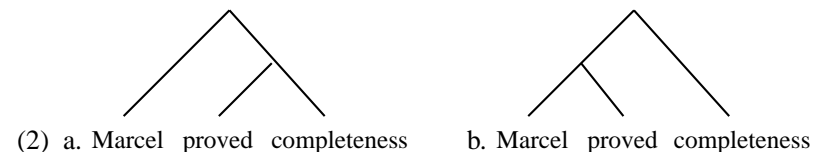


LCTG—Notes Part 8: Intonation and Information Structure

Is Intonation Orthogonal to Syntax?

- It certainly looks that way.
 - (1) a. (The absent minded professor)(was avidly reading)(about the latest biography)(of Marcel Proust).
 - b. (Marcel proved) (completeness).
- Intonation appears to regard subject and transitive verb as equally phrasal to the traditional transitive predicate:



1

2

Is Intonation Orthogonal to Syntax (Contd.)?

- But intonational restructuring remains subject to some syntactic restrictions:
 - (3) a. *(Three mathematicians)(in ten derive a lemma).
 - b. *(Seymour prefers the nuts)(and bolts approach).
 - c. *(Gilbert met a woman who prefers)(corduroy).
- Hence the “Sense Unit Condition” on intonational phrases of Halliday (1967) and Selkirk (1984).

Apparent Consequences for Theoretical Architecture

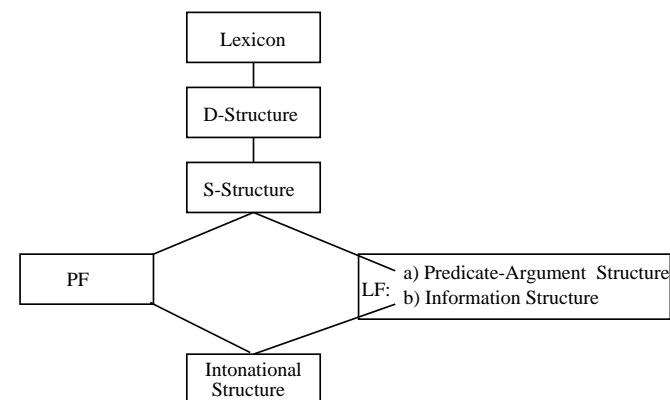


Figure 1

Architecture of a GB theory of Prosody (adapted from Selkirk 1984)

3

4

Intonation and Information Structure

- A minimal pair of contexts and contours:

(4) Q: I know who proved soundness. But who proved COMPLETENESS?

A: (MARCEL) (proved COMPLETENESS).

H*L L+H* LH%

(5) Q: I know which result Marcel PREDICTED. But which result did Marcel PROVE?

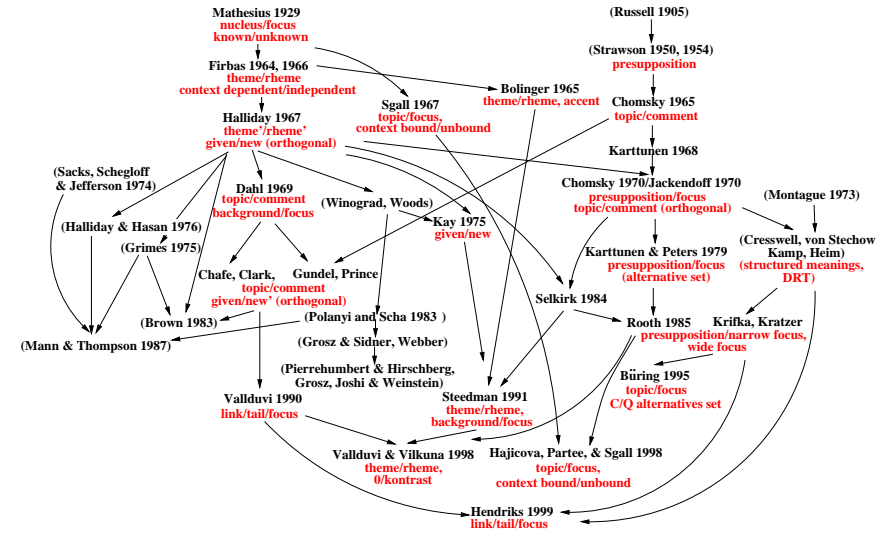
A: (Marcel PROVED) (COMPLETENESS).

