few. The general design is discussed in more detail in sections 4 and 4.2. The prediction made was that if a drive towards greater efficiency plays a role in production, including referring expression choice, there should be a preference for simpler and shorter forms, when the referent is more predictable.

It was found that, as predicted, writers are more likely to use shorter referring expressions (pronouns and proper names) when preceding context is more informative as to the identity of the upcoming NP – i.e., when the referent is more predictable in context. Writers are also more likely to use pronouns than names when entropy is higher, or when the correct referent is the only one with a relatively high probability of being mentioned⁴, while names are used when incorrect guesses are concentrated on fewer (and therefore more high-probability) competing referents. There was no separate effect of predictability on length of the NP, after the pronoun/proper name/description distinction is taken into account. The results suggest that shorter or less informative forms are preferentially used when the referent is more predictable and has fewer high-probability competitors, which is consistent with both efficiency-based accounts of production (e.g., Jaeger, 2006, 2010), and expectation-based models of reference production/interpretation (e.g., Arnold, 2008).

2.2.2 Pro-Drop

Here I briefly review theoretical accounts and descriptions of *pro*-drop, or omission of referential arguments of verbs, cross-linguistically. The question of whether subject omission in Russian can be counted as an instance of *pro*-drop is somewhat contentious, but it appears likely for a number of reasons, some recounted below, that Russian at the least shares a number

⁴Definite descriptions, as mentioned, are used when comprehenders are unable to accurately guess the referent at all.

of important properties with other *pro*-drop languages.

A number of theoretical accounts of *pro*-drop have been proposed, and progressively refined or altered. The latter has been necessitated by the discovery or more careful description of languages with referential argument omission which did not meet previously posited licensing or identification (recoverability) criteria (for an overview, see Huang, 2000). Definitions of *pro*-drop have also altered, but primarily it has been defined as omission of referential arguments of verbs, including subjects, direct objects, and indirect objects, licensed by the grammar of any particular language (what exactly constitutes licensing of verb argument omission, from a theoretical standpoint, is a separate topic which will not be discussed much here). More subtle definitions of *pro*-drop, or distinctions between e.g. different types of null subject languages, have also been proposed, although these have, at least to an extent, been theory-internal, though some have captured important empirical generalizations (Huang, 2000). In general, languages differ in whether they allow null subjects; which types of null subjects they allow (expletive only, or both expletive and referential); whether they allow omission of other arguments, such as verb objects (e.g. Chinese); and whether omission of subjects or other verb arguments occurs more-or-less freely, or is restricted to, for example, certain persons or tenses (e.g. Hebrew; Borer, 1989).

It was initially proposed that *pro*-drop was licensed only in languages with rich verb-argument morphological agreement (Chomsky, 1982), allowing for morphological recoverability of an omitted pronoun. This, for example, would account for why *pro*-drop is licensed in languages such as Italian and Spanish, but not in e.g. (modern) French or English. However intuitively appealing, although this appeared to provide good descriptive coverage, this generalization ran into a problem when it was observed that languages with no morphological agreement, such as Chinese and Japanese, also license *pro*-drop, with frequent omission of both subjects and verb objects. Further, languages with relatively restricted morphological agreement, such as Portuguese, allow omission of subjects, while languages with more extensive/uniform morphological agreement, such as German, do not typically allow subject omission. In short, there is no clear or stable correlation between morphological recoverability,

and licensing or rate of pronoun omission - it is neither the case that referential arguments are omitted only in languages where they are morphologically recoverable, nor that morphological agreement/recoverability is sufficient to license subject omission.

As a result, it was proposed that *pro*-drop is licensed only in languages with “morphologically uniform” inflectional paradigms (Jaeggli & Safir, 1989) – i.e. the language either has only underived, or only derived inflectional forms. This seems to have reasonably good coverage descriptively, but there is no obvious reason why morphological uniformity should be a precondition for licensing of *pro*-drop, although a number of theory-internal principles have been (mostly stipulatively) proposed to account for this (Huang, 2000). Furthermore, even the descriptive adequacy of this principle has come to be questioned, as both languages with uniform morphological agreement (e.g. German), and languages without (e.g. Old French, and possibly Brazilian Portuguese), exhibit contrary trends. In addition, Butt (2001) observed, in a survey of South Asian languages, that degree of morphological agreement does not appear to correlate with presence of verb argument omission; similarly, null subject languages with no verb-object morphological agreement may nevertheless permit omission of objects (Huang, 2000). A number of other proposals, discussed in Huang (2000) have also been on the market, but are less relevant, and will not be discussed here.

Franks (1995) argues that, while subjects are frequently omitted, Russian can not be considered a *pro*-drop language, as it is not morphologically uniform, and shows more restricted and context-dependent omission of referential subjects (but not expletive, which are uniformly omitted) - particularly main clause subjects - than other Slavic languages. Other Slavic languages, he argues, are both morphologically uniform (but see Huang, 2000) and *pro*-drop, providing somewhat weak support for the argument that their apparent morphological uniformity is the reason why they behave more like canonical null subject languages than does Russian (i.e. unlike Russian, they are *pro*-drop; Russian, in contrast, does not actually license null subjects, though subjects are frequently dropped). However, given that the principle of morphological uniformity appears to be more descriptive than independently motivated, and lacks strong explanatory or predictive power otherwise, the argument that Russian is not *pro*-drop because it

is not morphologically uniform seems like a somewhat circular determination. Nonetheless, Russian does demonstrate a variety of other features associated with subject omission in other languages - for example, elided (particularly embedded) subjects often preferentially, if not obligatorily, resolve towards subject-position antecedents (Kravtchenko, 2008), as seen in example (3). Similar trends are observed in e.g. Japanese (Ueno & Kehler, 2010, 2011) and Italian (Carminati, 2002).

- (8) Taro-wa/ga Jiro-ni hon-o watashita/watashi-te-iru tokoro-datta.
 Taro-TOP/NOM Jiro-to book-ACC handed/hand-INF-ASP scene-was
 shugo-shoryaku/ kare-wa/ jiyu -----
 subject-omission(*Null*)/ he-TOP(*Overt*)/ free(*Free*)
 “Taro handed/was handing a book to Jiro. [He] ...”

(Ueno & Kehler, 2010)

Ueno and Kehler (2010, 2011) find that in passage completion tasks using three different types of prompts (sentence-initial “null element”, overt pronoun, and free prompt, as above), preferred coreference resolution for overt pronouns, towards subjects, showed sensitivity to pragmatic manipulations – verb aspect (in transfer-of-possession contexts), and use of implicit causality verbs, which shift focus away from the subject referent. Null pronouns, however, were uniformly biased towards subjects, showing no sensitivity to aspect manipulation, and limited sensitivity to implicit causality manipulation – with a numerical trend in the case of ‘null element’ prompts, but a significant difference in the case of free prompts. Preferred reference resolution in this task was determined by disambiguating continuations. The potential relevance of this finding is discussed further in section 6.4.

Whether Russian is viewed as a *pro*-drop language or not, for the purposes of the work discussed here, is maybe important only to the extent where it makes accurate predictions, and accounts adequately, for both similarities and differences between Russian and other languages that allow subject (or other argument) omission. The precise theoretical status of subject omission in Russian is somewhat irrelevant, as long as it can be argued that the choice between an overt and a null subject in Russian constitutes a viable and genuine case of alternation between

(sufficiently) meaning-equivalent variants⁵.

2.3 Omission of Subjects in Russian

The frequency and optionality of subject omission in Russian (~20-35% in casual speech; Zdorenko, 2009) makes for a good test case of whether a drive towards communicative efficiency plays a role in production above the lexical level. The omission or reduction of functional elements has clearly been shown to be sensitive to these pressures, but to date there is only indirect evidence (Brown & Dell, 1987; Resnik, 1996) that contextual predictability affects the omission of non-functional elements, and few relevant investigations of languages other than English. The methodology used in Tily and Piantadosi (2009) was adapted, with minimal modification, to test whether subject omission in Russian is subject to these pressures, as would be predicted by UID. Further description of the task can be found in section 3.1.

The main question this work addresses is whether optional subjects in Russian are more likely to be omitted when they're redundant. The hypothesized reason for why omission may be favored in some contexts is an aim for efficient and low-error communication (Jaeger, 2010). The prediction is that more contextually predictable subjects are preferentially omitted. To test this, the contextual predictability of candidate subjects is measured experimentally, and a mixed logit model is used to test the effect of contextual predictability on subject omission.

The first experiment aims to answer the following questions:

1. Does the redundancy of subject referents correlate with their likelihood of being omitted?
I.e., if preceding context is more informative as to the identity of the subject referent, is it more likely to be omitted?
2. Is contextual predictability still a significant predictor of subject omission likelihood, after other factors known to influence referential expression choice are controlled for?

Arnold (2008) hypothesizes that other factors known to influence referential expression choice, or interpretation, do so because they allow the comprehender to estimate the probability

⁵As noted/discussed by Degen and Jaeger (2011), alternates may not need to be precisely meaning-equivalent.

that a certain referent will be mentioned next. More expected referents are then referred to with less informative expressions, possibly in the interest of efficiency. This assumes a role of audience design, although it is uncertain whether or not speakers explicitly make choices that facilitate comprehension in their listeners, or whether this occurs ‘incidentally,’ through listeners attending to cues in speech that are predictive of specific referent mention, regardless of the speaker’s reasons for producing them. Likewise, it is unclear if speakers attempt to construct a model of the discourse from the listener’s perspective, or simply consult their own internal model of the same discourse. This is discussed further in section 6.2.

It was hypothesized above that optional subjects are more likely to be omitted when their referents are more predictable given preceding context. Although it is difficult to make a clear prediction in the case of forward context, partially because there is no similar theoretical framework incorporating this information, it is hypothesized that:

1. If speakers take into account forward context in planning utterances, possibly in consideration of the likelihood of listeners recovering preceding information, then the predictability of subject referents based on *following* context (‘backward-predictability’) should be a significant predictor of subject omission, even after controlling for predictability based on preceding context (as well as other relevant factors).
2. If speakers take into account only preceding context in planning utterances, then predictability based on following context should not be a significant predictor of subject omission, after preceding context and other relevant factors are controlled for.

3 General Methods

Two experiments were designed to answer the questions outlined in the previous section. The design, with modifications detailed in sections 4.1.1 and 5.1.1, was based on that used in Tily and Piantadosi (2009). Shannon (1951) observed that readers’ accuracy in guessing upcoming concealed material can provide an estimate of the contextual predictability of this material (see

also: Manin, 2006). The information obtained using this measure can be viewed as similar to that obtained using a Cloze task (Taylor, 1953)⁶.

In order to obtain an estimate of referent predictability, participants were asked to read through a text progressively revealed to them on a computer screen, and guess the identity of concealed referents, based on their preceding and/or immediately following context (described in more detail below). The accuracy of readers' guesses, averaged across participants for each data point (clause subject), was therefore used as a proxy for predictability. In other words, referents that readers are more likely to guess accurately are more predictable in context, and those that are more difficult to guess accurately are less so. The first experiment was designed to determine if the predictability of a subject referent, conditioned on the text preceding the immediate clause, correlated with whether the writer/speaker elided the subject. To test this, participants were asked to guess the identity of successive concealed referents, while shown only preceding context. The experimental design is outlined in further detail in section 4.1.1.

The second experiment, using the same text materials, was designed to answer the following question: is the predictability of subject identity, conditioned on following context⁷, predictive of subject omission, after the effect of preceding context (as estimated by the first experiment) is taken into account? To this end, participants were asked to guess the identity of concealed referents, given both preceding and (limited) following context. This would make it possible to determine if referent predictability, given following context, makes an independent contribution to a speaker's choice to omit a clause subject. The design is described in more detail in section 5.1.1.

⁶An estimate of contextual predictability would otherwise be difficult to obtain, as the typical alternative method of estimating the contextual predictability of linguistic events is through use of large corpora (e.g. n-gram models or probabilities conditioned on specific elements in preceding discourse; see Jaeger (2011)), which were not publicly available at the time of experiment design.

⁷The following context was typically the verb, but occasionally adverb, or other material obligatorily adjacent to, and appearing to the left of the verb.

3.1 Materials

In order to adopt the method in Tily and Piantadosi (2009), it was first necessary to obtain a suitable corpus of Russian text. The use of informal writing or speech was considered crucial, given that subject omission in casual speech has been observed to occur $\sim 33\%$ of the time, but has an occurrence rate below 5% in more formal registers (Zdorenko, 2009). At the time of experiment design, no publicly available corpus of informal speech, or annotated for register, was available⁸. The use of a publicly available subtitle corpus (Tiedmann, 2009) was briefly considered, but it was found that the text was functionally uninterpretable outside of context. In addition, use of typical speech corpora would not be feasible, as transcribed speech disfluencies, errors, and false starts would be make the text excessively difficult for people to read. There are, similarly, no coreference-annotated corpora of informal Russian available⁹.

