Intonation, Grammar, and Spoken Language Processing

Mark Steedman: *Informatics, U. of Edinburgh*

http://www.iccs.informatics.ed.ac.uk/~steedman/papers.html

Toronto, Mar 2002

Content-to-Speech-Generation for Spoken Language Technology

- Intonation is Not a Toy Problem: Translating Telephones (etc.)

SPEAKER1: “How about on the Monday afternoon?”

TT1: *Ginge es am Montag nachmittag?*

SPEAKER2: “Montag nachmittag habe ich leider keine Zeit.”

TT2: *I unfortunately have no time on Monday afternoon.*

- **Passed through Standard Text-to-Speech:**

#I unfortunately have no time on Monday AFTERNOON

H*LL%

Translating Telephones (etc.) (contd.)

Quick fix #1: use Previous Mention Heuristic (Hirschberg) on the source/translation of the previous turn.

I	unfortunately	have	no	TIME	on	Monday	afternoon
				H*			LL%

- But there may BE no previous literal mention:

  "An diesem Tag habe ich leider keine Zeit"

  I unfortunately have no time on that day

#I	unfortunately	have	no	time	on	that	DAY
							H*LL%

Translating Telephones (etc.) (contd.)

Quick fix #2: Preserve the pattern of pitch accents from the German input (Stöber & Wagner).

- But there may be a "topic" pitch accent (Büring) on "diesem Tag":

An	DIESEM	Tag	habe ich leider	keine	ZEIT
	L*+H	H			H+L* LL%

- The L*+H H sequence on the (syntactically defined) topic *An diesem Tag* should be realised in English as an L+H* LH% "theme" tune (or as an unmarked theme with no pitch accent):

I	unfortunately	have	no	TIME	on	THAT	day
				H*L		L+H*	LH%

Translating Telephones (etc.) (contd.)

- This amounts to an analysis of **Information Structure** via the grammar. It seems possible that one might do quite well with shallow translation of information structure, at least for English and German.
- Komogata's 1999 Penn Thesis on topic and focus in Japanese suggests that this might generalize.
- *But it needs a rather strange kind of grammar.*

The Problem: Intonation seems to be Independent of Syntax . . .

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

A: (MARCEL) (proved COMPLETENESS).

H*L L+H* LH%

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

A: (Marcel PROVED) (COMPLETENESS).

L+H* LH% H* LL%

- —Hence Halliday and Selkirk's introduction of an autonomous level of **Intonation Structure** with attendant **Sense-Unit Condition**.

. . . But Intonation is Not All *That* Independent from Syntax

(3) a. *(Three mathematicians)(in ten derive a lemma).
b. *(Seymour prefers the nuts)(and bolts approach).

- —Hence Halliday and Selkirk's introduction of a *Sense Unit Condition* on phrasal constituents of Intonation Structure.

Coordination Induces Similar Fragments

(4) a. I will buy, and you will cook, the biggest turkey we can find.
b. I gave Deadeye Dick a sugar stick, and Mexican Pete a bun.
c. Deadeye Dick got a sugar stick, and Mexican Pete a bun.

- This also is not a toy problem: the Wall Street Journal corpus is full of this sort of thing:

(5) New England Electric System bowed out of the bidding for Public Service Co. of New Hampshire, saying that *the risks were too high, and the potential payoff too far in the future, to justify a higher offer.*

(Section 00, file 13, `/corpora/treebank/combined/wsj/00/wsj_0013.mrg`)

Combinatory Categorial Grammar (CCG)

- CCG trades **categories** for PS rules, and **type-driven** combinatory rules for structure-dependent transformations

(6) $S \rightarrow NP \quad VP$

$VP \rightarrow TV \quad NP$

$TV \rightarrow \{proved,\ finds,\ \ldots\}$

(7) proved $:= (S\backslash NP)/NP$

(8) *The functional application rules*

a. $X/Y \quad Y \quad \Rightarrow \quad X$ $\qquad (>)$

b. $Y \quad X\backslash Y \quad \Rightarrow \quad X$ $\qquad (<)$

A Derivation

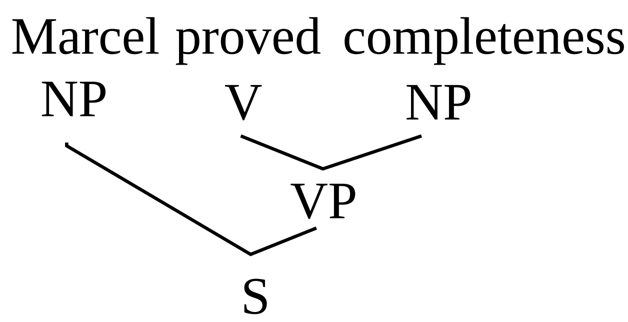
(9) a. Marcel proved completeness

$$\dfrac{\text{Marcel}}{NP} \quad \dfrac{\text{proved}}{(S\backslash NP)/NP} \quad \dfrac{\text{completeness}}{NP}$$