L+H* LH% H* LL%

- Claim: L+H* LH% marks **theme**. H* L and H* LL% mark **rheme**.
- Crossing contexts and responses yields complete incoherence.

5

A Brief History of Information Structure



6

The Two Dimensions of Information Structure

- A slightly more complex example:

(6) Q: I know that Marcel likes the man who wrote the musical.

But who does he ADMIRE?

A: (Marcel ADMIRES) (the woman who DIRECTED the musical)

$\underbrace{\quad \quad \quad L+H* LH% \quad \quad \quad}_{\text{Background}} \quad \underbrace{\quad \quad \quad H* \quad \quad \quad}_{\text{Focus}} \quad \underbrace{\quad \quad \quad LL% \quad \quad \quad}_{\text{Background}}$
 $\underbrace{\quad \quad \quad \text{Theme} \quad \quad \quad}_{\text{Background}} \quad \underbrace{\quad \quad \quad \text{Rheme} \quad \quad \quad}_{\text{Focus}} \quad \underbrace{\quad \quad \quad \text{Background} \quad \quad \quad}_{\text{Background}}$

- Theme and Rheme are from Firbas and Bolinger, rather than Halliday.
- Background and Focus are from Dahl, equivalent to Hallidean given and new
- We make the Bolingerian assumption that all focus is contrastive in nature and applies to both theme and rheme (equivalently, “contrastive focus” is just “focus”).

7

An Aside on the Unmarked Theme

- Such richly intoned examples are exceptional. But the unmarked default contour is simply ambiguous as to information structure.

(7) Q: Which result did Marcel prove?

A: (Marcel proved) (COMPLETENESS).

H* LL%

(8) Q: What do you know about Marcel?

A: (Marcel) (proved COMPLETENESS).

H* LL%

(9) Guess what? (Marcel proved COMPLETENESS!)

H* LL%

8

Theme and Rheme as λ -terms

- The interpretation of the theme can be thought of as the following λ -abstract:

(10) $\lambda x.admire'x marcel'$

- This theme reduces with the rheme to yield a predicate-argument structure which we can think of extensionally as reducing to the following canonical proposition:

(11) $admire'woman_1'marcel'$

9

Alternative Semantics: The Rheme Alternative Set

- The reduction of theme with rheme can be thought of as contrasting the stated proposition with a Rheme Alternative Set of alternative propositions about Marcel admiring contextually accessible people, which we can represent extensionally as follows:

(12) $\left\{ \begin{array}{l} admires'woman_1'marcel' \\ admires'woman_2'marcel' \\ admires'man_1'marcel' \end{array} \right\}$

- cf. Rooth's C

10

Alternative Semantics: The Theme Alternative Set

- The L+H* pitch accent in the theme marks it as contrasted with other contextually available themes:

(13) $\lambda x.*admires'x marcel'$

(14) $\lambda x.likes'x marcel'$

The set of available themes can also be represented as an alternative set, the Theme Alternative Set over—cf. Rooth and Büring's Q:

(15) $\left\{ \begin{array}{l} \lambda x.admires'x marcel' \\ \lambda x.likes'x marcel' \end{array} \right\}$

- As in the case of the Rheme, the position of the pitch accent determines the dimensions over which the Alternative set may vary.

11

Discontinuous Themes and Rhemes

- If we think of the process of restriction over alternative sets as intersection, then multiple themes like the following have the effect of a single discontinuous theme:

(16) Q: I know what Marcel SOLD to HARRY. But what did he GIVE to FRED?

A: (Marcel GAVE) (a BOOK) (to FRED).

L+H* LH% H* L L+H* LH%

- The two themes are the following lambda abstracts:

(17) $\lambda x.\lambda y.*give'y x marcel'$

(18) $\lambda p.\lambda x.p *fred'x$

12

Discontinuous Themes and Rhemes (Contd.)