In the absence of a coreference-annotated corpus of colloquial Russian writing or speech, a set of annotated passages was created using publicly available texts. I selected and annotated passages across several genres that could be considered representative of casual Russian writing or speech, based on corpus data regarding subject omission in a variety of registers (Grenoble, 2001; Seo, 2001; Zdorenko, 2009). Passages from multiple texts, from three genres, were selected: 8 plays, for informal personal dialogue; 8 personal blogs, for informal written monologue; and 8 transcribed magazine interviews, for slightly more formal (largely one-sided) dialogue. This provided a total of 24 passages. Each passage was selected from the beginning of the text, in order to avoid discontinuity effects or additional difficulty in interpretation, and was truncated after approximately 30 sentences. Sources are described in more detail in Appendix D. A small corpus, annotated for coreference, was therefore created for the purposes of these experiments.

As shown in Table 1, although the source selection was non-randomized, and made with some consideration of the impressionistic rate of subject omission, the rate of subject omission

⁸A publicly available web corpus, RuWac (<http://corpus.leeds.ac.uk/mocky/>) was considered, but as the text segments were not annotated for register, and were no more than a few sentences long, this was ruled out.

⁹Coreference annotation, or at least a method of determining a referent's identity, is necessary, given that it would otherwise be impossible to determine whether participants had accurately guessed a referent's identity.

observed across passages is comparable to that observed in casual speech (~20-35%). All material selection and annotation tasks were completed by me¹⁰. A summary table of the passage data can be seen below:

	Plays (8)	Interviews (8)	Blogs (8)	Total (24)
Length (Words)	374	426	438	413
Length (Sentences)	55.4	40.6	37	44.3
# Judgements	44.9	42.4	48.5	45.3
Subject Omission	37%	19%	29%	29%

Table 1: Passage summary statistics by genre

3.2 Annotation

All animate and prominent inanimate referents in the 24 passages – i.e., those that would be plausible candidates for subsequent subjecthood or potential omission – were annotated for coreference. The data points of interest included every clause subject in the passage set that could plausibly be omitted (or, conversely, pronounced). Each passage contained an average of 45 candidate subjects or data points, ranging from 26 to 79, for a total of 1086 data points.

The passages were annotated for several suspected predictors of subject omission, as well as other variables of (potential) interest. Each clause subject was coded for whether it was “Overt” (with no distinction made between pronouns, names, or definite descriptions), or “Zero” (silent/null). Data points were additionally coded for passage number, genre, and other descriptive variables. A further list of predictors were considered on more theoretical grounds; among others: accessibility-based factors (factors known or expected to affect the salience of a referent), such as number of previous mentions (Ariel, 1990); factors such as form or function of last mention (Gordon, Grosz, & Gilliom, 1993); and other factors known or suspected to influence subject omission in Russian (Kravtchenko, 2008; Zdorenko, 2009). In most cases, it was expected that the same contexts that facilitate use of pronouns, over other types of referential

¹⁰In several cases, participants in the experiment indicated during its course that they disagreed with some coreference annotation decisions (coreference was indicated by color highlighting of coreferring referential arguments, therefore readily ascertainable to readers). These cases were corrected if noted prior to the beginning of the experiment, and indeed reflecting a mistake or notable ambiguity, or excluded from analysis if they were noted after.

expressions, also facilitate their omission. For example, use of referring expression types higher on the following hierarchy is expected to be more preferred, or acceptable, as the salience of the referent increases: {zero ('null pronoun') ?(>) pronoun} > {proper name, definite description} (Ariel, 1990; Arnold, 2008). A full list of predictors that the texts were annotated for, including their relation to the variables of interest, can be found in Appendix C. An abbreviated list of those of most interest can be found below, including many of those found to be significant predictors of referential expression choice in Tily and Piantadosi (2009)

Form: [Overt/Zero/Uncertain]; Whether the form of the subject was overt, zero, or uncertain.

Only cases which could be annotated with reasonable confidence were included in the model. A separate predictor, marking all cases of coordinated or possibly coordinated subjects (which were optionally excluded when modeling the data), was also coded for.

Number: [Singular/Plural/Collective/Uncertain]; Whether the number of the subject was singular, plural, collective, or uncertain. Uncertain and collective cases (2.3% of the data) were immediately marked for exclusion.

Case: [Nominative/Dative/Accusative/Genitive/Uncertain]; The case of the target subject. Nominative and Dative subjects are the canonical variants. Accusative, genitive, or uncertain 'subjects' (0.8% of the data) were immediately marked for exclusion.

Form of Last Mention (*Last Mention/LM Form*): [Overt/Zero]; Whether the form of the most recent coreferring instance of the referent was overt or zero.

Coreference with the Preceding Subject (*Coreferent w/ Preceding S*): [Yes/No]; Whether the target referent was coreferent with the subject in the preceding clause.

Last Mention Function (*LM Function*): [Subject/Other]; Whether the grammatical function of the most recent NP coreferent with the target referent appeared in subject position, or not.

Distance in Words to Last Mention (*Distance in Words/Distance in Words to LM*): A continuous measure of the number of intervening words between the subject and the most recent

coreferring instance of the target referent.

Number of Previous Mentions (*Previous Mentions*): A continuous measure of the number of times the referent was previously mentioned in discourse (not counting the last subject, if coreferent).

Number of Referents Introduced (*Number of Referents*): A continuous measure of the number of referents introduced so far in the discourse. Roughly, the number of competing referents.

4 Experiment 1

The purpose of the first experiment was to determine whether the predictability of a subject referent, conditioned on preceding context, makes an independent contribution to a speaker/writer's choice to omit a clause subject, once other likely predictors of subject omission are controlled for. The experimental design is directly modeled after Tily and Piantadosi (2009). The code used for running the original experiment in English was obtained from Harry Tily upon request, and used for the current study with minimal modification. The task requires participants to guess the identity of upcoming referents, while reading progressively through a passage of Russian text. Participants' average accuracy in guessing the identity of the upcoming referent, for each individual subject/data point, is thus used as an estimate of the referents' predictability, or probability in context. The prediction is that subject omission is more likely when participants are, on average, more accurate in guessing the identity of the concealed subject referent. An additional, more exploratory measure of Entropy was also calculated, and is described further in section 4.2.

As mentioned in the section 2.1.1, Uniform Information Density predicts that optional elements in language are more likely to be reduced or omitted when their mention is more contextually predictable, given preceding context. The specific hypothesis tested is that referents that are more predictable (redundant) are more likely to be omitted in the selected texts. Given free choice in pronouncing or eliding a subject, UID would make the prediction that the choice to

produce or omit it, provided doing so does not¹¹ alter meaning, should reflect these pressures.

The purpose of the experiment is therefore the following:

1. Directly test the predictions of UID, as well as other theories of efficient communication, with respect to omission of a non-functional linguistic element, which to date has not been done.
2. Determine whether omission of subjects in Russian, which has likewise not been studied extensively to date, may in part be accounted for by a pressure for more efficient communication.

4.1 Methods

4.1.1 Design

Each passage was designed to be progressively revealed to participants, as they read it. Participants would initially see the text preceding the immediate clause, containing the first candidate subject¹² (which was concealed), make a guess as to its identity, and would then see the candidate subject itself and all following text up to the *next* candidate subject, and so forth. Participants were explicitly asked to answer the question: “Who or what is the agent of the upcoming action?”, rather than guessing the specific upcoming word. This was framed as a ‘game,’ with participants receiving both immediate feedback on their accuracy, and points for accurate guesses. As the ‘Someone or something new’ button shifted with the revealed text (always immediately following the end of the revealed passage), participants would presumably have less incentive to simply click on it repeatedly. All potential referents preceding the candidate subject were highlighted, with all coreferring arguments sharing a distinct color. An example trial can be seen below:

¹¹Substantially; see Degen and Jaeger (2011), for a case where speakers appear to structure their message with less regard for the precise semantic meaning of a variant, when the alternative would get the message across more efficiently.

¹²The clause was made the cutoff point, given word order flexibility in Russian, and the impossibility of determining the ‘position’ of an omitted subject (cf., MacWhinney, Bates, & Kliegl, 1984).

Maša dumala pro Petiu, kogda uslyšela, čto ... Kto-to ili čto-to novoe

(Masha was-thinking about Petia), when [she] heard that ... Someone or something new)

Participants were only asked to guess animate, and salient inanimate subjects – those that would plausibly be candidates for omission – although all (in principle) plausible referents in the preceding discourse were highlighted as potential antecedents. This is in contrast to Tily and Piantadosi (2009), where participants were asked to guess the identity of each referential NP (regardless of grammatical function). Each referent in the discourse was highlighted in a different color – so that, e.g., “Masha,” “she,” and “the girl” would all be colored identically if they referred to the same individual. For elided subjects, the main verb was highlighted instead (i.e. “Petia slept,” vs. “slept”), as shown above.

Only subsequent-mention trials, where the referent had been previously introduced, were considered, as it isn’t possible to claim that a person making a correct ‘Someone or something new’ guess has accurately guessed the upcoming referent’s identity. Of an initial 1085 utterances, 168 first-mention utterances (15.5%) were excluded. Cases involving annotation error, non-singular subjects, and non-canonical (with respect to case marking) subjects were excluded¹³. This left 745 cases total.

4.1.2 Participants

70 native Russian speakers were recruited online, via email. Speakers were asked to participate only if they were native speakers of Russian, able read Russian freely, and considered Russian their dominant language (i.e. the language they spoke most often, and felt most comfortable speaking). No other personal information was collected, although some participants chose to use their email addresses as identification. Participants were recruited by word-of-mouth. They were not told the purpose of the experiment prior to completing it, nor the phenomenon of interest. Data from one participant who indicated that they knew most of the answers in the text, due to close familiarity, was excluded. A total of 5933 individual trials were left.

¹³See the list of predictors in section 3.2 for more details.

4.1.3 Procedure

Each participant was presented with two passages from different genres. The combination and order of passage presentation was counterbalanced across subjects; 5-7 people read each text. The participants were instructed to read through the passages, guessing the identity of the “agent of the upcoming action” (either any one of referents in the preceding discourse, or “Someone or something new”), by clicking on it with a mouse. As previously mentioned, the texts were presented to participants progressively, with each participant first seeing all text preceding the immediate clause in which the first subject was located. After each guess, the participant would immediately get feedback on whether their guess was accurate, as well as 1 point for each correct ‘Something new’ guess, and 2 points for each correct guess involving a referent in the preceding discourse (to discourage repeatedly guessing ‘Something new’). Both the overall number of points, and the percent earned out of the total number of points that could have been collected, were displayed at the bottom of the screen.

After participants were finished reading the text, they were asked whether they were previously familiar with the text, and how familiar they were with it; i.e., whether they knew some, most, or all of the answers (or otherwise if they had no, or only passing familiarity). They were then able to submit the results. As a consequence of the design, there were no partial submissions, although a fairly high non-completion rate (39%).

4.2 Results

To determine whether predictability affects the likelihood of subject omission, beyond the effect of the primary ‘control’ predictors described in section 3.2, the choice between an overt and null subject referent was modeled as an outcome of both the control predictors in question, and the likelihood of readers guessing the identity of the concealed (upcoming) referent. Readers’ average (negative log-transformed) accuracy in guessing referents was thus used as a measure of contextual referent predictability. The issue of whether some of the control predictors may matter precisely (or partially) because speakers/comprehenders use this information in computing subjective probabilities (in this case, likelihood of referent mention), and therefore to make

the decision of whether to omit or pronounce an optional subject, will be discussed in section 6.2.1.

Continuous control predictors were log-transformed, centered and standardized. Each data point in the model represented a single clause subject from the 24-passage data set. Two measures of predictability were computed for each data point, based on guessing accuracy data. Surprisal, or the negative \log_2 probability of a correct guess (i.e. average accuracy) was included as the primary measure of predictability (Manin, 2006; Shannon, 1951):

$$-\log_2 P(\text{Response} = \text{correct})$$

This measure reflects a word’s information content - here, how predictable a word or phrase is, given preceding context. If a word or other element is perfectly predictable in context (i.e. given preceding context, comprehenders are always able to guess with perfect accuracy that it is coming up next), it carries zero information, or Surprisal. A word or other element maximally unpredictable in context (i.e. comprehenders never accurately guess that it is coming up next) carries ‘infinite’ information or Surprisal. Here, it is important to note that participants guess referent identity, instead of specific words, forms, or phrases (the same predictions, however, otherwise apply). Given the limited number of readers (trials) for each data point, clause subjects with ‘infinite’ Surprisal were assigned an arbitrary cutoff value of 4^{14} , under the assumption that most ‘unpredictable’ subjects would fall into the range of $\sim 3.25 \cdot 5^{15}$ (the rough range seen in Tily and Piantadosi (2009)).

A related measure computed was Entropy, or uncertainty – whether guesses are concentrated on a few highly predictable referents, or ‘spread out’ over multiple candidate referents:

$$H(\text{Response}) = - \sum_{r \in \text{Response}} P(r) \log P(r)$$

This reflects whether readers’ guesses were primarily concentrated on few (correct or incorrect) referents (low Entropy), or spread out over multiple candidate referents (high Entropy). It’s plausible, for example, that writers may be more likely to omit subjects when readers do not perceive the intended referent as having other likely competitors. Each ‘Something new’

¹⁴A maximum value of 3.5, which otherwise ranged from 0 to 2.81 in exp. 1, and 0 to 3.17 in exp. 2, is used in the density plots included in this paper, for better visual representation of the data.

