Title
Syntactic category disambiguation within an architecture of human language processing

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN
1069-7977

Author
Bauman, Peter

Publication Date
2013

Peer reviewed

Syntactic category disambiguation within an architecture of human language processing

Peter Baumann (baumann@u.northwestern.edu)
Northwestern University
Department of Linguistics, 2016 Sheridan Road
Evanston, IL 60208, USA

Abstract

Syntactic category ambiguities are very frequent in natural languages, and all architectures of language processing need a mechanism for disambiguating syntactic category ambiguities. Corley and Crocker (2000) suggested that syntactic category disambiguation can be assigned its own module within a modular architecture. We will show that the model defined by Corley and Crocker can account for a considerable amount of variance in reading times of naturally occurring texts. In addition, we provide evidence that syntactic category disambiguation may be independent of syntactic top-down expectations, emphasizing the important role of bottom-up processes within an architecture of human language processing.

Keywords: sentence processing; reading; eye-tracking; ambiguity; lexical access.

Introduction

Successful language processing requires the integration of bottom-up information extracted from the current input and top-down expectations generated from what has been processed so far. When and how bottom-up and top-down processes interact has been a distinguishing feature of different processing architectures. On the one hand, there are constraint-based models (e.g. MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994; Tabor, Juliano, & Tanenhaus, 1997), which assume one single processing unit, in which all available information is considered simultaneously. Modular architectures, on the other hand, consist of several distinct processing modules (e.g. Frazier, 1987; Frazier & Clifton, 1996; Corley & Crocker, 2000). These modules are restricted to each having its own internal representation, and they are independently predictive and informationally encapsulated (Crocker & Corley, 2002). Assuming this definition of modules in terms of information flow, bottom-up processes are more likely to be modular than top-down processes (Appelbaum, 1998; Fodor, 1983).

One particular process, for which constraint-based and modular models make contradicting predictions, is syntactic category assignment or disambiguation: constraint-based models assume that rich contextual information is utilized to determine the syntactic category (i.e part of speech) of a word, while modular architectures only allow context-independent information. Although previous research may seem to have provided evidence for both positions, Corley and Crocker (2000) (see also Gibson, 2006) have shown that most of the evidence for constraint-based models can also be accounted for under a modular architecture with a module for bottom-up syntactic category assignment. In this paper, we follow Corley and Crocker's proposal and provide further evidence for the existence of a syntactic category module by showing that Corley and Crocker's model of syntactic category disambiguation is a significant predictor of reading times in naturally occurring texts. In addition, we provide evidence that syntactic category disambiguation may be independent of syntactic top-down expectations, emphasizing the critical role of bottom-up processes within a modular architecture of human language processing.

Syntactic Category Ambiguity

Many words in English (and presumably all other languages) are ambiguous, they can have different senses and/or belong to different syntactic categories or part-of-speech (i.e. noun, verb, adjective, etc.). The following example (from Boland, 1997) illustrates these ambiguities:

(1) I saw her *duck* ...
   a. ... under the porch to eat some potato chips.
   b. ... under the porch eat some potato chips

In (1), the word *duck* is ambiguous between its verb and noun readings, and only the following context can disambiguate between the two syntactic categories and senses. Syntactic category ambiguity and lexical ambiguity (in terms of different senses) need not come together like in (1). Lexical ambiguity often occurs within the same syntactic category as in the word *cabinet*, which as a noun can denote either a group of advisors or a closet. Syntactic category ambiguity, on the other hand, does not require lexical ambiguity, as evidenced by the English verbal system, where for all regular verbs there is only one form for the past-tense and the past-participle. This ambiguity is crucial to many garden-path sentences.

(2) The horse raced past the barn fell.

(3) The horse ridden past the barn fell.

While example (2) is a classical garden-path sentence, which upon first encounter may be nearly impossible to understand, example (3) is unambiguous and relatively easy to process. The fact that example (2) is derived from (3) only by replacing the ambiguous word *raced* with the unambiguous *ridden* demonstrates the important role of syntactic category disambiguation in language processing (cf. Chomsky & Lasnik, 1977).