- The two rheme alternative sets are events of Marcel giving things to people and events of Fred receiving things. If these sets are intersected then they yield the set of events of Marcel giving things to Fred, just as if there was the following single theme:

(19) $\lambda x.*give' *fred' x marcel'$

(20) $\left\{ \begin{array}{l} give' fred' book' marcel' \\ give' fred' record' marcel' \\ give' fred' biscuit' marcel' \end{array} \right\}$

13

Is L+H*LH% Always a Theme?

- Examples like (22) seem very similar to those like (21). In particular both implicate a reply via world knowledge about opera-lovers and likers of musicals, rather than stating it directly:

(21) Q: Does Marcel love opera?

A: Marcel likes MUSICALS.

H* LL%

(22) Q: Does Marcel love opera?

A: Marcel likes MUSICALS.

L+H* LH%

- The latter have sometimes been described as rhematic, differing from (21) only in degree of commitment.

14

Is L+H*LH% Always a Theme? (Contd.)

- **Either** affirmation **or** denial may be implicated by both (21) and (22): which it is depends on your beliefs about the relation between opera-loving and musical-liking.
- In the case of affirmation, clearly neither (21) nor (22) entails denial of alternatives to liking musical.
- Any lack of commitment in (22) concerns the implicature, not the liking of musicals.
- Therefore it is equally possible that (22) states an **isolated theme**, and thereby implicates a rheme without stating it. The lack of commitment is implicated by their leaving the crucial inference to be drawn by the hearer.

15

H* Consistently Marks Rheme

- Claim: H* always marks rheme, whatever the boundary.

(23) There's ORANGE juice, and APPLE juice.

H* LH% H* LL%

(24) a) Your LUNCH is ready! b) Your LUNCH is ready!

H* LH% H* LL%

16

L+H* Consistently Marks Theme

- Claim: L+H* always marks theme, whatever the boundary.
(25) a. Q: Why did Marcel want to prove completeness?
A: THAT was the whole POINT of the exercise!
H* L L+H* LL%
- b. Q: I know that one result of the exercise was to prove completeness.
But what was the whole point of the exercise?
A: THAT was the whole POINT of the exercise!
H* L L+H* LH%

17

The Boundary

- The contribution of the boundary seems to be to do with speaker's claims concerning the "ownership" of or "commitment to" (Gunlogson 2001) the theme or rheme by speaker or hearer.
- LH% in (25b) seems to mean "I got this theme from the *hearer*." In (25a) LL% seems to mean "this theme is *mine* (and not this stupid hearer's).
- This distinction needs firming up, but appears to be particularly marked in US English:

18

The Boundary Contd.

- (26) BARTENDER: What can I get you?
CLIENT: #Can I get a Sam Adams?
H* H*LH%
- (27) BARTENDER: What can I get you?
CLIENT: Can I get a Sam Adams?
H* H*HL%
- The implication of H% in the first of these that the rheme is already the bartender's makes it impatient and therefore impolite in US (but not British) English (Ladd 1996). an L% boundary is appropriate.

19

Conflicts Between Contexts and Tones

- We have already noted that swapping contexts and replies in (4) and (5) yields incoherence. Many other conflicts are possible
- (28) Q: I know who proved soundness. But who proved COMPLETENESS?
A: #(MARCEL) (proved COMPLETENESS).
H*LH% L+H* LL%
- (29) Q: I know which result Marcel PREDICTED. But which result did Marcel PROVE?
A: #(Marcel PROVED) (COMPLETENESS).
L+H* L H* LH%
- Both of these fail because the theme is marked as speaker's when it is in fact hearer's, and because the rheme is marked as hearer's when by definition it cannot be.

20

What About The Other Four Pitch Accents?

- L* is a rheme marker parallel to H*.

(30) Do PRUNES have FEET? (Surely not!)

L* L* LH%

(31) Do PRUNES have FEET? (How wonderful!)

H* H* LH%

(32) There's ORANGE juice, and APPLE juice.

L* LH% H* LL%

- L* seems to differ from H* along a dimension of *denial*.

21

What About The Other Four Pitch Accents?

- H*+L, and H+L* (which may not be distinct from H* and L*) are also rheme markers:

(33) JEREMY! (Stop that!)

H*+L LH%

(34) Do PRUNES have FEET? (I really cannot believe this.)

