Dependencies and Hierarchical Structure in Sentence Processing

Peter Baumann (baumann@u.northwestern.edu)
Department of Linguistics
Northwestern University, Evanston, IL, USA

Abstract

In memory-based models of human sentence processing it is assumed that the completion of a dependency between a syntactic head and its dependents is a major source of processing difficulty in non-ambiguous sentences, and that this *integration cost* is a function of the distance between the two elements. However, it remains open how to measure the distance between two dependent elements. While many current models employ a linear distance measure, we instead propose to measure the distance between a head and its dependents as the path in the phrase structure tree connecting the two elements. We evaluate this *structural distance* as a measure of dependency integration and show that it is a better predictor of human reading times than other measures. Moreover, we find that evaluated on reading data from naturally occurring texts, dependency integration is not actually a cost, as higher dependency integration distances led to lower reading times.

Keywords: sentence processing; syntax; reading; eye-tracking; modeling; memory

Introduction

How do we understand language and what are the cognitive mechanisms reflected in measures of human sentence comprehension? Many attempts to answer this big question can be subsumed under two broad categories: memory-based models and experience-based models. One particular aspect, in which these two modeling approaches differ, is the role they assign to hierarchical syntactic structure: while experience-based models of syntactic expectations are often built on phrase-structure grammars modeling hierarchical sentence structure, memory-based models tend to be agnostic about it, as they are usually formulated in terms of non-hierarchical dependency structures.

Within the latter group of models it is a common assumption that the completion of a dependency between a syntactic head and its dependents (such as e.g. a verb and its argument) is a major source of processing difficulty in non-ambiguous sentences (Gibson, 1998), and that this *integration cost* is – at least in part – a function of the distance between the two elements.

However, specific models differ with regard to the question of how to measure the distance between two dependent linguistic units: some models emphasize the role of working memory decay and thus assume the distance to be measured in time (Wanner & Maratsos, 1978; King & Just, 1991), others, like the Dependency Locality Theory (DLT; Gibson, 1998, 2000) emphasize the general capacity limitations of working memory and measure the distance in terms of intervening discourse referents, which are supposed to occupy memory slots. A third possible alternative, which has not been implemented in memory-based models, is *structural distance* within the phrase structure (see e.g. O'Grady, 1997), i.e. the length of the path in the phrase-structure tree to 'travel' from one element to the other.

In this paper, we evaluate these three measures of distance between linguistic dependents. We will assume that the number of words intervening between two dependent elements is a good approximation of the time of memory decay so that the three distance measures can be summarizes as follows:

- LINEAR DISTANCE: number of words between two dependents
- DLT DISTANCE: number of discourse referents[1] between two dependents
- STRUCTURAL DISTANCE: number of syntactic nodes crossed in the syntactic phrase-structure tree between two dependents

Following a recent trend in psycholinguistics (e.g. Pynte, New, & Kennedy, 2008; Demberg & Keller, 2008), we do not focus on specific constructions or sentence types to evaluate the three distance measures, but instead employ a regression analysis on an 'eye-tracking corpus', i.e. eye-tracking data of people reading naturally occurring texts, to determine if dependency integration cost, when determined by one of the three distance measures is a significant predictor of reading times.

We will show that only structural distance between two dependent elements is a significant predictor of reading times, and that this effect can only be found on verbs, but not on nouns. More importantly, our analyses show that a *higher* distance between two dependent elements leads to *lower* reading times. This result is the exact opposite of what most memory-based models predict, but it is not unprecedented: in a self-paced reading experiment conducted in German, Konieczny (2000) observed that verbs were read faster when the number of intervening words between them and their arguments was higher. Similar *anti-locality effects* have since been observed in controlled experiments in Japanese (Nakatani & Gibson, 2008) and Hindi (Vasishth & Lewis, 2006). And in an analysis very similar to ours, Demberg and Keller (2008) found a similar negative effect of DLT distance, i.e. higher distance led to lower reading times.