¹⁵A subject with a Surprisal value of 4 has a 1/16 likelihood of being guessed accurately. One with a Surprisal value of 5 has a 1/32 likelihood.

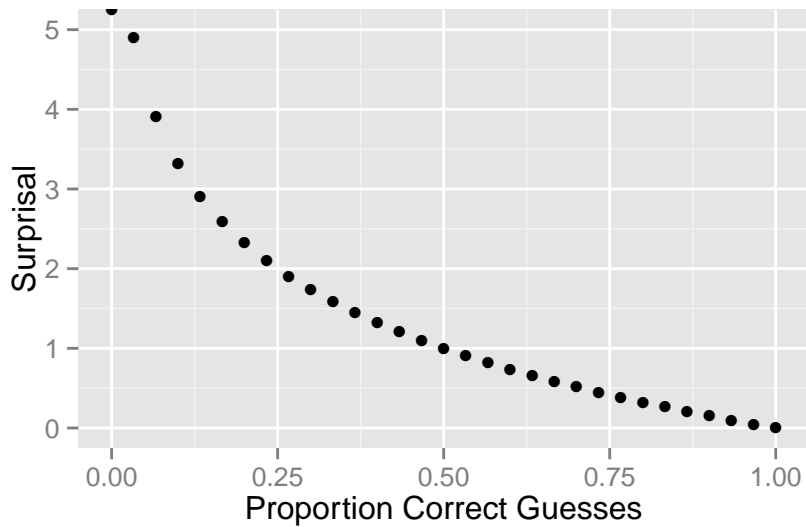


Figure 1: Surprisal vs. average accuracy (contextual probability)

guess was treated as a different referent, as it would be problematic to assume that ‘Something new’ guesses, by different readers, referred to the same referent, or reflected lower uncertainty/Surprisal. Plural referents were excluded, due in part to the difficulty of determining (for example) whether or not a plural subject referring to both of two previously mentioned referents was first- or subsequent-mention, as well as the question of whether plural subject referents would be subject to the same pressures as singular referents. Apparent non-nominative or dative subjects were likewise excluded, as mentioned earlier. Since Entropy and Surprisal are closely related, in order to avoid collinearity, Entropy was residualized against Surprisal¹⁶ before being entered into the model as a predictor.

It’s important to note that readers guessed the identity of an upcoming referent without seeing it. Therefore, any correlation seen between referent predictability and subject omission is not driven by the the readers’ inference of the subject’s identity based on its overt form, but rather stems from perceived probability of referent mention. A density plot looking at the distribution of overt vs. null subjects, by Surprisal value, can be seen in Figure 2.

As can be observed, while both overt and zero subjects cluster towards the low-intermediate Surprisal range, zero subjects occur more often in environments where the Surprisal is low (i.e.

¹⁶An inverse transformation was applied to the Surprisal variable first, given that a linear relationship between these two variables was then more apparent.

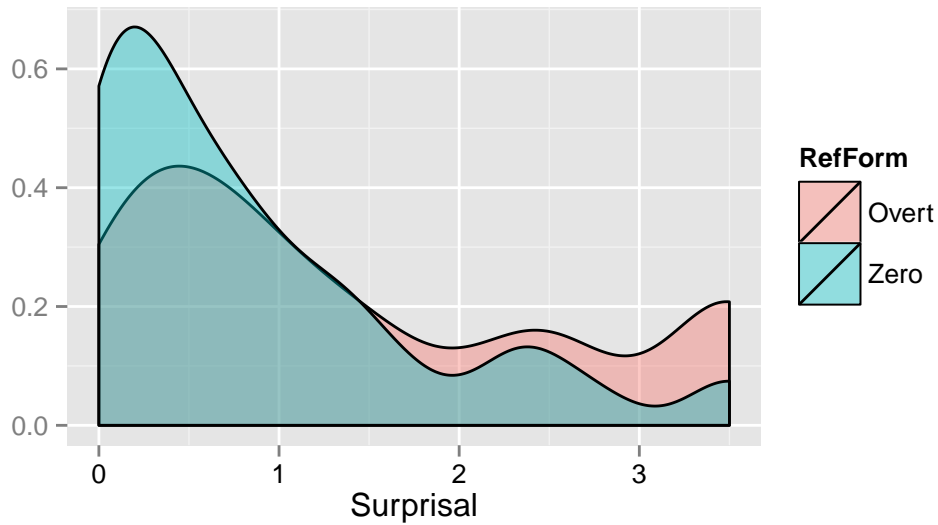


Figure 2: Overt/null subject density by Surprisal (exp. 1)

the referent is more predictable), and less often in environments where Surprisal is high (i.e. the referent is less predictable), as expected. Overt subjects, in contrast, appear relatively more evenly distributed across the range of Surprisal values, and have higher density at high Surprisal values than zero subjects.

A mixed effects (logit) model analysis (a subtype of generalized linear mixed model) was used to test the hypothesis that contextual predictability, or Surprisal, exerts an independent influence on likelihood of subject omission, after controlling for other predictors. The binary outcome variable was the realization – overt or zero (null/silent) – of the clause subject. The model was fitted using χ^2 -test model comparison. Predictors showing high collinearity with other predictors, such as Genre, were not included in the final model. Those that were determined to be a measure of the same factor as other predictors were also removed, leaving distance in words from last mention (Distance in Words), number of previous mentions (Previous Mentions), and number of referents introduced (Number of Referents) as Saliency measures¹⁷. Entropy, number of Previous Mentions, and Number of Referents introduced were not significant predictors, and did not improve model fit. They were therefore dropped. There were, notably, no higher-order interactions between Surprisal and any control predictors. Passage number was included

¹⁷Although other predictors, such as Form of Last Mention, also arguably measure Saliency, these may be an influence on referring expression choice or omission for independent reasons.

as a random effect.

	Coef β	SE(β)	z	p
Intercept	-1.34	0.27	-5.0	<.0001
Surprisal	-0.17	0.08	-2.1	<.05
Last Mention: Zero	0.42	0.19	2.2	<.05
Coreferent w/ Preceding S	1.33	0.22	6.1	<.0001
Last Mention Function: S	0.62	0.28	2.2	<.05
Distance in Words	-0.36	0.11	-3.3	<.001

Table 2: Mixed logit model - experiment 1

A bar plot, more clearly showing the direction of influence and effect sizes, can be seen in Figure 3. As can be observed, Surprisal exerts a relatively weak, but significant effect on subject omission, once other predictors are taken into account ($\beta = -0.17, z = -2.1, p < .05$), with speakers preferentially omitting more predictable (less surprising) subjects. All other influences are in the expected direction - i.e., speakers are more likely to omit subjects when the form of last mention is zero, when the target subject is coreferent with the preceding subject, and when the grammatical function of the last mention is subject. Speakers are also less likely to omit subjects, the farther the last mention is from the referent, in words.

Plotting mean observed proportions vs. predicted probabilities indicates a reasonably good model fit ($R^2 = .98$), seen in Figure 4 (this should not be considered a measure of model quality). Nagelkerke $R^2 = 0.33$, when compared to an ordinary logit model with the intercept as baseline, although this measure of model quality should be viewed with caution (*FAQ: What are pseudo R-squareds?*, 2011). Roughly, however, it can be viewed as analogous to the R^2 for linear models, or a measure of how much variation in the dependent variable is accounted for by the model, compared to baseline.

It's important to note here that coordinated, or possibly coordinated subjects were included in this data set, although they are likely different in status from other subjects. This was done partially due to the sparsity of data, and the difficulty of determining, in many instances, whether a subject was in fact coordinated – given the somewhat ambiguous structuring of sentences in

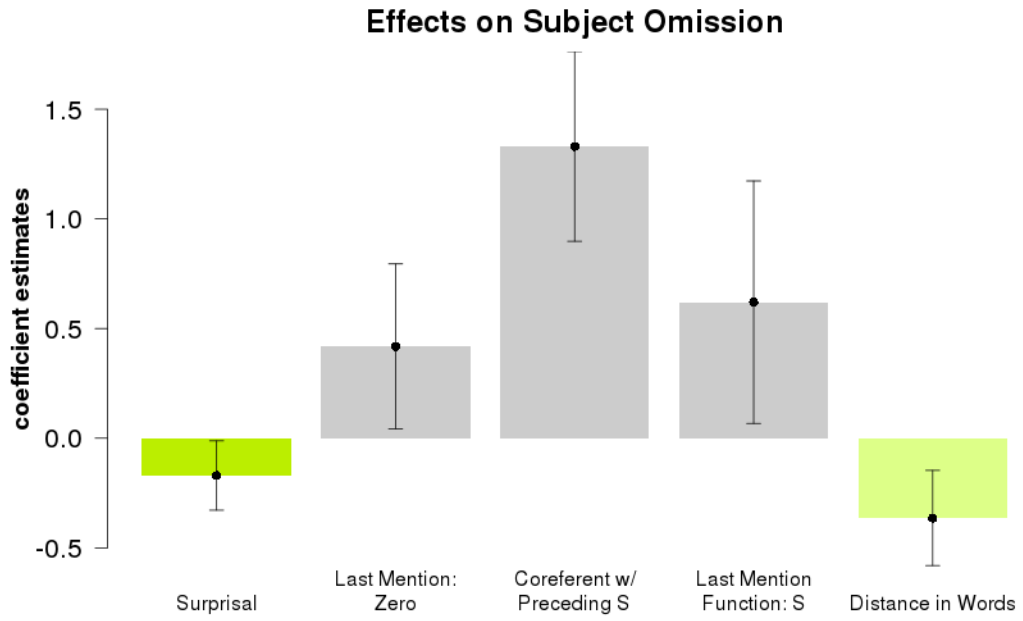


Figure 3: Coefficient bar plot - mixed logit model (exp. 1)

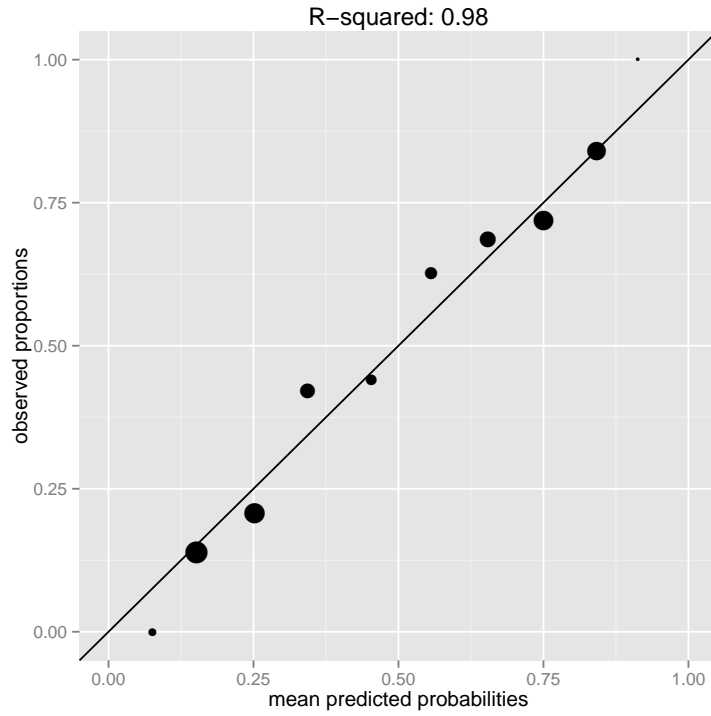


Figure 4: Experiment 1 model fit

informal writing/speech, and clausal ‘chains’ where it was not clearly apparent whether subjects were conjoined or not. Furthermore, in most instances, coordinated subjects can optionally be pronounced, though perhaps producing an awkward utterance. This may arguably be no more awkward, in some cases, than other instances of omission where a subject can optionally be pronounced, and it’s plausible that coordinated (coreferent) subject elision may be subject to the same pressures as elision of other subjects (e.g. they are more likely to be omitted because they are redundant in context, parallel and coreferent with the preceding subject, etc.). This affects 196 cases in the first experiment.

It’s important to note that in either case, the main effect of Surprisal holds, with the same set of control predictors, when the same statistical analysis is performed on a more conservative subset of the data, in which there is no question of whether the subject might be coordinated with the preceding subject - i.e. only main (sentence-initial) and embedded clause subjects are included. Form of Last Mention in this more conservative model does, however, lose significance ($\beta = 0.21$, $z = 0.99$, $p = .32$), and model quality, as measured by Nagelkerke R^2 , is decreased: 0.26 (vs. 0.33; Nagelkerke $R^2 = 0.25$ if Form of Last Mention is also dropped as a predictor). Likewise, model fit is decreased ($R^2 = 0.95$; see Appendix B.1 for plot). Crucially, Surprisal exerts the same effect ($\beta = -0.18$, $z = -2.12$, $p < .05$). Further information on the resulting model can be found in Appendix A.1.

In summary, the expected influence of Surprisal on likelihood of subject omission is observed, with speakers being more likely to omit more predictable subjects. The effect is statistically significant ($p < .05$) after controlling for other likely determinants of subject omission choice, but it is somewhat weak, with partial Nagelkerke $R^2 = .01$ (in both models). This suggests that only about 2.8% of the model quality (4.5% in the more conservative model), at least with respect to this specific measure, can be attributed to the effect of Surprisal, once other predictors are taken into account.