Previous Research

One particular type of syntactic category ambiguity, which has received considerable attention in research is the noun-verb ambiguity. Based on three experiments, Frazier (1987) suggested that the processor delays resolving the ambiguity until disambiguating information is encountered, as readers spent less time on ambiguous words and more time on disambiguating context than on unambiguous words and their respective contexts. These results were put into question by MacDonald (1993) (see also MacDonald, 1993), who in turn argued that different statistical measures and biases such as semantic biases, syntactic context and word co-occurrences could influence syntactic category disambiguation. Similarly, Tabor et al. (1997) showed that readers are sensitive to syntactic context when resolving syntactic category ambiguities between the determiner and complementizer readings of *that*: a reading time delay occurred when *that* following a verb was disambiguated as a determiner or sentence-initial *that* was disambiguated as a complementizer.

While the results cited so far suggest that syntactic category disambiguation is – at least to some degree – dependent on syntactic or discourse context, Boland (1997) and Boland and Blodgett (2001) demonstrated in a series of experiments that when reading a syntactic category ambiguous word like *duck*, readers are sensitive to its *lexical bias*, i.e. the relative frequencies of the lexical entries for this word, independent of the syntactic or discourse context it appears in. In a similar vein, Stolterfoht, Gese, and Maienborn (2010) showed that for German adjectival passives (e.g. closed), whose forms are ambiguous between passive participle and adjective, there is an increase in reading times when preceded by an adjective-copula auxiliary as compared to the passive auxiliary, and as compared to unambiguous adjectives. This suggests that syntactic category disambiguation has a strong bottom-up component, which cannot be overridden by any top-down information. It is thus rather uncontroversial that lexical bias plays an important role in syntactic category disambiguation (cf. e.g. Gibson, 2006).

However, it remains open to what extent additional contextual information is used in this process: Gibson (2006) proposed that in addition to the context-independent lexical bias syntactic category disambiguation is also affected by context-dependent syntactic expectations, which he broadly formalizes as the probability of a syntactic category in a given 'syntactic environment'. A more restrictive notion of sufficient contextual information in syntactic category disambiguation, which was proposed by Corley and Crocker (2000), will be introduced in the next section and forms the basis of this paper.

The Statistical Lexical Category Module

One curious fact about syntactic category disambiguation is that computers seem to be nearly as good at it as humans are: unlike many other tasks in natural language processing, part-of-speech tagging has been an area in which rather simple models can achieve near-ceiling accuracy (Charniak, 1993). Inspired by this observation, Corley and Crocker (2000) assumed that syntactic category disambiguation is distinct from syntactic parsing. Reasons for this assumption are that syntactic category disambiguation happens extremely locally, that the relevant statistics are different from syntactic parsing, and that syntactic category disambiguation does not involve structure building. This means that syntactic category disambiguation can have its own internal representation, be informationally encapsulated and independently predictive, thus constituting the requirements for a separate module, the *Statistical Lexical Category Module*[1] (Corley & Crocker, 2000).

Corley and Crocker's model for the Statistical Lexical Category Module (SLCM) is based on a simple bigram statistical part-of-speech tagger defined by Equation 1, which expresses the assumption that the joint probability $P(t_0, \ldots, t_n, w_0 \ldots w_n)$ of all part-of-speech tags $t_0, \ldots, t_n$ and all words $w_0 \ldots w_n$ read so far can be reasonably approximated by the product of the lexical bias (i.e. the probability of word w_i given tag t_i) and the category bigram transitional probability.

$$P(t_1, \ldots, t_n, w_0 \ldots w_n) \approx \prod_{i=1}^{n} P(w_i|t_i)P(t_i|t_{i-1}) \quad (1)$$

Since lexical bias $P(w_i|t_i)$ is a property of the word, the category bigram transitional probability $P(t_i|t_{i-1})$ is the only means to capture context-dependence in this model of syntactic category disambiguation, implying that syntactic context-dependence is in fact only a dependence on the syntactic category of the preceding word.