H+L* H+L* LH%

22

What About The Other Four Pitch Accents?

- L*+H is a theme marker, like L+H* (which it may or may not be distinct from)

- The following (from Ward and Hirschberg) seems comparable in every way to the “all theme” utterance (22).

(35) A: Harry's such a klutz.

B: He's a good BADMINTON player.

L*+H LH%

- the effect of a low boundary seems comparable to that in (25a):

(36) A: Harry's such a klutz.

B: He's a good BADMINTON player.

L*+H LL%

23

- Compare:

(37) A: Harry is such a klutz.

B: He's a good BADMINTON player.

L+H* LH%

(38) A: Harry is such a klutz.

B: He's a good BADMINTON player.

L+H* LL%

24

- As in the opposition of other H*/L* pairs, L*+H seems to carry connotations of denial or negative polarity:

(39) A: Who ate the whole thing?

B: FRED ate MOST of it.

H*L L+H* LH%

(40) A: I can't believe you ate the whole thing!

B: FRED ate MOST of it.

H*L L*+H LH%

25

Combinatory Prosody

- The above alternative derivations provide exactly what is needed to support the phonologists' notion of intonation structure.
- In particular, the derivations deliver lambda terms that correspond semantically to units of Information Structure.
- Thus we can bring notions like Topic (here, Theme) and Comment/Focus (here, Rheme) into compositional semantics and competence grammar.
- Here is one way to do this (work with Prevost, and see Steedman, 2000b).

27

Surface Structure in CCG

$$\begin{array}{c}
 (41) \quad \frac{\frac{\text{Marcel}}{NP : marcel'}}{\frac{S/(S \setminus NP) : \lambda f.f \text{ marcel}'}{S/NP : \lambda x.prove' x \text{ marcel}'}} \xrightarrow{T} \frac{\frac{\text{proved}}{(S \setminus NP)/NP : prove'}}{S \setminus (S/NP) : \lambda p.p \text{ completeness}'} \\
 \xrightarrow{B} \frac{}{S : prove' \text{ completeness}' \text{ marcel}'} \leftarrow
 \end{array}$$

$$\begin{array}{c}
 (42) \quad \frac{\frac{\text{Marcel}}{NP : marcel'}}{\frac{S/(S \setminus NP) : \lambda f.f \text{ marcel}'}{S \setminus NP : \lambda y.prove' \text{ completeness}' y}} \xrightarrow{T} \frac{\frac{\text{proved}}{(S \setminus NP)/NP : prove'}}{(S \setminus NP) \setminus ((S \setminus NP)/NP) : \lambda p.p \text{ completeness}'} \\
 \xrightarrow{} \frac{}{S : prove' \text{ completeness}' \text{ marcel}'} \leftarrow
 \end{array}$$

26

Combinatory Prosody: Pitch-Accents

- (43) θ -markers: L+H*, L*+H
 ρ -markers: H*, L*, H*+L, H+L*
- (44) $\text{proved} := (S_\rho \setminus NP_\rho)/NP_\rho : \lambda x.\lambda y. * \text{prove}' xy$
H*
- (45) $\text{proved} := (S_\eta \setminus NP_\eta)/NP_\eta : \lambda x.\lambda y. \text{prove}' xy$

28

Combinatory Prosody: Boundaries

$$(51) L, H := S\$_t \backslash S\$_\eta : \lambda f. \eta' f$$

$$(52) L\% := (S\$_\phi \backslash S\$_\eta) \backslash (S\$_t \backslash S\$_\eta) : \lambda f. \lambda g. [S](fg)$$

$$H\% := (S\$_\phi \backslash S\$_\eta) \backslash (S\$_t \backslash S\$_\eta) : \lambda f. \lambda g. [H](fg)$$

