A Step towards Natural-Sounding Speech Generation

Maarika Traat
Tartu University
Institute of Computer Science

Contents

- Natural Speech vs. Computer Speech
  - Intonation
- What is Information Structure?
  - Relation to Intonation
- Discourse Representation Theory (DRT)
- Information Structure in DRT
- Categorial Grammars
  - Combinatory Categorial Grammar
  - Unification Categorial Grammar
- Unification-based Combinatory Categorial Grammar (UCCG)
- Information Structure in UCCG

Natural Speech vs. Computer Speech

- Text-to-speech vs. content-to-speech
- Formant synthesis (rule-based synthesis)
  - an acoustic model.
  - parameters: fundamental frequency, voicing, and noise levels varied over time to create a waveform of artificial speech (Bell Labs 63) 
- Articulatory synthesis
  - computational models of the human vocal tract; the articulation processes occurring there
  - mostly of academic interest
- Concatenative Synthesis
  - concatenation (or stringing together) of segments of recorded speech
  - generally the most natural sounding synthesized speech
  - but sometimes audible glitches in the output, detracting from the naturalness of the synthesized speech

Natural Speech vs. Computer Speech

- Text-to-speech vs. content-to-speech
- Formant synthesis (rule-based synthesis)
  - an acoustic model.
  - parameters: fundamental frequency, voicing, and noise levels varied over time to create a waveform of artificial speech (Bell Labs 63)
- Articulatory synthesis
  - computational models of the human vocal tract; the articulation processes occurring there
  - mostly of academic interest
- Concatenative Synthesis
  - concatenation (or stringing together) of segments of recorded speech
  - generally the most natural sounding synthesized speech
  - but sometimes audible glitches in the output, detracting from the naturalness of the synthesized speech

Natural Speech vs. Computer Speech

- Examples of concatenative speech
  - AT&T: US  UK  IN 

I like all kinds of ice cream.

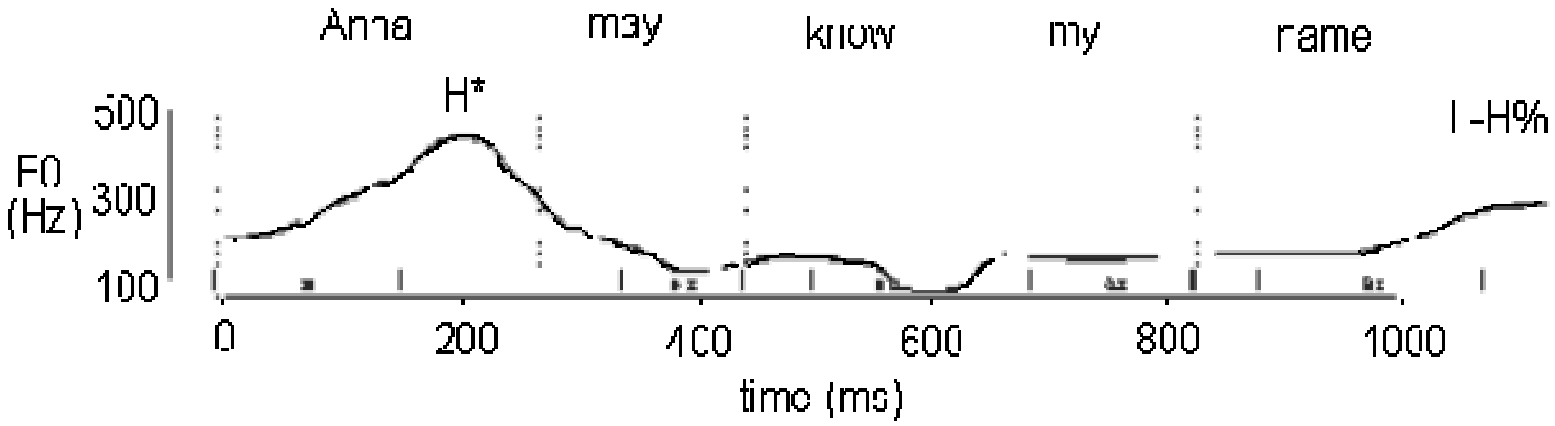
Usually chocolate flavour is my favourite, but sometimes simple vanilla ice cream is even better.

Which one do you like better: chocolate or vanilla ice cream?

