Combining Simplicity and Likelihood in Language and Music

Rens Bod (rens@science.uva.nl)
Cognitive Science Center Amsterdam, University of Amsterdam
Nieuwe Achtergracht 166, Amsterdam, The Netherlands

Abstract

It is widely accepted that the human cognitive system organizes perceptual input into complex hierarchical descriptions which can be represented by tree structures. Tree structures have been used to describe linguistic, musical and visual perception. In this paper, we will investigate whether there exists an underlying model that governs perceptual organization in general. Our key idea is that the cognitive system strives for the simplest structure (the "simplicity principle"), but in doing so it is biased by the likelihood of previous experiences (the "likelihood principle"). We will present a model which combines these two principles by balancing the notion of most likely tree with the notion of shortest derivation. Experiments with linguistic and musical benchmarks (Penn Treebank and Essen Folksong Collection) show that such a combination outperforms models that are based on either simplicity or likelihood alone.

Introduction

It is widely accepted that the human cognitive system organizes perceptual input into complex, hierarchical descriptions which can be represented by tree structures. Tree structures have been used to describe linguistic perception (e.g. Chomsky 1965), musical perception (e.g. Lerdahl & Jackendoff 1983) and visual perception (e.g. Marr 1982). Yet, there seems to be little or no work which emphasizes the commonalities between these different forms of perception and which searches for a general, underlying mechanism which governs all perceptual organization (cf. Leyton 2001). This paper aims to study exactly that question: acknowledging the differences between linguistic, musical and visual information, is there a general, unifying model which can predict the perceived tree structure for sensory input? In studying this question, we will use a strongly empirical methodology: any model that we might hypothesize will be tested against benchmarks such as the linguistically annotated Penn Treebank (Marcus et al. 1993) and the musically annotated Essen Folksong Collection (Schaffrath 1995). While we will argue for a unified model of language, music and vision, we will carry out experiments only with linguistic and musical benchmarks, since no benchmarks of visual tree structures are currently available, to the best of our knowledge.

Figure 1 gives three simple examples of linguistic, musical and visual input with their corresponding tree structures given below.

Thus a tree structure describes how parts of the input combine into constituents and how these constituents combine into a representation for the whole input. Note that the linguistic tree structure is labeled with syntactic categories, whereas the musical and visual tree structures are unlabeled. This is because in language there are syntactic constraints on how words can be combined into larger constituents, while in music (and to a lesser extent in vision) there are no such restrictions: in principle any note may be combined with any other note.

List the sales of products in 1973

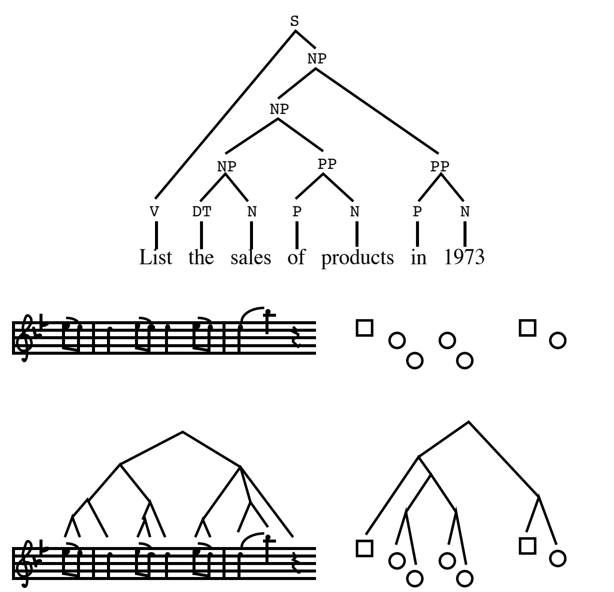
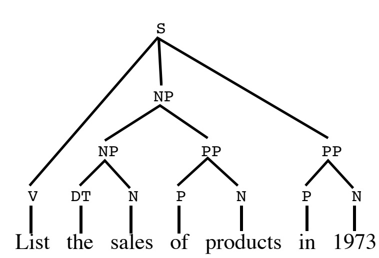


Figure 1: Examples of tree structures.