5 Experiment 2

Experiment 2 was designed to determine whether following context contributes separately to choice of referential expression, beyond the contribution of preceding context (which was tested in Experiment 1). The hypothesis this experiment tests is the following. If following, and not only preceding, context plays a role in the choice to omit redundant speech elements, then (all other things being equal) those elements – in this case, clause subjects – that are more predictable, given following context, should be more frequently omitted. This possibility is not directly addressed by UID, or other dominant theories of efficient production. Although UID (for example) does not rule out the possibility that following context also plays a role, the most straightforward prediction of the theory, as previously stated, would be that only preceding context plays a role in the choice to omit an optional element, and that following context should make no separate contribution.

As in the previous experiment, participants' average accuracy in guessing the identity of each concealed subject was used as a measure of its predictability. The prediction is therefore that if participants are more accurate in guessing the identity of concealed subjects, given following material, there should be an increased likelihood of those subjects being omitted by the writer/speaker. Crucially, this effect should not be subsumed by that of the preceding context (which, of course, participants also see in this experiment).

5.1 Methods

5.1.1 Design

The methodology used for the second experiment was identical to the first, except that participants saw the subjects of interest replaced with visible black 'boxes' of equal length, and varying amounts of following context (shown below). Participants were instructed to guess the identity of the agent in the last clause of the text presented, concealed by the black box. Following this, the box was replaced by either the clause subject, or main verb (in case of subject omission), and the text up to the next guessing point was revealed. As in the previous experi-

ment, all coreferring subjects were highlighted in the same color. Given that it's impossible, as previously mentioned, to determine the 'position' of an omitted subject in sentences with free word order, two different trial designs were used in this experiment:

(1) For clauses with typical (subject-initial) word order¹⁸, the black box was displayed in place of the overt or omitted subject, and either the verb or material immediately left-adjacent to the verb were displayed following it. If the following word was a relatively semantically content-less word such as a preposition, or a discourse/negation particle, the word after it was also displayed. This is intended to determine whether the informativity of immediately following context, with respect to the subject's identity, contributes to subject omission. Further following context was not provided, given that it made the subject so easily recoverable that it was judged likely that near-perfect accuracy would be obtained, rendering the results uninformative¹⁹.

Maša dumala pro Petiu , kogda uslyšela , čto [?] prišel ... Kto-to ili čto-to novoe
Masha was-thinking about Petia , when [she] heard that [?] came ... Someone or something new)

(2) For clauses with non-canonical (non-subject-initial) word order – i.e., those not meeting the criteria above²⁰ – all material between the clause onset and the verb was revealed, with a black box displayed, as in the previous experiment, in place of the overt subject, or left-adjacent to the verb if the subject was elided. It is unclear what can be made of this data, although it

¹⁸For clauses with omitted subjects, this was assumed if (a) the clause began with either a verb, or an obligatorily verb-adjacent modifier such as an adverb or a prepositional phrase (e.g. location); and (b) an overt subject was judged to be more likely pronounced preceding, rather than following, the verb (or other verb-adjacent material).

¹⁹As will be seen in the Results section, this concern is arguably justified, as the Surprisal of concealed subjects in this experiment was overall sufficiently low, compared to that of the first experiment, to make the results difficult to interpret.

²⁰Here, it was assumed that the subject was left-adjacent to the verb (when the subject was omitted). This is a somewhat problematic assumption, as overt subjects can, or sometimes must, occur in a different position when pronounced. This problem is however not addressed further, as these cases were not included in any analyses. The degree to which this may have affected the material in trial design (1) is uncertain, although it is impressionistically lesser. This may need to be explored further in the future.

may be informative as to how much topicalized or fronted material plays a role in the choice to pronounce or omit the subject. It may be difficult, however, to disentangle this from the more general role of preceding context.

Maša dumala pro Petiu, kogda uslyšela, što gde-to ... Kto-to ili što-to novoe

(Masha was-thinking about Petia, when [she] heard that somewhere [?] ...
Someone or something new)

Subjects in clauses with non-canonical word order, as described in the second scenario, were omitted from the analysis, as the word order/information structure alternation were judged to potentially introduce additional confounds, and make the results more difficult to interpret. Further, no following context was shown in these cases - as there was additional context preceding the subject, not shown in the first experiment, this would make a comparison with the preceding experiment difficult.

Of an initial 1079 utterances, 168 first-mention utterances (15.6%) were excluded. 320 cases (29.7%) were excluded because no following context was displayed, and cases involving annotation error, non-canonical subject case, as well as non-singular subjects, were again excluded (see the list of predictors in section 3.2 for more details). This left 532 cases total to be analyzed. The same materials and overall design as in the previous experiment were otherwise used.

5.1.2 Participants

102 native Russian speakers were recruited online. Participants were given the same information, regarding requirements for being able to participate in the experiment, as in the first experiment. Most participants were recruited through public postings in LiveJournal or odnoklassniki.ru communities dedicated to language, linguistics, or psychology experiments. They were not told the purpose of the experiment prior to completing it, or the phenomenon of interest. 6-11 participants saw each passage, with an average of 8/passage. None of the participants'

data was excluded. This left a total of 8490 trials.

5.2 Results

To recap, this experiment was run to determine whether ‘recoverability²¹,’ given following context, contributes to the likelihood of subject omission, independently of the effect of preceding context. As in the previous experiment, the choice between an overt and null subject was modeled as an outcome of both the primary control predictors, and the likelihood of readers accurately guessing the concealed referent’s identity (Surprisal_{exp2}). Surprisal from the previous experiment was also included in a final comparison model (Table 4), in order to determine whether any effect of Surprisal in the second experiment is subsumed by Surprisal in the first experiment (i.e. predictability given only preceding context). The implicit assumption made is that speakers ‘take into account’ listeners’ ability to infer preceding speech elements (but see section 6.2), given the material that follows them, and base their decision to reduce or omit optional elements in part on this information.

Continuous control predictors were log-transformed, centered and standardized. As in the previous experiment, two (new) measures of predictability were computed for each data point in the 24 passages. As discussed in section 5.1.1, only those trials where forward context was shown were included (those with apparent canonical subject-initial word order), leaving 532 data points, out of the 745 (70.2%) included in the analysis for the first experiment. Surprisal_{exp2} , the negative \log_2 -transformed probability of a correct guess given both preceding and following context, was again included as a primary measure of predictability. As in the previous experiment, target subjects with ‘infinite’ Surprisal were assigned a cutoff value of 4. All other procedures for computing the value were identical. Entropy_{exp2} was likewise computed, again reflecting how spread out readers’ guesses were over candidate referents (see 4.2 for further discussion of these measures).

Since there was no linear relationship between Entropy_{exp2} and Surprisal_{exp2} (even if stan-

²¹This term is used somewhat loosely, and should not be considered synonymous with morphological recoverability, which in any case is not always possible in Russian (Franks, 1995).

standard transformations were applied to the latter), $Entropy_{exp2}$ was entered into the model without being residualized against Surprisal. A density plot of overt vs. zero subject occurrence, by $Surprisal_{exp2}$ value, can be seen in Figure 5. It can be observed here that the frequency of both zero and overt subjects is pushed heavily towards the low- $Surprisal_{exp2}$ range. Although zero subjects clearly occur most frequently at the very low end of the Surprisal distribution, a similar trend is observed for overt subjects, and overt subjects are only slightly more likely to occur than zero subjects at other $Surprisal_{exp2}$ ranges.

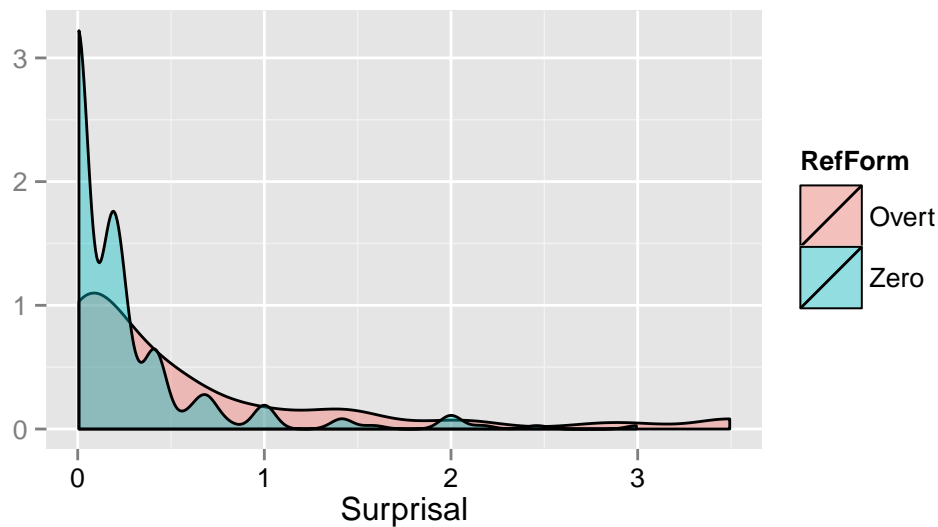


Figure 5: Overt/null subject density by Surprisal (exp. 2)

A mixed effects logit model analysis was again used to test the hypothesis that ‘backwards’ contextual predictability, or $Surprisal_{exp2}$, exerts an influence on subject omission, beyond the effect of Surprisal conditioned only on preceding context. The binary outcome variable was the realization (overt, or null/silent) of the clause subject. The model was fitted using χ^2 -test model comparison. The same predictors as in the analysis for Experiment 1 were entered into the model, as well as any significant higher-order interactions between $Surprisal_{exp2}$ and the control predictors²². Passage number was included as a random effect.

²²Number of Previous Mentions and Number of Referents introduced were again dropped as predictors, as they exerted no significant effect.

	Coef β	SE(β)	z	p
Intercept	-1.40	0.30	-4.6	<.0001
Surprisal	-0.38	0.22	-1.8	0.076
Last Mention - Zero	0.88	0.28	3.1	<.01
Coreferent w/ Preceding S	1.40	0.26	5.3	<.0001
Last Mention Function - S	0.58	0.34	1.7	0.087
Distance in Words	-0.42	0.13	-3.2	<.01
Surprisal : LM - Zero	-1.11	0.54	-2.0	<.05

Table 3: Mixed logit model - experiment 2

A bar plot, more clearly showing the direction of influence and effect sizes, can be seen in Figure 3. Entropy_{exp2} does not exert a main effect if it is added into the model on top of Surprisal_{exp2} and the Surprisal_{exp2} : LM - Zero interaction. Further, as the primary purpose of this experiment is to assess whether predictability conditioned on following context contributes anything to subject omission likelihood, and Entropy in the previous experiment was not a significant predictor of subject omission choice, Surprisal_{exp2} is retained as the only measure of predictability.

Plotting mean observed proportions vs. predicted probabilities indicates a somewhat decreased model fit ($R^2 = .97$), shown in Figure 7. This again should not be considered a measure of model quality. Nagelkerke $R^2 = 0.38$, when compared to an ordinary logit model with the intercept as baseline.

As can be observed in Figure 6, Surprisal_{exp2} exerts no main effect on subject omission, once other predictors are controlled for ($\beta = -0.38$, $z = -1.8$, $p = 0.08$), although there is a trend in the expected direction. All other main effects, as in the previous experiment, are in the expected direction. Speakers are more likely to omit subjects when the form of last mention is zero; when the target subject is coreferent with the preceding subject; and, marginally, when the grammatical function of the last mention is subject. Speakers are likewise less likely to omit subjects, the farther the last mention is from the target subject in words. Speakers crucially preferentially omit more predictable subjects only when the Form of Last Mention is zero, possibly indicating that Surprisal only matters in contexts where subject omission is particularly

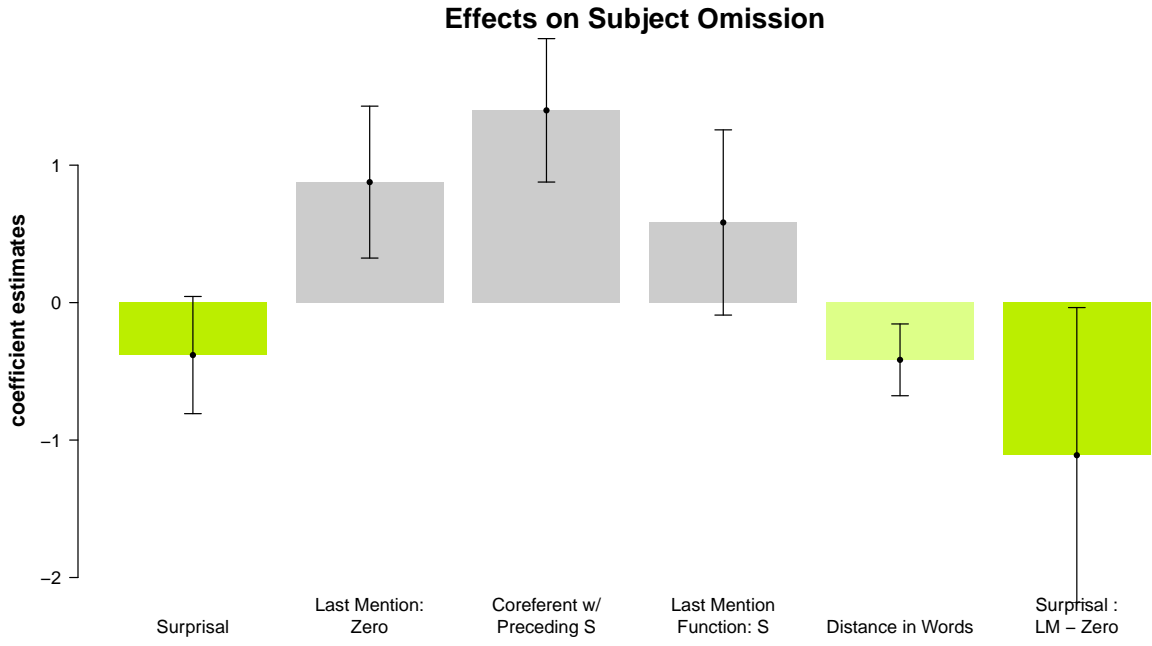


Figure 6: Coefficient bar plot - mixed logit model (exp. 2)