One may object that limiting context-dependence to the category of only the preceding word is a too restrictive assumption, but Corley and Crocker (2000) (see also Crocker & Corley, 2002) showed that it is enough to model the results reported by MacDonald (1993) and Tabor et al. (1997).

However, the aim of this paper is not to try to explain all psycholinguistic evidence involving syntactic category disambiguities. Instead, we will evaluate Corley and Crocker's SLCM model on a larger scale as a predictor of reading times in naturally occurring text. While Corley

[1] Corley and Crocker (2000) refer to syntactic category ambiguity as 'lexical category ambiguity'.

and Crocker assume a direct link between the probabilities derived from Equation 1 and human processing difficulties, we follow common practice (e.g. Demberg & Keller, 2008; Pynte, New, & Kennedy, 2008) and take the logarithm as the linking function between probabilities and reading times.

We thus obtain the following measure $\log P_{SLCM}$ for a word w_i given its tag t_i and the tag t_{i-1} of the previous word:

$$\log P_{SLCM} = \log P(w_i|t_i) + \log P(t_i|t_{i-1}) \qquad (2)$$

This measure is evaluated in Experiment 1, where we show that it is a significant predictor of reading times in naturally occurring texts. In Experiment 2, we evaluate both terms in Equation 2 separately and show that lexical bias and category bigram transitional probabilities make independent contributions to the model fit observed in Experiment 1. In the final experiment, we provide evidence that syntactic category disambiguation may be independent of syntactic top-down expectations as measured by surprisal (Hale, 2001) based on a probabilistic context-free grammar.

Experiments

In recent years, it has become standard to evaluate computational models of language processing on 'eye-tracking corpora', i.e. on eye-tracking data of people reading naturally occurring texts (Pynte et al., 2008; Demberg & Keller, 2008). The basic idea is to fit two regression models to a measure of readings times. One regression model (baseline model) includes as predictors control variables, which are known to have an influence on reading times. The second regression model includes all those predictors as well, but in addition it also includes our computational model of language processing as a predictor. To test whether our computational model of language processing is a significant predictor we compare the fit of the two regression models to the data by means of a log-likelihood test.

Methods

In this section we describe the methodological detail common across all three experiments.

Data and Dependent Variable All three experiments use the Dundee Corpus (Kennedy & Pynte, 2004), a collection of eye-movement data from 10 participants reading 51,501 words each of the British newspaper *The Independent*. We approximated lexical categories by part-of-speech (PoS) tags, which were obtained by tagging the Dundee Corpus with the CLAWS tagger (Garside, 1987). Since syntactic category disambiguation is assumed to happen 'early' in processing, we chose first-pass reading times as our dependent variable. First-pass reading times are calculated for a given word and participant as the sum of all eye fixations on that word in the first pass, i.e. before leaving the word either to the right or to the left. Data points were removed if a word was not fixated, appeared as the first or last word in a line, or contained any non-letter symbol.

Control Variables All regression models included the following control variables, which are known to have an influence on reading times (c.f. Demberg & Keller, 2008): number of characters per word, position of word in a sentence, an indicator variable whether the previous word was not fixated, and indicator variable whether the following word was not fixated, the frequency of the word, the frequency of the previous word, the forward transitional probability, i.e. bigram probability $P(w_i|w_{i-1})$, and the backward transitional 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.

Estimating Probabilities in the SLCM Model The probabilities in Equation 2 were estimated from a corpus obtained by concatenating the CLAWS-tagged versions of the British National Corpus and the Dundee Corpus. The lexical bias $P(w_i|t_i)$ was estimated as is, i.e. without any smoothing. For estimating the the category bigram transitional probability $P(t_i|t_{i-1})$ we again used a bigram model with modified Kneser-Ney smoothing.

