

# Parsing Spoken Language Using Combinatory Grammars\*

Mark Steedman  
Computer and Information Science, U.Penn.

Combinatory Grammars are a generalisation of Categorical Grammars to include operations on function categories corresponding to the combinators of Combinatory Logic, such as functional composition and type raising. The introduction of such operations is motivated by the need to provide an explanatory account of coordination and unbounded dependency. However, the associativity of functional composition tends to engender an equivalence class of possible derivations for each derivation permitted by more traditional grammars. While all derivations in each class by definition deliver the same function-argument relations in their interpretation, the proliferation of structural analyses presents obvious problems for parsing within this framework and the related approaches based on the Lambek calculus (Moortgat).

This problem has been called the problem of "spurious ambiguity", (although it will become apparent that the term is rather misleading). A number of ways of dealing with it have been proposed, including compiling the grammar into a different form (Wittenburg), "normal form"-based parsing (Hepple and Morrill, Koenig), and a "lazy" chart parsing technique which directly exploits the properties of the combinatory rules themselves to provide a unified treatment for "spurious" ambiguities and "genuine" attachment ambiguities (Pareschi and Steedman).

Recent work suggests that the very free notion of syntactic structure that is engendered by the theory is identical to the notion of structure that is required by recent theories of phrasal intonation and prosody. Intonational Structure is notoriously freer than traditional syntactic structure, and is commonly regarded as conveying distinctions of discourse focus and propositional attitude. It is argued that the focussed entities, propositions, and abstractions that are associated with a given intonational structure can be identified with the interpretations that the grammar provides for the non-standard constituents that it allows under one particular derivation from an equivalence class. The constituent interpretations corresponding to each possible intonational tune belong to the same equivalence class, and therefore reduce to the same canonical function argument relations. However, it is apparent that the ambiguity between derivations in the same equivalence class is not spurious at all, but meaning-bearing.

Of course, not *all* structural ambiguities are resolved by distinctions of intonation. (An example is PP attachment ambiguity). It follows that some of the techniques proposed for written parsing must be implicated as well. However, the theory opens the possibility of unifying phonological and syntactic processing, as well as simplifying the architecture required for integrating higher-level modules in spoken language processing.

---

\*I am grateful to Julia Hirschberg, Aravind Joshi, Mitch Marcus, Janet Pierrehumbert, and Bonnie Lynn Webber for comments and advice. The research was supported by DARPA grant no. N0014-85-K0018, ARO grant no. DAAG29-84-K-0061, and NSF grant no. CER MCS 82-19198.

## Structure and Intonation

Phrasal intonation is notorious for structuring the words of spoken utterances into groups which frequently violate orthodox notions of constituency. For example, the normal prosody for the answer (b) to the following question (a) imposes the intonational constituency indicated by the brackets (stress is indicated by capitals):

- (1) a. I know that brassicas are a good  
source of minerals, but what are  
LEGumes a good source of?  
b. (LEGumes are a good source of)  
VITamins.

Such a grouping cuts right across the traditional syntactic structure of the sentence. The presence of two apparently uncoupled levels of structure in natural language grammar appears to complicate the path from speech to interpretation unreasonably, and to thereby threaten a number of computational applications.

Nevertheless, intonational structure is strongly constrained by meaning. Contours imposing bracketings like the following are not allowed:

- (2) # Three doctors (in ten prefer cats)

Halliday [5] seems to have been the first to identify this phenomenon, which Selkirk [16] has called the "Sense Unit Condition", and to observe that this constraint seems to follow from the *function* of phrasal intonation, which is to convey distinctions of focus, information, and propositional attitude towards entities in the discourse. These entities are more diverse than mere nounphrase or propositional referents, but they do not include such non-concepts as "in ten prefer cats."

One discourse category that they *do* include is what E. Prince [15] calls "open propositions". Open propositions are most easily understood as being that which is introduced into the discourse context by a Wh-question. So for example the question in (1), *What are legumes a good source of?* introduces an open proposition which it is most natural to think of as a functional *abstraction*, which would be written as follows in the notation of the  $\lambda$ -calculus:

- (3)  $\lambda x[good'(source' x) legumes']$

(Primes indicate interpretations whose detailed semantics is of no direct concern here.) When this function or concept is supplied with an argument *vitamins'*, it *reduces* to give a proposition, with the same function argument relations as the canonical sentence:

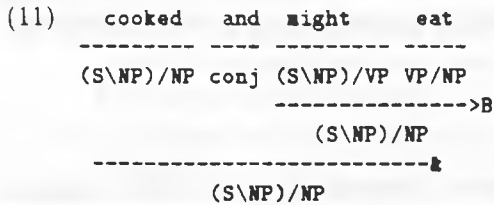
- (4)  $good'(source' vitamins')legumes'$

It is the presence of the above open proposition rather than some other that makes the intonation contour in (1) felicitous. (I am not claiming that its presence uniquely *determines* this response, nor that its explicit mention is necessary for interpreting the response.)