$$(53) \begin{array}{c} \text{Marcel} \quad \text{PROVED} \quad \text{L} \quad \text{H\%} \quad \text{COMPLETENESS} \\ \text{L+H*} \quad \text{L} \quad \text{H*} \quad \text{L} \quad \text{L\%} \\ \hline S/(S \backslash NP) \quad (S_\theta \backslash NP_\theta) / NP_\theta \quad S\$_t \backslash S\$_\eta \quad (S\$_\phi \backslash S\$_\eta) \backslash (S\$_t \backslash S\$_\eta) \\ \lambda p.p \text{ marcel}' : \lambda x. \lambda y. *prove' xy \quad : \lambda f. \eta' f \quad : \lambda f. \lambda g. [H](fg) \\ \hline S_\theta / NP_\theta : \lambda x. *prove' x \text{ marcel}' \quad S\$_\phi \backslash S\$_\eta : \lambda f. [H](\eta' f) \\ \hline S_\phi / NP_\phi : [H](\theta'(\lambda x. *prove' x \text{ marcel}')) \end{array}$$

33

Combinatory Prosody: Theme and Rheme

$$(54) \lambda x. \text{predict}' x \text{ marcel}'$$

$$(55) \begin{array}{c} \text{Marcel} \quad \text{PROVED} \quad \text{COMPLETENESS} \\ \text{L+H*} \quad \text{L} \quad \text{H\%} \quad \text{H*} \quad \text{L} \quad \text{L\%} \\ \hline S_\phi / NP_\phi \quad S_\rho \backslash (S_\rho / NP_\rho) \quad S\$_t \backslash S\$_\eta \quad (S\$_\phi \backslash S\$_\eta) \backslash (S\$_t \backslash S\$_\eta) \\ : [H](\theta'(\lambda x. *prove' x \text{ marcel}')) \quad \lambda p.p *completeness' \quad : \lambda f. \eta' f \quad : \lambda f. \lambda g. [S](fg) \\ \hline S_\phi \backslash S\$_\eta : \lambda f. [S](\eta' f) \\ \hline S_\phi \backslash (S_\phi / NP_\phi) : [S](\rho'(\lambda p.p *completeness')) \\ \hline S_\phi : [S](\rho'(\lambda p.p *completeness'))([H](\theta'(\lambda x. *prove' x \text{ marcel}')))) \\ \hline S_\phi : *prove' *completeness' \text{ marcel}' \end{array}$$

34

Alternative Semantics Revisited

$$(56) \lambda x. \text{prove}' \text{ soundness}' x$$

$$(57) \begin{array}{c} \text{MARCEL} \quad \text{proved} \quad \text{COMPLETENESS} \\ \text{H*} \quad \text{L} \quad \text{L+H*} \quad \text{L} \quad \text{H\%} \\ \hline S_\rho / (S_\rho \backslash NP_\rho) \quad S\$_t \backslash S\$_\eta \quad S_\theta \backslash NP_\theta \quad S\$_t \backslash S\$_\eta \quad (S\$_\phi \backslash S\$_\eta) \backslash (S\$_t \backslash S\$_\eta) \\ \lambda p.p *marcel' : \lambda f. \eta' f \quad : \lambda x. \text{prove}' *completeness' x \quad : \lambda f. \eta' f \quad : \lambda f. [H]f \\ \hline S_t / (S_t \backslash NP_t) \quad S\$_\phi \backslash S\$_\eta \\ : \rho'(\lambda p.p *marcel') \quad : \lambda f. [H]\eta' f \\ \hline S_\phi \backslash NP_\phi \\ : [H](\theta'(\lambda x. \text{prove}' *completeness' x)) \\ \hline S_\phi : \rho'(\lambda p.p *marcel')([H](\theta'(\lambda x. \text{prove}' *completeness' x))) \\ \hline S_\phi : \text{prove}' *completeness' *marcel' \end{array}$$

35

The Sense Unit Condition Follows

$$(58) \text{ a. } *(Three \text{ MATHEMATICIANS}) (\text{in ten derive a LEMMA}).$$

$$\text{L+H*} \quad \text{LH\%} \quad \text{H*} \quad \text{LL\%}$$

$$\text{ b. } *(Seymour prefers the NUTS) (\text{and bolts APPROACH}).$$

$$\text{L+H*} \quad \text{LH\%} \quad \text{H*} \quad \text{LL\%}$$

$$\text{ c. } *(They only asked whether I knew the woman who CHAIRS) (\text{the ZONING board}).$$

$$\text{L+H*} \quad \text{LH\%} \quad \text{H*} \quad \text{LL\%}$$

$$(59) \text{ a. } *Three \text{ mathematicians in ten derive a lemma and in a hundred can cook a boiled egg.}$$

