

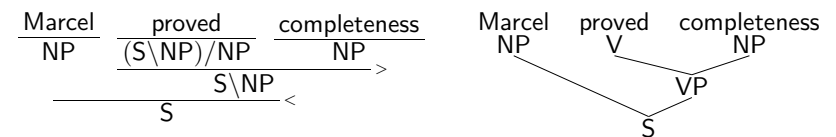
Day 3: CG approaches to information structure

- Combinatory Categorial Grammar (CCG; Steedman 2000a,b)
 - CCG in a nutshell
 - Structure, intonation, and information structure
 - The two dimensions of information structure
 - Combinatory Prosody
- Other Categorial Grammar approaches:
 - Multi-Modal Combinatory Categorial Grammar (Kruijff and Baldrige 2004)
 - Dependency Grammar Logic (Kruijff 2001)

1/33

Rules and derivations

- Functor categories can combine with their arguments by the following rules:
 - (2) Forward application ($>$)
 $X/Y \ Y \Rightarrow X$
 - (3) Backward application ($<$)
 $Y \ X \backslash Y \Rightarrow X$
- Derivations are written as shown below, on the left side. Note the direct correspondence to the upside-down constituency tree shown on the right.



CCG in a nutshell

3/33

CCG in a nutshell

- Syntactically potent elements such as verbs are associated with a syntactic category that identifies them as *functions* and specifies the type and directionality of their arguments and the type of their result.
 - A “result leftmost” notation is used:
 - α/β is a rightward-combining functor over a domain β into a range α
 - $\alpha \backslash \beta$ is the corresponding leftward-combining functor.
 - α and β may themselves be functional categories.
- (1) proved := $(S \backslash NP) / NP$

CCG in a nutshell

2/33

Semantics and Principle of Type Transparency

- The lexical categories can be augmented with an explicit identification of their semantic interpretation and the rules of functional application are accordingly expanded with an explicit semantics.
 - (4) proved := $(S \backslash NP) / NP : prove'$
 - (5) Forward application ($>$)
 $X/Y : f \ Y : a \Rightarrow X : fa$
- The semantic interpretation of all combinatory rules is fully determined by the *Principle of Type Transparency*:

All syntactic categories reflect the semantic type of the associated logical form, and all syntactic combinatory rules are type-transparent versions of one of a small number of semantic operations over functions including application, composition, and type-raising.

CCG in a nutshell

4/33

Example derivation with semantics

$$\frac{\frac{\text{Marcel}}{\text{NP} : \text{marcel}'} \quad \frac{\text{proved}}{(S \backslash \text{NP}) / \text{NP} : \text{prove}'} \quad \frac{\text{completeness}}{\text{NP} : \text{completeness}'}}{\text{S} \backslash \text{NP} : \text{prove}' \text{ completeness}' >} >$$

$$\frac{}{\text{S} : \text{prove}' \text{ completeness}' \text{ marcel}' <} <$$

Combinators

- In order to account for coordination of contiguous strings that do not constitute traditional constituents, CCG allows certain operations on functions called “combinators”, including the rule of functional composition in (7).

(7) Forward composition (>**B**)
 $X/Y : f \quad Y/Z : g \Rightarrow X/Z : \lambda x.f(gx)$

- CCG includes type-raising rules, which turn arguments into functions over functions-over-such-arguments.

(8) Forward type-raising (>**T**) (9) Backward type-raising (<**T**)
 $X : a \Rightarrow T / (T \backslash X) : \lambda f.f a$ $X : a \Rightarrow T \backslash (T / X) : \lambda f.f a$

X ranges over argument categories (e.g., NP and PP). The rules are order-preserving, e.g., (8) can turn an NP into a rightward-looking function over leftward functions, preserving the linear order of subjects and predicates.

More rule schemata

CCG includes linguistically motivated rule schemata such as the one for coordination of constituents of like type shown below:

(6) Coordination (< & >)
 $X \text{ conj } X \Rightarrow X$

Non-standard surface structures

- Complement-taking verbs like *think*, VP/S, can compose with fragments like *Marcel proved*, S/NP, which accounts for right-node raising (10), and also provides the basis for an analysis of unbounded dependencies (11).