Our results can be interpreted in different ways: one may argue that that dependency integration can lead to integration costs, but that dependencies may also help to predict and thus facilitate the upcoming head, and that the latter process is more common than the former in everyday language comprehension. A somewhat weaker conclusion is that current memory-based models, such as DLT, may be too narrow in

[1] We follow (Gibson, 1998) and assume that nouns and verbs introduce discourse referents.

scope, in that they can only account for effects that arise from a very limited set of dependencies.

Memory-based Models and Relative Clauses

In sentence processing research it is well-established that object-extracted relative clauses (ORC) are harder to process than comparable subject-extracted relative clauses (SRC), not only in English (King & Just, 1991), but in may other languages.

Accounting for this processing difference has been one of the major motivations for memory-based models (King & Just, 1991; Gibson, 1998, 2000). One particular memory-based processing model, the Dependency Locality Theory (DLT; Gibson, 1998, 2000) explains SRC/ORC processing difference in terms of an *integration cost* occurring at the heads of linguistic dependencies: one component of the integration cost is a distance-based cost, which is monotonely increasing in the number of discourse referents intervening between the head and its dependents, e.g. a verb and its arguments.[2]

In the case of English relative clauses (1), integration cost of DLT makes the right predictions for processing differences on the embedded verb *attacked*.

(1) a. The journalist$_i$ [who$_i$ attacked the senator] is famous.
b. The journalist$_i$ [who$_i$ the senator attacked] is famous.

In an SRC (1a), the embedded verb needs to integrate its preceding subject, the relative pronoun *who*, which is co-indexed with the noun phrase *the journalist*[3]. Since no discourse referent occurs between the subject and the verb, there is no integration cost on the verb in SRCs.

In an ORC (1b), on the other hand, two integrations need to take place at the embedded verb *attacked*: the embedded subject *the senator* needs to be integrated with the verb. Since no discourse referent occurs between subject and verb, there is no cost associated with this integration. Second, the object *the journalist* (or the co-indexed relative pronoun *who*) needs to be integrated with the verb. This integration incurs a cost of 1, as there is the discourse referent *the senator* intervening between the object and the verb. So overall, there is an integration cost of 1 on the embedded verb in an ORC, which is higher than the integration cost of 0 in an SRC.

A qualitatively equivalent prediction is obtained when integration cost is not measured in terms of DLT distance, but in terms of linear distance or structural distance: for linear distance, one can see that in an SRC (1a) the relative pronoun *who* and the verb are adjacent, while in an ORC (1b) they are separated by the intervening subject. The predictions of structural distance are illustrated in the phrase-structure trees in Figure 1: for the SRC (top) the nodes crossed between the verb and the head noun *the journalist* are VP, S' and NP, while for ORC (bottom) there is an extra S node, which is crossed in addition to VP, S' and NP. This yields a greater structural distance for the ORC than for the SRC.

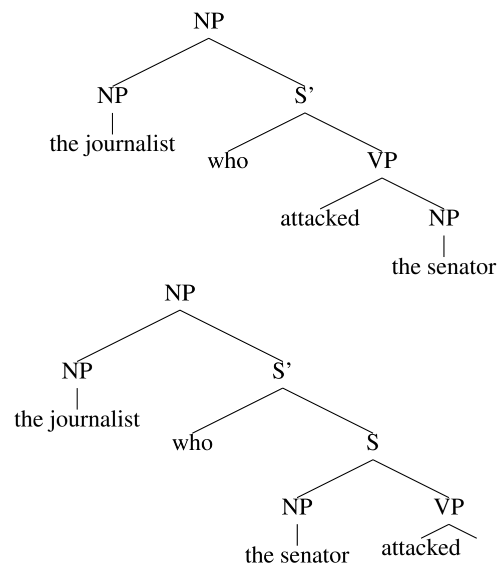

Figure 1: Phrase-structure tree of English relative clauses: SRC (top) and ORC (bottom)

Korean Relative Clauses

While DLT makes the correct predictions for relative clauses in English, this is not the case for all other languages. One particularly interesting case is Korean, a head-final language with prenominal relative clauses. Despite being significantly different from English, all available empirical evidence from self-paced reading and eye-tracking experiments indicates that in Korean, like in English, ORCs are more difficult to process than SRCs (Kwon, Polinsky, & Kluender, 2006; Kwon, Lee, Gordon, Kluender, & Polinsky, 2010).