Apart from these differences, there is also a fundamental commonality: the perceptual input undergoes a process of hierarchical structuring which is not found in the input itself. The main problem is thus: how can we derive the perceived tree structure for a given input? That this problem is not trivial may be illustrated by the fact that the inputs above can also be assigned the following, alternative tree structures in figure 2.



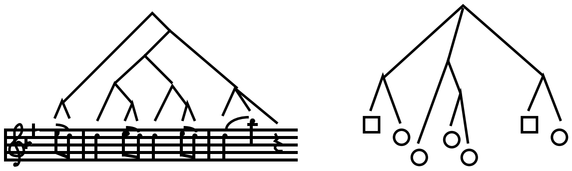


Figure 2: Alternative tree structures for figure 1.

These alternative structures are possible in that they *can* be perceived. But while the alternative tree structures are all possible, they are not plausible: they do not correspond to the structures that are actually perceived by the human perceptual system.

The phenomenon that the same input may be assigned different structural organizations is known as the *ambiguity problem*. This problem is one of the hardest problems in modeling human perception. Even in language, where a phrase-structure grammar may specify which words can be combined into constituents, the ambiguity problem is notoriously hard. Charniak (1997: 37) argues that almost every sentence from the Wall Street Journal has many, often more than one million different parse trees. The ambiguity problem for musical and visual input is even harder. Talking about rhythm perception in music, Longuet-Higgins and Lee (1987) note that "Any given sequence of note values is in principle infinitely ambiguous, but this ambiguity is seldom apparent to the listener."

Two principles: likelihood and simplicity

How can we predict from the set of all possible tree structures the tree that is actually perceived by the human cognitive system? In the field of visual perception, two competing principles have traditionally been proposed to govern perceptual organization. The first, initiated by Helmholtz (1910), advocates the *likelihood principle*: sensory input will be organized into the most probable organization. The second, initiated by Wertheimer (1923) and developed by other Gestalt psychologists, advocates the *simplicity principle*: the perceptual system is viewed as finding the simplest perceptual organization (see Chater 1999 or Van der Helm 2000 for an overview). These two principles are not only relevant for visual perception, but also for linguistic and musical perception. In the following, we briefly discuss these principles for each modality, after which we go into the question of how the two principles can be integrated.

Likelihood

The likelihood principle is particularly influential in the field of natural language processing (see Manning and Schütze 1999, for a review). In this field, the most appropriate tree structure of a sentence is assumed to be its most likely structure. The likelihood of a tree is usually computed from the probabilities of its parts (e.g. phrase-structure rules) taken from a large annotated language corpus (a *treebank*). A widely used treebank for testing and comparing probabilistic natural language parsers is the Penn Wall Street Journal Treebank (Marcus et al. 1993). State-of-the-art probabilistic parsers such as Collins (2000), Charniak (2000) and Bod (2001a) obtain around 90% precision and recall on the Wall Street Journal. Also in the field of psycholinguistics, the likelihood principle is widely used: Jurafsky (1996), Crocker and Brantz (2000) and Hale (2001) are examples of psycholinguistically inspired probabilistic parsers.

The likelihood principle has also been applied to musical perception, e.g. in Raphael (1999) and Bod (2001b/c). As in probabilistic natural language processing, the most probable musical tree structure can be computed from the probabilities of rules or fragments taken from a large annotated musical corpus, for instance from the Essen Folksong Collection (Bod 2001b).

In visual perception psychology and vision science, there has recently been a resurgence of interest in probabilistic models (e.g. Hoffman 1998; Kersten 1999). Mumford (1999) has seen fit to declare the Dawning of Stochasticity.