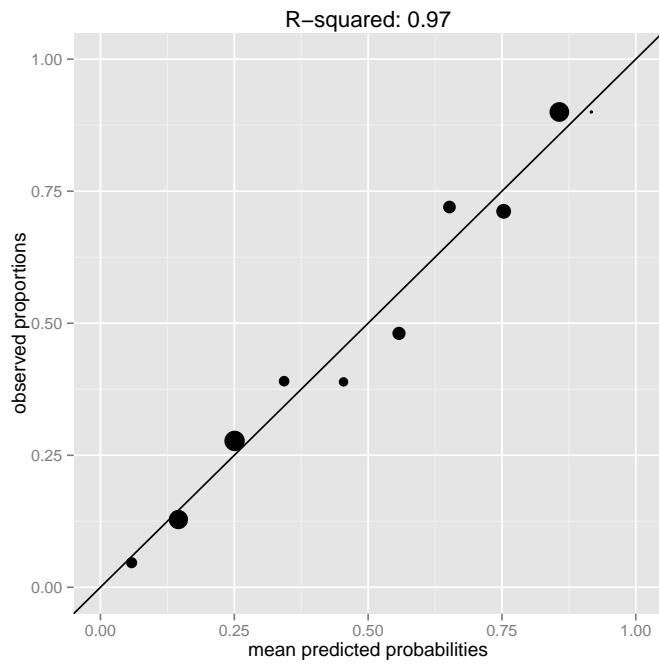


Figure 7: Experiment 2 model fit

likely for independent reasons ($\beta = -1.11$, $z = -2.0$, $p < .05$). This interaction, however, is somewhat difficult to interpret, and is discussed further below.

As in the previous experiment, a separate analysis, including only subjects that were not coordinated (or possibly coordinated) with preceding subjects, was carried out (see Appendix A.2 for summary tables, and Appendix B.2 for plots). Notably, the Surprisal : LM - Zero interaction is no longer significant, if the analysis is run on only the subset of the data where conjoined or possibly conjoined subjects are excluded. Loss of significance may however be due to loss of statistical power, given the relatively large number of predictors and small number of data points (415/532, or 78% of those included in the less conservative model).

In the less conservative model (coordinated or possibly coordinated subjects included), it's unclear why the main effect of Surprisal disappears after the interaction with Form of Last Mention is added in, with Surprisal decreasing likelihood of omission only when the Form of Last Mention is zero. It's possible, given that participants on average found upcoming referents very predictable in this experiment, that this is simply an artifact of a ceiling effect, with respect to participants' ability to accurately guess concealed referents. In general, Surprisal decreases significantly once readers have some forward context, which may account for the lack of a main effect (there is less variation in referring form, at most Surprisal ranges, overall). However, it appears that in contexts where subject omission is more likely to happen overall (where the Form of Last Mention is zero), Surprisal differentiates, to an extent, between contexts where subject omission is more or less likely to occur. A boxplot of Surprisal values, by subject form and Form of Last Mention, is shown in Figure 8.

A more straightforward explanation of this effect, given the absence of a significant interaction in the more conservative model, may be that the interaction is driven by subjects of clausal 'chains' – where the form of previous mention is zero, and the subject is possibly, though not necessarily, coordinated with the subject of the last clause.

It should be noted that, in the more conservative model, the Form of Last Mention effect disappears if the marginal interaction is dropped, and a main effect of Surprisal_{exp2} is observed ($\beta = -0.58$, $z = -2.70$, $p < .01$); see Appendix A.2. However, as this is somewhat difficult to

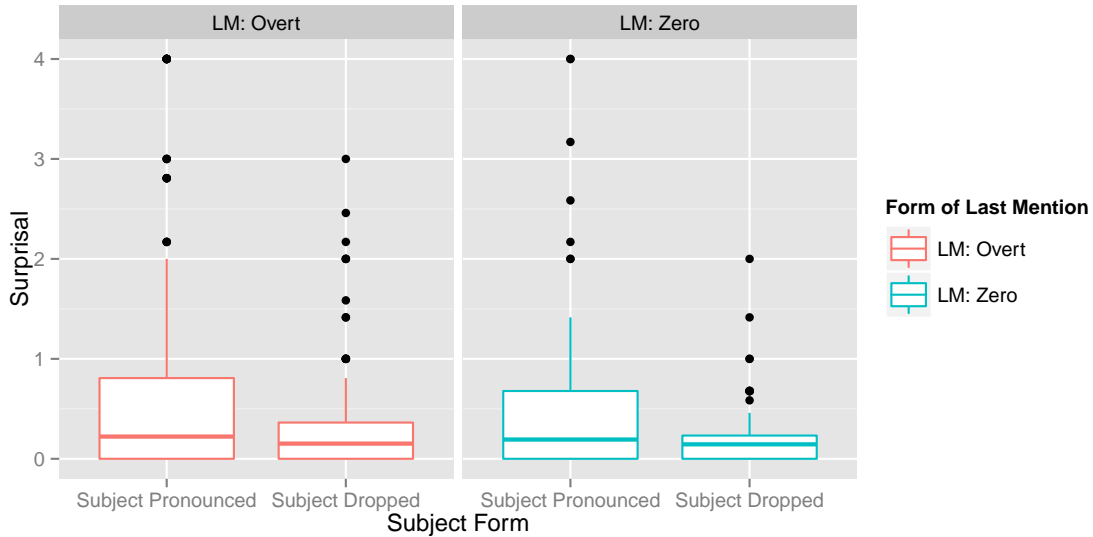


Figure 8: Surprisal by subject form and Form of Last Mention

interpret, given the substantially reduced size of the data set, the marginally significant (in the less conservative model) interaction is tentatively retained (further description of the models can be found in Appendix A.2).

In summary, in contrast to Surprisal conditioned on preceding context, Surprisal_{exp2} , or predictability conditioned on both preceding and following context, exerts no main effect on likelihood of subject realization once other predictors, and higher-order interactions that Surprisal_{exp2} is involved in, are controlled for. Since only Surprisal was a significant predictor in the first experiment, with respect to measures of predictability, all following discussion is restricted to the two ‘Surprisals.’ The distribution of Entropy_{exp2} , further, is distinctly abnormal (see Appendix B.2), likely due to constraints on possible ranges of its values, given the limited number of observations per data point – making any effects of this variable more difficult to interpret.

Since Surprisal in the second experiment does not appear to have a main effect on likelihood of subject omission (as in the first), and there appears to be a floor effect for Surprisal values, compared to the first experiment, further attempt at determining whether following context has an independent effect appears difficult. A model including Surprisal measures from both experiments was therefore run on the larger, and less conservative subset of the data. Since the two

Surprisal measures are extremely closely related, with one effectively subsuming the other (glm: $\beta = 0.25$, $z = 8.63$, $p < .0001$), Surprisal_{exp2} was first residualized against Surprisal_{exp1} . A summary of the model can be seen in table 4. The following model suggests that Surprisal_{exp2} does not significantly improve on a Surprisal_{exp1} -only model, except when Form of Last Mention is Zero - this marginally significant interaction is again difficult to interpret for same the reasons mentioned above. The same measures applied previously indicate a fair model fit ($R^2 = 0.95$ when mean predicted probabilities are plotted against observed proportions), and reasonable model quality (Nagelkerke $R^2 = 0.37$).

	Coef β	SE(β)	z	p
Intercept	-1.28	0.31	-4.1	<.0001
Exp. 1 Surprisal	-0.24	0.10	-2.3	<.05
Exp. 2 Surprisal (Resid.)	-0.36	0.23	-1.6	0.1
Last Mention - Zero	0.46	0.24	1.9	0.06
Coreferent w/ Preceding S	1.37	0.27	5.1	<.0001
Last Mention Function - S	0.58	0.34	1.7	0.089
Distance in Words	-0.40	0.13	-3.0	<.01
Exp. 2 Surprisal (R.) : LM - Zero	-0.82	0.48	-1.7	0.086

Table 4: Mixed logit model - experiments 1 and 2

This does not rule out the possibility that Surprisal_{exp2} exerts a general effect on subject omission, beyond that of Surprisal in the first experiment. This appears difficult, however, to determine from the data at hand, given that the ‘floor’ effect on Surprisal_{exp2} values makes more direct comparison with the data from experiment 1 difficult²³. As Surprisal_{exp2} in the model above (Table 4) has no main effect on subject omission likelihood, and the marginal interaction with Form of Last Mention is difficult to interpret, the cautious conclusion is that Surprisal based on both preceding and following context is no more predictive of subject omission than is Surprisal based on preceding context only. In other words, while it is still possible that

²³In addition, excluding utterances with non-canonical word order, as was done in experiment 2, eliminates a non-random subset of the data, and it is difficult to tell how this might impact the analysis. It is plausible, for instance, that contexts in which subject omission is more likely to occur may overlap with contexts in which word order alternations are more (or less) likely to occur.

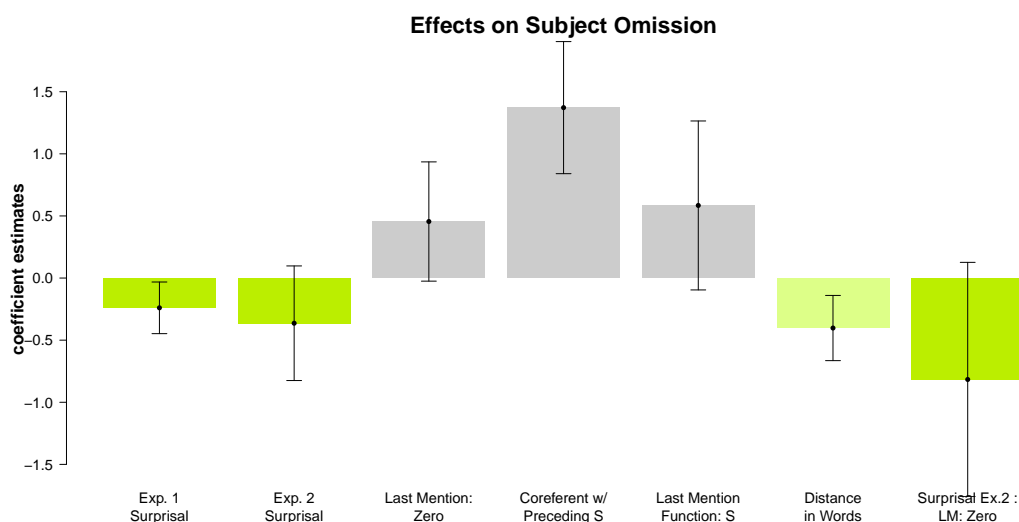


Figure 9: Coefficient bar plot - mixed logit model (exp. 1 and 2)

following context aids ‘recoverability’ of the intended referent, there is insufficient evidence that it’s a significant factor in the choice to pronounce or omit a subject, in the first place.

6 Discussion

The two experiments described in this paper test the hypothesis that the choice to omit optional subjects in Russian is made with considerations of communicative efficiency. The first experiment provides preliminary evidence that this is in fact the case: speakers are more likely to omit those subjects that readers find more predictable (easier to guess accurately), or redundant, given preceding material. The second experiment, in contrast, provides no clear evidence that speakers additionally consider a referent’s predictability given following material in making the choice to omit optional subjects. In this section, I discuss the results of the experiments briefly, as well as how they relate to theories of efficient communication, and, to a lesser degree, audience design, particularly in processing of referential expressions. I also discuss whether the results provide evidence that language production is, at least to an extent, probabilistic; further considerations that may need to be taken into account given the task used here; and possible follow-up work that could elucidate or build up on these findings.

6.1 Experiments 1 and 2

The results of the first experiment demonstrate that greater subject referent predictability, depending on preceding context, contributes significantly to the likelihood that the subject will be omitted in speech or text. In other words, as hypothesized, and consistent with the predictions of UID, more predictable subject referents are more likely to be omitted in Russian. The main implication of this result is that referent predictability contributes to realization of referring expressions, crucially beyond what is predicted solely by referent salience, parallelism effects, or other predictors discussed in section 3.2. It provides further support for UID as a (partial) theory of language production, demonstrating that its effects, above the lexical level, extend beyond those seen in omission or reduction of functional elements such as complementizers, relativizers, or contractions. In addition, rather than looking at the predictability of certain words or syntactic structures, as is typically done (Jaeger & Tily, 2011), it tests whether omission correlates with the predictability of discourse *referents*, regardless of their precise form.