(2) a. [uywon-ul kongkyekha-n] enlonin-i ...
[senator-ACC attack-ADN] journalist-NOM ...
'The journalist who attacked the senator is famous'
b. [uywon-i kongkyekha-n] enlonin-i ...
[senator-NOM attack-ADN] journalist-NOM ...
'The journalist who attacked the senator is famous'

However, integration costs derived from linear and DLT distance predict that there should be no processing difference between Korean SRCs and ORCs (2). In both the SRC (2a) and the ORC (2b), only one integration occurs when processing the embedded verb *kongkyekha-n*: in an SRC, it is the

[2] In DLT, there is a second component to integration cost for the introduction of discourse referents, but since this component makes the same contribution in all cases considered here and we are mainly interested in the contribution of distance-based costs, we will ignore this component.

[3] In this presentation, we adopt a 'gap-free' or direct association based processing model (e.g. Pickering & Barry, 1991), in which extracted elements, such as relative or interrogative pronouns, directly associate with their subcategorizer, i.e. the verb, without positing any 'gaps'. This choice was made purely for the purpose of an accessible presentation. All claims hold (and could be made even stronger, as in the case of Korean below) when assuming a more traditional filler-gap processing mechanism.

embedded object *uywon-ul* that can be integrated and in an ORC the embedded subject *uywon-i*. Crucially, in both cases there is no material intervening between the embedded verb and the integrated element and so the integration costs are the same for SRCs and ORCs.

Integration cost measured in terms of structural distance, on the other hand, predicts the observed pattern, as illustrated in Figure 2: like in English, the nodes crossed between the verb and the head noun *the journalist* in an SRC (top) are VP, S' and NP, while for the ORC (bottom) there is an extra S node, which is crossed in addition to VP, S' and NP. This yields the correct prediction that ORCs should be harder to process than SRCs.

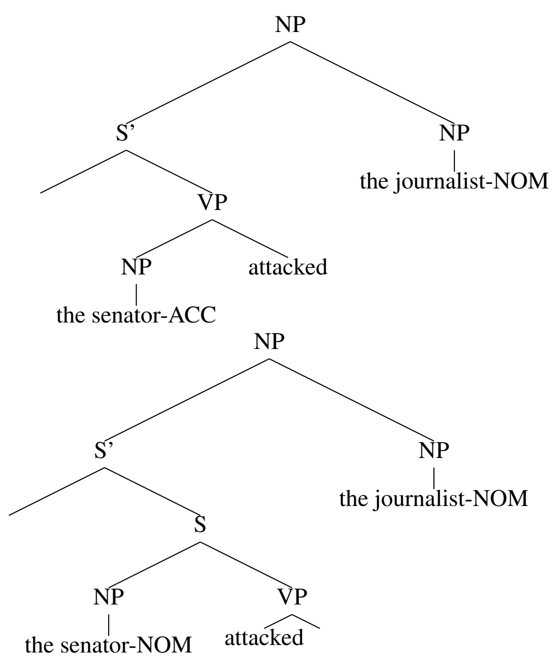

Figure 2: Phrase-structure tree of Korean relative clauses: SRC (top) and ORC (bottom)

Anti-Locality Effects

Another challenge for DLT and memory-based models in general, comes from *anti-locality effects*, which were first reported from German by Konieczny (2000) and Konieczny and Döring (2003): in self-paced reading and eye-tracking experiments, Konieczny observed that – contrary to the predictions of DLT – verbs were read faster when the number of intervening words between them and their arguments was higher. Similar effects have since been observed in controlled experiments in Japanese (Nakatani & Gibson, 2008) and Hindi (Vasishth & Lewis, 2006).