Simplicity

The simplicity principle has a long tradition in the field of visual perception psychology (e.g. Restle 1970; Leeuwenberg 1971; Simon 1972; Buffart et al. 1983; van der Helm 2000). In this field, a visual pattern is formalized as a constituent structure by means of a "visual coding language" based on primitive elements such as line segments and angles. Perception is described as the process of selecting the simplest structure corresponding to the "shortest encoding" of a visual pattern.

The notion of simplicity has also been applied to musical perception. Collard et al. (1981) use the coding language of Leeuwenberg (1971) to predict the metrical structure for four preludes from Bach's *Well-Tempered Clavier*. More well-known in musical perception is the theory proposed by Lerdahl and Jackendoff (1983) who use a system of preference rules based on the Gestalt-preferences identified by Wertheimer (1923), and which can therefore also be seen as an embodiment of the simplicity principle.

Notions of simplicity also exist in language processing (e.g. Frazier 1978; Gorrell 1995; Osborne 2000). Bod (2000a) defines the simplest tree structure of a sentence as the structure generated by the smallest number of subtrees from a given treebank.

Combining the two principles

The key idea of the current paper is that both principles play a role in perceptual organization: the simplicity principle as a general cognitive preference for economy, and the likelihood principle as a probabilistic bias due to previous perceptual experiences. Informally stated, our working hypothesis is that the human cognitive system strives for maximal economy (the simplest structure), but that in doing so it is biased by the likelihood of previous experiences (in the last section we will discuss some other combinations of simplicity and likelihood that have been proposed). To formally instantiate our working hypothesis, we need a parsing model to start with which can incorporate these principles. In principle any parsing model might do, as long as it can assign tree structures to perceptual input according to some criterion. For the current paper, we have chosen to start with the Data-Oriented Parsing model (Bod 1998) because (1) it has several other models as special cases, such as context-free parsing models and lexicalized models, and (2) it has been quite successful in predicting tree structures for both linguistic input (Bod 2001a) and musical input (Bod 2001b).

The basic idea of DOP is that it learns a grammar by extracting subtrees from a given treebank and uses these subtrees to analyze fresh input. Suppose we are given the following linguistic treebank of only two trees (we will come back to musical treebanks in the next section),

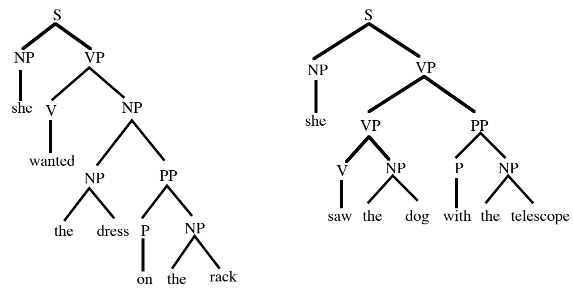


Figure 3: An example treebank

then the DOP model can parse a new sentence, e.g. *She saw the dress with the telescope*, by combining subtrees from this treebank by means of a *node-substitution operation* (indicated as ∘):

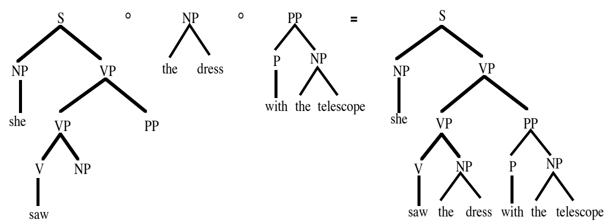


Figure 4: Parsing a sentence by combining subtrees

Thus the node-substitution operation combines two subtrees by substituting the second subtree on the leftmost nonterminal leaf node of the first subtree. Since DOP uses subtrees of arbitrary size, there are typically several derivations, involving different subtrees, that produce the *same* parse tree; for instance:

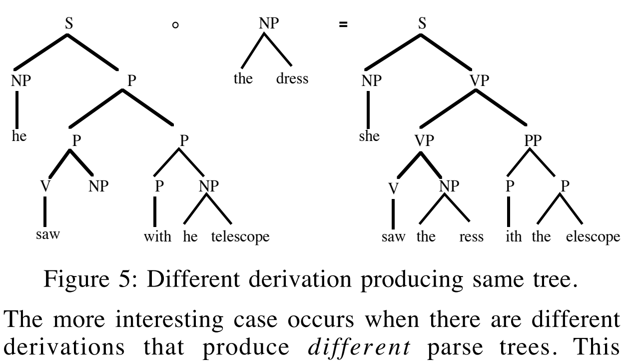


Figure 5: Different derivation producing same tree.