  - Institute of Cybernetics at TTU: EST 

Intonation

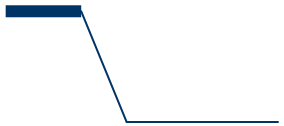
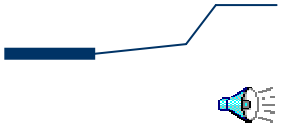
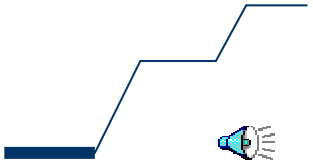
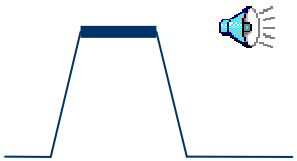
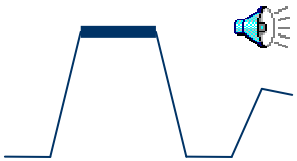
- Intonation, prosody, speech melody
- Intonation contour: F0 contour
  - The fundamental tone: the lowest frequency in a harmonic series. In addition to F0 there are overtones (faster waves).



Pierrehumbert's System of Prosodic Elements (mod.)

- Pitch accents
  - On words: makes words stand out, emphasises them
    - H*, L*, L+H*, L*+H, H+L*, H*+L, H*+H
- Boundary tones
  - At the end of prosodic phrases
    - LL%, LH%, HH%, HL%

Prosodic Elements Contd.

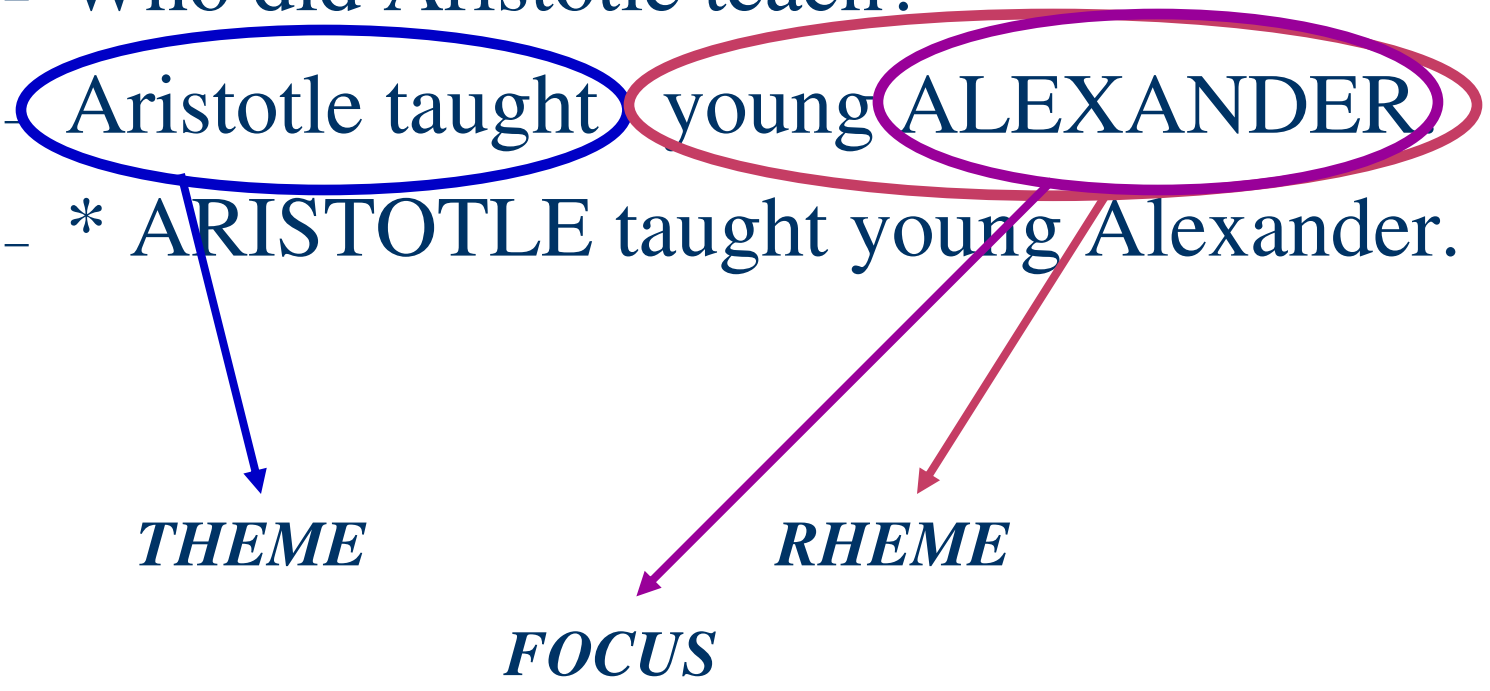
	LL%	LH%	HL%	HH%
H*				
L*				
L*+H				
L+H*				