Related Work

In recent years, it has become standard to evaluate computational models of language processing on 'eye-tracking corpora': Demberg and Keller (2008) used the Dundee Reading Corpus (Kennedy & Pynte, 2004) to evaluate computational models of sentence processing. In particular, they compared the experience-based model of surprisal (Hale, 2001) and the memory-based integration cost of Dependency Locality Theory (Gibson, 1998, 2000). For surprisal, the authors found that it is a good predictor of human reading times in naturalistic texts. For integration cost, on the other hand, the situation was a bit more complicated: when evaluated over the whole data set, integration cost was a statistically significant predictor of reading times (Demberg & Keller, 2008, Table 1, p. 199). However, the direction of the effect of integration cost was the opposite of what was predicted by DLT: a *higher* distance between two dependent elements led to *lower* reading times. While Demberg and Keller (2008) interpret their findings as a failure of DLT to serve as a "broad-coverage theory of syntactic complexity", their finding of a negative effect of integration cost on reading times (i.e. higher integration costs lead to lower reading times) has since been replicated on the same data set, but with different parsers (Baumann, 2012; Schuler & van Schijndel, 2014).

Analyses

In the following two statistical analyses we evaluate the relative importance of dependency integration based on three measures of distance as predictors of reading times in an on 'eye-tracking corpus', i.e. eye-tracking data of people reading naturally occurring texts (cf. Pynte et al., 2008; Demberg & Keller, 2008). In Analysis 1, we evaluate the distance measures across syntactic heads of all word classes, while in Analysis 2, we differentiate between nominal and verbal heads, as some memory-based models predict dependency integration only for verbal heads (e.g. Gibson, 2000).

Data and Dependent Variable

We used the English portion of the Dundee Corpus (Kennedy & Pynte, 2004), an eye-tracking corpus based on texts from the British newspaper *The Independent*. The corpus consists of 51,502 word tokens from 20 texts, which were read by ten English native speakers, while their eye-movements were recorded using a Dr. Boise eye-tracker. Following common practice (e.g. Pynte et al., 2008; Demberg & Keller, 2008) we preprocessed the data and removed cases (i.e., word-participant pairs), in which the word was not fixated, was presented as the first or last on a line, was attached to punctuation, contained more than one capital letter (likely to be an acronym) or any non-alphabetic symbol.

Since we are interested in how to measure distance in determining the cost of dependency integration on syntactic heads, we restricted our data set to syntactic heads, which are preceded by at least one dependents. In all other cases, i.e. on words, which are no syntactic heads, or heads which precede their dependents, memory-based models of sentence processing do not predict any dependency integration cost. While such a restriction of the data set was not performed by Demberg and Keller (2008), it seems mandatory from a statistical perspective, as a high number of data points with an irrelevant value in the predictor of interest can have adverse effects on the fit of a linear regression model.

The sequence of fixations obtained from eye-tracking experiments can be analyzed by calculating a range of eye-tracking measures. We chose first-pass reading times as our dependent variable, as it is supposed to be indicative of early syntactic processing. The first-pass reading time on a given word is the sum of all eye fixations on that word in the first pass reading, i.e. before leaving the word either to the right or to the left.

Procedure

The three measures of distance and several other well-established control variables, which are known to have an influence on reading times, were calculated for each word and annotated to the reading time data of the Dundee corpus. On this data set we performed regression analyses using linear mixed-effects regressions with PARTICIPANT, WORD and TEXT NUMBER as random effects, as a generalization of the common by-subject and by-item analyses. All models were fit in R using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2013).