The more interesting case occurs when there are different derivations that produce *different* parse trees. This happens when a sentence is structurally ambiguous; for example, DOP also produces the following alternative parse tree for *She saw the dress with the telescope*:

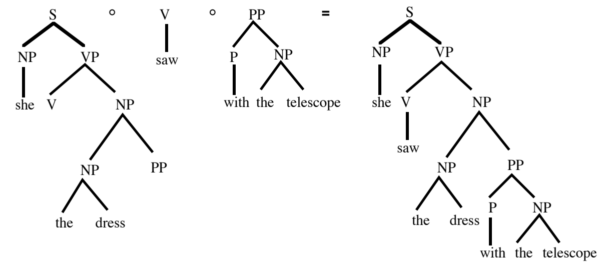


Figure 6: Different derivation producing different tree.

The original DOP model in Bod (1993) uses the likelihood principle to predict the perceived tree structure. We will refer to this model as *Likelihood-DOP*. Likelihood-DOP selects the most likely tree structure from among all possible tree structures on the basis of the probabilities of its subtrees. The probability of a subtree *t* is estimated as the number of occurrences of *t* seen in the corpus, divided by the total number of occurrences of corpus-subtrees that have the same root label as *t*. The probability of a derivation is computed as the product of the probabilities of the subtrees involved in it. Finally, the probability of a parse tree is equal to the sum of the probabilities of all distinct derivations that produce that tree. In Bod (2001a) and Goodman (2002), efficient algorithms are given that compute for an input string the most probable parse tree.

Likelihood-DOP does not do justice to the preference humans display for the simplest structure, e.g. the one that is generated by the shortest derivation consisting of the fewest subtrees. This is what we will call *Simplicity-DOP*. Instead of producing the most probable parse tree for an input, Simplicity-DOP produces the parse tree generated by the fewest corpus-subtrees, independent of the probabilities of these subtrees. For example, given the corpus in Figure 3, the simplest parse tree for *She saw the dress with the telescope* according to Simplicity-DOP is given in Figure 5, since that parse tree can be generated by a derivation of only two corpus-subtrees, while the parse tree in Figure 6 (and any other parse tree) needs at least three corpus-subtrees to be generated. In Bod (2000a) it is shown how the shortest derivation can be efficiently computed by means of a best-first bottom-up chart parsing algorithm. Simplicity-DOP obtains quite impressive results on the WSJ, though its results are lower than Likelihood-DOP (Bod 2000a). Yet, the set of correctly predicted parse trees of Simplicity-DOP is *not* a subset of the set of correctly predicted parse trees of Likelihood-DOP. This suggests that we may expect an accuracy improvement if simplicity and likelihood are combined into a new model, which we will call *Combined-DOP*.

The underlying idea of Combined-DOP is that the human perceptual system searches for the shortest derivation (i.e. the simplest tree structure), but that in doing so it is biased by the "weights" of the subtrees. The length of a derivation is then not defined simply as the sum of the derivation steps (as in Simplicity-DOP), but as the sum of the weights of these steps, where a low weight should be seen as an easy step and a heavy weight as a difficult step. As a measure for weight of a subtree, we have worked out various proposals that were experimentally tested over the last few years. Most of these proposals have been reported in the literature (e.g. Bod 2000a; Cormons 1999). The best measure for subtree weight so far is based on the *rank* of a subtree. First, all subtrees are grouped with respect to their root label. Next, for each root label the weight of a subtree is defined as its rank in the frequency ordering in the corpus. Thus, the most frequent subtree in each ordering gets a weight of 1, the second most frequent subtree gets a weight of 2, etc. The weight of a derivation is then defined as the sum of the weights of the subtrees in the derivation. The derivation with the lowest weight is taken as the "best" derivation producing the perceived parse tree. Thus, the best derivation is not determined by the smallest sum of the subtrees (as in Simplicity-DOP), but by the smallest sum of the *weights* of the subtrees.