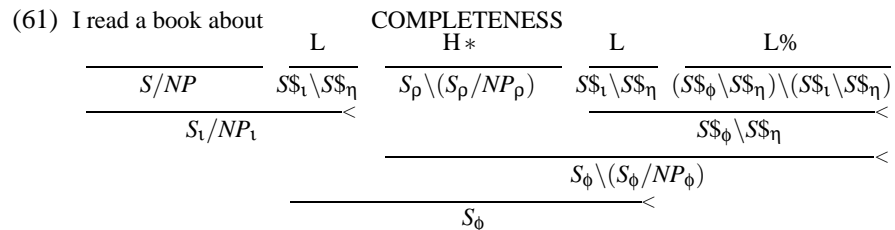
$$\text{ b. } *The \text{ nuts which Seymour prefers and bolts approach}$$

$$\text{ c. } *Which \text{ boards did they ask whether you knew the woman who chairs?}$$

36

Unmarked Theme Revisited

- (60) a. (I read a book about)_{Theme}(COMPLETENESS)_{Rheme}
 b. (I read)_{Theme}(a book about COMPLETENESS)_{Rheme}
 c. (I)_{Theme}(read a book about COMPLETENESS)_{Rheme}
 d. (I read a book about COMPLETENESS)_{Rheme}



The Sense Unit Condition Still Follows

- (62) a. *[Three mathematicians]_{Theme} [in ten derive a LEMMA]_{Rheme}
 b. *[Seymour prefers the nuts]_{Theme} [and bolts APPROACH]_{Rheme}
 c. *[They only asked whether I knew the woman who chaired]_{Theme}
 [the ZONING board]_{Rheme}

There are More Unmarked Themes Than You Think

- (63) (MARCEL) (proved completeness) .

H* L LL%

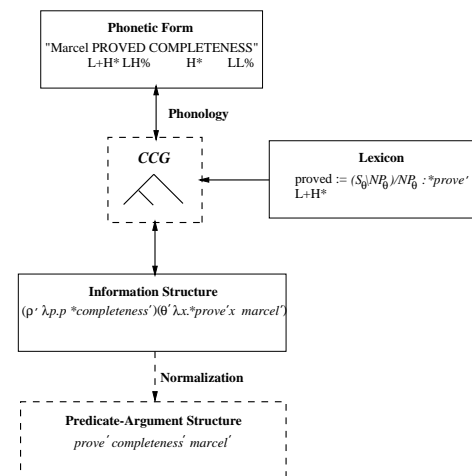
- (64) (Marcel) (PROVED) (COMPLETENESS) , and won a PRIZE .

LL+H* LH% H* LL% H* LL%

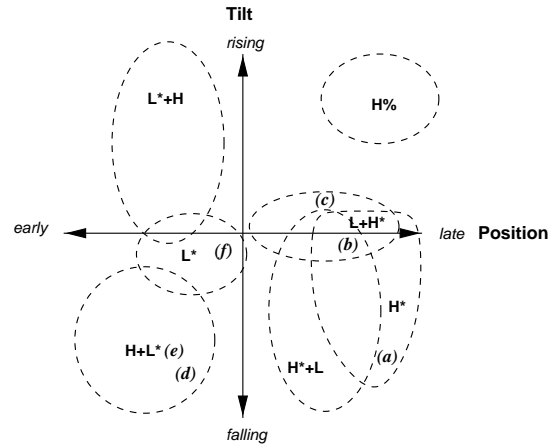
- (65) Your TROUSERS are on fire!