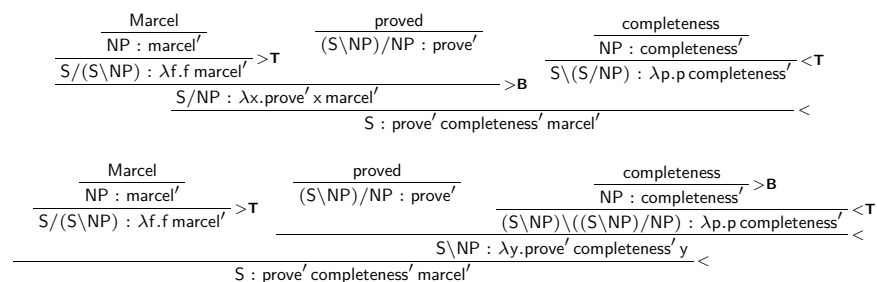
(10) [I disproved]_{S/NP} and [you think that Marcel proved]_{S/NP} completeness.

(11) the result that [you think that Marcel proved]_{S/NP}

Strings such as *you think that Marcel proved* are taken to be surface constituents of type S/NP.

Non-standard surface structures are licensed throughout

- Steedman assumes that the non-traditional constituents motivated for right-node raising and similar coordinations are also possible constituents of non-coordinate sentences like *Marcel proved completeness*.



- The Principle of Type Transparency guarantees that all such non-standard derivations yield identical interpretations.

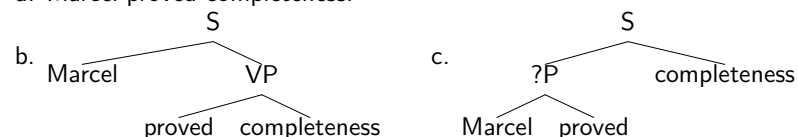
Structure and intonation

Steedman's claims:

- Surface structure* and *information structure* coincide, the latter simply consisting in the interpretation associated with a constituent analysis of the sentence.
- Intonation* coincides with *surface structure* (and hence information structure) in the sense that all intonational boundaries coincide with syntactic boundaries.

As a result, fragments such as *Marcel proved* in (12c), are not only prosodic constituents but surface syntactic constituents, complete with interpretations.

(12) a. Marcel proved completeness.



Motivating non-standard surface structures

- According to Steedman (2000a), the non-standard surface structures are not spurious ambiguities but relevant since they subsume the intonation structures needed to explain the possible intonation contours for sentences of English.
- Intonational boundaries contribute to determining which of the possible combinatory derivations is intended.
- The interpretations of the constituents that arise from these derivations are related to semantic distinctions of information structure and discourse focus.
- Steedman's claims:
 - Where intonational boundaries are present, they contribute to disambiguation.
 - Conversely, any such boundaries must be consistent with *some* syntactic derivation, or ill-formedness will result.

Intonation and Information Structure

- A sequence of one or more pitch accents followed by a boundary is referred to as an *intonational phrasal tune*.
 - Claim*: phrasal tunes in this sense are associated with specific discourse meanings distinguishing information type and/or propositional attitude.
- (13) Q: I know who proved soundness. But who proved COMPLETENESS?
 A: (MARCEL) (proved COMPLETENESS).
 H* L L+H* LH%
- (14) Q: I know which result Marcel PREDICTED. But which result did Marcel PROVE?
 A: (Marcel PROVED) (COMPLETENESS).
 L+H* LH% H* LL%
- Evidence*: Exchanging the answer tunes between the two contexts in (13) and (14) yields complete incoherence.

The two dimensions of Information Structure

- *Theme and Rheme*:
 - The *theme* is to be thought of as that part of an utterance which connects it to the rest of the discourse.
 - The *rheme* is that part of an utterance that advances the discussion by contributing novel information.
- *Focus and Background*:
 - The information marked by the pitch accent is called the *focus*, distinguishing theme focus and rheme focus, where necessary.
 - The term *background* is used for the part unmarked by pitch accent or boundary.