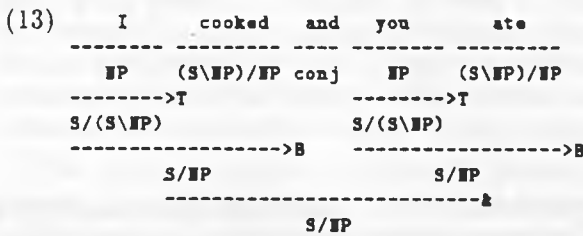
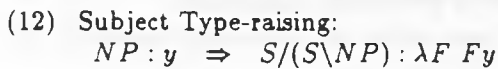
All natural languages include syntactic constructions whose semantics is also reminiscent of functional abstraction. The most obvious and tractable class are Wh-constructions themselves, in which exactly the same fragments that can be delineated by a single intonation contour appear as the residue of the subordinate clause. But another and much more problematic class are the fragments that result from coordinate constructions. It is striking that the residues of wh-movement and conjunction reduction are



The most important single property of combinatory rules like this is that they have an invariant semantics. This one composes the interpretations of the functions that it applies to, as is apparent from the right hand side of the rule.<sup>1</sup> Thus sentences like *I cooked, and might eat, the beans* can be accepted, via the following composition of two verbs (indexed as B, following Curry's nomenclature) to yield a composite of the same category as a transitive verb. Crucially, composition also yields the appropriate interpretation, assuming that a semantics is also provided for the coordination rule.



Combinatory grammars also include type-raising rules, which turn arguments into functions over functions-over-such-arguments. These rules allow arguments to compose, and thereby take part in coordinations like *I cooked, and you ate, the legumes*. They too have an invariant compositional semantics which ensures that the result has an appropriate interpretation. For example, the following rule allows the conjuncts to form as below (again, the remainder of the derivation is omitted):



## Intonation in a CCG.

Inspection of the above examples shows that Combinatory grammars embody an unusual view of surface structure, according to which strings like *Betty might eat* are constituents. In fact, according to this view, surface structure is a much more ambiguous affair than is generally realised, for they must also be possible constituents of non-coordinate sentences like *Betty might eat the mushrooms*, as well. (See [11] and [19] for a discussion of the obvious problems that this fact engenders for parsing written text.) An entirely unconstrained combinatory grammar would in fact allow more or less any bracketing on a sentence. However, the actual grammars we write for configurational languages like English are heavily constrained by local conditions. (An example would be a condition on the composition rule that is tacitly assumed here, forbidding the variable Y in the composition rule to be instantiated as NP, thus excluding constituents like *\*[eat the]<sub>VP/N</sub>*).

The claim of the present paper is simply that particular surface structures that are induced by the specific combinatory grammar that was introduced to explain coordination in English are identical to the intonational structures that are required to specify the possible intonation contours for those

<sup>1</sup>The rule uses the notation of the  $\lambda$ -calculus in the semantics, for clarity. This should not obscure the fact that it is functional composition itself that is the primitive, not the  $\lambda$  operator.

same sentences of English.<sup>2</sup> More specifically, the claim is that that in spoken utterance, intonation largely determines *which* of the many possible bracketings permitted by the combinatory syntax of English is intended, and that the interpretations of the constituents are related to distinctions of focus among the concepts and open propositions that the speaker has in mind. Thus, whatever problems for parsing written text arise from the profusion of equivalent alternative surface structures engendered by this theory, these "spurious" ambiguities seem to be to a great extent resolved by prosody in spoken language. The theory therefore offers the possibility that phonology and parsing can be merged into a single unitary process.

The proof of this claim lies in showing that the rules of combinatory grammar can be annotated with intonation contour schemata, which limit their application in spoken discourse, and to showing that the major constituents of intonated utterances like (1)b, under the analyses that these rules permit, correspond to the focus structure of the context to which they are appropriate, such as (1)a.

