Prediction in SVO and SOV languages: Processing and typological considerations

Paul E. Engelhardt[1], Luna Filipović[2], and John A. Hawkins[2,3]

[1]University of East Anglia, U.K.

[2]University of California Davis, U.S.A

[3]University of Cambridge, U.K.

Address for correspondence:
John A. Hawkins,
Department of Linguistics,
270 Kerr Hall,
One Shields Avenue,
California 95616,
USA
Email: jhawkins@ucdavis.edu

Abstract

In this study, we tested the possibility that different word orders engender different processing preferences. Our key hypothesis was that a head-initial language like English (SVO) allows more prediction compared to a head-final language like Japanese (SOV). In Experiment 1, English and Japanese native speakers completed a cloze task in which they heard a sentence fragment (SV_ in English and SO_ in Japanese) and had to complete it with the word they thought best. We assessed cloze probability of the words produced and voice onset times. Experiment 2 examined written completions in English, in which we compared the cloze probabilities in SV_ fragments versus OS_ fragments. Following the central hypothesis, this experiment allowed us to determine (within language) whether there is more prediction from a noun and verb compared to two nouns. Finally, in Experiment 3, we compared written completions in English and Japanese, in which the stimuli given to participants were identical (SV_ in English and S_V in Japanese). Results across all three experiments were consistent with greater prediction in English. We argue that prediction is one important factor in processing, that it is relied on more in English than in Japanese, and that prediction will be especially favored in languages like English in which the verb regularly precedes its direct object.

Keywords: cross-linguistic, prediction, language typology, cloze probability, voice onset time

Prediction in SVO and SOV languages: Processing and typological considerations

Introduction

A good deal of psycholinguistic research has been devoted to determining the kinds of information that guide parsing decisions. For example, Frazier (1987) examined the effects of proximity and structural complexity, and Altmann and Steedman (1988) addressed discourse context. Ueno and Polinsky (2009) have shown how headedness may affect processing in languages with different word orders, teasing apart some apparently universal processing biases (i.e. the pro-drop bias) from those pertaining to specific language type (e.g. the intransitive bias in OV languages). Of special interest in the present study is research that has dealt with the effects of head information on processing, and in particular, prediction.

A number of linguistic theories have advocated that the head constrains and determines the kind of information that will be available in the sentence because of its power to determine relationships with the other constituents (e.g. Chomsky 1965; Grimshaw 1990; Jackendoff 1972).[1] Psycholinguistic research has tried to determine empirically whether the information supplied by the head is crucial for parsing, and a number of head-driven parsers have been proposed, some more and some less radically head-driven (e.g. Pritchett 1991; 1992; Trueswell and Tanenhaus 1994a, 1994b). On the other hand, other studies have pointed out that the structural (incremental) analysis of a phrase may be initiated before its head is reached. That is, it has been proposed that processing of verb-argument structure can happen before the verb itself is reached (e.g. Konieczny 1996; Mazuka and Ito 1995). For example, Bornkessel et al. (2004) reported that morphological case can be used to assign a theta-role interpretation to a noun phrase (NP) before the verb is encountered. Kamide and Mitchell (1999) have shown that some pre-head incremental processing occurs in Japanese (i.e. in

[1] For a concise introduction to the theory of heads in grammatical theory, and a discussion of some of the issues and disagreements that arise between different grammatical theories, see Corbett, Fraser, and McGlashan (1993).

early dative attachment disambiguation), and that head-driven accounts of parsing do not quite capture the data. However, they conclude that their pre-head parsing account cannot be seen as any more generally applicable than head-driven accounts, and they speculate that parsing mechanisms may differ across languages as a consequence of the specific structural configuration of the language. Thus, the comprehension system will be configured as "a head-driven system when it is needed to handle head-initial languages, whereas it takes the alternative pre-head driven form when dealing with the head-final languages, like Japanese" (Kamide and Mitchell 1990: 657).

The verb has often been argued to play a crucial role in incremental parsing and also in production (see Melinger et al. 2009 for a summary of relevant studies). The general assumption is that verbs are better predictors of their arguments than vice versa. This is because objects are listed in lexical co-occurrence frames for each verb, but a co-occurring verb, or set of verbs, is not listed for each noun (cf. Gentner 1981). Noun phrases and their thematic roles, for example, agent, patient, instrument, etc., do not in general select for, and predict, particular predicates, because they are compatible with too many alternatives. A verb, by contrast, activates a finite set of co-occurrence frames along with frequency-based or contextual preferences. If the verb occurs before one or more of its arguments, then the activated frames can be gradually reduced and any preferences confirmed or disconfirmed as the input continues. If the arguments occur before the verb, then alternative co-occurrence frames will not generally be activated and reduced, but instead, the intended co-occurrence frame for the verb will be selected at the verb and in a way that takes account of the preceding arguments. Some pre-head parsing may occur before the head is reached, but full assignment of all constituents must happen at the verb and will be determined by the specifications of that verb. In short, verbs are assumed to be better predictors of nouns than nouns are of verbs.

Ueno and Polinsky (2009) point out that this asymmetry in verb-argument processing is fully captured by Pritchett's (1992) head-driven parsing model, where it is proposed that syntactic attachment happens at the head. As the parser moves along the string in an SOV language, it encounters different items that must be held in memory before it reaches the verb. For an SVO language, like English, only the subject argument comes before the verb. With SOV both subject and object and possibly other complements and adjuncts precede the verb. SOV constructions should therefore exert an extra (memory-based) processing cost prior to integrating the pre-verbal constituents with the verb, when the verb is encountered. However, speakers of SVO languages also burden their working memory, not incrementally, but immediately as soon as they encounter the verb because they need to hold predicted items (including co-occurrence frames) in memory, which adds to the online processing cost.

Gibson (1998) appeals to this added processing cost resulting from online prediction in working memory in his explanation for the dispreferred ordering of direct objects before subjects in German and Finnish, two languages with rich case marking and with considerable word order flexibility. Object before subject is highly dispreferred, and Gibson argued that this is because a preceding case-marked object NP predicts a subject in these languages, whereas a subject does not predict an object, resulting in a memory-based preference for SO over OS (see also Levy 2013). By this logic, the English SVO pattern would be better for prediction and worse for working memory, with respect to the verb-object relationship, while the Japanese SOV would involve less working memory demands prior to the verb but would permit fewer online predictions. However, this view may be overly simple, and as Hawkins (2022) argues, there are also other factors that encourage or discourage the positioning of one category prior to another in addition to prediction and working memory load, including (1) the agent versus patient theta-role assigned to a noun phrase, and (2) its animacy or inanimacy.

The precise role of online prediction in Japanese is complicated further by an important typological consideration documented in Polinsky and Magyar (2020). They point out that verb-final languages have much lower ratios of verbs to nouns in their lexicons than verb-early languages, and more generally that the verb-to-noun ratio declines proportionately as the verb moves further to the right in the clause across languages. This means, they argue, that more information content is given earlier in processing about the event in question, by a rich and differentiated set of nouns in SOV languages, whereas verb-early languages describe the event early on the basis of their richer set of verbs and the co-occurrence frames they activate.[2]

Nouns in SOV languages can accordingly be predictive of further details of an event, and even of upcoming verb tokens selected from the smaller set available in this language type (cf. further Strunk et al. 2015). In the extreme case Polinsky and Magyar (op.cit.) point out that there are many instances in Japanese of semantically rich nouns co-occurring with semantically empty "light" verbs like *suru* (do), which add little or nothing semantically to an event description following the lexical content of the accompanying noun, as in *guuguuru suru* (i.e. 'to Google', literally google do) (Polinsky & Magyar, 2020). Light verbs are found in many SOV languages (cf. Amberber et al. 2010), though they do also occur in certain non-SOV languages (including English) for structural and processing reasons that may be independent of online prediction (cf. Hawkins 2019).

Prediction and incremental argument interpretations have been much discussed in Surprisal Theory (Hale 2001; Levy 2008). Surprisal is measured by the degree to which a word is *un*predictable due to its inconsistency with previous assignments of structure and meaning that have been made at earlier points in the string. There is evidence that the

[2] Note that both language types can be said to adhere to the efficiency principle of Maximize Online Processing efficiency given in Hawkins (2004, 2014) in these respective ways.

processing load in the interpretation of a word is often proportional to its surprisal (e.g. Frank et al. 2015; Smith and Levy 2013). Most of the work done in this theoretical framework has been done on English or similar languages (cf. Hörberg 2016). The role of prediction (and its close cousin surprisal) may have been overestimated because of the type of language examined, and when structurally different languages are considered, the extent of prediction/surprisal may need to be reconsidered.

Indeed, a study on German (a mixed VO/OV language) demonstrated that both prediction and memory-load restrictions in integration play a role in the processing of the same syntactic structure within the language (Levy and Keller 2013). The results were framed within the debate over locality (which claims that additional material that needs to be integrated with a head increases processing effort through distant and non-local dependencies; see Gibson 1998, 2000) versus anti-locality (which claims that additional material facilitates processing; see Surprisal Theory and related models; Konieczny 2000; Konieczny and Döring 2003; Levy 2008). In Levy and Keller's study, locality and anti-locality effects were examined in verb-final constructions in German, while controlling for lexical identity, plausibility, and sentence position. Clear anti-locality effects were found in the main clause, where presence of a preceding dative argument facilitated processing at the final verb (contra the predictions of Gibson 1998).

