INTONATION AND SYNTAX IN SPOKEN LANGUAGE SYSTEMS

Mark J. Steedman

MS-CIS-89-20
LINC LAB 146

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104

Revised
December 1989

ACKNOWLEDGEMENTS:
This research was supported in part by DARPA grant N00014-85-K-0018, NSF grants MCS-8219196-CER, IRI84-10413-A02 and US Army grants DAA29-84-K-0061, DAA29-84-9-0027.

Intonation and Syntax in Spoken Language Systems*

Mark Steedman
Computer and Information Science, U.Penn.

Phrasal intonation is notorious for a tendency to perceptually segment the word-string of a spoken utterance into groups which may violate orthodox syntactic 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, marked in this case by raised pitch, 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 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 in speech recognition and and speech synthesis.

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 [6] 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 Wilson and Sperber and E. Prince [15] have termed "open propositions". Open propositions are most easily understood as being that which is introduced into the discourse context by a Wh-question. 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 λ-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. (That is not to say 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

*The present paper is an expansion, including an entirely novel rule system, of unpublished presentations to the AAAI workshop on Spoken Language Systems, Stanford CA, 1989, and the Workshop on Parsing Technologies, CMU, August 1989.

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. Another and much more problematic class of fragments results from coordinate constructions. It is striking that the residues of wh-movement and conjunction reduction are also subject to something like a "sense unit condition". For example, strings like "in ten prefer cats" are not conjoinable:

(5) *Three doctors in ten prefer cats,
and in twenty eat carrots.

Since coordinate constructions have constituted another major source of complexity for theories of natural language grammar, it is tempting to think that this conspiracy between syntax and prosody might point to a unified notion of structure that is somewhat different from traditional surface constituency.

Combinatory Grammars.

Combinatory Categorial Grammar (CCG, [18]) is an extension of Categorial Grammar (CG). Elements like verbs are associated with a syntactic "category" which identifies them as *functions*, and specifies the type and directionality of their arguments and the type of their result:

(6) *eats* :- (S\NP)/NP: eat'

The category can be regarded as encoding the semantic type of their translation, which in the notation used here is identified by the expression to the right of the colon. Such functions can combine with arguments of the appropriate type and position by functional application:

(7)
```
Harry      eats      apples
------  ---------  ------
  NP    (S\NP)/NP     NP
        ------------->
             S\NP
      ---------------<
              S
```

Because the syntactic functional type is identical to the semantic type, apart from directionality, this derivation also builds a compositional interpretation, *eats'apples'harry'*, and of course such a "pure" categorial grammar is context free. Coordination might be included in CG via the following rule, allowing any constituents of like type, including functions, to form a single constituent of the same type:

(8) $X \quad conj \quad X \Rightarrow X$

(9)
```
I    cooked   and     ate    a frog
-- --------- ---- --------- ------
NP (S\NP)/NP conj (S\NP)/NP   NP
   ----------------------&
          (S\NP)/NP
```

(The rest of the derivation is omitted, being the same as in (7).) In order to allow coordination of contiguous strings that do not constitute constituents, CCG generalises the grammar to allow certain operations on functions related to Curry's combinators [4]. For example, functions may *compose*, as well as apply, under the following rule

(10) Forward Composition:
$X/Y : F \quad Y/Z : G \Rightarrow X/Z : \lambda x \; F(Gx)$

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.[1] 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 for the composite verb *might eat*:

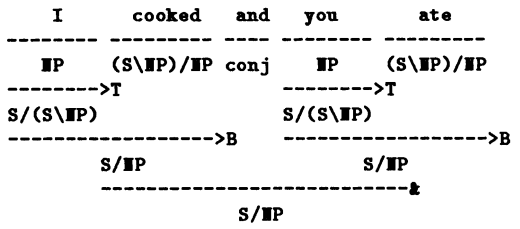
(11)
```
  cooked   and  might     eat
--------- ---- --------- -----
(S\NP)/NP conj (S\NP)/VP VP/NP
                ---------------->B
                   (S\NP)/NP
          ------------------------&
                (S\NP)/NP
```

[1] The rule uses the notation of the λ-calculus in the semantics, for clarity. This should not obscure the fact that it is functional composition itself that is the primitive, not the λ operator.

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):

(12) Subject Type-raising:
$NP : y \Rightarrow S/(S\backslash NP) : \lambda F\ Fy$

(13)



This apparatus has been applied to a wide variety of coordination phenomena (cf. [5], [17]).

Intonation in a CCG.

Examples like the above show that combinatory grammars embody a view of surface structure according to which strings like *Betty might eat* are constituents. In fact, according to this view, they must also be possible constituents of non-coordinate sentences like *Betty might eat the mushrooms*, as well. (See [11] and [20] for a discussion of the obvious problems that this fact engenders for parsing written text.) An entirely unconstrained combinatory grammar would in fact allow any bracketing on a sentence, although the grammars we actually write for configurational languages like English are heavily constrained by local conditions. (An example might 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 $*[eat\ the]_{VP/N}$).