Since our goal is to determine the relative importance of the three different measures of distance, we first fitted a baseline model with all control variables to the first-pass reading times in our data set. We then calculated three new regression models for our three measures of distance, which included all baseline predictors and one of our three measures. These three models were then compared to the baseline model through a log-likelihood test, which follows a χ^2-distribution. If one of the models was a significantly better fit to the reading time data (as determined by the log-likelihood test), we conclude that the predictor added to the model is a significant predictor of reading times (cf. Gelman & Hill, 2007).

Control Variables

All regression models included the following control variables, which are known to have an influence on reading times (cf. Demberg & Keller, 2008): the number of characters per word (WORD LENGTH), the position of word in a sentence (WORD POSITION), an indicator variable whether there was no fixation on the previous (PREV NOFIX) the next (NEXT NOFIX) word, the frequency of the word (WORD FREQ), the frequency of the previous word (FREQ PREV WORD), the forward transitional probability (FORW TRANS PROB), i.e. the bigram probability $P(w_i|w_{i-1})$, and the backward transitional probability (BACKW TRANS PROB), i.e. the bigram probability $P(w_i|w_{i+1})$.

All frequencies and transitional probabilities were obtained by fitting a unigram or bigram model with modified Kneser-Ney smoothing (Chen & Goodman, 1998) to the British National Corpus (100 million words) using the SRILM toolkit (Stolcke, 2002). All continuous variables were centered and scaled to two standard deviations to minimize collinearity. In addition, all frequencies and transitional probabilities were log-transformed before scaling.

In addition to including these control variables as single predictors, we also included all binary interaction terms between them, which improved the model fit in a log-likelihood test. The full set of predictors used in all models is listed in Table 1.

Measures of Dependency Distances

The three measures of dependency distance were calculated based on dependency relations and syntactic tree structures obtained from parsing the Dundee Corpus with the Stanford Parser (de Marneffe & Manning, 2006; Klein & Manning, 2003). For each dependency relation, the distance between a head and its dependent was calculated as follows:

- LINEAR DISTANCE: number of words between dependent and head
- DLT DISTANCE: number of nouns and verbs between dependent and head
- STRUCTURAL DISTANCE: number of non-terminal nodes crossed when traversing the syntactic tree structure from dependent to head

Since the dependency distances are claimed to cause integration costs at the heads of a dependency (Gibson, 1998, 2000), we annotated all syntactic heads with their respective dependency distances. If a head had more than one dependency, its distance measure is the sum of distances of the individual dependencies.

Analysis 1

In our first analysis, we evaluate the three distance measures across syntactic heads of all word classes by fitting one model for each distance measure and comparing it to the baseline model via log-likelihood tests.

Results

The model comparisons via log-likelihood tests showed that integration cost as measured by STRUCTURAL DISTANCE is a significant predictor of readings times, as adding it to the regression model significantly improved the model fit over the baseline ($\chi^2 = 10.49, p < .01$). Integration cost based on the other two distance measures, however, turned out not to be significant predictors of reading times, as neither of them led to a significantly better model fit (LINEAR DISTANCE: $\chi^2 = 0.76$, n.s.; DLT DISTANCE: $\chi^2 = 1.70$, n.s.).

The coefficients and standard errors of the model with structural distance are listed in Table 1. It can be seen that STRUCTURAL DISTANCE is negatively correlated with reading times ($\beta = -3.17, t = -3.24$), i.e. higher dependency distances lead to lower reading times.