I shall use a notation which is based on the theory of Pierrehumbert [12], as modified in more recent work by Selkirk [16], Beckman and Pierrehumbert [2], [13], and Pierrehumbert and Hirschberg [14]. I have tried as far as possible to take my examples and the associated intonational annotations from those authors.

I follow Pierrehumbert in assuming two abstract pitch levels, and three types of tones, as follows. There are two phrasal tones, written H and L, denoting high or low "simple" tones — that is, level functions of pitch against time. There are also two boundary tones, written H% and L%, denoting an intonational phrase-final rise or fall. Of Pierrehumbert's six pitch accent tones, I shall only be concerned with two, the H\* accent and the L+H\*. The phonetic or acoustic realisation of pitch accents is a complex matter. Roughly speaking, the L+H\* pitch accent that is extensively discussed below in the context of the L+H\* L% melody generally appears as a maximum which is preceded by a distinctive low level, and peaks *later* than the corresponding H\* pitch accent when the same sequence is spoken with the H\* L melody that goes with "new" information, and which is the other melody considered below.

In the more recent versions of the theory, Pierrehumbert and her colleagues distinguish *two* levels of prosodic phrase that include a pitch accent tone. They are the intonational phrase proper, and the "intermediate phrase". Both end in a phrasal tone, but only intonational phrases have additional boundary tones H% and L%. Intermediate phrases are bounded on the right by their phrasal tone alone, and do not appear to be characterised in  $F_0$  by the same kind of final rise or fall that is characteristic of true intonational phrases. The distinction does not play an active role in the present account, but I shall follow the more recent notation of prosodic phrase boundaries in the examples, without further comment on the distinction.

There may also be parts of prosodic phrases where the fundamental frequency is merely interpolated between tones, notably the region between pitch accent and phrasal tone, and the region before a pitch accent. In Pierrehumbert's notation, such substrings bear no indication of abstract tone whatsoever.

A crucial feature of this theory for present purposes is that the position and shape of a given pitch accent in a prosodic phrase, and of its phrase accent and the associated right-hand boundary, are essentially invariant. If the constituent is very short — say, a monosyllabic nounphrase — then the whole intonational contour may be squeezed onto that one syllable. If the constituent is longer, then the pitch accent will appear at its left edge, the phrasal tone and boundary tone if any will appear at its right edge, and the intervening pitch contour will merely be interpolated. In this way, the tune can be spread over longer or shorter strings, in order to mark the corresponding constituents for the particular distinction of focus and propositional attitude that the melody denotes.

Consider for example the prosody of the sentence *Fred ate the beans* in the following pair of discourse

---

<sup>2</sup> There is a precedent for the claim that prosodic structure can be identified with the structures arising from the inclusion of associative operations in grammar in the work of Moortgat [9] and Oehrle [10], and in [?]

settings, which are adapted from Jackendoff [7, pp. 260]:

(14) Q: Well, what about the BEAns?  
Who ate THEM?

A: FRED ate the BEA-ns.  
H\*L L+H\*LE%

(15) Q: Well, what about FRED?  
What did HE eat?

A: FRED ate the BEAns.  
L+H\* LE% H\* LL%

In these contexts, the main stressed syllables on both *Fred* and *the beans* receive a pitch accent, but a different one. In (14), the pitch accent contour on *Fred* is H\*, while that on *beans* is L+H\*. (I base these annotations on Pierrehumbert and Hirschberg's [14, ex. 33] discussion of this example.)

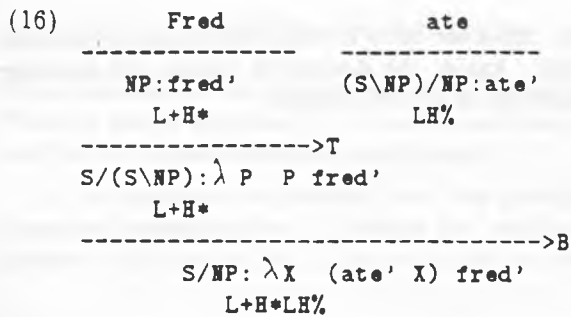
In the second example (15) above, the pitch accents are reversed: this time *Fred* is L+H\* and *beans* is H\*. The assignment of these tones seem to reflect the fact that (as Pierrehumbert and Hirschberg point out) H\* is used to mark information that the speaker believes to be *new to the hearer*. In contrast, L+H\* seems to be used to mark information which the current speaker knows to be given to the hearer (because the current hearer asked the original question), but which constitutes a novel topic of conversation for the speaker, standing in a contrastive relation to some *other* given information, constituting the previous topic. (If the information were merely given, it would receive *no* tone in Pierrehumbert's terms — or be left out altogether.) Thus in (15), the L+H\* LE% phrase including this accent is spread across the phrase *Fred ate*.<sup>3</sup> Similarly, in (14), the same tune is confined to the object of the open proposition *ate the beans*, because the intonation of the original question indicates that eating beans *as opposed to some other comestible* is the new topic.