The claim of the present paper is simply that particular surface structures that are induced by the specific combinatory grammar that are postulated to explain coordination in English subsume the intonational structures that are postulated by Pierrehumbert *et al.* to explain the possible intonation contours for sentences of English.[2] More specifically, the claim is that that in spoken utterance, intonation helps to determine *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.

The proof of this claim lies in showing that the rules of combinatory grammar can be made sensitive to intonation contour, which limit their application in spoken discourse. We must also show that the major constituents of intonated utterances like (1)b, under the analyses that are permitted by any given intonation, correspond to the focus structure of the context to which the intonation is appropriate are appropriate, as in (a) in the example (1) with which the paper begins. This demonstration will be quite simple, once we have established the folowing notation for intonation contours.

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. The theory proposed below is in principle compatible with any of the standard descriptive accounts of phrasal intonation. However, a crucial feature of Pierrehumberts theory for present purposes is that it distinguishes two subcomponents of the prosodic phrase, the *pitch accent* and the *boundary*.[3] The first of these tones or tone-sequences coincides with the perceived major stress or stresses of the prosodic phrase, while the second marks the right-hand boundary of the phrase. These two compo-

[2] 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 work by Moortgat [9] and Oehrle [10], and in [17, p. 540]

[3] For the purposes of this abstract, I am ignoring the distinction between the intonational phrase proper, and what Pierrehumbert and her colleagues call the "intermediate" phrase, which differ in respect of boundary tone-sequences.

nents are essentially invariant, and all other parts of the intonational tune are interpolated. Pierrehumberts theory thus captures in a very natural way the intuition that the same 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. It will help the exposition to augment Pierrehumberts notation with explicit prosodic phrase boundaries, using brackets. These do not change her theory in any way: all the information is implicit in the original notation.

Consider for example the prosody of the sentence *Fred ate the beans* in the following pair of discourse 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* LH% )`

(15) `Q: Well, what about FRED?`
`What did HE eat?`
`A: FRED ate the BEAns.`
`( L+H* LH% )( H* LL% )`

In these contexts, the main stressed syllables on both *Fred* and *the beans* receive a pitch accent, but a different one. In the former example, (14), there is a prosodic phrase on *Fred* made up of the pitch accent which Pierrehumbert calls `H*`, immediately followed by an `L` boundary. There is another prosodic phrase having the pitch accent called `L+H*` on *beans*, preceded by null or interpolated tone on the words *ate the*, and immediately followed by a boundary which is written `LH%`. (I base these annotations on Pierrehumbert and Hirschberg's [14, ex. 33] discussion of this example.)[4] In the second example (15) above, the two tunes are reversed: this time the tune with pitch accent `L+H*` and boundary `LH%` is spread across a prosodic phrase *Fred ate*, while the other tune with pitch accent `H*` and boundary `LL%` is carried by the prosodic phrase *the beans* (again starting with an interpolated or null tone.[5] Pierrehumbert and Hirschberg point out that the latter tune seems to be used to mark information that the speaker believes to be *new to the hearer*. In contrast, the `L+H* LH%` tune 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* LH%` phrase including this accent is spread across the phrase *Fred ate*.[6] 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* LH%` 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 $\lambda x[(ate'\ x)\ fred']$. The combinatory theory thus offers a way to derive such intonational phrases, using only the independently motivated rules of combinatory grammar, and under the control of appropriate intonation contours like `L+H* LH%`.

It is extremely simple to make make the existing combinatory rules do this. We need only specify two quite general principles that will govern the application of all combinatory rules to all intonated categories.

The first principle is so obvious as to hardly need stating. It simply says that the phonology borne by the result of applying a combinatory rule to two phonologically specified categories bears the concatenation of the two input phonological char-

[4] I continue to gloss over Pierrehumbert's distinction between "intermediate" and "intonational" phrases.

[5] The reason for notating the latter boundary as `LL%`, rather than `L` is again to do with the distinction between intonational and intermediate phrases.

[6] 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 too is accepted by the rules below.

acterisations (including their intonation contours).

The second principle is more interesting. It means that, while combinatory rules can apply to prosodic constituents at all levels, they can only apply to *complete* prosodic constituents. The condition is imposed by the following general rule:

(16) **The Prosodic Constituent Condition:** If a rule combining two categories applies across a prosodic phrase boundary at any level, then the categories must be complete prosodic phrases at that level.

It follows from this rule that if the leftmost intonational tune ends in an intonational or intermediate phrase boundary, then the leftmost category must be a complete phrase including the *left* boundary, and that the rightmost combining category must also be a complete prosodic phrase. The rule therefore has the interesting effect of making intonational/intermediate phrase boundaries block combinations that would otherwise be allowed, and of only permitting the derivations which deliver interpretations that are appropriate to the intonation contours in question.