But in subject-extracting relative clauses with identical linear ordering of verbal dependents, Levy and Keller (2013) found locality effects in addition to anti-locality effects. Specifically, sentence processing was facilitated when the verb was preceded by a dative argument alone and hindered when both the dative argument and an adjunct preceded the verb. Apparently, under some circumstances locality effects can partially override expectation-based facilitation. Levy and Keller (2013: 215) offered the explanation that "native speakers of verb-final languages are simply more practiced, and therefore, more skilled at comprehending non-local syntactic configurations". They also noted other studies that appear to offer empirical support for this interpretation. For instance, average total-sentence dependency lengths have been shown to be considerably longer in German than in English (Gildea and Temperley 2010; Park and Levy 2009). Second, Vasishth et al. 2010 found that German speakers are better than English speakers at tracking multiple incomplete noun-verb dependencies induced by multiple center embedding. Levy and Keller (2013: 215) concluded that "it appears that theories of syntactic complexity may need to posit memory costs which are *a function of a speaker's linguistic experience rather than fixed and universal*".

The proposal that parsing may be quite different in languages with different word orders was first made in Hawkins (1995), using a different kind of theory and data, namely distinct grammatical patterns in SVO and SOV languages, for which he hypothesized a distinct set of processing routines that could make sense of the grammatical differences (see Hawkins 2014, 1995).[3] Specifically, Hawkins argued that speakers of SVO and SOV languages use distinct parsing strategies deriving from the different verb positions of the verb and from the verb's activation of possible and preferred co-occurrences. In an SVO language, these co-occurrences will gradually be selected from when post-verbal material is parsed, and the preferences will be confirmed or disconfirmed. For SOV both activation and selection will occur simultaneously at the clause-final verb, and the verb will gather in its arguments by looking backwards rather than forwards in the string.[4]

---

[3] Hawkins (2004, 2014) proposes a general Performance-Grammar Correspondence Hypothesis. This hypothesis makes many predictions for grammars and grammatical variation on the basis of processing ease and efficiency that are well-supported empirically and that reinforce the kind of explanation that he offered in Hawkins (1995) for different verb positions and the processing of argument structure.

[4] Hawkins (1995) argued that a number of key grammatical and lexical differences between VO and OV languages, first observed in Müller-Gotama (1994) and Hawkins (1986), can be linked to different processing routines in these languages. A final verb must immediately and correctly select the intended co-occurrence frame on the basis of previous items. It does not have the luxury of SVO languages to gradually eliminate, throughout the post-verbal string, possibilities that have been activated and sometimes erroneously selected at or just after the verb. Verbs in SOV languages generally have tighter and less ambiguous strict subcategorization and selectional restrictions compared with VO languages, which makes recognition of their co-occurrences

The parsing theory of Hawkins (1995) did not use the language of prediction (or surprisal) versus integration, since these were developed much later in the psycholinguistic literature, but the activation potential of the verb which he was assuming, and the possible time lag he proposed between activation and selection of a verb's co-occurrences in SVO languages versus their simultaneity in SOV, meant that what he was proposing in more modern psycholinguistic parlance was essentially this: an SVO language activates and *predicts* one or more co-occurrence frames at the verb, then selects and confirms or disconfirms any preferred one *subsequently*; an SOV language activates its co-occurrence frames at the verb and uses them to select and *integrate* preceding items *simultaneously* with the processing of the verb. If we think in terms of Ferreira and Chantavarin's (2019) proposed synthesis of integration and prediction models in psycholinguistics, the co-occurrences activated at the verb in SVO languages predict possibilities and preferences into which subsequent material can be integrated; in SOV these possibilities and preferences are not predictive at the verb, but instead serve to integrate the verb with its co-occurring arguments when the verb has rich semantic content.

Each of these VO and OV strategies has been argued to have advantages and disadvantages for processing (Hawkins 2014). An SVO language permits more prediction of possible and likely co-occurrences, but there is then additional prediction-related memory cost prior to the integration of post-verb material, and mis-assignments (and garden paths) can and do occur. An SOV language avoids these by integrating its verbs with already encountered arguments and by immediately removing unintended co-occurrence possibilities, but in the process there is an online delay in assigning arguments to the verb (i.e. there is an

easier, e.g. they have less ambiguity between transitive and intransitive readings (more "predicate frame differentiation"). They also have more "argument differentiation" (through e.g. rich case marking), and less "argument trespassing" (less movement of a subject or object argument into a clause where it does not belong semantically with its most immediate verb, in so-called "raising" constructions). SVO languages, by contrast, are more variable in these respects and have less predicate frame and argument differentiation overall and more argument trespassing, consistent with the claim that verb-argument processing is less time-constrained and less demanding when the verb occurs first, before its arguments. See Hawkins (2014: 139-146) for details.

"un-assignment" of argument structure contra the efficiency principle of Maximize Online Processing; see Hawkins 2014), plus there is additional argument-related memory load prior to the verb. These competing efficiencies in processing form part of the explanation for why VO and OV languages are both productive and roughly equally frequent across the globe.[5]

In addition, they can have equally minimal and efficient domains for the processing of basic phrase structure (Hawkins 2004), or, using the theory of Dependency Grammar (Hays 1984; Hudson 1984), they can have equal Dependency Length Minimizations (Futrell et al. 2015).

The fundamental question for prediction and integration that has not yet been sufficiently addressed empirically is whether it is only the verb that can activate and predict its co-occurring phrases, or whether and to what extent noun phrases can also predict the verb, prior to its occurrence at the end of the clause. In an SOV order, any such prediction would have to be based on pre-verbal information, such as the rich lexical content of preceding nouns mentioned above (Polinsky and Magyar 2020), or the case marking and associated theta-role of an NP (see Bornkessel 2002; Bornkessel and Schlesewsky 2006), which may be compatible with the selectional and subcategorization restrictions of only a limited set of verbs. There is also the issue of how/when the discourse context could make a particular verb predictable, as could speakers' real-world knowledge about the probabilities of certain events and linguistic knowledge about the frequencies of noun-verb co-occurrences in their previous linguistic experience. In other words, SOV language users are still plausibly engaging in some prediction and attempting to anticipate the verb, and if so, we might expect similar

[5] See Greenberg (1963), Hawkins (1983), Dryer (1992) and Dryer (2013a) for precise figures, and Hawkins and Cutler (1988) for comparison of numerous word order samples in the typological literature. The proportions of basic word orders in the Expanded (Greenbergian) Sample of Hawkins (1983), with 336 entries, were: SOV = 174 (52%), SVO = 109 (32%), and V-1 (VSO or VOS) = 53 (16%); i.e. 52% OV (head-final) to 48% VO (head-initial). The most recent sample of Dryer (2013b), with 1,173 entries for languages with a dominant word order of one of these types, gives: SOV = 565 (48%), SVO = 488 (42%), and V-1 (VSO or VOS) = 120 (10%); i.e. 48% OV (head-final) to 52% VO (head-initial).

processing outcomes in VO and OV languages, suggesting that sentence processing patterns (and prediction) may not be primarily guided by the position of the verb.

The traditional head-driven parsing assumption is that only the verb activates and predicts the co-occurrences of that verb, and that object NPs or other strictly subcategorized phrases do not activate and predict the set of verbs with which they co-occur. But it may often be possible for semantically rich nouns or real-world knowledge and previous linguistic experience, or discourse context, to activate predictions regarding the appropriate actions that apply to certain entities that are encountered. On-line predictions may then look similar to the grammatically and lexically based parsing predictions that derive from the asymmetrical structuring of lexical entries for verbs and nouns.

Our key hypothesis is that we should nonetheless see a greater and more systematic role for on-line prediction in an SVO language like English, compared with SOV Japanese, because these languages differ in the relative order in which object nouns and verbs are processed. Both sets of speakers will have prior real-world knowledge of events and of noun-verb co-occurrence frequencies, and both will presumably have similar discourse processing skills that enable them to integrate the current sentence into its discourse context. Both will also, we assume, have asymmetrical lexical listing for nouns and verbs, whereby nouns are listed alongside verbs in the mental lexicon but not vice versa. The crucial, and plausibly only, difference between English and Japanese therefore involves the relative order of access to these two asymmetrically listed categories, and this should, we hypothesize, result in a more productive and systematic role for prediction in English than in Japanese.

The empirical questions addressed in this study are therefore: how predictive are object NPs of their following verbs, and does prediction actually work in both directions, from verb to object NPs, and from object NPs to verb, and to what extent does it do so in contexts with different co-occurrence constraints (high vs. low)?

The Current Study

Our key hypothesis was that prediction will be a more robust and regularly used parsing strategy in English than in Japanese, since the most predictive category, the verb, appears earlier in English than in Japanese. It is important to note that the current study does not manipulate discourse context. Instead, we examined only decontextualized sentences in order to test whether there is evidence for the differential use of prediction in the online processing of basic transitive sentences within these two language types.