Discussion

Our results show that dependency integration as measured by structural distance is a significant predictor of reading times, while the other two distance measures do not make dependency integration a significant predictor of reading times. This result is in contrast to the one obtained by Demberg and Keller (2008), who found DLT's dependency integration cost

Table 1: Coefficients and standard errors of the regression model for structural distance

Predictor	Coef.	Std.Error
(INTERCEPT)	213.36	8.45
WORD LENGTH	39.74	1.36
FREQ PREV WORD	-28.82	1.36
PREV NOFIX	28.17	1.41
NEXT NOFIX	10.15	1.24
WORD POSITION	-6.23	0.92
FORW TRANS PROB	-11.33	1.78
WORD FREQ	-9.20	1.66
BACKW TRANS PROB	-1.97	1.19
WORD LENGTH : WORD FREQ	-28.89	2.55
FREQ PREV WORD : PREV NOFIX	16.22	1.93
NEXT NOFIX : PREV NOFIX	9.37	1.72
WORD FREQ : FREQ PREV WORD	8.31	1.94
WORD POSITION : FREQ PREV WORD	-6.47	1.79
FORW TRANS PROB : NEXT NOFIX	-6.17	1.74
WORD LENGTH : BACKW TRANS PROB	6.60	2.38
WORD POSITION : WORD FREQ	-4.39	2.00
WORD POSITION : WORD LENGTH	-5.02	2.11
STRUCTURAL DISTANCE	-3.17	0.98

to be a significant predictor of reading times in the Dundee Corpus. One explanation for the two differing results may be that we restricted our data set to include only syntactic heads with non-zero distance values, i.e. syntactic heads preceded by at least one of their dependents, and did not include the extra component of DLT's integration cost for the introduction of the head in our calculations. In addition, we obtained our dependency parses of the Dundee Corpus from the Stanford Parser, which outputs both phrase-structure and dependency parses, while Demberg and Keller used a dedicated dependency parser.

More importantly, however, like Demberg and Keller (2008) we also obtained a negative effect of dependency integration on reading times, i.e. i.e. higher dependency distances lead to lower reading times. This results is the exact opposite of what most memory-based models, in particular DLT and its integration cost, predict, but given that the direction of the effect is stable across different evaluations based on different parsers, we are willing to accept that dependency integration is negatively correlated with reading times of natural texts.

Analysis 2

Our first analysis has shown that dependency integration is a significant predictor of reading times but only when measured by structural distance. Since some memory-based models predict dependency integration only for verbal heads (e.g. Gibson, 2000), we now we differentiate between nominal and verbal heads and repeat the previous analysis for nouns and verbs separately.

Data

We constructed two subsets of the data set used in Analysis 1: one subset containing only syntactic heads that are verbs and one subset containing only heads that are nouns. The decision of the whether a word is a noun or a verb was based on the part-of-speech tag of each word obtained from the Stanford parser.

Results Nouns

For the subset of nominal heads, none of the three distance measures made dependency integration a significant predictor of reading times (STRUCTURAL DISTANCE: $\chi^2 = 2.06$, n.s.; LINEAR DISTANCE: $\chi^2 = 0.0001$, n.s.; DLT DISTANCE: $\chi^2 = 0.70$, n.s.).

Results Verbs

For the subset of verbal heads, the results are very similar to the full data set: the model comparisons via log-likelihood tests showed that dependency integration as measured by STRUCTURAL DISTANCE is a significant predictor of readings times on verbs ($\chi^2 = 7.62, p < .01$), while dependency integration based on the other two distance measures turned out not to be significant predictors of reading times on verbs (LINEAR DISTANCE: $\chi^2 = 0.01$, n.s.; DLT DISTANCE: $\chi^2 = 1.57$, n.s.). The effect of STRUCTURAL DISTANCE on reading times was again negative ($\beta = -4.03, t = -2.76$), i.e. higher dependency distances lead to lower reading times.

Discussion

The results of Analysis 2 are a refinement of the results of Analysis 1, as they show that only for verbs dependency integration as measured by structural distance is a significant predictor of reading times, while for nouns no such effect could be found. While our results are in line with the assumptions of some memory-based models (e.g. Gibson, 2000), which predict dependency integration only for verbal heads, the direction is again the exact opposite of what these models predict. However, the fact that the results on verbs were similar to the ones on the whole data set, is in line with the (psycho)linguistic intuition that dependency integration should mainly affect verbs.