For example, consider the derivations that it permits for example (15) above. The rule of forward composition is allowed to apply to the words *Fred ate ...*, because there is no intonational/intermediate phrase boundary at the end of *Fred*:[7]

(17)
```
        Fred                ate
---------------    --------------
   NP:fred'          (S\NP)/NP:ate'
   ( L+H*                 LH% )
----------------->T
S/(S\NP):λP  P fred'
   ( L+H*
---------------------------------->B
       S/NP: λX  (ate' X) fred'
        ( L+H*LH% )
```

The prosodic constituent condition also allows 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:

(18)
```
   Fred      must     have    eaten
--------  ---------  -------  -------
   NP     (S\NP)/VP VP/VPen VPen/NP
( L+H*                          LH% )
-------->T
( L+H*
------------------>B
   ( L+H*
     --------------------->B
          ( L+H*
            --------------------->B
                 ( L+H*LH% )
```

The presence of a boundary at the end of the sequence *Fred ate ...* in (15) implies the presence of a left boundary at the start of *the beans*. The Prosodic constituent condition therefore allows the derivation of (15) to be completed as follows:

(19)
```
  Fred         ate           the      beans
--------- -------------- ---------- --------
 NP:fred' (S\NP)/NP:ate' NP/N: the' N:beans'
( L+H*         LH%     )(            H* LL% )
--------->T              ------------------>
S/(S\NP):                  NP:the' beans'
λP P fred'                     H* LL% )
( L+H*
------------------------>B
 S/NP:λX (ate' X) fred'
     ( L+H* LH% )
       --------------------------------->
          S: ate' (the' beans') fred'
            ( L+H* LH%  H* LL% )
```

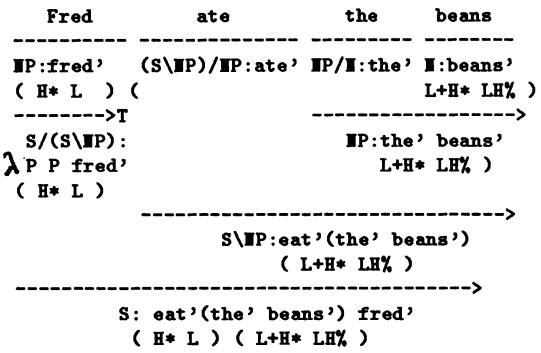
The division into contrastive/given open proposition versus new information is appropriate. Moreover, the prosodic constituent condition permits no *other* derivation for this intonation contour. Repeated application of the composition rule, as in (18), would allow the L+H* LH% contour to spread further, as in *(FRED must have eaten) the BEANS.*

In contrast, the parallel derivation is forbidden

[7] Again, the semantic annotations simply identify interpretations that are implicit in the categories themselves. 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.

by the prosodic constituent condition for the alternative intonation contour on (14). Instead, the following derivation, excluded for the previous example, is now allowed:

(20)



No other analysis is allowed for (20). Again, the derivation divides the sentence into new and given information consistent with the context given in the example. The effect of the derivation 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 given, or theme. 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 sequences are associated with words at the lowest level of the derivation does not mean that they are *part of* the word, or specified in the lexicon, 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.

All the other possibilities for combining these two contours considered by Jackendoff can be shown also to yield unique and contextually appropriate interpretations.

The full paper will also discuss sentences bearing only a single intonational phrase, such as the following:

(21)

They're a good source of PROtein
H* LL%

Such sentences are notoriously ambiguous as to the open proposition they presuppose. They therefore require a generalisation of the theory presented above, to allow syntactic and information structural boundaries that are not explicitly marked in phonology, in association with unmarked given contextual information. This generalisation is spelled out in the full paper. With the generalisation, we are in a position to make the following claim:

(22) The structures demanded by the theory of intonation and its relation to contextual information are the same as the surface syntactic structures permitted by the combinatory grammar.

A number of predictions concerning the relation of intonation structures and coordinate structures are shown to follow.

Conclusion.

According to the present theory, the pathway between phonological form and interpretation is much simpler than has been thought hitherto. Phonological form maps directly onto surface structure, annotated with abstract intonation contours identifying their discourse function, via the rules of combinatory grammar. Surface structure therefore subsumes intonational structure. It also subsumes information structure. Focussed and backgrounded entities and open propositions are represented by the functional abstractions and arguments which the grammar associates with the top-level of surface constituents as their interpretations. These reduce to yield canonical function-argument structures. The proposal thus in a sense 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.

The full paper concludes by discussing the implications of the theory for discourse-model-driven synthesis and analysis of spoken language by machine.

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] Dowty, David: 1988, Type raising, functional composition, and non-constituent coordination, in Richard T. Oehrle, E. Bach and D. Wheeler, (eds), *Categorial Grammars and Natural Language Structures*, Reidel, Dordrecht, 153–198.

[6] Halliday, Michael: 1967, *Intonation and Grammar in British English*, Mouton, The Hague.

[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: 1985a. Dependency and Coordination ... Language 61.523-568.

[18] Steedman, Mark: 1987. Combinatory grammars and parasitic gaps. NL&LT, 5, 403-439.

[19] Steedman, Mark: 1989, Structure and Intonation, ms. U. Penn.

[20] 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.