We performed one additional adjustment to the weight of a subtree. This adjustment consists in a smoothing technique which averages the weight of a subtree by the weights of its own *sub*-subtrees. That is, instead of taking only the rank of a subtree as its weight, we compute the weight of a subtree as the (arithmetic) mean of the weights of all its sub-subtrees (including the subtree itself). The effect of this smoothing technique is that it redresses a very low-frequency subtree if it contains high-frequency sub-subtrees.

The Test Domains

Our linguistic test domain consists of sections 02-21 of the Wall Street Journal portion of the Penn Treebank, which contains approx. 40,000 phrase-structure trees. Since the Penn Treebank has been extensively described in the literature (e.g. Marcus et al. 1993; Manning & Schütze 1999), we will not go into it any further here.

The musical test domain consists of the European folksongs in the Essen Folksong Collection (Schaffrath 1995), which correspond to approx. 6,200 musical grouping structures. The Essen Folksong Collection has been previously used by Bod (2001b) and Temperley (2001) to test their musical parsers. The musical coding language used in the Essen Folksong Collection is based on the Essen Associative Code (ESAC). The pitch encodings in ESAC resemble "solfege": scale degree numbers are used to replace the movable syllables "do", "re", "mi", etc. Thus 1 corresponds to "do", 2 corresponds to "re", etc. Chromatic alterations are represented by adding either a "#" or a "b" after the number. The plus ("+") and minus ("-") signs are added before the number if a note falls resp. above or below the principle octave (thus -1, 1 and +1 refer al to "do", but on different octaves). Duration is represented by adding a period or an underscore after the number. A period (".") increases duration by 50% and an underscore ("_") increases duration by 100%; more than one underscore may be added after each number. If a number has no duration indicator, its duration corresponds to the smallest value. A pause is represented by 0, possibly followed by duration indicators. No loudness or timbre indicators are used in ESAC. The only extra information we (automatically) added to the grouping structures in the Essen Folksong Collection consists of the label "S" for each top node of each whole song and the label "P" for each underlying phrase. In this way, we obtained conventional parse trees that can directly be used by our DOP models to parse new input strings (see also Bod 2001b). The Essen Folksong Collection is freely available via http://www.esac-data.org.

As mentioned in the introduction, no visual treebank is currently available, to the best of our knowledge. We are currently developing a treebank of analyzed architectural plans, and will report on experiments with that treebank in due time.

Experimental Evaluation

To evaluate our DOP models, we used the *blind testing* method which dictates that a treebank be randomly divided into a training set and a test set, where the strings from the test set are parsed by means of the subtrees from the training set. We applied the PARSEVAL metrics of *precision* and *recall* to compare a proposed parse tree P with the corresponding correct test set parse tree T (see Black et al. 1991):

$$\text{Precision} = \frac{\text{\# correct constituents in } P}{\text{\# constituents in } P} \qquad \text{Recall} = \frac{\text{\# correct constituents in } P}{\text{\# constituents in } T}$$

A constituent in P is "correct" if there exists a constituent in T of the same label that spans the same elements (i.e. words or notes). To balance precision and recall into a single measure, we will employ the widely used F-score: *F-score = 2*Precision*Recall / (Precision+Recall).*

We will use this F-score to quantitatively evaluate our models on the Wall Street Journal and the Essen Folksong treebanks. We divided both treebanks into 10 training/test set splits, where 90% of the trees was used for training and 10% for testing. These splits were random, except for one constraint: that all the primitive elements (i.e. words and notes) in the test set also occurred in the training set. In this way, we did not have to worry about unknown words or unknown notes (the latter being actually inexistent for our musical treebank). Although there are various statistical ways to cope with unknown words, we wanted to rule out this problem as it might obscure our comparison.