Regression Models For the regression models we used linear 'mixed-effects' models (Pinheiro & Bates, 2000; Gelman & Hill, 2007) of first-pass reading times with *participant*, *word* and *text number* as random effects, as a generalization of the common by-subject and by-item analyses, thus taking into account that the different words and texts read by the participants are random samples in the same sense as the participants are (cf. Clark, 1973). All models were fit in R (R Development Core Team, 2011) using the *lme4* package (Bates, 2005).

Baseline Model Results

The coefficients and standard errors of the baseline model are shown in Table 1. The coefficients are as expected based on prior research: e.g. reading times decreases with increasing position in the sentence and increasing word frequency, and increase with an increasing number of characters in a word.

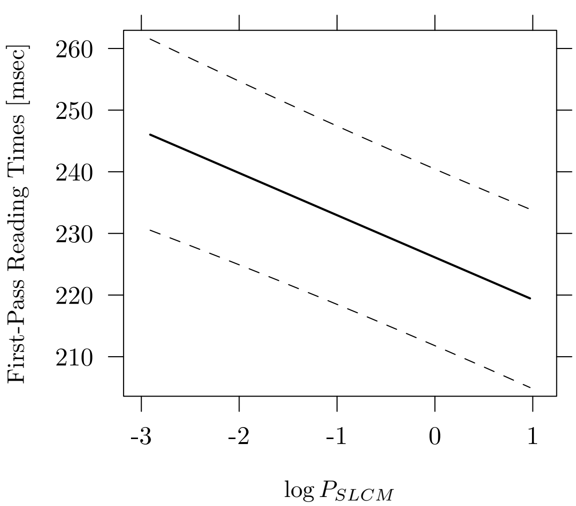
Table 1: Baseline model coefficients

Predictor	Coeff.	Std.Error	t
(Intercept)	206.34	7.31	28.22
Position in Sentence	-6.02	0.51	-11.76
Number of Characters	51.68	1.16	44.45
Frequency of Word	-23.89	1.58	-15.15
Freq. of Prev. Word	-12.84	0.61	-20.90
Forward Trans. Prob	-10.24	0.94	-10.94
Backward Trans. Prob.	-1.95	0.70	-2.77
No Fixation Next	10.14	0.49	20.61
No Fixation Previous	27.84	0.52	53.70

Experiment 1

The objective of Experiment 1 is to evaluate Corley and Crocker's model of the SLCM as a predictor of reading times. The predictor to be evaluated is the full model as stated in Equation 2.

Figure 1: Partial effect of full SLCM model with all other predictors held constant



Results The coefficient and standard error of the full tagger-based model of syntactic category disambiguation (Equation 2) are shown in Table 2[2]. A log-likelihood test between the regression model with the predictor $\log P_{SLCM}$ and the baseline model confirmed that Equation 2 is a significant predictor of reading times ($\chi^2 = 29.955, p < .0001$). The relation between $\log P_{SLCM}$ and reading times is plotted in Figure 1, which shows the partial effect of $\log P_{SLCM}$ with all other predictors held constant at their respective means. It can be seen that reading times increase as $\log P_{SLCM}$ or P_{SLCM} decrease.

[2] Coefficients for the control variables are not listed as they are qualitatively similar to the ones reported for the baseline model.

Table 2: Model coefficient of full SLCM model

Predictor	Coeff.	Std.Error	t
$\log P_{SLCM}$	-6.85	1.05	-6.54

Discussion The above result shows that the simple model of syntactic category disambiguation in Equation 2 cannot only account for many empirical results in psycholinguistic experiments, but is also a significant predictor of reading times in naturally occurring text. The direction of the effect is in line with experimental evidence modeled by Corley and Crocker (2000) in the sense that a lower probability in Equation 2 leads to higher reading times.

Experiment 2

In Experiment 2 we investigate whether lexical bias and category bigram transitional probability are also independently significant as predictors of reading times. To test this hypothesis, we fitted three models, one with only lexical bias (log-transformed $P(w_i|t_i)$), one with only category bigram transitional probability (log-transformed $P(t_i|t_{i-1})$), and a third one with both terms as additional predictors to the baseline model.