The results of the second experiment suggest that any effects of predictability are primarily limited to preceding context, and that following context plays minimal, if any role in choice of referential expression, at least when considering this measure of efficiency.

6.2 Communicative Efficiency and Audience Design

These findings are most straightforwardly compatible with theories of efficient communication that make reference to omission or reduction of highly probable, or redundant, elements in language production. Here, as in the rest of the paper, UID is used as a proxy for such theories, except where explicitly noted otherwise, or where the theories might make substantially different predictions²⁴. They are also compatible with theories of efficient reference production specifically (Arnold, 2008), which argue that speakers use less informative or reduced referential 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). Functionally

²⁴For instance, most theories (e.g., Aylett & Turk, 2004; Jurafsky, Bell, Gregory, & Raymond, 2001) do not make reference to production beyond the phonetic/phonological level.

the latter might be seen as a subset of the former, although e.g. Arnold's Expectancy hypothesis only makes specific predictions for reference production²⁵.

As mentioned in Jaeger (2010), the view of UID as a theory of audience design is intuitive and straightforward, but may not necessarily hold in its more stringent form. Although there is robust evidence that speakers produce utterances in a manner that facilitates comprehension, it is less clear in many cases whether speakers explicitly design utterances with listeners in mind, or whether listeners independently make use of speech cues informative regarding upcoming material to facilitate comprehension – but with speakers producing this information incidentally, for unrelated processing reasons (Arnold, 2008; Jaeger, 2010). In addition, there is the question of whether, for instance, in reference production, speakers attempt to consider their addressee's discourse model, or simply use their own discourse model, which presumably matches the listener's closely (except for speakers' awareness of upcoming material), for making decisions about how explicit to make a reference. As Arnold (2008) summarizes, there is evidence for all of these perspectives, and most would not need to be considered mutually exclusive.

The question of whether predictability contributes anything additional to production choices, rather than whether it is the primary driving factor in any instance of alternation, then, is perhaps appropriate – it is implausible that it is the single contributor to referential expression choice, given the other linguistic constraints and production pressures that speakers encounter. A finding that it is one of many pressures or constraints among many, aside from providing more information regarding how speakers choose between referential expression types, would also provide a valuable test (one way or the other) of just how far predictability effects on reduction/omission extend - particularly in the case of choices between variants that do not, or may arguably not, have identical status semantically or syntactically (cf., Degen & Jaeger, 2011).

As noted in section 4.2, after the addition of control predictors, Surprisal exerts a relatively weak effect on subject omission - i.e. when the effects of other predictors are held constant, Surprisal does not substantially impact the likelihood of subject omission, despite having a significant effect in the expected direction. This may imply that Surprisal is in fact a relatively

²⁵Arnold does, however, base her hypothesis on findings of probability-/expectancy-based reduction in other domains, e.g. Jurafsky et al. (2001).

poor predictor of subject omission. For example, information carried by the complement cause onset is the single largest contributor, in terms of effect size, to *that*-omission, among variables testing the predictions of multiple competing processing accounts (Jaeger, 2010). In contrast, the contribution of Surprisal to referring form choice/omission in this study, as well as Tily and Piantadosi (2009), is relatively small. This may speak to which linguistic events are more (or less) susceptible to similar predictability influences. In essence, it may suggest that there are more constraints on referential expression choice/omission than on other production choices (Jaeger & Tily, 2011), and that efficiency, at least as defined in this framework, plays less of a role. This could for example, be due to the variants being less meaning-equivalent than alternations involving functional elements.

However, UID would also make the prediction that any properties which increase the likelihood of subsequent mention of a particular referent – such as parallelism of grammatical function, recency of mention, and so forth (Arnold, 2008) – should correlate with higher rates of subject omission (or, in e.g. English, pronoun use). Arnold (2008) similarly hypothesizes that these properties facilitate comprehension of expected referents, precisely because they are associated by listeners with increased likelihood of upcoming referent mention. In this respect – even if these effects partially subsume the effect of predictability – their influence may well be due to readers using them in order to accurately estimate the probability of certain referents being mentioned in upcoming discourse. Further work would be needed to answer this question, however.

6.2.1 Speakers' Use of Subjective Probabilities

Redundancy is defined probabilistically in UID and similar accounts, and may be most easily formalized as such, in order to empirically test theories of efficient communication that make reference to redundancy avoidance. However, redundancy can arguably also be defined non-probabilistically, in terms of (non-probabilistic) information carried by linguistic elements in discourse (Jaeger, 2010). In this case, the findings do not necessarily imply that speakers have access to, or make use of, probability distributions associated with possible continuations (here,

upcoming referents) when making production choices. As such, the results only provide suggestive evidence that speakers keep track of the relative probabilities of possible continuations during incremental sentence production, and make production choices on their basis.

In summary, UID predicts that speakers preferentially pronounce optional subjects when Surprisal is high, or when their mention is relatively unpredictable, but prefer to omit subjects when the Surprisal is so low that their explicit mention would be uninformative (i.e. redundant). These findings are also compatible, as mentioned, with Arnold (2008)'s Expectancy hypothesis, which likewise predicts that speakers use attenuated or reduced referring expressions/forms for those referents that are highly predictable in context. It may arguably be consistent with other accounts where speakers are assumed to have access to probability distributions of various linguistic events, and make use of them when making the choice to reduce or omit redundant information, although many of these theories (e.g., Aylett & Turk, 2004) do not make specific predictions above the lexical level. Arguably they may be compatible with theories of efficient communication that are not cast in terms of probabilities, but this leaves it somewhat unclear what (kind of) information speakers use, or keep track of – or how they do so – in order to make communication more efficient. The tracking of subjective probabilities by speakers and listeners does, at the least, provide a plausible and more straightforwardly quantifiable account of this.

6.3 Task

Following 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. Possibly improving on corpus methods for estimating these probabilities in at least one respect, this task may provide a more accurate estimate of the probabilities comprehenders subjectively assign to various linguistic events (which may differ from the probabilities that producers assign to the same events). For example, there appear to be systematic differences between probability estimates obtained using corpus and cloze studies – it is, however, still unclear which provides a better representation of speakers' subjective probabilities, in either production or comprehen-

sion (Smith & Levy, 2011). Further, the method used here may be a more feasible option for testing similar hypotheses in languages where no large corpora are available.

The two experiments described in this paper also alter the task used in Tily and Piantadosi (2009) in at least one (possibly crucial) respect - in contrast with the use of newspaper texts, the materials used may be more representative of the processing of ‘everyday’ text used in an interactive context, or dialogue²⁶. With respect to dialogue, in some addressee-oriented accounts, choice of referential expression is understood as a ‘conceptual pact’ between people engaged in the same conversation (Brennan & Clark, 1996). However, as previously mentioned, the use of typical transcribed speech, as found in most corpora, would have been problematic with this experimental design, as natural speech is difficult to read due to disfluencies and false starts (etc.).

It should be noted, finally, that readers may have different goals when guessing intended reference in an task such as the one used here, and when processing reference in natural discourse (spoken or written). For example, the time course of resolution is entirely arbitrary: readers have as long as they wish to arrive at an answer, and may consult cues in the text that may not normally play a role in processing or reference resolution. Further, the explicit goal of the task is to accurately guess the identities of upcoming referents (or to accurately indicate that an upcoming referent hasn’t been mentioned yet), while speakers’ immediate goals during processing may be less clear, at least at a fine-grained level. To this end, readers may be more motivated to guess that an upcoming referent is ‘Something new,’ for example, rather than a previously mentioned referent, particularly if they perceive the ‘Something new’ option as the default one in case of doubt. In this case, they may be disproportionately likely to select it when they are particularly uncertain of the upcoming referent’s identity, regardless of whether they think the upcoming referent is something that hasn’t been mentioned yet.

Summing up, this appears to be a promising paradigm for obtaining predictability-based measures, which could supplement corpus work and controlled experiments. However, it may need to be validated further, in order to determine its particular strengths and weaknesses in this

²⁶It should be kept in mind, however, that the dialogue found in plays is arguably artificial, and should differ systematically, though perhaps not in relevant aspects, from natural dialogue.

respect.

6.4 Follow-Up

To follow up on these findings, it may be prudent to first replicate them with a larger number of sources, or data points. This would also allow for the use of additional control predictors, which may help confirm whether Surprisal plays a unique role in subject omission, which would be the strongest form of evidence for the UID hypothesis. There are likely other pressures on production unaccounted for in this model, such as ambiguity avoidance or effects of availability; as well as other constraints on reference processing and production. These could be included in a model of a larger set of data than is presented here. A larger number of data points may also allow for investigating whether there are relevant differences in the use of different types of overt subjects (pronouns, names, definite descriptions) in Russian, as were found for English by Tily and Piantadosi (2009). Doing so was not possible in this study due to the low number of subsequent-mention trials with non-pronominal overt subjects (74/741; 10%, collapsing across names and definite descriptions). Likewise, more observations per data point (i.e. readers) may provide a more fine-grained and stable measure of referents' predictability in context, as well as the overall uncertainty with respect to a referent's identity (Entropy). Overall, however, the findings do lend preliminary support to the hypothesis that referential expression choice, or omission, is guided in part by considerations of efficiency in message transmission, as predicted by e.g. UID, and Arnold's Expectancy hypothesis.

I end this paper with an outline of possible follow-up work on subject (pronoun) omission, cross-linguistically. Although the findings described here are limited to online processing, it is also possible that the average predictability of referential expressions such as pronouns, in context, may play a role in the licensing and frequency of pronoun omission, or *pro*-drop, cross-linguistically. As this could, on average, increase communicative efficiency, by allowing for the more systematic or grammaticized omission of information that is (on average) more redundant, this would also be predicted by UID (cf., Cohen Priva, 2008). The primary hypothesis in this case would be that pronouns are more likely to be elided cross-linguistically if they contribute

little information generally. For example, if the referent of any pronoun (overt or null), in any given language, is more likely than not to have the same grammatical function as its antecedent, it follows that the intended referent of any given pronoun is more predictable, in context, than in a scenario where antecedent grammatical function does not correlate with the grammatical function of a subsequent mention. Conversely, the same would be predicted for a language that more consistently has a non-parallel pattern of coreference. In this scenario, overt pronominal reference contributes relatively little information regarding a referent's identity, relative to a situation where the intended referent of a pronoun is less predictable. Assuming that less informative elements are preferentially omitted, the more predictable a particular co-reference pattern is, the less information an overt form of a pronoun contributes on average towards disambiguating the intended referent, and the more likely it should be to be omitted. Put differently, I would predict that in a language which permits *pro*-drop, the most predictable pattern of co-reference should also be the one that null pronouns are more likely to be used for. This may, for example, help account for why null subjects are more strongly biased towards subject antecedents, in reference resolution, than are overt pronouns (see section 2.2.2). Further, the more predictable the patterns of coreference in a language, the more likely the language may be to license subject omission in the first place (all other relevant factors taken into account). This is an empirically testable hypothesis that may be reserved for future work.

If it is found that average information is predictive of rates of pronoun omission cross-linguistically, beyond the effects of local information (local contextual predictability; cf. Cohen Priva, 2008), it would also need to be determined whether average information is correlated with other constraints on *pro*-drop that have been proposed in the literature. Presumably, even if there is a correlation with contextual predictability that is subsumed by other variables relevant to pronoun omission, it may be suggestive of a relatively dominant role of information density, or Surprisal, in licensing of omission, particularly if the other factors in question are more complex, or have more parameters associated with them. One possible account of the predicted findings, or an avenue that may need to be explored further, would be that information-theoretic constraints on language production have, through a historical process, become grammaticized -

this would point towards biases that operate mainly during language acquisition. If, alternately, it's found that a large amount of the variability is determined by the information that a referent carries in a particular context, this would argue for a more direct effect, and would be indicative of a more continuous process of learning stretching into adulthood (see Fedzechkina, Jaeger, and Newport (2011), for a similar discussion of functional biases in language acquisition).

Summarizing, the theory of Uniform Information Density would predict that the probability of omission of an optional linguistic element increases, as the information it carries in a given context (or, its Surprisal) decreases. The findings described in this paper support this hypothesis. This account would similarly predict that linguistic elements which carry little information on average are more likely overall to be reduced or omitted. Since pronouns, in some languages, effectively 'replicate' the grammatical information provided by the verb, and carry little referential information (as compared to lexical NPs, for instance), they appear to be likely candidates for elision. Overall, the eventual aim of this line of work may be to arrive at a model of *pro*-drop (or similar phenomena) that can, in concert with existing accounts, more effectively, or in a more principled manner, predict what licenses subject omission cross-linguistically.