## Syntax-driven Prosody.

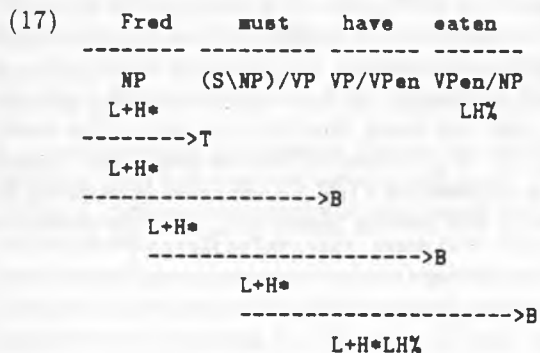
The L+H\* LE% intonational melody in example (15) belongs to a phrase *Fred ate ...* which corresponds under the combinatory theory of grammar to a grammatical constituent, complete with a translation equivalent to the open proposition  $\lambda x[(\text{ate}' x) \text{fred}']$ . The combinatory theory thus offers a way to assign contours like L+H\* LE% to such novel constituents, entirely under the control of independently motivated rules of grammar. For example, the rule of forward composition should be made subject to a restriction which is in the terms of Pierrehumbert's theory an extremely natural one, amounting to the straightforward injunction "Don't apply this rule across an intonational phrase or intermediate phrase boundary". The modified rule allows the following derivation for *Fred ate ...*, in which for once the semantic interpretation is included:<sup>4</sup>

<sup>3</sup> An alternative prosody, in which the contrastive tune is confined to *Fred*, seems equally coherent, and may be the one intended by Jackendoff. I believe that this alternative is informationally distinct, and arises from an ambiguity as to whether the topic of this discourse is *Fred* or *What Fred ate*. It is accepted by the present rules.

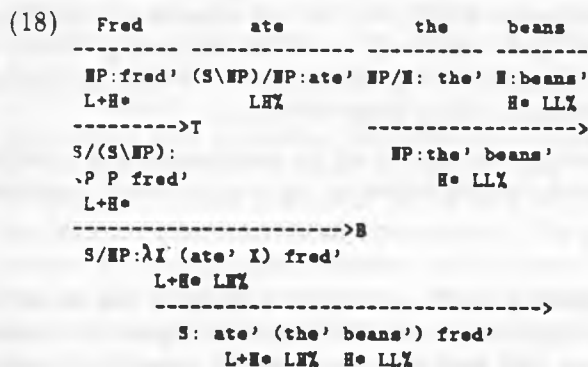
<sup>4</sup> Again primes indicate interpretations whose details are of no concern here. It will be apparent from the derivations that the assumed semantic representation is at a level prior to the explicit representation of matters related to quantifier scope.



The options incorporated in the tonal annotations of the rule allow the L+H\* LH% tune to spread across any sequence that can be composed by repeated applications of the rule. For example, if the reply to the same question *What did Fred eat?* is *FRED must have eaten the BEANS*, then the tune will typically be spread over *Fred must have eaten ...*, as in the following derivation, in which much of the syntactic and semantic detail has been omitted in the interests of brevity:



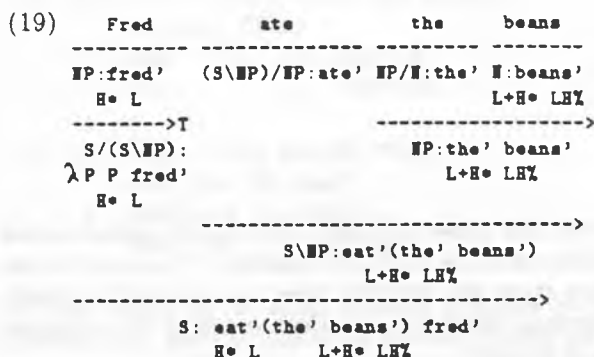
On the assumption that forward functional application bears a complementary restriction, and can combine any intonation contours to yield their concatenation, except when the leftmost is a bare phrasal tone or phrasal tone and boundary tone, the derivation of (15) can be completed as follows:



The division into contrastive/given open proposition versus new information is appropriate, and no other derivation is allowed, given this intonation contour. Repeated application of the composition rule, as in (17), would allow the L+H\* LH% contour to spread further, as in (*FRED must have eaten*) the BEANS.