In our experiments we were first of all interested in comparing the three DOP models (Likelihood-DOP, Simplicity-DOP and Combined-DOP) on the two domains. For computational reasons, we limited the maximum size of the subtrees to depth 14, as in Bod (2001a). Table 1 shows the average F-scores for each of the models.

Table 1: F-scores obtained by the three DOP models

	Likelihood-DOP	Simplicity-DOP	Combined-DOP
Language	90.4%	88.1%	91.7%
Music	86.0%	84.3%	86.9%

The table shows that Likelihood-DOP outperforms Simplicity-DOP, but that Combined-DOP outperforms Likelihood-DOP. According to paired *t*-testing, the improvement of Combined-DOP over Likelihood-DOP was statistically significant both for language ($p<.0001$) and for music ($p<.04$).

We also performed a series of experiments where we restricted the size of the subtrees. Recall that by restricting the subtrees to depth 1, Likelihood-DOP becomes equivalent to a probabilistic context-free grammar, while Simplicity-DOP would just return the smallest possible tree structure. While Likelihood-DOP still obtained relatively good results at depth 1 for both language and music (resp. 75.1% and 76.6%), Simplicity-DOP scored very badly for language (22.5%) though still reasonably for music (70.0%). Interestingly, Combined-DOP scored worse than Likelihood-DOP at depth 1 (resp. 68.2% vs. 74.6%). Only after subtree depth 6 for language and subtree depth 2 for music, Combined-DOP outperformed Likelihood-DOP. The highest F-scores were obtained with the "unrestricted" subtrees (in table 1).

Elsewhere we have shown that virtually *any* constraint on the subtrees results in an accuracy decrease (Bod 2001a/b). This is because in language, almost any relation between words (including between so-called non-headwords) can be important for predicting the perceived parse tree of a sentence. The same counts for music, where there is a continuity between "jump-phrases" and "non-jump-phrases", which can only be captured by large subtrees (see Bod 2001b/c for an extensive discussion).

Discussion: Other Combinations of Simplicity and Likelihood

We have seen that our combination of simplicity and likelihood is quite rewarding for linguistic and musical perception, suggesting a deep parallel between the two modalities. Yet, we should raise the question whether a model which massively stores and re-uses previously perceived structures has any cognitive plausibility. Interestingly, there is quite some evidence that people store various kinds of previously heard fragments, both in music (Saffran et al. 2000) and language (Jurafsky 2002). But do people store fragments of *arbitrary* size, as proposed by DOP? In his overview article, Jurafsky (2002) reports on a large body of psycholinguistic evidence showing that people not only store lexical items and bigrams, but also frequent phrases and even whole sentences. For the case of sentences, people not only store idiomatic sentences, but also "regular" high-frequency sentences. Thus, at least for language it seems that humans store fragments of arbitrary size provided that these fragments have a certain minimal frequency. However, there seems to be no evidence that people store *all* fragments they hear, as suggested by DOP. Only high-frequency fragments seem to be memorized. However, if the human perceptual faculty needs to *learn* which fragments will be stored, it will initially need to keep track of all fragments (with the possibility of forgetting them) otherwise frequencies can never accumulate. This results in a model which continuously and incrementally updates its fragment memory given new input, which is in correspondence with the DOP approach.