Appendix A. Tables

A.1 Experiment 1

	Coef β	SE(β)	z	p
Intercept	-1.36	0.28	-4.8	<.0001
Surprisal	-0.18	0.09	-2.1	<.05
Last Mention - Zero	0.21	0.21	1.0	0.3
Coreferent w/ Preceding S	0.84	0.24	3.4	<.001
Last Mention Function - S	0.67	0.29	2.3	<.05
Distance in Words	-0.41	0.12	-3.5	<.001

Table 5: Mixed logit model - experiment 1 (subset)

A.2 Experiment 2

	Coef β	SE(β)	z	p
Intercept	-1.44	0.32	-4.6	<.0001
Surprisal	-0.43	0.24	-1.8	0.073
Last Mention - Zero	0.57	0.31	1.9	0.063
Coreferent w/ Preceding S	0.88	0.30	2.9	<.01
Last Mention Function - S	0.65	0.35	1.8	0.067
Distance in Words	-0.46	0.15	-3.1	<.01
Surprisal : LM - Zero	-0.60	0.53	-1.1	0.3

Table 6: Mixed logit model - experiment 2 (subset) with interaction

	Coef β	SE(β)	z	p
Intercept	-1.37	0.31	-4.4	<.0001
Surprisal	-0.58	0.22	-2.7	<.01
Last Mention - Zero	0.37	0.26	1.4	0.1
Coreferent w/ Preceding S	0.85	0.30	2.9	<.01
Last Mention Function - S	0.64	0.35	1.8	0.071
Distance in Words	-0.45	0.15	-3.1	<.01

Table 7: Mixed logit model - experiment 2 (subset)

Appendix B. Figures

B.1 Experiment 1

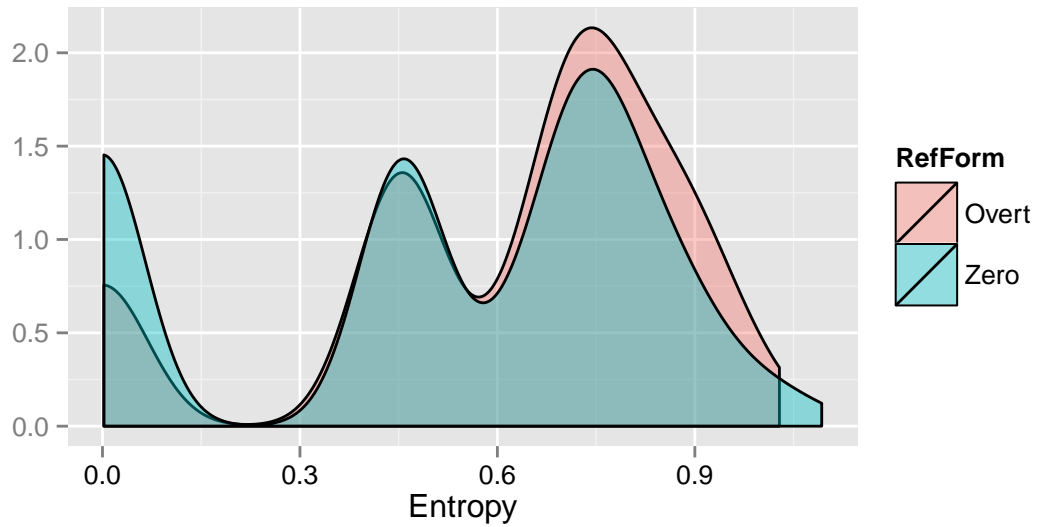


Figure 10: Overt/null subject density by Entropy (exp. 1)

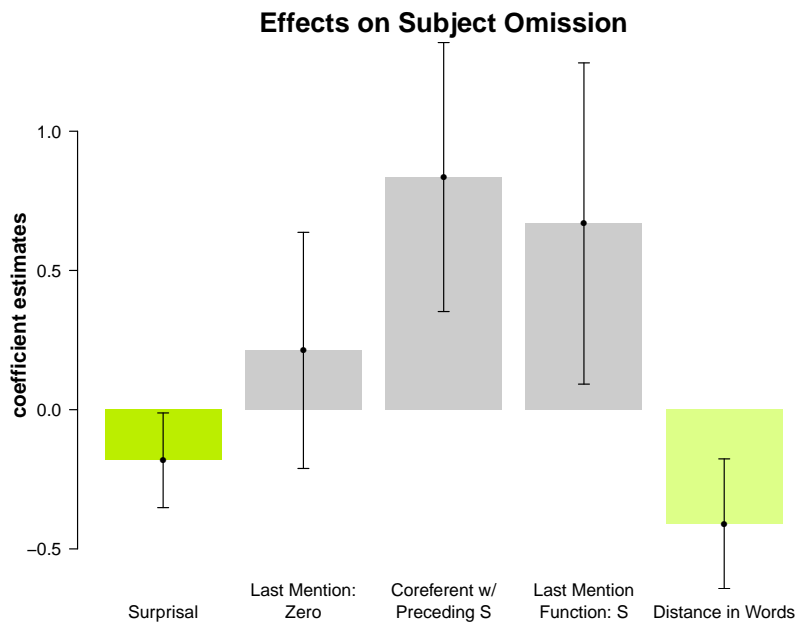


Figure 11: Coefficient bar plot - mixed logit model (exp. 1) subset

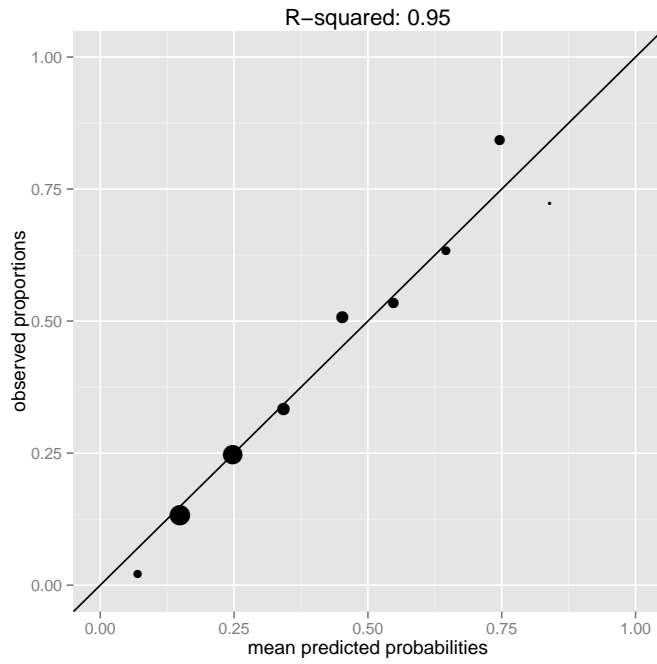


Figure 12: Experiment 1 (subset) model fit

B.2 Experiment 2

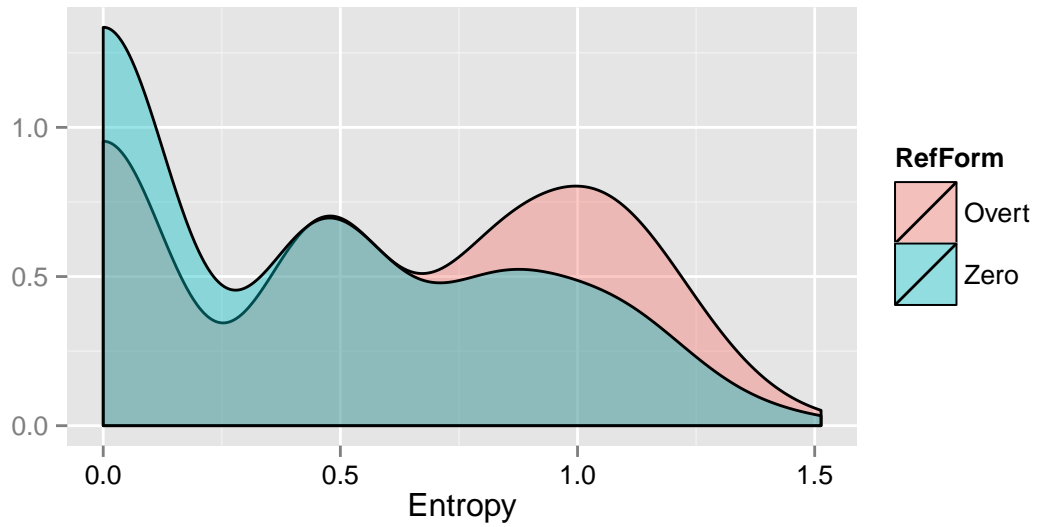


Figure 13: Overt/null subject density by Entropy (exp. 2)

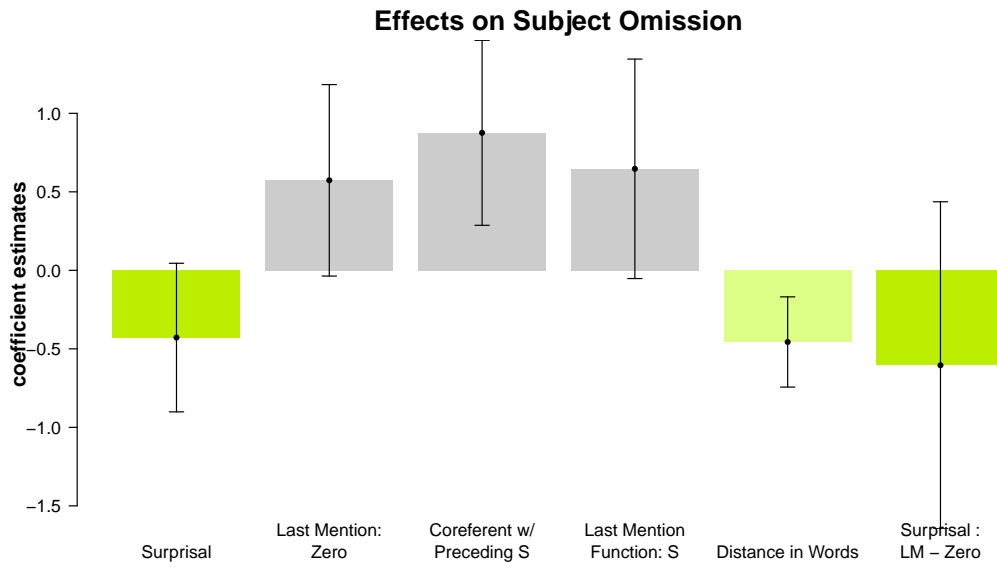


Figure 14: Coefficient bar plot - mixed logit model (exp. 2) subset with interaction

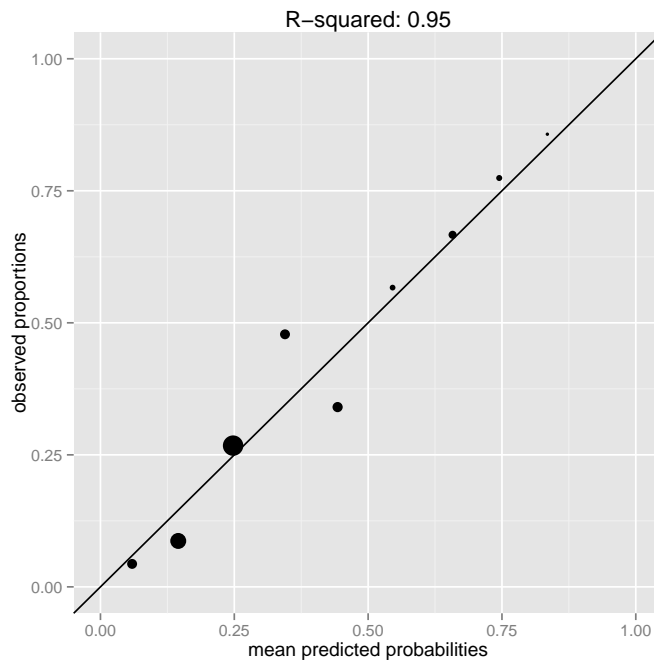


Figure 15: Experiment 2 (subset) model fit with interaction

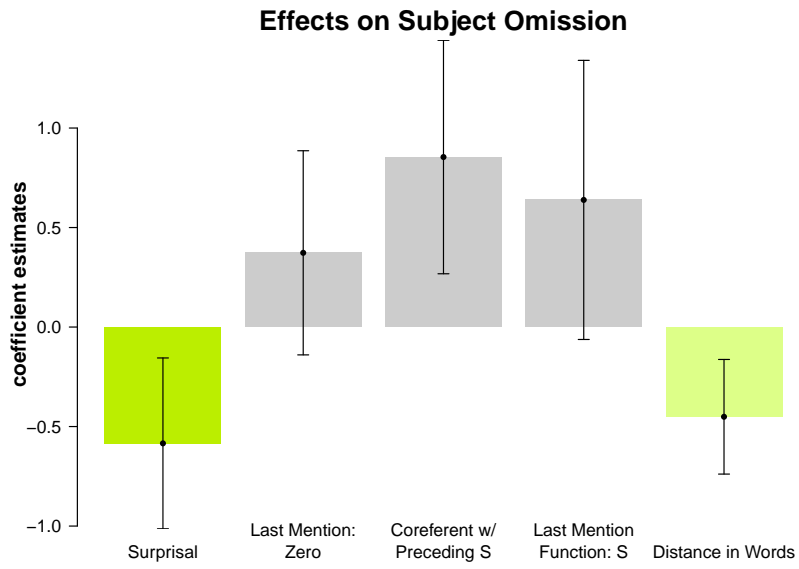


Figure 16: Coefficient bar plot - mixed logit model (exp. 2) subset

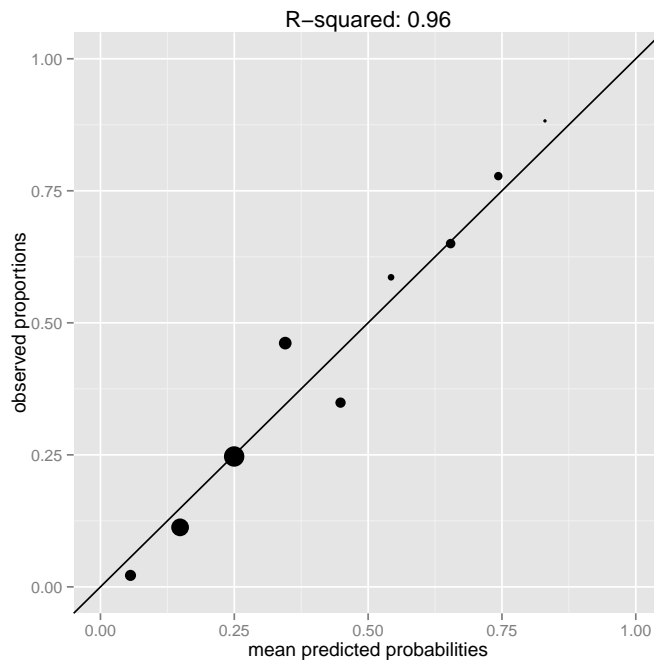


Figure 17: Experiment 2 (subset) model fit

Appendix C. Predictors

This sections lists all predictors coded for that were looked at when modeling the data (at initial or later stages), with those included in the final model enclosed in frame boxes.