Intonationally unmarked themes/rhemes

- There also are *intonationally unmarked themes*:
 - (15) Q: Which result did Marcel prove?
A: (Marcel proved) (COMPLETENESS).
H* LL%
 - (16) Q: What do you know about Marcel?
A: (Marcel) (proved COMPLETENESS).
H* LL%
- The same contour can also occur with an *all-rheme* utterance:
 - (17) Guess what? (Marcel proved COMPLETENESS!)
H* LL%

Theme and Rheme and their intonational realization

Steedman observes the following relationship for English:

- The L+H* LH% tune is associated with the *theme*.
- The H* L and H* LL% tunes (among others) are associated with the *rheme*.

Semantic characterization of theme and rheme

- Following Jackendoff (1972), the *theme* is characterized semantically via functional abstraction, using the notation of λ -calculus, as in (18), corresponding to the theme of (14) and (15).
 - (18) $\lambda x. prove' x marcel'$
- When such a function is supplied with an argument in the form of the rheme, it reduces to give a proposition, with the same predicate-argument relation as the canonical sentence.
 - (19) $prove' completeness' marcel'$

Semantic characterization of theme and rheme (cont.)

- The λ -abstraction operator is closely related to the existential quantifier \exists
- (20) $\exists x. \text{prove}' x \text{ marcel}'$
- The *theme* can be associated with the *rheme alternative set*: the set of propositions that could instantiate the corresponding existentially quantified proposition.

$$(21) \left\{ \begin{array}{l} \text{prove}' \text{ decidability}' \text{ marcel}' \\ \text{prove}' \text{ soundness}' \text{ marcel}' \\ \text{prove}' \text{ completeness}' \text{ marcel}' \end{array} \right\}$$

- The theme tune and the rheme tune can be specified in semantic terms:
- (22) Theme tunes *presuppose* a rheme alternative set.
Rheme tunes *restrict* the rheme alternative set.

Themes, pitch accents, and the theme alternative set

- The significance of the presence or absence of primary pitch accents within a theme lies in the prior existence of a theme differing in its translation only in those elements corresponding to the accented items.
- The presence of pitch accents in the translation of themes is marked by distinguishing the corresponding constant with an asterisk.

$$(24) \exists x. * \text{ admires}' x \text{ marcel}'$$

- The set of alternative themes is called the *theme alternative set*.

$$(25) \left\{ \begin{array}{l} \exists x. \text{ admires}' x \text{ marcel}' \\ \exists x. \text{ likes}' x \text{ marcel}' \end{array} \right\}$$

Focus and Background

- Within both theme and rheme, those words that contribute to distinguishing the theme and the rheme of an utterance from other alternatives made available by the context may be marked via a pitch accent.

(23) 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).
 $\underbrace{\hspace{1.5cm}}_{\text{background}} \quad \underbrace{\text{L+H* LH\%}}_{\text{focus}} \quad \underbrace{\hspace{1.5cm}}_{\text{background}} \quad \underbrace{\text{H*}}_{\text{focus}} \quad \underbrace{\text{LL\%}}_{\text{background}}$
 $\underbrace{\hspace{3.5cm}}_{\text{theme}} \quad \underbrace{\hspace{3.5cm}}_{\text{rheme}}$

Combinatory Prosody: Pitch Accents

- Six pitch accents are distinguished as markers either of theme (θ) or rheme (ρ).

$$(26) \theta\text{-markers: L+H*}, \text{L*+H}$$

$$\rho\text{-markers: H*}, \text{L*}, \text{H*+L}, \text{H+L*}$$

- Pitch accents affect both the syntactic category and the interpretation of the words they occur on.
 - With basic types, such as NP, the effect of a θ - or ρ -marking accent is to associate with the category a value of θ or ρ on a feature INFORMATION, which is notated as NP_ρ .
 - With function types, such as $S \setminus NP$, the effect of a θ - or ρ -marking accent is to θ - or ρ -mark the domain and range of the function, as in $S_\rho \setminus NP_\rho$.
 - Any argument that combines with such a marked function has to be compatible with its theme- or rheme-hood.

Combinatory Prosody: The Boundaries (cont.)