There have been other proposals for integrating or reconciling the principles of simplicity and likelihood. Chater (1999) argues that the principles are identical in the context of Kolmogorov's complexity theory (Kolmogorov 1965). And in the context of Information Theory the simplicity principle can be defined in terms of bit length, such that maximizing likelihood corresponds to minimizing bit length (cf. Rissanen 1978). First note that the likelihood principle aims at maximizing the probability of a structure given an input, $p(structure \mid input)$. Next, define the simplicity principle as minimizing the informatic-theoretical notion of bit length, which is the (negative) logarithm of the probability of a structure given an input: $-log\, p(structure \mid input)$. Now it is easy to see that maximizing $p(structure \mid input)$ leads to the same *structure* as minimizing $-log\, p(structure \mid input)$. Thus the two principles lead to the same result.

However, in the context of DOP we defined the simplest structure as the one generated by the shortest derivation consisting of the smallest number of subtrees (reflecting the smallest number of steps needed to parse an input). And this notion of simplest structure is provably different from the most probable structure given an input. Although it is possible to redefine our notion of simplest structure in terms of bit length, it would not lead to any new model, and to no improved result. By conceptually separating between simplicity and likelihood in DOP and by combining them in a novel way, we have shown that an improved model can be obtained.

What we have not done in this paper is to isolate the perceptual properties for which *no* prior expectations are needed. Even Simplicity-DOP, albeit non-probabilistic, is heavily based on previously perceived data. It is very likely that there are perceptual grouping properties for which no prior expectations are necessary. DOP does not contribute to the discovery of such properties, but it does neither neglect them, as they are implicit in the treebank. Bod (2001b) shows that Wertheimer's Gestalt principles are reflected in about 85% of the phrases in the Essen Folksong Collection (where phrases have boundaries that fall on large time or pitch intervals). DOP automatically takes these principles into account by subtrees that contain such phrases, but DOP also takes into account phrases whose boundaries do *not* fall on large intervals (so-called "jump-phrases"). By using all subtrees, DOP mimics the preferences humans have used in analyzing the perceptual data, whatever these preferences may have been.

References

Black, E. et al. (1991). A Procedure for Quantitatively Comparing the Syntactic Coverage of English, *Proceedings DARPA Speech and Natural Language Workshop*, Pacific Grove, Morgan Kaufmann.

Bod, R. (1993). Using an Annotated Language Corpus as a Virtual Stochastic Grammar, *Proceedings AAAI-93*, Menlo Park, Ca.

Bod, R. (1998). *Beyond Grammar: An Experience-Based Theory of Language*, Stanford: CSLI Publications.

Bod, R. (2000a). Parsing with the Shortest Derivation. *Proceedings COLING-2000*, Saarbrücken, Germany.

Bod, R. (2001a). What is the Minimal Set of Fragments that Achieves Maximal Parse Accuracy? *Proceedings ACL'2001*, Toulouse, France.

Bod, R. (2001b). Memory-Based Models of Music Analysis. *Proceedings International Computer Music Conference* (ICMC'2001), Havana, Cuba.

Bod, R. (2001c). Memory-Based Models of Melodic Analysis: Challenging the Gestalt Principles. *Journal of New Music Research,* 30(3), in press.

Bod, R., J. Hay and S. Jannedy (eds.) (2002a). *Probabilistic Linguistics*. Cambridge, The MIT Press. (in press)

Bod, R., R. Scha and K. Sima'an (eds.) (2002b). *Data-Oriented Parsing*. Stanford, CSLI Publications. (in press)

Buffart, H., E. Leeuwenberg and F. Restle (1983). Analysis of Ambiguity in Visual Pattern Completion. *Journal of Experimental Psychology: Human Perception and Performance*. 9, 980-1000.

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

Charniak, E. (1997). Statistical Techniques for Natural Language Parsing, *AI Magazine*, Winter 1997, 32-43.

Charniak, E. (2000). A Maximum-Entropy-Inspired Parser. *Proceedings ANLP-NAACL'2000*, Seattle, Washington.

Chater, N. (1999). The Search for Simplicity: A Fundamental Cognitive Principle? *The Quarterly Journal of Experimental Psychology*, 52A(2), 273-302.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*, Cambridge, The MIT Press.