In contrast, the intonation contour on (14) will not permit the annotated composition rule to apply, because *Fred* end with a L boundary intonation, so the bracketing imposed in (15) (and the formation of

the corresponding open proposition) is simply not allowed. However, since forward functional application is unrestricted, the following derivation of (14) is allowed. Again, the derivation divides the sentence into new and given information consistent with the context given in the example:



The effect of the rules is to annotate the entire predicate as an L+H\* LH%. It is emphasised that this does *not* mean that the *tone* is spread, but that the whole constituent is marked for the corresponding discourse function — roughly, as contrastive. The finer grain information that it is the object that is contrasted, while the verb is given, resides in the tree itself. Similarly, the fact that boundary tones are associated with words at the lowest level of the derivation does not mean that they are *part of* the word, nor that the word is the entity that they are a boundary *of*. It is prosodic phrases that they bound, and these also are defined by the tree. No other analysis is allowed for (19). Other cases considered by Jackendoff are considered in a more extended companion to the present paper [19], and are shown to yield only contextually appropriate interpretations.

## Conclusions.

The problem of so-called “spurious” ambiguity, or multiple semantically equivalent derivations, now appears in a quite different light. While the semantic properties of the rules (notably the associativity of functional composition that engenders the problem in the first place) do indeed guarantee that these analyses are semantically equivalent at the level of Argument Structure, they are nonetheless meaning-bearing at the level of Information Structure. To call them “spurious” is rather misleading. What is more, while there are usually a great many different analyses for any given sequence of words, intonation contour often limits or even eliminates the non-determinism arising from this source.

The significance of eliminating non-determinism in this way should not be under-estimated. Similar intonational markers are involved in coordinate sentences, like the following ‘right-node-raised’ example:

(20) I will, and you won't, eat mushrooms

In such sentences the local ambiguity between composing *won't* and *eat* and applying the latter to its argument first is a *genuine* local ambiguity, equivalent to a local attachment ambiguity in a more traditional grammar, for only *one* of the alternatives will lead to a parse at all. And the correct alternative is the one that is selected by the restriction against forward composition across prosodic phrase boundaries.

However, the extent to which intonation alone renders parsing deterministic should also not be overstated. There still are sources of non-determinism in the grammar, which must be coped with somehow. Most obviously, there are sources common to all natural language grammars, such as the well-known PP-attachment ambiguities in the following example:



(21) Put the block in the box on the table.

While intonation *can* distinguish the two analyses, they do not seem to be *necessarily* so distinguished. There is also a residuum of so-called spurious ambiguity, because function categories bearing *no* tone are free to forward compose *and* to apply.

It is important to observe that this ambiguity is widespread, and that it is a true ambiguity in discourse interpretation. Consider yet another version of the example with which the paper began, uttered with only an H\* LL% tune on the last word:

(22)

Legumes are a good source of Vitamins.  
H\* LL%

Such an intonation contour is compatible with *all* the analyses that the unannotated CCG would allow. However, such an utterance is also compatible with a large number of contextual open propositions. For example, it is a reasonable response to the question *What can you tell me about legumes?* But it is similarly reasonable as an answer to *What are legumes?*, or to *What are legumes a good source of?* The ambiguity of intonation with respect to such distinctions is well-known, and it would simply be incorrect not to include it. (See discussion in [1] and [8] for alternative proposals for ways of resolving it that are compatible with the present proposal.)

According to the present theory, the pathway between phonological form and interpretation is much simpler than has been thought up till now. Phonological Form maps directly onto Surface Structure, via rules of combinatory grammar annotated with abstract intonation contours. Surface Structure is identical to intonational structure, and maps directly onto Focus Structure, in which focussed and backgrounded entities and open propositions are represented by functional abstractions and arguments. Such structures reduce to yield canonical Function-Argument Structures. The proposal thus represents a return to the architecture proposed by Chomsky [3] and Jackendoff [7]. The difference is that the concept of surface structure has changed. It now really is *only* surface structure, supplemented by “annotations” which do nothing more than indicate the information structural status and intonational tune of *constituents* at that level.