Form: [Overt/Zero/Uncertain]; Whether the form of the subject was overt, zero, or uncertain.

Only cases which could be annotated with reasonable confidence were included. A separate predictor, marking all cases of coordinated or possibly coordinated subjects (which were optionally excluded when modeling the data), was also coded for.

Number: [Singular/Plural/Collective/Uncertain]; Whether the number of the subject was singular, plural, collective, or uncertain. Uncertain and collective cases (2.3% of the data) were immediately marked for exclusion.

Case: [Nominative/Dative/Accusative/Genitive/Uncertain]; The case of the target subject. Nominative and Dative subjects are the canonical variants. Accusative, genitive, or uncertain ‘subjects’ (0.8% of the data) were immediately marked for exclusion.

Form of Last Mention (*Last Mention/LM Form*): [Overt/Zero]; Whether the form of the most recent coreferring instance of the referent was overt or zero.

Coreference with the Preceding Subject (*Coreferent w/ Preceding S*): [Yes/No]; Whether the target referent was coreferent with the subject in the preceding clause.

Last Mention Function (*LM Function*): [Subject/Other]; Whether the grammatical function of the most recent NP coreferent with the target referent appeared in subject position, or not.

Distance in Words to Last Mention (*Distance in Words/Distance in Words to LM*): A continuous measure of the number of intervening words between the subject and the most recent coreferring instance of the target referent.

Number of Previous Mentions (*Previous Mentions*): A continuous measure of the number

of times the referent was previously mentioned in discourse (not counting the last subject, if coreferent).

Number of Referents Introduced (*Number of Referents*): A continuous measure of the number of referents introduced so far in the discourse. Roughly, the number of competing referents.

Additional Predictors:

Genre: [Plays/Blogs/Interviews]; The passage genre. This was not included in the final models due to collinearity with other predictors of interest, and unclear theoretical predictions.

Animacy: [Animate/Inanimate/Uncertain]; Whether the subject is animate, inanimate, or uncertain. This was not modeled, due to relatively small number of 'omittable' inanimate subjects (both animate and inanimate subjects were included in the model, with no distinction between the two).

Person: [1st/2nd/3rd]; This was coded for given some evidence (Zdorenko, 2009; Kravtchenko, 2008) that referent person influences likelihood of omission (generally, 1st/2nd person subjects are more likely to be dropped than 3rd - or the dropping of 1st/2nd person subjects may be perceived as marginally more acceptable). This was not included in the final models due to unclear theoretical importance.

Previous Right Guesses: A continuous measure of the number of times the referent was previously the accurate guess. This was not included due to unclear theoretical import.

Distance in Referents to LM: A continuous measure of the distance in referents to the point of last mention. This was not included, due to an inability to normalize the distribution, and given that **Distance in Words to Last Mention**, arguably measuring the same thing, was also coded as a predictor, and did not demonstrate the same problems.

Coreference with the Preceding Object: [Yes/No]; Whether the target referent was coreferent with the object in the preceding clause. This was not included due to reduced theoretic-

cal import, compared with **Coreference with the Preceding Subject**, and low correlation with subject omission.

Referring Expression Form: [Noun/Pronoun/Zero]; Whether the subject is a noun, a pronoun, or silent. This was not modeled due to the low proportion of non-pronominal overt subjects, and limited amount of data.

Last Mention Speaker Same: [Yes/No]; Whether the speaker at the point of last mention is the same or not. This was not included due to reduced theoretical import.

Guess Number: A continuous measure of the number of guesses made (in the passage) so far + 1. This was ultimately not included in the final models, due to unclear theoretical predictions/significance.

Percent New: The proportion of 'Something new' guesses made by readers for the data point in question. This was not included in the final models due to unclear theoretical predictions/significance.

Sentence Number: A continuous measure of the number of the sentence in the passage – roughly, the amount of discourse context so far. This was not included in the final models due to unclear theoretical importance, and lack of an initial correlation.

Average RT: A continuous measure of the average reaction time, from the last guess, for each data point. This was not included due to unclear significance.

Length in Characters to Last Guess: A continuous measure of the length of the text, in characters, from the current point in the text to the point when the last guess was made. This was coded for to control for the effects of text length on reaction times.

Appendix D. Sources

Blogs

	Author	Source
1	miu mau	LiveJournal (http://miumau.livejournal.com/949066.html)
2	Alika	LiveJournal (http://rikki-t-tavi.livejournal.com/1008550.html)
3	Prekrasnaya Jelena	LiveJournal (http://lenk.livejournal.com/20676.html)
4	Taoky	LiveJournal (http://taoky.livejournal.com/549791.html)
5	miu mau	LiveJournal (http://miumau.livejournal.com/936374.html)
6	Rina	LiveJournal (http://rina-caa.livejournal.com/118690.html)
7	tanyant	LiveJournal (http://tanyant.livejournal.com/75432.html)
8	_kukushka_	LiveJournal (http://_kukushka_.livejournal.com/396247.html)

Table 8: Blog sources

Plays

	Author	Title
1	Evgeny Shvarts	A Common Miracle (Obyknovennoe chudo)
2	Evgeny Shvarts	Two Maples (Dva klena)
3	Nikolai Gogol	Marriage (Zhenit'ba)
4	Maxim Gorky	The Lower Depths (Na dne)
5	Anton Chekhov	Uncle Vanya (Dyadya Vanya)
6	Mikhail Bulgakov	Flight (Beg)
7	Alexander Ostrovsky	The Forest (Les)
8	Anton Chekhov	The Cherry Orchard (Vishnevyi sad)

Table 9: Play sources

Interviews

	Interviewee	Source
1	Boris Yefimov	Vestnik (http://www.vestnik.com/issues/98/0707/win/nuzov.htm)
2	Yuri Lyubimov	Vestnik (http://www.vestnik.com/issues/98/0414/win/nuzov.htm)
3	Anatoly Rybakov	Vestnik (http://www.vestnik.com/issues/98/0106/win/nuzov.htm)
4	Yevgeny Pashkovsky	Vestnik (http://www.vestnik.com/issues/1999/1123/win/nuzov.htm)
5	Stanislav Govorukhin	Vestnik (http://www.vestnik.com/issues/98/0331/win/nuzov.htm)
6	Igor Guberman	Vestnik (http://www.vestnik.com/issues/2000/0411/win/nuzov.htm)
7	Alexander Bovin	Vestnik (http://www.vestnik.com/issues/97/1223/koi/nuzov.htm)
8	Maksim Dunayevsky	Vestnik (http://www.vestnik.com/issues/97/0930/koi/nuzov.htm)

Table 10: Interview sources

References

- Ariel, M. (1990). *Accessing noun phrase antecedents*. London: Routledge.
- 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.
- Borer, H. (1989). Anaphoric agr. In O. Jaeggli & K. J. Safir (Eds.), *The null subject parameter* (p. 69-109). Dordrecht: Kluwer Academic Publishers.
- Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology*, 22, 1482-1493.
- Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, 19(4), 441-472.
- Butt, M. (2001). *Case, agreement, pronoun incorporation and pro-drop in South Asian languages*. Talk handout for the Workshop on The Role of Agreement in Argument Structure, Utrecht, August, 2001.
- Carminati, M. (2002). *The processing of Italian subject pronouns*. Doctoral Dissertation, UMass, Amherst.
- Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, Mass.: The MIT Press.
- Clark, H. H., & Fox Tree, J. E. (2002). Using “uh” and “um” in spontaneous speech. *Cognition*,

84, 73-111.

- Cohen Priva, U. (2008). Using information content to predict phone deletion. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*.
- Degen, J., & Jaeger, T. F. (2011). Speakers sacrifice some (of the) precision in conveyed meaning to accommodate robust communication. In *The 85th annual meeting of the Linguistic Society of America*. Pittsburgh, PA.
- FAQ: What are pseudo r-squareds? (2011). Retrieved from http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Pseudo_RSquareds.htm (accessed January 13, 2013)
- Fedzechkina, M., Jaeger, T. F., & Newport, E. (2011, July). Functional biases in language learning: Evidence from word order and case-marking interaction. In *The 33rd annual meeting of the cognitive science society (cogsci11)*. Boston, MA.
- Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4), 296-340.
- Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th annual meeting of the cognitive science society (cogsci08)* (p. 939-944). Washington, D.C..
- 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).
- Gómez Gallo, C., Jaeger, T. F., & Smyth, R. (2008). Incremental syntactic planning across clauses. In *The 30th annual meeting of the cognitive science society (cogsci08)* (p. 1294-1299). Washington, D.C..
- Gordishevsky, G., & Avrutin, S. (2003). Subject and object omissions in child Russian. In *Proceedings of IATL 19*.
- Gordishevsky, G., & Avrutin, S. (2004). Optional omissions in an optionally null subject language. In J. Van Kampen & S. Baauw (Eds.), *Proceedings of GALA 2003* (Vol. 1, p. 187-198). LOT, The Netherlands.

- 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.
- Grenoble, L. (2001). Conceptual reference points, pronouns and conversational structure in Russian. *Glossos*, 1. Retrieved from <http://seelrc.org/glossos/>
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Huang, Y. (2000). *Anaphora: a cross-linguistic approach*. Oxford University Press.
- Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. Doctoral Dissertation, Stanford University.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23-62.
- Jaeger, T. F. (2011). Corpus-based research on language production: Information density and reducible subject relatives. In E. M. Bender & J. E. Arnold (Eds.), *Language from a cognitive perspective: Grammar, usage, and processing* (p. 161-197). CSLI Publishers.
- Jaeger, T. F., Levy, R., & Ferreira, V. (2010). Context-sensitive information density affects syntactic production (even in the lab). *23rd CUNY Conference on Human Sentence Processing*, 20 Mar. 2010.
- 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. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In *Probabilistic linguistics* (p. 39-96). MIT Press.
- 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.
- Kravtchenko, E. (2008). *The roles of antecedent salience and syntactic position in the interpre-*

- tation of Russian pro*. Unpublished manuscript, UMass Amherst.
- Kurumada, C., & Jaeger, T. F. (2012). Communicatively efficient language production and case-marker omission in Japanese. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society (CogSci12)*.
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schkoppf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems 19* (p. 849-856). Cambridge, MA: MIT Press.
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal Of Verbal Learning And Verbal Behavior*, 23(2), 127-150.
- Manin, D. (2006). Experiments on predictability of word in context and information rate in natural language. *Journal of Information Processes*, 6(3), 229-236.
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13, 329-336.
- Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62, 146-159.
- Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, 61, 127-159.
- Seo, S. (2001). *The frequency of null subject in Russian, Polish, Czech, Bulgarian, and Serbo-Croatian: an analysis according to morphosyntactic environments*. Doctoral Dissertation, Indiana University at Bloomington.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(4), 623-656.
- Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, 3(4), 50-64.
- 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 (cogsci11)*.

- Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.
- Tiedmann, J. (2009). News from OPUS - a collection of multilingual parallel corpora with tools and interfaces. In N. Nicolov, K. Bontcheva, G. Angelova, & R. Mitkov (Eds.), *Recent advances in natural language processing* (Vol. V, p. 237-248). John Benjamins, Amsterdam.
- 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).
- Ueno, M., & Kehler, A. (2011). Implicit causality biases in Japanese pronoun interpretation (poster). *24th CUNY Conference on Human Sentence Processing*.
- Weischedel, R., Pradhan, S., Ramshaw, L., & Micciulla, L. (2008). *Ontonotes release 2.0*. (Linguistic Data Consortium).
- 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.