Results The coefficients and standard errors for lexical bias and category bigram transitional probability are shown in Table 3. The negative coefficients indicate that increasing the lexical bias (i.e. making the 'correct' category more likely) and increasing the category bigram transitional probability both lead to shorter reading times. A log-likelihood test confirmed that a model with either lexical bias ($\chi^2 = 7.37, p < .001$) or category bigram transitional probability ($\chi^2 = 22.97, p < .0001$) yields a significantly better fit to the data than the baseline model, and that a model with both predictors significantly improves over a model with only one.

Discussion Our results show that both lexical bias and category bigram transitional probability are significant predictors of reading times. For lexical bias this is in line with the results of Boland (1997) and Boland

Table 3: Model coefficient for lexical bias and category bigram probabilities

Predictor	Coeff.	Std.Error	t
Lexical Bias	-4.86	1.57	-3.09
Category Bigram	-4.28	0.69	-6.18

and Blodgett (2001), who also found a significant effect of lexical bias on reading times. The effect of category bigram transitional probabilities shows that the immediately preceding category contains information beyond what is contained in the corresponding preceding word, as including category bigram transitional probabilities improves over a baseline model, which already contained word bigram transitional probabilities.

Experiment 3

In Experiment 3 we test whether the effects of syntactic category disambiguation accounted for by the SLCM model can be ascribed to syntactic top-down expectations. If this were the case, it would provide strong evidence against any modular approach to syntactic category disambiguation. Syntactic top-down expectations are often measured by surprisal (Hale, 2001), which can be calculated from a probabilistic context-free grammar.

We calculated unlexicalized surprisal values for all words in the Dundee Corpus using the top-down parser described in (Roark, 2001) and (Roark, Bachrach, Cardenas, & Pallier, 2009) and included it as an additional predictor in our baseline model. We than compared this enriched baseline model to a regression model, which contained both surprisal and the log-probabilities of the tagger-based model of syntactic category disambiguation (Equation 2).

Results The coefficients and standard errors for surprisal and the tagger-based model of syntactic category disambiguation are shown in Table 4. As in Experiment 1, the coefficient of the tagger-based model is negative coefficients indicating that increasing the probability in Equation 1 leads to shorter reading times. The coefficient of surprisal is positive. This is expected as higher surprisal is associated with longer reading times (Hale, 2001; Demberg & Keller, 2008). A log-likelihood test confirmed that a model with the tagger-based model and surprisal improves significantly over a baseline model with only surprisal ($\chi^2 = 13.18, p < .001$).

Discussion The above results show that SLCM model is a significant predictor of reading times even if surprisal is included in the baseline regression model. Although this does not rule out the hypothesis that the effects of syntactic category disambiguation accounted for by the SLCM model may be reduced to syntactic top-down expectations, it provides strong evidence against such a hypothesis, and suggests instead that syntactic top-down expectations and bottom-up syntactic category disambiguation may be independent processes, as suggested by Gibson (2006) and Corley and Crocker (2000).

Table 4: Model coefficient for surprisal and the full SLCM model

Predictor	Coeff.	Std.Error	t
Surprisal	2.17	0.69	3.13
$\log P_{SLCM}$	-5.67	1.11	-5.10

General Discussion

In our experiments, we have shown that the model of a Statistical Lexical Category Module as formulated by Corley and Crocker (2000) is a significant predictor of reading times in naturally occurring texts. While our results do not necessarily imply that syntactic category disambiguation is a separate module, they provide further evidence for modular models relying on simple context-independent statistics for lexical category disambiguation. The observation that SLCM model is a significant predictor of reading times in addition to syntactic expectations as measured by surprisal indicates that Corley and Crocker's model may indeed account for bottom-up processes in reading, while surprisal accounts for top-down processes.