$$\dfrac{(S\backslash NP)/NP \quad NP}{S\backslash NP}>$$

$$\dfrac{NP \quad S\backslash NP}{S}<$$

b.



Semantics

(10) proved := $(S\backslash NP)/NP : \lambda x.\lambda y.prove'xy$

(11) *Functional application*

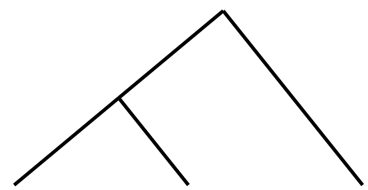
a. $X/Y : f \quad Y : a \quad \Rightarrow \quad X : f\ a$ $\quad (>)$

b. $Y : a \quad X\backslash Y : f \quad \Rightarrow \quad X : f\ a$ $\quad (<)$

(12)

$$
\begin{array}{ccc}
\text{Marcel} & \text{proved} & \text{completeness} \\
\hline
NP : marcel' & (S\backslash NP)/NP : \lambda x.\lambda y.prove'xy & NP : completeness' \\
\hline
 & \multicolumn{2}{c}{S\backslash NP : \lambda y.prove'completeness'y} \quad > \\
\hline
\multicolumn{3}{c}{S : prove'completeness'marcel'} \quad <
\end{array}
$$

Left-Associativity Convention

(13) a. $(prove' \, completeness')\, marcel'$ b. 

prove' completeness' marcel'

- A nonordered form of the traditional VP is reflected at the level of propositional logical form. **Binding etc. must be defined at this level.**

Coordination

(14) *Simplified coordination rule* (<Φ>)

$$X \quad CONJ \quad X' \Rightarrow \quad X''$$

(15) Marcel conjectured and proved completeness

Marcel	conjectured	and	proved	completeness
NP	$(S\backslash NP)/NP$	$CONJ$	$(S\backslash NP)/NP$	NP

conjectured and proved: $(S\backslash NP)/NP$ (<Φ>)

conjectured and proved completeness: $S\backslash NP$ (>)

Marcel conjectured and proved completeness: S (<)

Composition

(16) *Forward composition* ($>\mathbf{B}$)

$$X/Y : f \quad Y/Z : g \quad \Rightarrow_{\mathbf{B}} \quad X/Z : \lambda x.f(gx)$$

(17)

Marcel	conjectured	and	might	prove	completeness
NP	$(S\backslash NP)/NP$	$CONJ$	$(S\backslash NP)/VP$	VP/NP	NP
$: marcel'$	$: conjecture'$	$: and'$	$: might'$	$: prove'$	$: completeness'$

might prove $\Rightarrow_{>\mathbf{B}}$

$$(S\backslash NP)/NP : \lambda x.\lambda y.might'(prove'x)y$$

conjectured and might prove $\Rightarrow_{<\Phi>}$

$$(S\backslash NP)/NP : \lambda x.\lambda y.and'(might'(prove'x)y)(conjecture'xy)$$

conjectured and might prove completeness $\Rightarrow_{>}$

$$S\backslash NP : \lambda y.and'(might'(prove'completeness')y)(conjecture'completeness'y)$$

Marcel conjectured and might prove completeness $\Rightarrow_{<}$

$$S : and'(might'(prove'completeness')marcel')(conjecture'completeness'marcel')$$

Type-Raising

(18) *Subject type-raising* $(>\mathbf{T})$

$NP : a \quad \Rightarrow_{\mathbf{T}} \quad S/(S\backslash NP) : \lambda f.fa$

(19) Marcel proved and I disproved completeness

Marcel	proved	and	I	disproved	completeness
NP	$(S\backslash NP)/NP$	$CONJ$	NP	$(S\backslash NP)/NP$	NP
$S/(S\backslash NP)$ $>\mathbf{T}$			$S/(S\backslash NP)$ $>\mathbf{T}$		
S/NP $>\mathbf{B}$			S/NP $>\mathbf{B}$		
S/NP $<\Phi>$					
S $>$					

- Type raising is restricted to primitive argument categories, NP, PP etc.