Note also that we have chosen English and Japanese as representatives of the SVO (head-initial) and SOV (head-final) language types that are commonly assumed in language typology in order to test our hypothesis about the crucial role of verb position in relation to prediction in processing. English is a prototypical SVO language; it exhibits some variability with respect to the positioning of (auxiliary) verbs and subjects (*Never have I seen such a thing*), but not with respect to the verb and direct object, and there is universal agreement that SVO is its "dominant" order typologically (Lehmann 1978; Dryer 2013b), as well as its "basic" order in different models of grammar (cf. Brown and Miller 1996). Japanese is a prototypical SOV language (Kuno 1973, 1978), indeed a prototypically "rigid" SOV language (Greenberg 1963), with grammatically permitted albeit relatively infrequent reorderings of SO to OS before V (Yamashita 2002), and also some observed postposings of elements to the right of the verb that occur in Japanese conversation (Clancy 1982; Whitman 2000; Ono 2006). These two languages are well-placed typologically, therefore, to form the basis for this initial test of whether different verb positions are indeed associated with differences in prediction in language processing. If our hypothesis is supported, more fine-tuned studies can then be conducted on languages that are less consistently SVO than English, for example Polish (Siewierska 1993) with its more variable word orders alongside a dominant SVO, and on languages that are less rigidly verb-final than Japanese and combine

OV and VO orders. The different grammatically permitted positionings of the verb within these languages would provide a rich set of stimuli for within-language comparison and for further testing of our hypothesis.

To examine our key hypothesis, we conducted three experiments on English and Japanese. The first compared native speakers of English to native speakers of Japanese, using a cloze task. English participants heard SV_ and had to produce what they thought was the best (direct object) completion. Japanese participants heard SO_ and had to produce what they thought was the best (verb) completion. Because this experiment compared the production of nouns to verbs cross-linguistically, we believed that two further experiments were necessary to augment Experiment 1’s findings. In Experiment 2, we examined native speakers of English, who also completed a cloze task. One group read SV_ and had to provide a best written (direct object) completion. The other group read OS_ (using an it-cleft construction, see below) and had to provide a best written (verb) completion. This allowed us to examine noun versus verb prediction within one of our languages. Following the key hypothesis, we expected to observe less prediction in the OS_ order (i.e. nouns are not (as) good predictors of verbs). Finally, in Experiment 3, we again compared native speakers of English and Japanese, but this time, unlike the cloze completions in Experiment 1 (which were different, i.e. nouns vs. verbs) we used direct translations, which meant that participants were presented with the exact same stimuli. In this case, English participants were presented with SV_ and had to produce a direct object. Japanese participants were presented with S_V and also had to produce a direct object. We expected that Japanese participants would show less evidence of prediction, even in this sentence type in which participants heard both S and V, since the processing routines associated with rigid verb-finality are not normally designed to use the verb as a predictor in online sentence comprehension but instead as an integrator of already encountered material.

Experiment 1

In Experiment 1, we utilized an online version of the cloze task in which participants heard a sentence fragment and had to complete the sentence as naturally as possible with the word that best completed the sentence. We refer to the task as "online" because in addition to assessing the cloze probability of the words produced, we also assessed the reaction time, that is, the voice onset time. Voice onset was defined from the time from the end of the sentence fragment to the onset of the word that the participant produced.

The cloze task has been the gold standard psycholinguistic task used to assess prediction. Staub et al. (2015) showed a relationship between the cloze probability of words produced and voice onset time.[6] In particular, there was a negative relationship, such that, higher cloze probability responses were associated with shorter voice onset times, indicating that more predictable responses were produced faster. Based on these findings, we hypothesized that Japanese speakers (SOV) would show differences compared to English speakers (SVO). More specifically, we expected that SOV speakers would produce either lower cloze probability responses with similar voice onset times or similar cloze probability responses but would be slower to do so. Both of these alternatives would be indicative of less prediction.

Method

Participants. Participants were 104 adults with a mean age of 23.75 years (*SD* = 7.94, range 18 – 65). Eighty-two were female, and 22 were male. The demographic information broken down by language is presented in Table 1. The groups were significantly different in terms of age, but not different in gender. Analyses investigating age as a possible confound are presented in Section B, Supplementary Materials. Native English speakers were primarily

[6] Our results showed a similar relationship in terms of significant negative correlations between reaction time and cloze probability (see Section A, Supplementary Materials).

undergraduate psychology students from the University of East Anglia, and native Japanese speakers were primarily University of East Anglia undergraduates from various departments, as well as a few community-recruited adults. Thus, all Japanese speakers were Japanese-English bilinguals. All participants had normal or corrected-to-normal vision and hearing. Each participant gave informed consent agreeing to participate, and ethical clearance for the study was granted by the School of Psychology Research Ethics Committee.

Table 1

Demographic characteristics for the two language samples

	English (50)	Japanese (54)	Significance
Demographic	Mean (SD)	Mean (SD)	
Age (years)	20.12 (2.44)	27.11 (9.64)	$t(102) = -4.98$, $p < .001$
Gender (% male)	16.0%	25.9%	$t(24) = -1.24$, $p = .219$

Materials. We selected 12 SVO English sentences from the published cloze norms by Arcuri et al. (2001), which served as the critical items. There were an additional ten items selected, which served as fillers. (Fillers consisted of a variety of different syntactic structures and required participants to produce, for example, adjuncts or indirect objects in ditransitive structures.) These 22 items ranged in cloze probability from 1.0 to .11, and thus, provided a full range of sentence predictabilities. The English versions were recorded by a female native speaker of British English. The Japanese versions were translated and recorded by a female native speaker of Japanese. The Japanese translations were created by using the highest cloze probability noun from the published English norms and placing it in the object position of the Japanese SO structure, corresponding to the English SVO. For both English and Japanese, each of the 22 sentences was first recorded with an anomalous word (with a voiceless plosive consonant) in sentence final position (e.g. *The hunter shot a large peak*). This ensured no coarticulation effects between the final and penultimate words in the sentence. The final word was digitally removed to create the actual stimuli for the experiment.

Table 2

Example stimuli, with cloze probabilities from current study

English

In the distance, they heard the _______.
Cloze Probabilities: noise (.32), thunder (.10), birds (.06), scream (.06)

The hunter shot a large _______.
Cloze Probabilities: deer (.78), lion (.06), bear (.06)

John wisely chose to pay the _______.
Cloze Probabilities: bill (.50), debt (.08), taxes (.04), man (.04)

At night, the old woman locked the _______.
Cloze Probabilities: door (.76), basement (.04)

Japanese

遠くで、 彼ら は、 列車の 音 を_______.
In the distance they は-sub train sound を-obj
Cloze Probabilities: heard (.72), listened (.28)

猟師 は、 大きな 鹿 を_______.
Hunter は-sub large deer を-obj
Cloze Probabilities: caught (.37), shot (.24), killed (.17), hunted (.09)

ジョン は、 賢明にも 勘定の 支払い を すること を_______.
John は-sub wisely bill payment を-obj doing を-obj
Cloze Probabilities: refused (.33), decided (.11), hesitated (.04)

夜、 その 年配の 女性 は、 ドアの 鍵 を_______.
At night old woman は-sub door key を-obj
Cloze Probabilities: locked (.39), forgot (.28), opened (.11)

Design. There was a single independent variable (i.e. language), which was between subjects. We assessed two dependent variables. The first was the cloze probability of the responses provided for each sentence. Cloze probabilities were calculated in the typical manner based on the responses provided by participants in the study. The second dependent variable was the reaction time (RT), which was defined as the time from the end of the last word in the recorded sentence to the voice onset time of participants' responses (i.e. from the end of the recording to when the participant began speaking).

Procedure. Participants were instructed to have a seat at the experiment computer. They were then asked to read the instructions, which indicated that they would hear an incomplete sentence and that their task was to complete it as naturally as possible using the word that they thought would fit best. Written instructions were provided in the native language of the participant. The experimenter was a native speaker of English, and thus, any requests for clarification to native Japanese speakers were provided in English. Participants were further told that there would be three practice trials and 22 regular session trials, and that if they could not think of a word, they were allowed to respond "I don't know". These trials were then excluded from all analyses. Participants pressed the space bar after each response to proceed to the next trial, and the entire task took approximately 10 minutes. All participants were tested individually and given an information sheet and consent form. After providing consent, participants completed a short demographic form, and they were debriefed following the study.

Data-Analytic Plan. The data-analytic plan had two main components. The first was to compare the groups (English vs. Japanese) on the mean cloze probability and reaction time for the full set of critical items, using linear mixed effects models in R. In addition, following the analysis procedures of Staub et al. (2015), we also examined the "modal" responses, again using linear mixed effects models in R. Modal responses are the most frequent response for each particular item in each language. We expected that the modal responses would have higher mean cloze probabilities and faster reaction times. Because there is a reasonably strong linear relationship between the cloze probability of words produced and reaction time, we followed up each of the reaction time analyses with a model that simultaneously included cloze probability. This allowed us to examine whether there was a significant group effect whilst removing variance due to the cloze probability of the words produced. If the group effect remains significant, then it suggests that reaction time differences are not simply due to differences in the cloze probability of the word produced (i.e. at issue here is whether language exerts an effect **beyond** the typical cloze-RT relationship).

The second component of the analysis focused on examining low-predictability items and high-predictability items, henceforth referred to as low- and high-constraint items. To do so, we split the critical trials based on their overall mean cloze probability. Items with a mean cloze probability of (< .50) were deemed "low" constraint and items with a mean (> .50) were deemed "high" constraint. This analysis allowed us to examine whether group differences were driven by items falling at one or the other end of the predictability continuum. For this set of analyses, we followed up as before with an additional linear mixed effects model which had reaction time as the DV, and also, included cloze probability.