H* LL%

Architecture of a CCG theory of Prosody



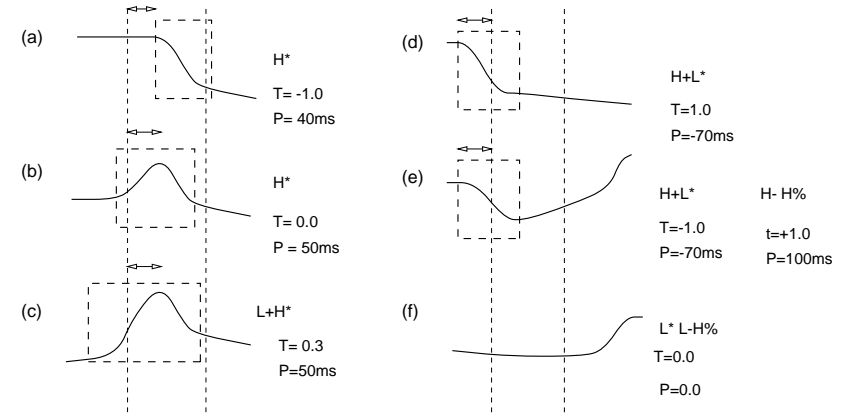
Can Tones be Distinguished for Practical Speech Recognition?



ToBI Hand Annotations of pitch contours characterized in Tilt Space (from Taylor 2000)

41

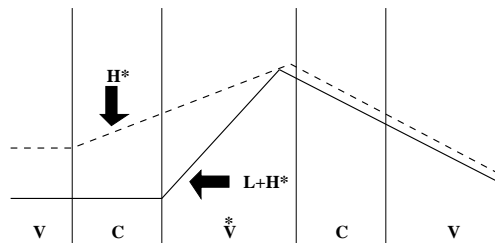
This is a Problem of Annotation, not of the Tone System



Examples of Tilt Categorization of results of TOBI hand-annot (from Taylor 2000)

42

Pilot Study of Elicited Theme and Rheme Contour



43

Can We Achieve Large Coverage Robust CCG Parsing?

- Rumors of Intractability arising from “spurious ambiguity” in CCG have been greatly exaggerated (Karttunen 1989; Komagata 1999, Hockenmaier et al. 2000).
- The currently most successful large coverage statistical parsers (Collins 1998; Charniak 1999) work by exploiting a statistical model based on dependencies between heads of predicates and heads of their arguments. CCG is a lexicalized grammar that is ideally suited to application of this technique using lexicons and models induced from labeled data.
- It is certainly possible in principle to apply such techniques to suitably labeled speech corpora, and realistic to expect that tones could be recognized, given segments or syllable boundaries.
- Pitch accents are highly ambiguous—but so is everything else. That’s what statistics is *for*.

44

References

Büring, Daniel: 1995, "The 59th Street Bridge Accent: On the Meaning of Topic and Focus," PhD Thesis, Universität Tübingen.

Hockenmaier, Julia, Gann Bierner and Jason Baldridge: 2000. Providing robustness for a CCG system <http://www.cogsci.ed.ac.uk/~julia/>

Gunlogson, Christine: 2001, "Rising Declarative Questions", <http://www.semanticsarchive.net>.

Hockenmaier, Julia: 2001, "Statistical Parsing for CCG with Simple Generative Models," *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, July, Toulouse, France. (To appear)

Pierrehumbert, Janet: 1980, *The Phonology and Phonetics of English Intonation*, Ph.D dissertation, MIT. (Dist. by Indiana University Linguistics Club, Bloomington, IN.)

Pierrehumbert, Janet, and Julia Hirschberg: 1990, "The Meaning of Intonational Contours in the Interpretation of Discourse," in Philip Cohen, Jerry Morgan, and Martha Pollack (eds.), *Intentions in Communication*, MIT Press Cambridge MA, 271-312.

Prevost, Scott and Mark Steedman: 1994, "Specifying Intonation from Context for Speech Synthesis," *Speech Communication*, 15:139-153.

Rooth, Mats: 1992, "A Theory of Focus Interpretation," *Natural Language Semantics*, 1:75-116.

Steedman, Mark: 1991a, Structure and Intonation, *Language*, 68, 260-296.

Steedman, Mark: 2000a, *The Syntactic Process*, MIT Press, Cambridge MA.

Steedman, Mark: 2000b, "Information Structure and the Syntax-Phonology Interface," *Linguistic Inquiry*, 31.4, 649-689.

Taylor, Paul: 2000, "Analysis and Synthesis of Intonation Using the Tilt Model", *Journal of the Acoustical Society of America*, 107, 1697-1714.