Information Structure

- Aristotle taught young ALEXANDER.
- * ARISTOTLE taught young Alexander.

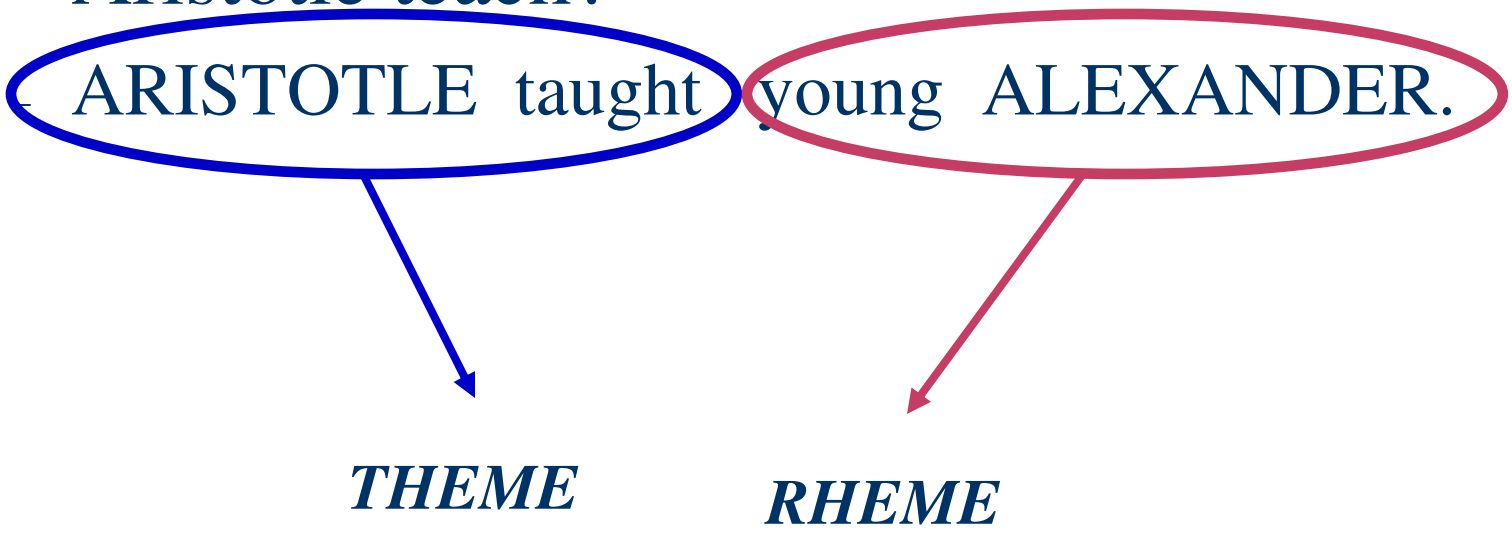
Information Structure

- Who did Aristotle teach?
- Aristotle taught young ALEXANDER.
- * ARISTOTLE taught young Alexander.