Since any architecture of language processing needs to integrate bottom-up and top-down processes, one may conclude that the combination of a restricted (or modular) model of bottom-up syntactic category disambiguation with a model of syntactic top-down expectations may ultimately lead to better models of the architecture of human language processing and, more specifically, to a better understanding of syntactic category disambiguation as a phenomenon at interface of lexical access and syntactic processing, as recent experiments have shown that syntactic category ambiguity also plays a crucial rule in lexical-semantic access and disambiguation (Jones, Folk, & Brusnighan, 2012).

Finally, our results may also contribute to the ongoing debate on lexicalized vs. unlexicalized measures of syntactic expectations and their reflections in reading times (for a review, see Roark et al., 2009): since bigram probabilities are the simplest form of syntactic expectations, our observation that category bigram probabilities are a significant predictor of reading times, even if controlled for word bigram probabilities, suggests that lexicalized and unlexicalized measures of syntactic expectations may have independent contributions to reading times.

Acknowledgments

The author would like to thank the four anonymous reviewers for their suggestions and Masaya Yoshida and Daniel Müller-Feldmeth for many fruitful discussions.

References

Appelbaum, I. (1998). Fodor, modularity, and speech perception. *Philosophical Psychology*, *11*(3), 317–330.

Bates, D. M. (2005). Fitting linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project*, *5*(1), 27–30.

Boland, J. E. (1997). Resolving syntactic category ambiguities in discourse context: Probabilistic and discourse constraints. *Journal of Memory and Language*, *36*, 588–615.

Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, *45*, 391–411.

Charniak, E. (1993). *Statistical Language Learning.* Cambridge: MIT Press.

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.

Chomsky, N., & Lasnik, H. (1977). Filters and control. *Linguistic Inquiry*, *8*(3), 425–504.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359.

Corley, S., & Crocker, M. W. (2000). The Modular Statistical Hypothesis: Exploring Lexical Category Ambiguity. In M. W. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and Mechanisms for Language Processing* (pp. 135–160). Cambridge: Cambridge University Press.

Crocker, M. W., & Corley, S. (2002). Modular Architectures and Statistical Mechanisms: The Case from Lexical Category Disambiguation. In P. Merlo & S. Stevenson (Eds.), *The Lexical Basis of Sentence Processing: Formal, computational and experimental issues* (pp. 157–180). Amsterdam: John Benjamins Publishing.

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

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge: MIT Press.

Frazier, L. (1987). Theories of Sentence Processing. In J. Garfield (Ed.), *Modularity in Knowledge Representation and Natural Language Processing* (pp. 291–308). Cambridge: MIT Press.

Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge: MIT Press.

Garside, R. (1987). The CLAWS Word-tagging System. In R. Garside, G. Leech, & G. Sampson (Eds.), *The Computational Analysis of English: A Corpus-based Approach.* London: Longman.

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

Gibson, E. (2006). The interaction of top–down and bottom–up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, *54*, 363–388.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 159–166). Pittsburgh, PA.

Jones, A., Folk, J., & Brusnighan, S. (2012). Resolving syntactic category ambiguity: An eye- movement analysis. *Journal of Cognitive Psychology*, *24*(6), 672–688.

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

MacDonald, M. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, *32*, 692–715.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676-703.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed Effects Models in S and S-Plus.* New York: Springer.

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.

R Development Core Team. (2011). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna.

Roark, B. (2001). Probabilistic top-down parsing and language modeling. *Computational Linguistics*, *27*(2), 249–276.

Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).* Singapore.

Stolcke, A. (2002). SRILM - an extensible language modeling toolkit. In *Seventh international conference on spoken language processing.*

Stolterfoht, B., Gese, H., & Maienborn, C. (2010). Word category conversion causes processing costs: Evidence from adjectival passives. *Psychonomic Bulletin & Review*, *17*(5), 651–656.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, *12*(211–272).

Trueswell, J., & Tanenhaus, M. (1994). Toward a lexical framework of constraint-based syntactic ambiguity resolution. In C. Clifton & L. Frazier (Eds.), *Perspectives on Sentence Processing* (pp. 155–179). Mahwah: Erlbaum.