Conclusion

Dependency integration is a central component of many memory-based models of sentence processing and it is generally assumed that the length of a dependency, i.e. the distance between a dependent and its head, influences processing of the head. We showed that measuring the length of a dependency in terms of the structural distance between a dependent and its head makes dependency integration a significant predictor of reading times.

More importantly, however, our analyses confirmed that a *higher* distance between two dependent elements leads to *lower* reading times (cf. Demberg & Keller, 2008; Baumann, 2012; Schuler & van Schijndel, 2014), which is the exact opposite of what most memory-based models predict.

Given the ample experimental evidence in support of memory-based models, our results can be interpreted in different ways: one may simply argue that current memory-based models, such as DLT, are too narrow in scope, in that they can only account for effects that arise from a very limited set of dependencies, such as e.g. dependencies resulting from extractions (or movement) like in relative clauses.

Taking a more positive stance, one may assume that dependency integration can lead to integration costs, but that dependencies may also help to predict and thus facilitate the upcoming head. Under this assumption, our results simply state that the latter process is more common than the former in everyday language comprehension. Since predictions of upcoming words is the main mechanism in experience-based models (e.g. Hale, 2001), this assumption may may point a way towards a model of sentence processing that integrates aspects of memory-based and experience-based models. We leave this task for future research.

Acknowledgments

The author thanks Masaya Yoshida and Brady Clark for their support, and the audience of the 25th Annual CUNY Conference on Human Sentence Processing, in particular Lars Konieczny, for valuable feedback and critical questions during a lively Q&A session.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4.*

Baumann, P. (2012). The role of hierarchical structure in syntactic dependency integration. Paper presented at the *25th Annual CUNY Conference on Human Sentence Processing*. CUNY, New York, NY.

Chen, S. F., & Goodman, J. (1998). *An empirical study of smoothing techniques for language modeling* (Tech. Rep.). Center for Research in Computing Technology, Harvard University.

de Marneffe, B. M., Marie-Catherine, & Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006).* Genova, Italy.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

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

Gibson, E. (2000). The Dependency Locality Theory: a distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain.* Cambridge: MIT Press.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Asssociation for Computational Linguistics (NAACL).* Pittsburgh, PA.

Kennedy, A., & Pynte, J. (2004). Parafoveal-on-foveal effects in normal reading. *Vision Reserach*, *45*, 153–168.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, *30*(5), 580-602.

Klein, D., & Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL '03)* (pp. 423–430). Sapporo, Japan.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627–645.

Konieczny, L., & Döring, P. (2003). Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. In *Proceedings of ICCS/ASCS Joint International Conference on Cognitive Science.* Sydney, Australia.

Kwon, N., Lee, Y., Gordon, P., Kluender, R., & Polinsky, M. (2010). Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of prenominal relative clauses in Korean. *Language*, *89*, 546–582.

Kwon, N., Polinsky, M., & Kluender, R. (2006). Subject preference in Korean. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics (WCCFL 25)* (pp. 1–14). Somerville: Cascadilla Press.

Nakatani, K., & Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: Evidence from Japanese. *Linguistics*, *46*(1), 63–86.

O'Grady, W. (1997). *Syntactic Development*. Chicago: University of Chicago Press.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, *6*(3), 229–259.

Pynte, J., New, B., & Kennedy, A. (2008). A multiple regression analysis of syntactic and semantic influences in reading normal text. *Journal of Eye Movement Research*, *2*(4), 1–11.

Schuler, W., & van Schijndel, M. (2014). Effects of integration in eye tracking. Poster presented at the *27th Annual CUNY Conference on Human Sentence Processing*. Ohio State University, Columbus, OH.

Stolcke, A. (2002). SRILM - an extensible language modeling toolkit. In *Proceedings of the Seventh International Conference on Spoken Language Processing (ICSLP '02).* Denver, CO.

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 767–794.

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic Theory and Psychological Reality.* The MIT Press.