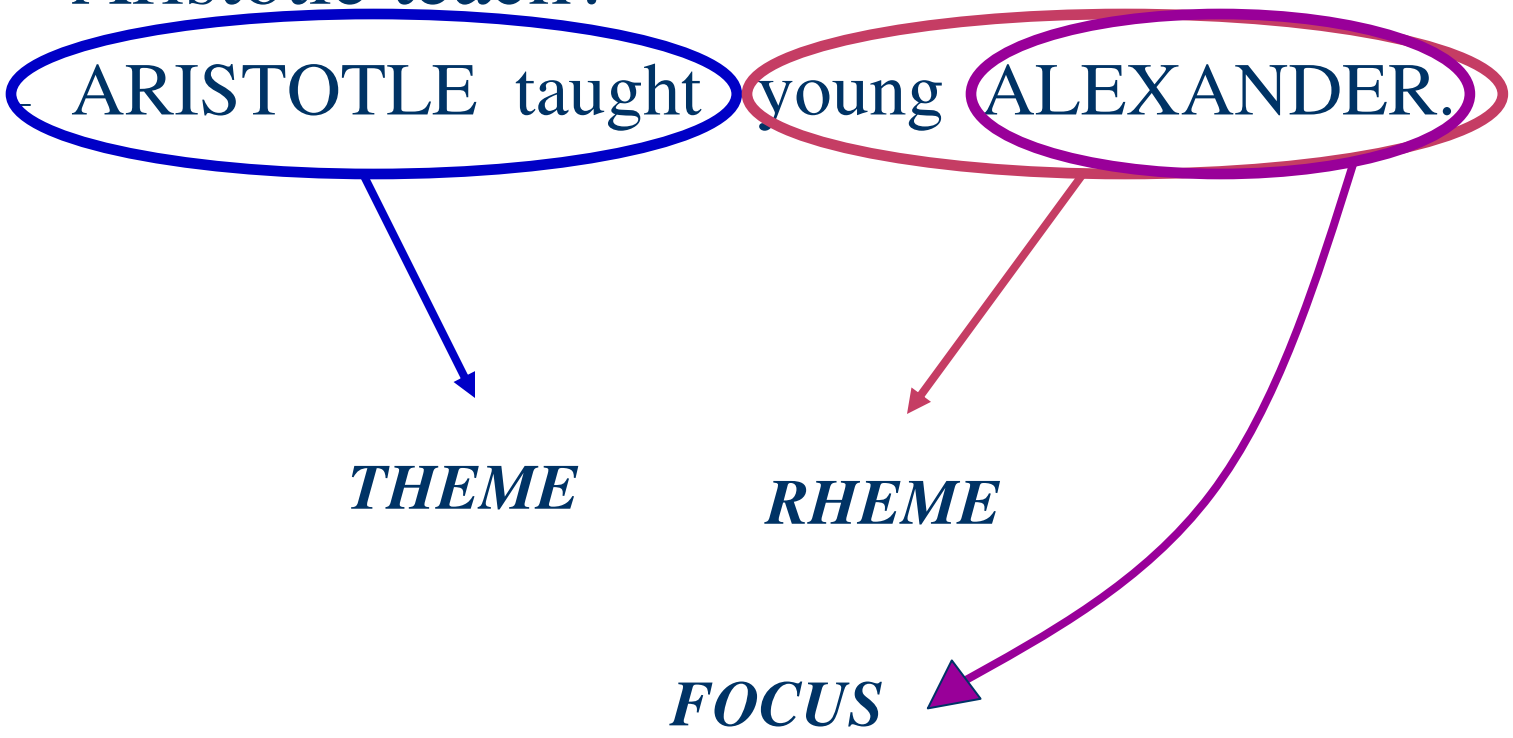
Information Structure

- I know who you teach, but who did Aristotle teach?
- ARISTOTLE taught young ALEXANDER.