Many Linguistic Predictions

(20) give a teacher an apple and a policeman a flower

$$
\begin{array}{cccccc}
\text{give} & \text{a teacher} & \text{an apple} & \text{and} & \text{a policeman} & \text{a flower} \\
\overline{DTV} & \overline{TV\backslash DTV}^{<\mathbf{T}} & \overline{VP\backslash TV}^{<\mathbf{T}} & \overline{CONJ} & \overline{TV\backslash DTV}^{<\mathbf{T}} & \overline{VP\backslash TV}^{<\mathbf{T}}
\end{array}
$$

$$\overline{VP\backslash DTV}^{<\mathbf{B}} \qquad \overline{VP\backslash DTV}^{<\mathbf{B}}$$

$$\overline{VP\backslash DTV}^{<\Phi>}$$

$$\overline{VP}^{<}$$

$VP = S\backslash NP$

$TV = (S\backslash NP)/NP$

$DTV = ((S\backslash NP)/NP)/NP$

Surface Structure in CCG

(21) Marcel proved completeness

$$\frac{\text{Marcel}}{NP : marcel'} \qquad \frac{\text{proved}}{(S\backslash NP)/NP : prove'} \qquad \frac{\text{completeness}}{S\backslash(S/NP) : \lambda p.p\ completeness'}$$

$$\frac{NP : marcel'}{S/(S\backslash NP) : \lambda f.f\ marcel'}{>}\mathsf{T}$$

$$\frac{S/(S\backslash NP) : \lambda f.f\ marcel' \quad (S\backslash NP)/NP : prove'}{S/NP : \lambda x.prove' x\ marcel'}{>}\mathsf{B}$$

$$\frac{S/NP : \lambda x.prove' x\ marcel' \quad S\backslash(S/NP) : \lambda p.p\ completeness'}{S : prove' completeness' marcel'}{<}$$

(22) Marcel proved completeness

$$\frac{\text{Marcel}}{NP : marcel'} \qquad \frac{\text{proved}}{(S\backslash NP)/NP : prove'} \qquad \frac{\text{completeness}}{(S\backslash NP)\backslash((S\backslash NP)/NP) : \lambda p.p\ completeness'}$$

$$\frac{NP : marcel'}{S/(S\backslash NP) : \lambda f.f\ marcel'}{>}\mathsf{T}$$

$$\frac{(S\backslash NP)/NP : prove' \quad (S\backslash NP)\backslash((S\backslash NP)/NP) : \lambda p.p\ completeness'}{S\backslash NP : \lambda y.prove' completeness' y}{<}$$

$$\frac{S/(S\backslash NP) : \lambda f.f\ marcel' \quad S\backslash NP : \lambda y.prove' completeness' y}{S : prove' completeness' marcel'}{<}$$

Intonation/Information Structure = Syntax/Semantics

- Clearly, such a theory is capable of capturing intonation structure and coordinate structure directly in the same “mildly context-sensitive” syntax that derives predicate argument relations.

- We do this by making pitch accents mark constituents as theme or rheme, and making boundaries limit combinatory derivation (as well as marking speech act, continuation etc.).

The Sense Unit Condition Follows

- The following is ruled out:

(23) a. *(Three MATHEMATICIANS) (in ten derive a LEMMA).

L+H* LH% H* LL%

b. *(Seymour prefers the NUTS) (and bolts APPROACH).

L+H* LH% H* LL%

—for the same reason as the following:

(24) a. *Three mathematicians in ten derive a lemma and in a hundred can cook a boiled egg.