- The syntactic category $S\$_\iota \backslash S\$_\eta$ maps θ and ρ -marked categories onto identically ι -marked categories, where ι will no longer unify with η , θ or ρ , i.e. this prevents further combination with anything except similarly complete prosodic phrases.
 - The intonational phrase boundary tones L% and H% are assigned the categories in (32). Intermediate phrase boundaries are mapped into intonational phrase boundaries.
- (32) $L\% := (S\$_\phi \backslash S\$_\eta) \backslash (S\$_\iota \backslash S\$_\eta) : \lambda f. \lambda g. [S](fg)$
 $H\% := (S\$_\phi \backslash S\$_\eta) \backslash (S\$_\iota \backslash S\$_\eta) : \lambda f. \lambda g. [H](fg)$
- ϕ is a value that again unifies only with itself and ι , preventing further combination with anything except similarly complete prosodic phrases.

Invisible boundaries

- The majority of themes in utterances are null themes, unmarked by explicit boundary tones.
 - The position of the theme-rheme boundary is usually ambiguous in these cases, as for example in (33).
- (33) 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}
- Steedman assumes that intermediate phrase L and H boundaries are indistinguishable from the null tone and may therefore be postulated anywhere there is no tone.

Two examples from Steedman (2000a)

(67) Marcel PROVED L H% COMPLETENESS L L%

L+H* H* L L%

$$\frac{\frac{S/(S \backslash NP)}{\lambda p.p \text{ marcel}' : \lambda x. \lambda y. *prove'xy} \quad \frac{(S_\phi \backslash NP_\theta) / NP_\theta}{S\$_\iota \backslash S\$_\eta} \quad \frac{L}{S\$_\iota \backslash S\$_\eta} \quad \frac{H\%}{(S\$_\phi \backslash S\$_\eta) \backslash (S\$_\iota \backslash S\$_\eta)}}{S_\theta / NP_\theta : \lambda x. *prove'x \text{ marcel}' \quad \frac{S\$_\phi \backslash S\$_\eta}{\lambda f. \eta'f} \quad \frac{H\%}{\lambda f. \lambda g. [H](fg)}} \leftarrow$$

$$\frac{S_\theta / NP_\theta : \lambda x. *prove'x \text{ marcel}' \quad \frac{S\$_\phi \backslash S\$_\eta}{\lambda f. [H](\eta'f)}}{S_\phi / NP_\phi : [H](\theta'(\lambda x. *prove'x \text{ marcel}'))} \leftarrow$$

(68) Marcel PROVED L H% COMPLETENESS L L%

L+H* H* L L%

$$\frac{\frac{S_\phi / NP_\phi}{[H](\theta'(\lambda x. *prove'x \text{ marcel}'))} \quad \frac{S_\rho \backslash (S_\rho / NP_\rho)}{\lambda p.p *completeness'} \quad \frac{L}{S\$_\iota \backslash S\$_\eta} \quad \frac{H\%}{(S\$_\phi \backslash S\$_\eta) \backslash (S\$_\iota \backslash S\$_\eta)}}{S_\phi \backslash (S_\phi / NP_\phi) : [S](\rho'(\lambda p.p *completeness')) \quad \frac{S\$_\phi \backslash S\$_\eta}{\lambda f. \eta'f} \quad \frac{L\%}{\lambda f. \lambda g. [S](fg)}} \leftarrow$$

$$\frac{S_\phi \backslash (S_\phi / NP_\phi) : [S](\rho'(\lambda p.p *completeness')) \quad \frac{S\$_\phi \backslash S\$_\eta}{\lambda f. [S](\eta'f)}}{S_\phi : [S](\rho'(\lambda p.p *completeness'))([H](\theta'(\lambda x. *prove'x \text{ marcel}')))} \leftarrow$$

Invisible boundaries (cont.)

- Invisible boundaries can act as an edge of an unmarked theme.
- Undetectable boundaries are also allowed in other positions where there is no tone - for example at the right-hand edge of an utterance-initial rheme followed by an unmarked theme.

(34) Q: Who proved completeness?

A: (MARCEL) (proved completeness).

H* L LL%

(35) I read a book about

$$\frac{\frac{S/NP}{S_\iota / NP_\iota} \quad \frac{L}{S\$_\iota \backslash S\$_\eta} \quad \frac{COMPLETENESS}{S_\rho \backslash (S_\rho / NP_\rho)} \quad \frac{L}{S\$_\iota \backslash S\$_\eta} \quad \frac{L\%}{(S\$_\phi \backslash S\$_\eta) \backslash (S\$_\iota \backslash S\$_\eta)}}{S_\phi \backslash (S_\phi / NP_\phi)} \leftarrow$$

S_ϕ

Information Structure in Multi-Modal Combinatory Grammar

- Kruijff and Baldridge (2004) extend CCG with a generalized notion of multi-dimensional signs, inspired by the types of representation found in constraint-based frameworks.
- The generalized sign allows multiple levels to share information, but only in a resource-bounded way through a very restricted indexation mechanism.
- Well-formedness of a linguistic expression remains entirely determined by the CCG derivation.