We ran linear mixed effects models in R (version 3.6.1), including language and high vs. low constraint as fixed effects and subjects and items as random effects (Bates et al. 2018; Jaeger 2008; R Core Team 2018). We also included random slopes (Barr et al. 2013). In cases where the model failed to converge, we simplified the random effects by eliminating random slopes (item and then subject) and then the item random effect. For significant interactions, we followed up with Tukey post-hoc tests comparing the two word orders. Model fit was by maximum likelihood and *t*-tests use Satterthwaite's method. The data, as well as analysis scripts, are provided in the Supplementary Materials (see Section F).

Results

Prior to the inferential analyses, the data were checked for outliers and to ensure that the data were normally distributed (i.e. that skew was less than two times the standard error across by-subject mean). Missing data (due to "I don't know" responses) constituted approximately 3% for English and 2% for Japanese.

Overall Means. We began the analysis by assessing language group differences in cloze probability and reaction times on the full set of critical items. Results showed nearly identical results for cloze probability, but English participants had significantly shorter voice onset times than did the Japanese participants (see Table 3). This is consistent with greater prediction in English. In order to control for differences in cloze probability, we conducted an additional analysis which included, language group and cloze probability as IVs and reaction time as DV. The results from that analysis showed that language group remained significant (t = 4.10, p < .001), as well as a significant effect of cloze probability (t = -11.70, p < .001). Thus, it is clear that the participants' offline behaviour was not different between the two groups, whereas the online behaviour showed that Japanese speakers took (159 msec) longer to issue their responses. This supports the hypothesis that Japanese speakers are less able to predict sentence-final verbs in online processing than English speakers can predict sentence-final direct objects. This is consistent with English activating the co-occurrence frames of the verb early, allowing prediction of what is to come, whereas in Japanese, activation of the frame occurs late and simultaneously with integration of what has come before the verb.

Table 3

Results of Tukey post-hoc comparisons for both cloze probability and voice onset time.

	English	Japanese	Significance
	Mean	Mean	
Overall Results			
Cloze response	.422	.403	z = -.98, p = .33
Cloze RT (msec)	1081	1240	z = 5.48, p < .001
Modal Responses			
Cloze response	.679	.618	z = -3.29, p < .01
Cloze RT (msec)	901	1164	z = 3.25, p < .01

We also calculated the mean reaction times and cloze probabilities for the modal responses produced (i.e. the most frequent response for each item). Results from this analysis are also provided in Table 3. As can be seen, Japanese participants produced lower cloze probability responses compared to English participants, and again, reaction times were significantly longer in Japanese. In order to control for the differences in cloze probability, we again conducted a follow up analysis including, language and cloze probability as IVs and reaction time as DV. The results from that analysis showed that language was again significant (t = 3.06 p < .01), as was cloze probability (t = -3.96, p < .001) This analysis enables us to rule out the possibility that the "overall means" reported above could be due to Japanese participants producing more variable "non-modal" responses.

High- vs. Low-Constraint Items. For the second set of analyses, we divided the dataset into high-constraint items and low-constraint items (defined by a .50 cutoff), and included this variable in the model. Thus, giving us a 2 × 2 (language: English vs. Japanese, and constraint: high vs. low) design. The means for this breakdown of the data are presented in Table 4. Results, examining cloze probability, showed a significant main effect of constraint (t = -25.55, p < .001). High constraint items produced higher cloze probability responses than did low constraint items, which is fully expected. There was also a significant main effect of language (t = -5.31, p < .001) and a significant interaction (t = 8.03, p < .001). Japanese speakers produced significantly lower cloze probability responses for the high constraint items, and higher cloze probability responses for low constraint items (see Table 4). The analyses of reaction times showed a similar pattern, with a significant main effects of constraint (t = 7.08, p < .001) and language (t = 5.16, p < .05), as well as a significant interaction (t = -7.13, p < .001). With respect to the interaction, Japanese speakers had significantly longer voice onset times for high-constraint items, and lower reaction times for low-constraint items. However, the reaction time difference for the low-constraint items was not significant.

We again conducted a follow up analysis in which we included the cloze probability of responses in the model, which ensures that group differences account for significant (unique) variance beyond that accounted for by the cloze probability of the responses produced. Results showed a significant main effect of language (t = 4.31, p < .001), and a significant interaction (t = -5.00, p < .001). The main effect of cloze probability was also significant (t = -9.21, p < .001).

Table 4

Results of Tukey post-hoc comparisons based on high and low cloze probability items.

	English	Japanese	Significance
	Mean	Mean	
High-Constraint			
Cloze response	.682	.570	z = -5.31, p < .001
Cloze RT	838	1209	z = 5.16, p < .001
Low-Constraint			
Cloze response	.147	.275	z = 6.07, p < .001
Cloze RT	1352	1265	z = -.75, p = .86

Discussion

To summarize the main findings, the overall mean cloze probability was virtually identical in English and Japanese. In other words, in this offline sentence-completion measure, the cloze probabilities for selecting relevant and appropriate completions within each language group did not differ significantly across the groups. This supports the idea discussed in the Introduction that nouns can predict verbs just as verbs can predict nouns. However, what did differ was the reaction times. We observed significantly longer reaction times (179 msec) among Japanese speakers. For the modal responses and high-constraint items, we observed **both** lower cloze probability responses and longer reaction times in Japanese speakers. All three of these analyses would suggest less efficiency-related prediction in Japanese. The low-constraint items elicited higher cloze probability responses in Japanese than in English but the reaction time analysis was not significant. It is important to note that these items afford less prediction anyway, and so, the mixed/non-significant differences are not unexpected.

One acknowledged limitation of the current study was the low number of critical items. There were only 12 critical items, which is modest by most standards. However, when we designed the study, we relied on the published cloze probability norms (for English) that were available, and those norms contained only twelve SVO structures. In order to ensure that the results of the current study were reliable, we conducted a replication study (see section C, Supplementary Materials). The cloze probabilities of the replication study were, on the whole, highly similar. This helps ensures reliability of the findings and addresses the main limitation of Experiment 1.

Experiment 2

Experiment 1 compared SO_ Japanese sentences to SV_ English sentences in a cloze task and showed that Japanese speakers engaged in less prediction. More specifically, across the full set of sentences, cloze probabilities were equal but Japanese speakers had significantly slower reaction times. When analyses focused on the most predictable sentences, Experiment 1 found **both** lower probability responses and longer reaction times. However, this study leaves open some alternative explanations (e.g. nouns may be easier to produce than verbs, cultural differences in background knowledge, etc.). Thus, in Experiments 2 and 3, we attempted to narrow down and reduce the number of differences in our comparisons so as to try and confirm our interpretation of the results of Experiment 1.

In Experiment 2 we accordingly set out to examine differences in prediction **within** one of our languages, English and we compared the production of nouns to the production of verbs in this language. In order to do so, we took advantage of the English it-cleft sentence structure (e.g. *It was the match the ailing team* ______), and compared it to (e.g. *The ailing team forfeited the* _______*)*. The rationale for the hypothesis (i.e. that less prediction would be observed in OS__) is based on the idea that nouns are not (as) good predictors of verbs, as verbs are of nouns. Second, we also obtained new materials from Staub et al. (2015), which gave us 50 critical items. Half of these were deemed by the original authors to be high constraint (cloze probability > .50) and half were low constraint (cloze probability < .50). The OS_ sentences were created by taking the highest cloze probability noun from the SV_ completions and inserting it into an it-cleft sentence and removing the verb. All sentences were simple transitive sentences, except that they had an adjective prior to the subject noun (see Section D, Supplementary Materials). We had intended to run the study in the laboratory, as in Experiment 1. However, very early in the data collection process, the global pandemic hit. Thus, the experiment was conducted on Qualtrics, which prevented the collection of reaction times. Thus, Experiments 2 and 3 only analysed mean cloze probabilities.

Methods

Participants. Participants were 81 adults with a mean age of 21.65 years (*SD* = 6.89, range 18 – 57). Sixty-eight were female, 12 were male, and 1 non-binary. The demographic information broken down by group is presented in Table 5. The groups were not significantly different in age or gender. Participants were primarily undergraduate psychology students from the University of East Anglia. All confirmed being native speaker of English.

Table 5

Demographic characteristics for the sample

	English SVO	English OSV	Significance
Sample Size	45	36	
Demographic	Mean (SD)	Mean (SD)	
Age (years)	22.42 (8.28)	20.69 (4.57)	$t(79) = 0.27$, $p = .27$
Gender (% male)	22.2%	5.6%	$t(79) = 0.23$, $p = .23$

Materials. We selected 50 SVO English sentences from materials used by Staub et al. (2015). Half were high constraint and half were low constraint. Three additional items were selected to serve as practice trials.

Design and Procedure. The study had 2 × 2 design. The word order variable (SV_ and OS_) was manipulated between subjects, and constraint (high vs. low) was manipulated within subject. The dependent variable was the mean cloze probability of the responses produced by each participant. Mean cloze probabilities were calculated in the typical manner. We also calculated the mean cloze probability of the modal responses.