b. *The nuts which Seymour prefers and bolts approach

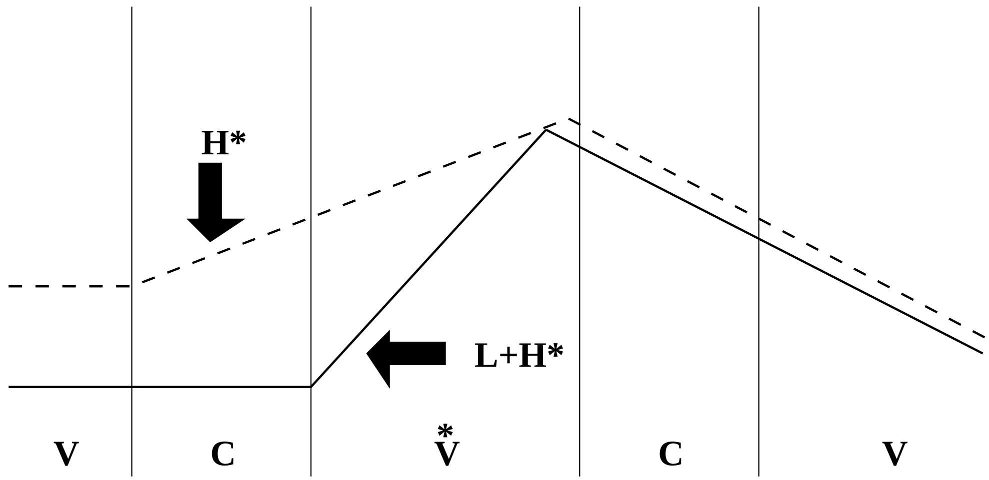
- That is, substrings like **in a hundred can cook a boiled egg* are not consituents in a CCG of English.

Aside on Expressive Power of BTS

- The Combinatory Calculus **BTS** is essentially equivalent to the simply-typed λ_I calculus.
- All linguistic restrictions stem from constraints which forbid rules like the following, which override the directionality specified in the lexicon:
  - $Y \quad X/Y \Rightarrow X$
  - $X/Y \quad Y/Z \Rightarrow X\backslash Z$
- Such restrictions imply that directionality is as much a heritable feature-value as S or N, and still yield a general reordering/rebracketing calculus that must be further constrained for languages like English.
- It is interesting in terms of the origins of language to speculate on whether individual combinators like **B** and **T** have independent cognitive uses, conferring an evolutionary advantage

Aside on a Pilot Study of Elicited Theme and Rheme Contour

- ToBI annotators failure to reliably distinguish L+H* and H* seems to be an artefact of the annotation instructions:



(Ladd & Steedman, in prep.)

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.

A Maximum Entropy CCG "Supertagger" (Clark)

- $p(t|h) = \frac{1}{Z(h)} e^{\sum \lambda_j f_j(t,h)}$ where f_j are the features
- Example features: current word, POS tag of current word, next word, previous word, previous two supertags...
- Trained a model on Sections 02-23 of the Treebank using Generalized Iterative Scaling over 377 CCG categories
- Current performance $\approx$ 90% per word accuracy and $\approx$ 98% with a $\pm$.01% beam (average cats/word 3.8) on a development set (Section 00) compares well with Chen 1999 using a similar sized ($>$ 300) set of TAG elementary trees and Chen and Vijayshanker 2000 for an automatically induced lexicon.

Dependency-based PCFG-style CCG Parser (Hockenmaier)

Model	NoParse	Cat	LP	LR	BP	BR	$\langle P,H,S \rangle$	$\langle S \rangle$	$\langle \rangle$	CM	≤2 ID
Baseline	6	87.7	72.8	72.4	78,3	77.9	75.7	81.1	84.3	23.0	51.1
HWDep (+POS)	8	92.0	81.6	81.9	85.5	85.9	84.0	87.8	90.1	37.9	69.2
HWDep (+ tagger)	7	91.7	81.4	81.8	85.6	85.9	83.6	87.5	89.9	38.1	69.1

Table 1: Cat = word categories correct; LP, LR, BP and BR = Parseval scores; $\langle H,C \rangle$, $\langle C \rangle$, and $\langle \rangle$ = completely labeled, complement-labeled and unlabeled dependencies. CM = complete match on $\langle \rangle$, and ≤2 ID = under 2 incorrect $\langle \rangle$.

- State of the art (Collins ’98): unlabeled dependencies $\langle \rangle$ 91.0%
- CCG does less well on parseval because of binary rules and size of category set (400 vs. 37), and the consequent problem of **unknown word-category pairs**.

References

Hockenmaier, Julia: 2001, “Statistical Parsing for CCG with Simple Generative Models,” *Proceedings of the Student Session of the 39th Annual Meeting of the Association for Computational Linguistics*, July, Toulouse, France. 7-12.

Hockenmaier, Julia and Gann Bierner and Jason Baldridge: 2002, Extending the Coverage of a CCG System, *Journal of Logic and Computation*, (to appear).

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

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.