Collard, R., P. Vos and E. Leeuwenberg, (1981). What Melody Tells about Metre in Music. *Zeitschrift für Psychologie*. 189, 25-33.

Collins, M. (2000). Discriminative Reranking for Natural Language Parsing, *Proceedings ICML-2000*, Stanford, Ca.

Cormons, B. (1999). *Analyse et désambiguisation: Une approche à base de corpus (Data-Oriented Parsing) pour les répresentations lexicales fonctionnelles*. PhD thesis, Université de Rennes, France.

Crocker, M. and T. Brants (2000). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research* 29, 647-669.

Frazier, L. (1978). *On Comprehending Sentences: Syntactic Parsing Strategies*. PhD. Thesis, University of Connecticut.

Goodman, J. (2002). Efficient Parsing of DOP with PCFG-Reductions. In R. Bod et al. 2002b.

Gorrell, P. (1995). *Syntax and Parsing*. Cambridge University Press.

Hale, J. (2001). A Probabilistic Earley Parser as a Psycholinguistic Model. *Proceedings NAACL'01*, Pittsburgh, PA.

van der Helm, P. (2000). Simplicity versus Likelihood in Visual Perception: From Surprisals to Precisals. *Psychological Bulletin*, 126(5), 770-799.

von Helmholtz, H. (1910). *Treatise on Physiological Optics* (Vol. 3), Dover, New York.

Hoffman, D. (1998). *Visual Intelligence*. New York, Norton & Company, Inc.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation, *Cognitive Science*, 20, 137-194.

Jurafsky, D. (2002). Probabilistic Modeling in Psycholinguistics: Comprehension and Production. In R. Bod et al. 2002a.

Kersten, D. (1999). High-level vision as statistical inference. In S. Gazzaniga (ed.), *The New Cognitive Neurosciences*, Cambridge, The MIT Press.

Kolmogorov, A. (1965). Three approaches to the quantitative definition of information. *Problems in Information Transmission* 1, 1-7.

Leeuwenberg, E. (1971). A Perceptual Coding Language for Perceptual and Auditory Patterns. *American Journal of Psychology*. 84, 307-349.

Lerdahl, F. and R. Jackendoff (1983). *A Generative Theory of Tonal Music*. Cambridge, The MIT Press.

Leyton, M. (2001). *A Generative Theory of Shape*. Heidelberg, Springer-Verlag.

Longuet-Higgins, H. and C. Lee, (1987). The Rhythmic Interpretation of Monophonic Music. In: *Mental Processes: Studies in Cognitive Science*, Cambridge, The MIT Press.

Manning, C. and H. Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, The MIT Press.

Marcus, M., B. Santorini and M. Marcinkiewicz (1993). Building a Large Annotated Corpus of English: the Penn Treebank, *Computational Linguistics* 19(2).

Marr, D. (1982). *Vision*. San Francisco, Freeman.

Mumford, D. (1999). The dawning of the age of stochasticity. Based on a lecture at the Accademia Nazionale dei Lincei. (http://www.dam.brown.edu/people/mumford/Papers/Dawning.ps)

Osborne, M. (1999). Minimal Description Length-Based Induction of Definite Clause Grammars for Noun Phrase Identification. *Proceedings EACL Workshop on Computational Natural Language Learning*, Bergen, Norway.

Raphael, C. (1999). Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4), 360-370.

Restle, F. (1970). Theory of Serial Pattern Learning: Structural Trees. *Psychological Review*. 86, 1-24.

Rissanen, J. 1978. Modeling by the shortest data description. *Automatica*, 14, 465-471.

Saffran, J., M. Loman and R. Robertson (2000). Infant Memory for Musical Experiences. *Cognition* 77, B16-23.

Schaffrath, H. (1995). The Essen Folksong Collection in the Humdrum Kern Format. D. Huron (ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Simon, H. (1972). Complexity and the Representation of Patterned Sequences os Symbols. *Psychological Review*. 79, 369-382.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, The MIT Press.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung* 4, 301-350.