An example from Kruijff and Baldridge (2004)

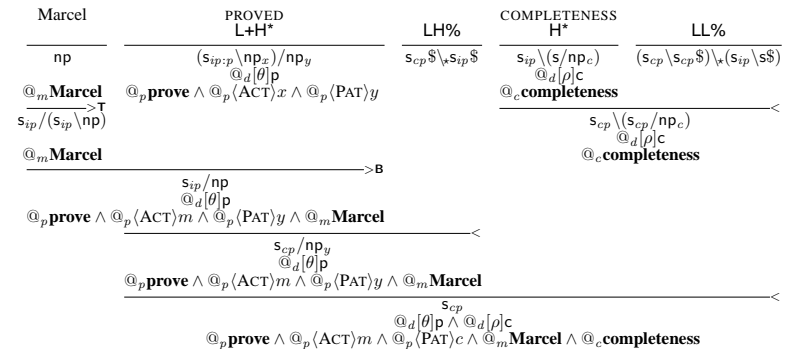


Figure 4: Information structure for derivation for (67)-(68) from (Steedman, 2000a)

Dimensions of signs used by Kruijff and Baldridge (2004)

The signs include the following dimensions:

- Phonemic representation
- Prosody
- Syntactic category: well-formed categories, combinatory rules
- Information structure:
 - hybrid logic formulas of the form $\textcircled{d}[in]_r$, with r a discourse referent that has informativity in (theme θ , or rheme ρ) relative to the current point in the discourse
- Predicate-argument structure:
 - hybrid logic formulas of the form as proposed in Baldridge and Kruijff (2002).

Dependency Grammar Logic and Information Structure

- Kruijff (2001) presents a theory of information structure within Dependency Grammar Logic (DGL) and provides specific points of criticism of the approach of Steedman (2000a):
 - Steedman only shows how prosody can be related to information structure; how word order can be related to information structure is not discussed.
 - It is unclear how Steedman's approach can handle multiple focus examples.
 - The approach makes extensive use of invisible categories for boundary tones.
- Kruijff proposes a reformulation of Steedman's description of intonation which
 - does not require extra categories for boundary tones and
 - shows how his DGL approach can handle multiple focus constructions (Kruijff 2001, p. 186)

Open issues for Steedman's approach

- Steedman's approach requires continuous constituents since only adjacent material can be combined. This seems to incorrectly predict that information structure units must be continuous.
- Steedman's account seems to lack a restrictive theory of focus projection. How is projection of the focus restricted, for example, from the subject onto the verb given this the two can form a constituent in his approach? How can word order changes restrict focus projection?
- How can multiple focus constructions be dealt with?
- Is there convincing motivation for the empty categories Steedman introduces for invisible boundary tones?

References

- Baldrige, Jason and Geert-Jan Kruijff (2002). Coupling CCG with Hybrid Logic Dependency Semantics. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02)*. Philadelphia, PA. <http://www.aclweb.org/anthology/P02-1041.pdf>.
- Jackendoff, Ray (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- Kruijff, Geert-Jan and Jason Baldrige (2004). Generalizing Dimensionality in Combinatory Categorical Grammar. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING-04)*. Geneva. <http://www.cogsci.ed.ac.uk/~jmb/KruijffBaldrige-coling04.pdf>.
- Kruijff, Geert-Jan M. (2001). A Categorical-Modal Architecture of Informativity: Dependency Grammar Logic and Information Structure. Ph.D. thesis, Charles University, Prague, Czech Republic.
- Steedman, Mark (2000a). Information Structure and the Syntax-Phonology Interface. *Linguistic Inquiry* 31(4), 649–689.
- Steedman, Mark (2000b). *The Syntactic Process*. Cambridge, MA: MIT Press. Bradford Books.