Participants were recruited from the UEA psychology participation pool, and participants were automatically redirected to Qualtrics. On the information sheet, participants were provided with background information about the study. Then, they were given written instructions for the task in their native language. Those instructions informed participants that they would be presented with a list of incomplete sentences, which they need to complete with the word they thought best fit. Participants were informed that even though there was no time limit, the task would take no more than 20 minutes. Participants then completed some demographic and language background questions, and then proceeded to the main task. Participants were debriefed at the end of the study, and permission to use their data was asked a second time. They were also provided with the researcher contacts for further queries regarding the study. Ethical clearance was granted by the School of Psychology Research Ethics Committee.

Data-Analytic Plan. The data-analytic plan again had two main sections. The first was to compare the two word orders (SV_ and OS_) for the overall mean responses and for the mean modal responses, similar to first section in Experiment 1. Second, we conducted two 2 × 2 linear mixed effects analyses, which in addition to word order also included the variable for item constraint (high vs. low). We ran a linear mixed effects model in R (version 3.6.1), including response category and language as fixed effects and subjects and items as random effects (Bates et al. 2018; Jaeger 2008; R Core Team 2018). We also included random slopes (Barr et al. 2013). In cases where the model failed to converge, we simplified the random effects by eliminating random slopes (item and then subject) and then item random effects. For significant interactions, we followed up with Tukey post-hoc tests comparing the two word orders. Model fit was by maximum likelihood and *t*-tests use Satterthwaite's method. The data, as well as analysis scripts, are provided in the Supplementary Materials (see Section F).

Results

Prior to the inferential analyses, the data were checked for outliers and to ensure that they were normally distributed (i.e. that the skew was less than two times the standard error). Missing data constituted approximately 4% for English SVO and none for English OSV.

Overall Means. We began the analysis by assessing word order differences in overall cloze probability on the full set of critical items (see Table 6). Results showed that SV_ order did not have higher cloze probability than did OS_ order, and the same effect held for the modal responses (see Table 6).

Table 6

Results of Tukey post-hoc comparisons for word order (response category), including high and low cloze probability items.

	English SVO	English OSV	Significance
	Mean	Mean	
Overall Responses	.372	.355	$z = 1.33$, $p = .187$
Modal Responses	.631	.615	$z = 1.50$, $p = .133$
High Constraint Items			
Overall response	.542	.410	$z = 10.06$, $p < .001$
Modal response	.734	.647	$z = 9.37$, $p < .001$
Low Constraint Items			
Overall response	.199	.299	$z = -7.99$, $p < .001$
Modal response	.426	.574	$z = -13.68$, $p < .001$

High- and Low-Constraint Items. Results of a 2 × 2 linear mixed effects analysis on the overall responses showed a significant main effect of constraint ($t = -2.44$, $p < .05$). This was expected, because items were pre-selected (by Staub et al., 2015) to show this difference. The main effect of word order was significant ($t = 10.06$, $p < .001$). Finally, there was a significant interaction between word order and constraint ($t = -14.61$, $p < .001$). The important simple effects comparisons are presented in Table 6. There were significant differences based on word order for both the high- and low-constraint items. For the high-constraint items, the SV_ group showed higher cloze probability and the reverse occurred for the low-constraint items (OS_ > SV_).

Results of a 2 × 2 linear mixed effects model on the modal responses showed a significant main effect of constraint ($t = -4.98$, $p < .001$), which again, was expected. The main effect of word order was significant ($t = 6.77$, $p < .001$), in which the OS_ order had higher cloze probability. This was not as expected. Finally, the interaction ($t = -11.46$, $p < .001$) was also significant. However, the pattern of that interaction was highly similar to the results of the full set of items. That is, there were significant differences in the expected direction for the high-constraint items and the reversed pattern for the low-constraint items.

Discussion

The rationale for this experiment was to examine whether, within a language, prediction was greater from a noun and verb, as compared to two nouns. The results showed this to be the case.[7] Specifically, in the items affording the most prediction (high constraint), we observed between .133 and .083 higher cloze probabilities in high-constraint items when comparing the word orders. This is consistent with the hypothesis that predicting a verb from two nouns is more difficult than predicting an object noun from the combination of a subject noun and a verb. In the low-constraint sentences, the reversed pattern was observed, namely that predicting a verb from two nouns was less difficult.

It is important to note that low-constraint items afford less prediction. In order to try to understand these "reversed" differences further, we took a close look at the verbs produced with OS_ structures and compared them to the verbs in the original items, particularly focusing on the modal responses. For the high constraint items, 17 items resulted in an exact match (e.g. *fixed* vs *fixed*) or synonym (e.g. *forfeited* vs. *lost*), three resulted in semantically related verbs (*cried* vs. *overwhelmed*), and only five were unrelated (*scolded* vs. *hit*, *threaded* vs. *used*) (for further details, see Supplementary Materials, Section E). In the low constraint items, a different pattern emerged. Only eight were exact matches or synonyms, four were

---

[7] A reviewer of this paper raised the possibility that there might be a confound in this experiment resulting from the pragmatics of the cleft construction. There is indeed both a semantic and a pragmatic difference between a cleft construction like *It was the apples John bought* and *John bought the apples*. The former logically presupposes that John bought something, the latter does not (cf. Keenan 1971), and in a discourse context the former would be uttered when it is mutually known to the interlocutors that he did buy something and the cleft serves to identify and clarify what exactly. But our stimuli contain no context beyond the sentence itself. Hence the only "context" for the missing final category here, the verb, is the previous two NPs within the sentence itself, and so this is a precise test of the very issue we are trying to address: do two nouns predict a verb better than a noun and a verb predict a following (object) noun. By removing any previous or larger discourse context, we focus on exactly the issue at hand, whether verbs are better predictors of nouns or nouns of verbs, and we remove the relevance or applicability of discourse-pragmatic considerations.

related, and 13 were different. This pattern is again likely based on the fact that the low constraint items do not afford much prediction. Also, the non-matches tended to result in shorter (*hit* vs. scolded) and “lighter’ verbs” (*made* vs. *kneaded*). These two points combined are the likely reason why the low constraint items resulted in different (reversed) results compared to expectations. However, it is very important to note that the high constraint items did conform to expectations and thus support our key hypothesis.

Experiment 3

In Experiment 3, we compared SV_ English sentences to S_V Japanese sentences. The rationale for doing so was to consider the extent to which cross-linguistic differences in processing are observed when the stimuli are the same. The key hypothesis underlying this study is that prediction is generally more difficult in head-final languages, primarily due to the position of the verb, and that the verb serves to integrate material that precedes it and that has already been processed rather than to predict what lies ahead and has not yet been encountered. Because of this typological difference, we expected that Japanese speakers will have developed general processing routines for the integration of prior material and engage in less prediction than English speakers, but that Japanese speakers do nonetheless engage in some prediction on the basis of pre-verbal material. Given these assumptions, it is an empirical question whether Japanese speakers will produce equal cloze probability responses to those found in English when they are given a subject noun and a verb (and ample time), or whether their usage of a language with general processing routines that have responded to the verb-final grammar and afford less prediction will mean that they do not engage in prediction to the same extent as English speakers, even when stimuli contain the same items for cloze probability completions. Accordingly we used the same materials as in Experiment 2, and translated them directly into Japanese. The experiment was again conducted on Qualtrics.

English speakers provided a noun for an SV_ sentence fragment and Japanese speakers provided a noun for an S_V fragment with its missing pre-verbal direct object.

Before presenting the details of this experiment, some further clarification and an analogy may be helpful about the logic that underlies it. It is becoming increasingly clear in cross-linguistic processing that speakers of structurally different languages adjust their general processing routines to the grammars of the relevant languages and they tolerate e.g. more versus fewer center embeddings and incomplete noun-verb dependencies (cf. Gildea and Temperley 2010 and Levy and Keller 2013 on German versus English), or, as we are arguing here, they develop different strategies for the efficient processing of the verb and its arguments. We believe that languages with early verbs in the clause encourage a general processing strategy in which speakers predict the different options that lie ahead and assign certain probabilities to them based on context, real-world knowledge, structural simplicity, and so on. Verb-final clauses, on the other hand, lead to different routines in which the co-occurring items are already there and what the verb needs to do is gather them in, integrate them with the relevant co-occurrence frame, check for compatibility, and so on. This same integration routine applies when a verb's argument occurs in an earlier sentence in the text and is "deleted" from the current clause being processed, which is grammatically possible in Japanese. There is no prediction of numerous unencountered possibilities here, therefore. To use an analogy, the English strategy is that of a fisherman sitting besides a muddy English pond and having to predict what kind of fish are likely to be in there and what kind of hook and bait and net might be needed to catch something. The Japanese strategy, by contrast, is that of a fisherman who can clearly see the koi swimming in the clear water before him and who then adjusts the size of his net and the length of his pole to make them compatible with gathering in the fish that he already sees and wants. If your verb is early, the processor adjusts to making regular predictions; if it is late, it adjusts to regular integrations of previous material, i.e. two very different processing routines, albeit both ultimately deriving from verbal co-occurrence frames in the mental lexicon and from the online activation of these when the verb is encountered. The issue in Experiment 3, then, is whether we can see any evidence, even when the stimuli for a cloze probability completion test are the same, for these two different general processing routines of English and Japanese.