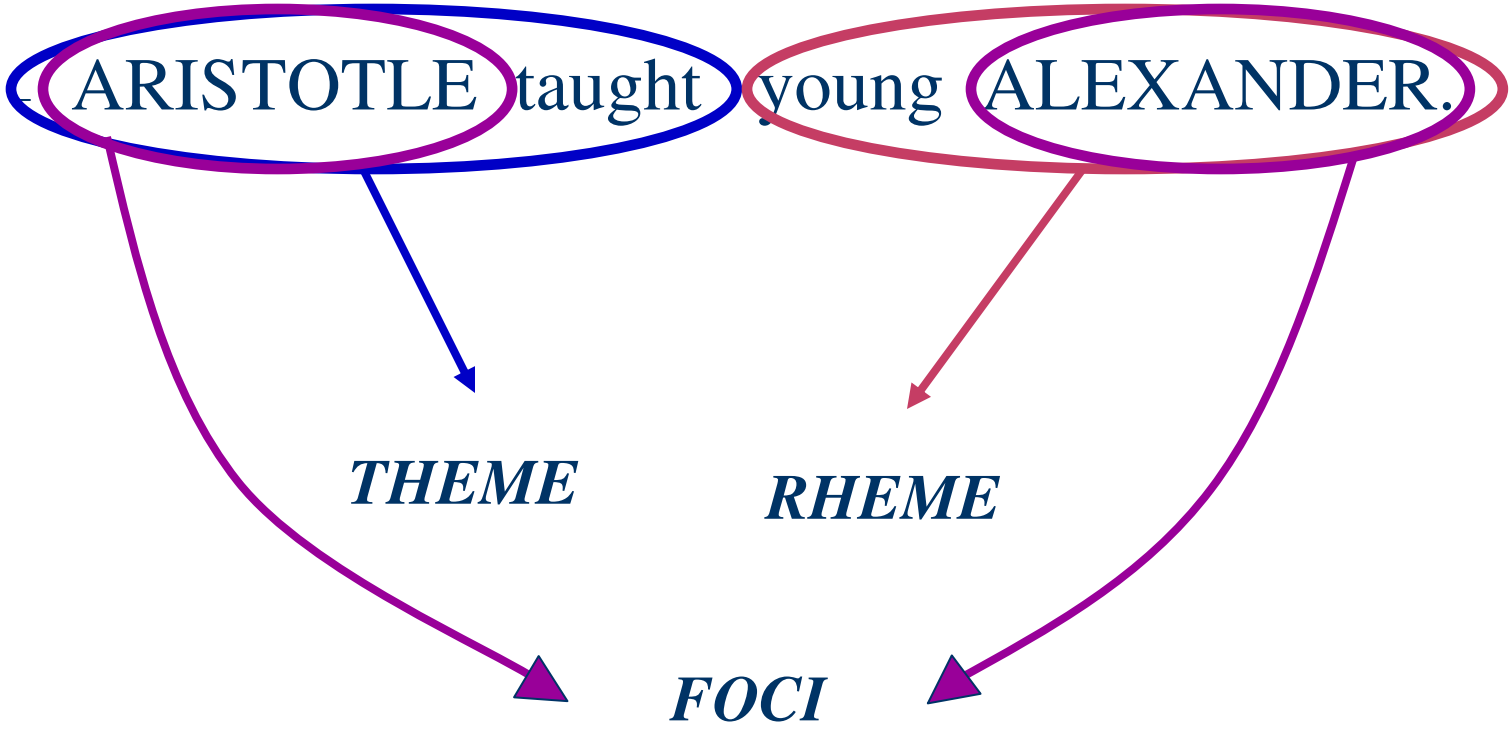
Information Structure

- I know who you teach, but who did Aristotle teach?
- ARISTOTLE taught young ALEXANDER.



Information Structure

- I know who you teach, but who did Aristotle teach?

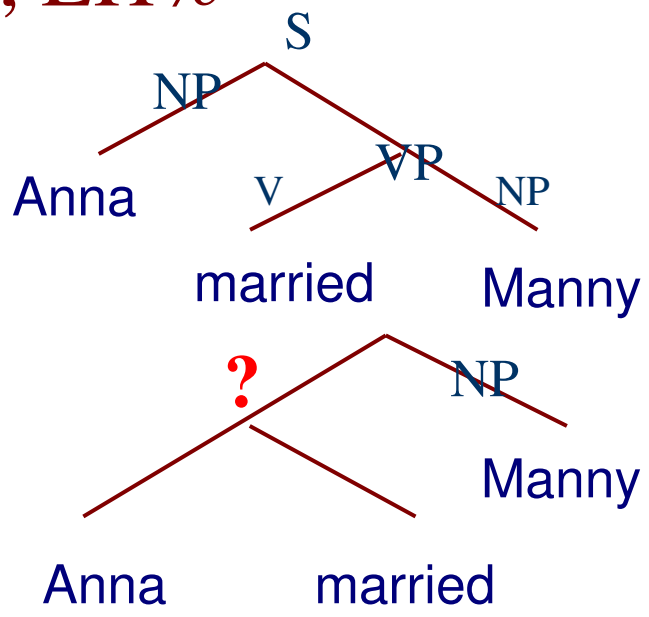


Steedman's Prosodic Approach to Information Structure

- Theme pitch accents, e.g. L+H*
- Rheme pitch accents, e.g. H*
- Boundary tones, e.g. LL%, LH%