While many problems remain, both in parsing written text with grammars that include associative operations, and at the signal-processing end, the benefits for automatic spoken language understanding are likely to be significant. Most obviously, where in the past parsing and phonological processing have delivered conflicting structural analyses, and have had to be pursued independently, they now are seen to be in concert. Processors can therefore be devised which use both sources of information at once, thus simplifying both problems. Furthermore, a syntactic analysis that is so closely related to the structure of the signal should be easier to use to “filter” the ambiguities arising from lexical recognition. What is likely to be more important in the long run, however, is that the constituents that arise under this analysis are also semantically interpreted. The paper has argued that these interpretations are directly related to the concepts, referents and themes that have been established in the context of discourse, say as the result of a question. The shortening and simplification of the path from speech to these higher levels of analysis offers the possibility of using those probably more effective resources to filter the proliferation of low level analyses as well.

## References

- [1] Altmann, Gerry and Mark Steedman: 1988, ‘Interaction with Context During Human Sentence Processing’ *Cognition*, 30, 191-238

- [2] Beckman, Mary and Janet Pierrehumbert: 1986, 'Intonational Structure in Japanese and English', *Phonology Yearbook*, 3, 255-310.
- [3] Chomsky, Noam: 1970, 'Deep Structure, Surface Structure, and Semantic Interpretation', in D. Steinberg and L. Jakobovits, *Semantics*, CUP, Cambridge, 1971, 183-216.
- [4] Curry, Haskell and Robert Feys: 1958, *Combinatory Logic*, North Holland, Amsterdam.
- [5] Halliday, Michael: 1967, *Intonation and Grammar in British English*, Mouton, The Hague.
- [6] Hepple, Mark, and Glyn Morrill: 1989, 'Parsing and Derivational Equivalence', Proceedings of the Fourth Conference of the European Chapter of the ACL, Manchester, April 1989, 10-18.
- [7] Jackendoff, Ray: 1972, *Semantic Interpretation in Generative Grammar*, MIT Press, Cambridge MA.
- [8] Marcus, Mitch, Don Hindle, and Margaret Fleck: 1983, D-theory: Talking about Talking about Trees, Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics, Cambridge Mass, June, 1983, 129-136.
- [9] Moortgat, Michael: 1988, *Categorial Investigations*, Foris, Dordrecht.
- [10] Oehrle, Richard T.: 1985, paper to the Conference on Categorial Grammar, Tucson, AR, June 1985, in Richard T. Oehrle, E. Bach and D. Wheeler, (eds), *Categorial Grammars and Natural Language Structures*, Reidel, Dordrecht, (in press).
- [11] Pareschi, Remo, and Mark Steedman. 1987. A lazy way to chart parse with categorial grammars, *Proceedings of the 25th Annual Conference of the ACL, Stanford*, July 1987, 81-88.
- [12] Pierrehumbert, Janet: 1980, *The Phonology and Phonetics of English Intonation*, Ph.D dissertation, MIT. (Distributed by Indiana University Linguistics Club, Bloomington, IN.)
- [13] Pierrehumbert, Janet, and Mary Beckman: 1989, *Japanese Tone Structure*, MIT Press, Cambridge MA.
- [14] Pierrehumbert, Janet, and Julia Hirschberg, 1987, 'The Meaning of Intonational Contours in the Interpretation of Discourse', ms. Bell Labs.
- [15] Prince, Ellen F. 1986. On the syntactic marking of presupposed open propositions. Papers from the Parasession on Pragmatics and Grammatical Theory at the 22nd Regional Meeting of the Chicago Linguistic Society, 208-222.
- [16] Selkirk, Elisabeth: *Phonology and Syntax*, MIT Press, Cambridge MA.
- [17] Steedman, Mark: 1987. Combinatory grammars and parasitic gaps. *NL&LT*, 5, 403-439.
- [18] Steedman, Mark: 1989, Structure and Intonation, ms. U. Penn.
- [19] Wittenburg, Kent: 1987, 'Predictive Combinators: a Method for Efficient Processing of Combinatory Grammars', *Proceedings of the 25th Annual Conference of the ACL, Stanford*, July 1987, 73-80.