Methods

Participants. Participants were 37 adults with a mean age of 28.33 years (*SD* = 13.79, range 18 – 81). (One participant did not report their age.) Twenty-four were female and 13 were male. The demographic information broken down by group is presented in Table 7. The groups were not significantly different in age or gender. English-speaking participants were primarily undergraduate psychology students from the University of East Anglia. Japanese speakers were students (various departments) from the University of East Anglia, and a few community-recruited adults. All confirmed being native speakers of English or native speakers of Japanese.

Table 7

Demographic characteristics for the sample

	English SVO	Japanese SOV	Significance
Sample Size	18	19	
Demographic	Mean (SD)	Mean (SD)	
Age (years)	26.89 (11.84)	29.78 (15.72)	$t(34) = 0.62$, $p = .54$
Gender (% male)	27.8%	42.1%	$t(35) = 0.90$, $p = .38$

Materials. Same as Experiment 2, except that Japanese versions were created.

Design and Procedure. The study had a 2 × 2 design. The language variable (English and Japanese) was manipulated between subjects, and constraint (high vs. low) was manipulated within subject. The dependent variable was the mean cloze probability of the responses produced by each participant. We also calculated the mean cloze probability of the modal responses. English participants were primarily recruited from the UEA psychology participation pool, and participants were automatically redirected to Qualtrics. The Japanese participants were sent a study invitation and link that directly took them to Qualtrics. Instructions were provided in the native language of the participant, and participants were instructed to type the word that they thought most naturally (or best) completed the sentence. The remainder of the procedure was the same as Experiment 2.

Data-Analytic Plan. Same as in Experiment 2.

Results

Prior to the inferential analyses, the data were checked for outliers and to ensure that they were normally distributed (i.e. that the skew was less than two times the standard error). Missing data constituted approximately .5% for English and 6.5% for Japanese. The latter was primarily due to one participant not finishing the questionnaire. Results for that participant were checked with reference to the full sample of Japanese participants, and because they were not otherwise atypical, we retained them for the sake of power.

Overall Means. We began the analysis by assessing language differences in overall cloze probability (see Table 8). Results showed that English had significantly higher cloze probability than did Japanese. This is consistent with a general tendency for more prediction in English. A significant difference was also found in mean cloze probability when considering only the modal responses (see Table 8). Again, this was consistent with greater prediction in English as compared to Japanese.

Table 8

Results of Tukey post-hoc comparisons based on high and low cloze probability items.

	English SVO	Japanese SOV	Significance
	Mean	Mean	

Overall Responses	.388	.292	z = -6.85, p < .001
Modal Responses	.621	.521	z = -6.01, p < .001
High Constraint Items			
Overall response	.563	.398	z = -9.22, p < .001
Modal response	.745	.615	z = -13.61, p < .001
Low Constraint Items			
Overall response	.216	.178	z = -2.11, p = .122
Modal response	.401	.356	z = -3.51, p < .01

High- and Low-Constraint Items. Results of a 2 × 2 linear mixed effects analysis on the overall responses showed a significant main effect of constraint (t = -7.88, p < .001). This is expected, because items were pre-selected (by Staub et al., 2015) to show this difference. The main effect of language was significant (t = -9.22, p < .001). English speakers showed higher cloze probability compared to Japanese speakers. Finally, there was a significant interaction between language and constraint (t = 6.36, p < .001). The important simple effects are presented in Table 8. There was a significant difference based on word order for the high--constraint items (English > Japanese), but not for the low-constraint items.

The results of a 2 × 2 linear mixed effects analysis on the modal responses showed a significant main effect of constraint (t = -18.20, p < .001), which again, is expected. The main effect of language was significant (t = -7.73, p < .001), in which the English speakers had higher cloze probabilities than did Japanese speakers. Finally, the interaction was also significant (t = 3.00, p < .01). The pattern of this interaction was highly similar to the results of the full set of items, with much greater differences with high constraint items. In sum, there were significant differences in the expected direction (English > Japanese) for both the high- and low-constraint items.

Discussion

This experiment showed the most consistent pattern of all three experiments. We observed clearly and significantly higher cloze probability responses for English speakers as compared to Japanese speakers, and the same finding held for all analyses (cf. overall responses – low constraint). Here we compared the production of object nouns across the two languages, based on the same subject noun phrase and verb (i.e. direct translations). In this experiment compared to Experiment 1, we hypothesized that the main difference between English and Japanese would be due to the general tendency of Japanese speakers to engage in less prediction, given their (dominant) experience with a head-final language. The results were consistent with this assumption.

There are two additional points that we think are important to raise here. First, in examining the results of this experiment, we observed two or three items that showed clear cross-cultural differences. One example was *The ailing team forfeited* _____ . The modal response in English was *the game* (cloze probability = .33) and the modal response in Japanese was *fighting spirit* (cloze probability = .37). Thus, the verb *forfeited* would seem to be more psychological in Japanese than in English. We do not view these differences as a major issue for our conclusions because (1) they affect a small number of items, approximately 10%, and (2) there is as much variability within language as between languages (see Section E, Supplementary Materials). The whole notion of cloze probability, and specifically higher cloze probability, is that the more consensus or consistency in responses, the more predictable the sentence fragment is. Our analyses average across items and participants to assess whether at a more coarse-grained level, there is more-or-less evidence for prediction. Nuanced differences are common even within a given language and are not central to the main arguments.

The second point concerns a slightly different way to analyze the data from this experiment. Up to this point, we have focused on mean cloze probabilities (averaged across participants) and the mean cloze probabilities of the modal responses (i.e. the most frequently issued response for a particular item), again averaged across participants. However, we also examined the cloze probabilities (by item) of the second most frequent responses, the third most frequent responses, the fourth most frequent responses, and a final category which contained only singleton responses (see Figure 1). What these item results nicely show is that in high-constraint items, Japanese participants produce significantly (1) fewer modal responses (i.e. fewer “first” responses) and (2) more singletons. In contrast, in low-constraint items, the only significant difference is in the number of “other singletons”.

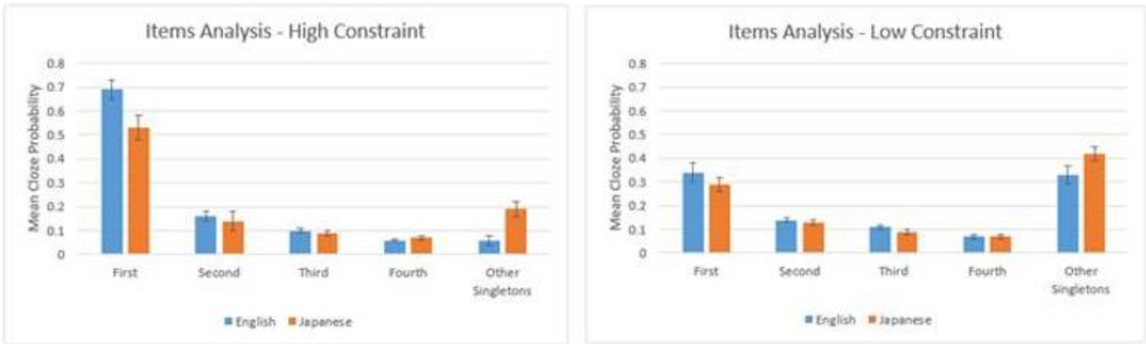

Figure 1: Mean cloze probability averaged across items. The left panel shows results from high-constraint items and the right panel shows results from low-constraint items. Error bars show the standard error of the mean.

General Discussion

The focus of the current paper has been on cross-linguistic differences in prediction, specifically with respect to verb-object relations. We hypothesized that there would be less evidence for prediction as a parsing strategy in head-final (SOV) Japanese than in SVO English since there is an asymmetry in the mental lexicon between verbs and nouns: the former always make explicit reference to co-occurring arguments, which consist regularly of nouns, whereas nouns do not generally make explicit reference to their accompanying verbs. In a verb-early language these co-occurrences will be activated early in online processing, whereas in an SOV language they will be activated late, when the verb is encountered. Nouns clearly do make some predictions for upcoming verbs in verb-final languages, and there is typological evidence for higher ratios of nouns to verbs in the lexicons of these languages, which has been argued to provide informationally richer and more predictive semantic processing online (Polinsky and Magyar 2020). But the predictive potential of this higher noun to verb ratio is, we argue, weaker and less systematic than that provided by the basic asymmetry between verbs and nouns and by their differences in linear ordering.

Our three experiments provided support for this key hypothesis that there would be more prediction in English compared to Japanese overall, despite the fact that there is clearly some prediction taking place from nouns to verbs in Japanese, as seen in the cloze probabilities for this language. Head-final languages place the verb at the end of the sentence and so activation and selection of verb co-occurrence frames occurs only after the arguments that select the relevant co-occurrence frame have already been processed. The verb then serves an integrating function gathering in this previous material, checking it for compatibility with its listed co-occurrences in the mental lexicon, and does not predict a range of possibilities that have not yet been encountered. By contrast, in English, the head comes earlier and so prediction of an upcoming object is routinely possible.

Another relevant factor in all of this is the speaker's experience actually speaking and hearing a head-initial or head-final language. If an individual is a native speaker of a head-final language (and prediction is not as robust a parsing strategy in that language), then these individuals will tend to show less prediction even when given a noun and a verb (as was shown between languages in Experiment 3). Future studies would now be valuable that examine languages with more variable orderings of S, V and O (e.g. Slavic languages like Polish) and with more variable verb-finality than Japanese, i.e. non-rigid SOV languages (Greenberg 1963), in order to see whether languages with less "dominant" SVO and SOV still show the same or similar differences with respect to prediction that we have shown here across English and Japanese and within English.

Moreover, in this study, we tested Japanese native speakers who were also bilingual in English and studying at a primarily English-speaking university (i.e. an immersion situation). At this juncture, we do not know exactly whether and how proficiency, as well as frequencies and usage patterns in a head-initial language affect prediction ability in a native head-final language, and vice versa (but see Filipović and Hawkins 2013, 2019 and Filipović, 2019 for a multi-factor model of bilingualism that proposes ways to address such questions). In the context of the current study, it would be worth testing monolingual Japanese speakers situated in Japan who do not have any knowledge of another, head-initial language. However, if Japanese bilinguals in an immersion situation show the effects we have reported in this paper, then monolinguals would almost certainly show even **greater** differences compared with native English monolinguals than we have reported here.

We believe that the results across all three experiments support our key hypothesis. In the remainder of the General Discussion, we focus particularly on the cloze probability data from the high-constraint items (both overall means and means from modal responses). The expectation was that English SV_ should show the highest cloze probabilities (i.e. the most evidence of prediction). This is what we found. The results across experiments are moderately variable (.54 – .75), but this is in part due to some clear differences between our experiments (e.g. different items and different modalities between Experiment 1 versus Experiments 2 and 3). When we examined the comparisons between English SV_ and the other three comparison groups, we observed lower mean cloze probabilities and the differences range from .09 to .16, and all resulted in large effect sizes (see Table 9 for a summary of the data).

Finally we pooled the data from Experiments 2 and 3, and ran a linear mixed effects model on them. We acknowledge issues involved in between experiment comparisons and primarily because in Experiment 2 one group of (native) English speakers produced verbs in OS_ structures and in Experiment 3 the (native) Japanese speakers produced objects in S_V structures. Thus, both groups were engaged in tasks that were atypical of their presumed dominant language experience. We combined the data (from Experiments 2 and 3) and coded one variable based on the category of the response provided (object vs. verb), and another based on language (English vs. Japanese). If there is an effect of response category, it would be driven by the English speakers (Experiment 2 OS_ order), and if there is an effect of language, it would be driven by Japanese speakers in which they provided an object in response to a subject and a verb (Experiment 3 S_V order). This is actually a strong test of our key hypothesis because (1) an effect of language would be driven by a situation in which Japanese speakers produced a direct object (based on a subject and a verb, which should lead to greater prediction), and (2) an effect of response category would be driven by English speakers, who produced a verb (based on a direct object and subject, which should lead to less prediction). We ran a linear mixed effects model in R (version 3.6.1), similar to the previous analyses. The results are shown in Table 10. Both variables are significant. However, the effect of language was larger than the effect of response category. Again, this result is consistent with our key hypothesis.

Table 9

Summary of means and effects sizes (Cohen's D) for high-constraint items across all three experiments

	English SV_	Japanese SO_	English OS_	Japanese S_V	Pred. Effect	Effect Size
All Responses High-Constraint						
Experiment 1	.68	.57			>.11	.92
Experiment 2	.54		.41		>.13	1.95
Experiment 3	.56			.40	>.16	2.74
Modal Responses High-Constraint						
Experiment 2	.73		.65		>.08	2.15
Experiment 3	.75			.62	>.13	3.74

Table 10

Results of the Linear Mixed Effects Model for 118 Participants. Random effects of subject

Fixed Effects	Estimate	Std. Error	*t*-value	*p*-value
Intercept	0.38	.006	64.69	$p < .001$
Response Category (Object)	-0.02	.010	-2.33	$p < .05$
Language (Japanese)	-0.08	.012	-6.92	$p < .001$

Before addressing further theoretical implications and conclusions from this study, there are some alternative explanations that can be excluded based on prior research. Staub et al. (2015) demonstrated that the relationship between cloze probability and reaction time was not due to (1) the frequency of the words produced or (2) semantic associations between words in the sentence and the words produced. Instead, the only variable that produced a significant effect in terms of lower reaction times and higher cloze probability responses was the length of the sentence fragment. In our study, this is not a concern because the sentences were direct translations. Therefore, if the main patterns of results are not due to low-level lexical factors or semantic associations, then as Staub et al. concluded, predictability in sentence contexts is derived from the combinatorial nature of incremental parsing, combined with “background” real-world knowledge.

Limitations

We see three main limitations with the current study. First, we were not able to collect reaction time data for Experiments 2 and 3, and thus, this is an important next step in this line of research. At the same time, it is important to note that use of “online” studies via the internet is a potential avenue forward, and greatly reduces the amount of work required for assessing prediction in monolingual speakers across the world, particularly if reaction time data could be also collected. Second, our Japanese speakers were all proficient in English as a second language, and it is currently unknown whether proficiency in a SVO language might impact prediction in an SOV language. This issue could be addressed by testing monolingual speakers of Japanese and comparing them to bilinguals. Third, our Japanese materials were translations from English, and therefore, there may be some unknown (lexical or cultural) factors in which the translation is not equivalent. We think that these potential issues are on the minor side, and we did carefully examine responses on a by-item level for anomalies. Finally, in the current study, we made no attempt to manipulate discourse context. The goal of this initial set of studies was to assess whether prediction operates differently in these two languages in decontextualized situations. It is an open question whether there are or are not differences in prediction between English and Japanese based on contextual factors.

Theoretical Implications and Conclusions

We have argued that SVO languages, like English, and SOV languages, like Japanese, have both advantages and disadvantages for processing. SVO permits more prediction of possible and likely co-occurrences at the verb, but there is then an additional working memory cost in online processing before possible co-occurrences from the mental lexicon are eliminated by post-verbal material, and there are also mis-assignments (and garden paths) that can result from these possibilities before they are eliminated. An SOV language like Japanese avoids these by integrating its verbs with already encountered arguments and by immediately removing unintended co-occurrence possibilities. But in the process there is an online delay in assigning arguments to their verbs, and there is less actual online prediction.[8]

These competing efficiencies in processing form part of the explanation that Hawkins (2004, 2014) has given for why SVO and SOV languages are both frequent across the globe (see fn.4). This also follows from his earlier calculations (see Hawkins 1990, 1994) that SVO and SOV have equally minimal and efficient domains for the processing of basic phrase structure and head ordering.

[8] It is interesting to recall in this connection the typological observation made by Ueno and Polinsky (2009) that SOV languages have an intransitive bias and permit fewer and shorter verbal co-occurrences preceding the verb than are found following it in VO languages.

Our argument, based on typological distributions for different verb positions (see again fn.4) and on the experimental results reported here, is that prediction is not necessarily the key factor that facilitates ease of processing in language. We have presented arguments that predictions add to working memory and may be incorrect, and that the world's languages show no evidence of favouring basic word orders that would optimize prediction by positioning verbs and other heads first or even early in their respective phrases, so bringing grammars more into line with the online probabilities and shorter reaction times seen in this study of English performance compared with Japanese. But prediction is certainly one of the factors that makes processing easier, and in particular it appears to be more relevant for the "look ahead" SVO type, like English, and less relevant for "look back" SOV languages like Japanese. More generally we advocate a model of competing motivations in both processing and grammars, along the lines of MacWhinney et al. (2014), within which we see predictive processing as one of several cooperating and competing forces that determine both ease of processing, as evidenced by experimental and corpus results, and also preferences for grammatical conventions and their cross-linguistic distributions (see the Performance-Grammar Correspondence Hypothesis of Hawkins 2004, 2014 and fn.3 above). The purpose of this paper was to show the greater role of prediction in an SVO language like English compared to Japanese.

References

Altmann, Gerry T. & Mark Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30. 191–238.

Amberber, Mengistu, Brett Baker & Mark Harvey (eds.). 2010. *Complex predicates: Cross-linguistic perspectives on event structure*. Cambridge: Cambridge University Press.

Arcuri, Silvia M., Sophia Rabe-Hesketh, Robin G. Morris & Phillip K. McGuire. 2001. Regional variation of cloze probabilities for sentence contexts. *Behavior Research Methods, Instruments & Computers* 33. 80–82.

Barr, D.J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278.

Bates, Douglas M., Martin Maechler, Ben Bolker & Steve Walker. 2018. Lme4: Linear Mixed-Effects Models Using S4 Classes; R Package Version 1.1-18-1.

Bornkessel, Ina. 2002. The Argument Dependency Model: A neurocognitive approach to incremental interpretation. Leipzig: Max Planck Institute of Cognitive Neuroscience PhD thesis.

Bornkessel, Ina & Matthias Shlesewsky 2006. The extended argument dependency model: Neurocognitive approach to sentence comprehension across languages. *Psychological Review* 113(4). 787-821.

Bornkessel, Ina, Brian McElree, Matthias Schlesewsky & Angela D. Friederici 2004. Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language* 51. 495-522.

Brown, Keith & Jim Miller. 1996. *Concise Encyclopedia of Syntactic Theories*. Oxford: Elsevier Science.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Clancy, Patricia M. (1982). Written and spoken style in Japanese narratives. In Deborah Tannen (ed.), *Spoken and written Language: Exploring orality and literacy*, 55-76. Norwood: Ablex Publishing Corporation.

Corbett, Greville G., Norman M. Fraser & Scott McGlashan (eds.). 1993. *Heads in grammatical theory.* Cambridge: Cambridge University Press.

Dryer, Matthew S. 1992.. The Greenbergian word order correlations. *Language* 68. 81-138.

Dryer, Matthew S. 2013a. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at: http://wals.info/chapter/81, accessed on 2018-12-30)

Dryer, Matthew S. 2013b. Determining dominant word order. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Plank Institute for Evolutionary Anthropology.
(Available online at: <http://wals.info/chapter/s6>Accessed on 2015-03-17)

Ferreira, Fernanda & Chantavari, Suphasiree. 2018. Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science* 27. 443-448.

Filipović, Luna. 2019. *Bilingualism in Action: Theory and Practice*. Cambridge: Cambridge University Press.

Filipović, Luna & John A. Hawkins. 2013. Multiple factors in second language acquisition: The CASP model. *Linguistics* 51(1). 145-176.

Filipović, Luna & John A. Hawkins. 2019. The Complex Adaptive System Principles model for bilingualism: Language interactions within and across bilingual minds. *International Journal of Bilingualism* 23(6). 1223-1248.

Frank, Stefan L., Leun J. Otten, Giulia Galli & Gabriella Vigliocco. 2015. The ERP response to the amount of information conveyed by words in sentences. *Brain and Language* 140, 1-11.

Frazier, Lyn. 1978. *On comprehending sentences: Syntactic parsing strategies.* University of Connecticut Indiana Linguistic Club PhD thesis.

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

Gentner, Dedre. 1981. Some interesting differences between verbs and nouns. *Cognition and Brain Theory 4*. 161-178.

Gibson, Edward. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68. 1-76.

Gibson, Edward. 2000. The dependency locality theory: a distance-based theory of linguistic complexity. In Alec P. Marantz, Yasushi Miyashita and Wayne O'Neil (eds.) *Image, language and brain*, 95-112. Cambridge, MA: MIT Press.

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

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73-113. Cambridge, MA: MIT Press.

Grimshaw, Jane. 1990. *Argument structure.* Cambridge, MA: MIT Press.

Hale, John T. 2001. A probabilistic early parser as a psycholinguistic model. In Proceedings of NAACL (Vol. 2), 159–166. Morristown NJ, USA: Association for Computational Linguistics.

Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.

Hawkins, John A. 1986. *A comparative typology of English and German: Unifying the contrasts*. Austin: University of Texas Press & London: Routledge.

Hawkins, John A. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21. 223-261.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, John A. 1995. Argument-predicate structure in grammar and performance: A comparison of English and German. In Irmengard Rauch & Gerald F. Carr (eds.), *Insights in Germanic linguistics 1*, 127-44. Berlin: Mouton de Gruyter.

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

Hawkins, J. A. 2014. *Cross-linguistic Variation and Efficiency*. Oxford: Oxford University Press.

Hawkins, J.A. 2022. Have grammars been shaped by working memory and if so, how? In John W. Schwieter & Edward Wen (eds.), *The Cambridge handbook of working memory and language*, 127-144. Cambridge: Cambridge University Press.

Hawkins, John A. & Anne Cutler. 1988. Psycholinguistic factors in morphological asymmetry. In John A. Hawkins (ed.), *Explaining language universals,* 280-317. Oxford: Basil Blackwell Ltd.

Hays, David G. 1964. Dependency theory: A formalism and some observations. *Language 40*. 511-525.

Hörberg, Thomas 2016. Probabilistic and prominence-driven incremental argument interpretation in Swedish. Stockholm University PhD thesis.

Hudson, Richard A. 1984. *Word grammar*. Oxford: Blackwell.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar.* Cambridge, MA: The MIT Press.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59. 434-446.

Kamide, Yuki & Don C. Mitchell. 1999. Incremental pre-head attachment in Japanese parsing. *Language and Cognitive Processes* 14. 631-662.

Keenan, Edward L. 1971. Two kinds of presupposition in natural language. In Charles J. Fillmore & D. Terence Langendoen (eds.), *Studies in linguistic semantics*, 44-52. New York: Holt, Rinehart & Winston.

Konieczny, Lars. 1996. Human sentence processing: A semantics-oriented parsing approach. University of Freiburg, Institut für Informatik und Gesellschaft PhD thesis.

Konieczny, Lars. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29. 627-45.

Konieczny, Lars & Philpp Döring. 2003. *Anticipation of clause-final heads: Evidence from eye tracking and SRNs*. In Proceedings of the ICCS/ASCS-2003 joint international conference on cognitive science. Sydney, Australia 13–17 July.

Kuno, Susumu. 1973. *The Structure of the Japanese language*. Cambridge, MA: MIT Press.

Kuno, Susumu. 1978. Japanese: A characteristic OV language. In Winfred P. Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*, 57-137Austin and London: University of Texas Press.

Kuperberg, Gina R. & T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition & Neuroscience* 31. 32-59.

Lehmann, Winfred P. 1978. English: A characteristic SVO language. In Winfred P. Lehmann (ed.), *Syntactic typology. Studies in the phenomenology of language*, 169-222. Austin and London: University of Texas Press

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–1177.

Levy, Roger P. 2013. Memory and surprisal in human sentence comprehension. In Roger P. G. van Gompel (ed.), *Sentence processing*, 78-114. Hove: Psychology Press.

Levy, Roger & Frank Keller. 2013. Expectation and locality effects in German verb-final structures. *Journal of Memory and Language* 68. 199-222.

MacWhinney, Brian, Andrej L. Malchukov & Edith Moravcsik (eds.). 2014. *Competing motivations in grammar and usage*. Oxford: Oxford University Press.

Mazuka, Reiko & Kenji Itoh. 1995. Can Japanese speakers be lead down the garden path? In Reiko Mazuka & Noriko Nagai (eds), *Japanese sentence processing*, 295-329. Hillsdale, NJ: Lawrence Erlbaum Associates.

Melinger, Alissa, Thomas Pechmann & Sandra Pappert. 2009. Case in language production. In Andrej L. Malchukov & Andrew Spencer (eds.), *Oxford handbook of case*, 384-401. Oxford: Oxford University Press.

Müller-Gotama, Franz. 1994. *Grammatical relations: A cross-linguistic perspective on their syntax and semantics.* Berlin: de Gruyter.

Ono, Tsuyoshi. 2006. Postpredicate elements in Japanese conversation: Nonmodularity and panchrony. In Timothy J. Vance & Kimberly Jones (eds.), *Japanese/Korean Linguistics Vol.14,* 381-391. Stanford: CSLI Publications & SLA.

Park, Y. Albert & Roger Levy. 2009. Minimal-length linearizations for mildly context-sensitive dependency trees. *Proceedings of the 10th Annual Meeting of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT) conference Seattle*, *Washington*, 335–343.

Polinsky, Maria & Lilla Magyar. 2020. Headedness and the lexicon: The case of verb-to-noun ratios. *Languages* 5(1). 9. https://doi.org/10.3390/languages5010009

Pritchett, Bradley L. 1991. Head position and parsing ambiguity. *Journal of Psycholinguistic Research* 20. 251–270.

Pritchett, Bradley L. 1992. *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

R Core Team (2018). R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing; R Core Team: Vienna, Austria, Available online: https://www.R-project.org/ (accessed on 12 December 2019).

Shaoul, Cyrus, R. Harald Baayen & Chris Westbury, C. 2014. N-gram probability effects in a cloze task. *The Mental Lexicon* 9(3). 437–472.

Siewierska, Anna. 1993. Syntactic weight vs information structure and word order variation in Polish. *Journal of Linguistics* 29(2). 233-265.

Smith, Nathaniel J. & Roger Levy. 2011. Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. *Proceedings of the Annual Meeting of the Cognitive Science Society 33.* 1637–1642.

Smith, Nathaniel J. & Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128. 302–319.

Staub, Adrian, Margaret Grant, Lori Astheimer & Andrew Cohen. 2015. The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language* 82. 1-17.

Strunk, Jan, Bathasar Bickel, Swintha Danielson, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Taras Zakharko & Frank Seifert. 2015. Noun-to-verb ratio and word order. Paper presented at the “Diversity Linguistics: Retrospect and Prospect” conference, Max Planck Instituter for Evolutionary Anthropology Leipzig, 1-3 May.

Trueswell, John C. & Michael K. Tanenhaus. 1994a. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33. 285-318.

Trueswell, John C. & Michael K. Tanenhaus. 1994b. Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. In Charles Clifton Jr., Lyn Frazier & Keith Rayner (eds.), *Perspectives on sentence processing*, 155-179. Hillsdale, NJ: Erlbaum.

Ueno, Mieko & Maria Polinsky. 2009. Does headedness affect processing? A new look at the VO-OV contrast. *Journal of Linguistics* 45(3). 675-710.

Vasishth, Shravan, Katja Suckow, Richard L. Lewis & Sabine Kern. 2010. Short-term forgetting in sentence comprehension: Cross-linguistic evidence from verb-final structures. *Language and Cognitive Processes* 25. 533–567.

Whitman, John. 2000. Right dislocation in English and Japanese. In Ken Ichi Takami, Akio Kamio & John Whitman (eds.), *Syntactic and functional explanations: In honor of Susumu Kuno*, 445-470. Tokyo: Kurosio Publishers.

Yamashita, Hiroko. 2002. Scrambled sentences in Japanese: Linguistic properties and motivation for production. *Text* 22. 597-633.