  - *Anna married Manny.*
    H* LL% L+H* LH%

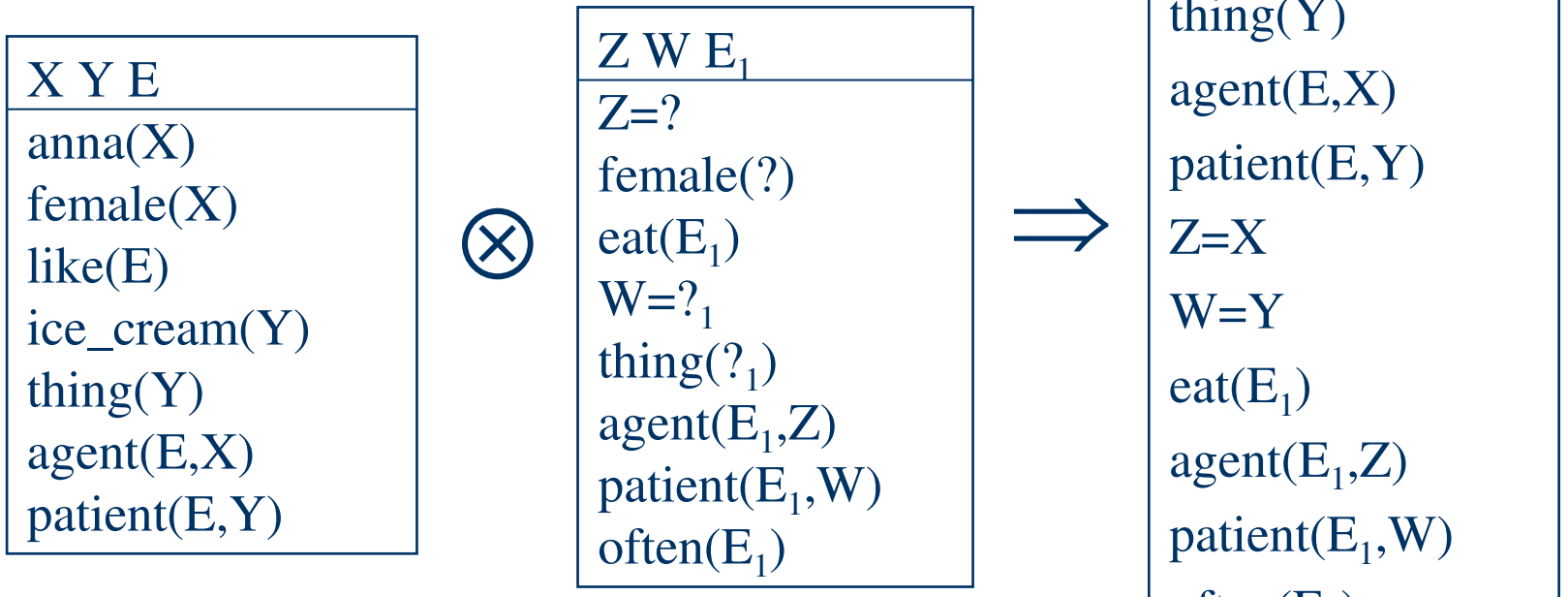
  - *Anna married Manny.*
    L+H* LH% H* LL%



Discourse Representation Theory (DRT)

- Discourse Representation Structure (DRS):
  - Universe: discourse referents
  - Discourse conditions

Anna likes ice cream. She eats it often.



Goals

- Combine
  - information structure
  - DRT semantics

} IS-DRS

- Provide a mechanism to get from
  - an intonationally annotated text to our semantic representation marked with information structure, and vice versa

Our approach: IS-DRS

- Information structure flags on discourse conditions
- Suitable both for natural language generation and automatic inference

Anna H* LL% married
Manny L+H* LH%.

theme θ
rheme ρ
focus +/–

X,Y,E	
anna(X)	ρ+
manny(Y)	θ+
marry(E)	θ-
agent(E,X)	
patient(E,Y)	

Categorial Grammars

- Grammar: lexicon and rules for constructing more complex structures from simpler ones
- Categorial Grammars:
  - Lexicalised theories of grammar
  - Category – functional type associated with an entry in the lexicon

    man :- n

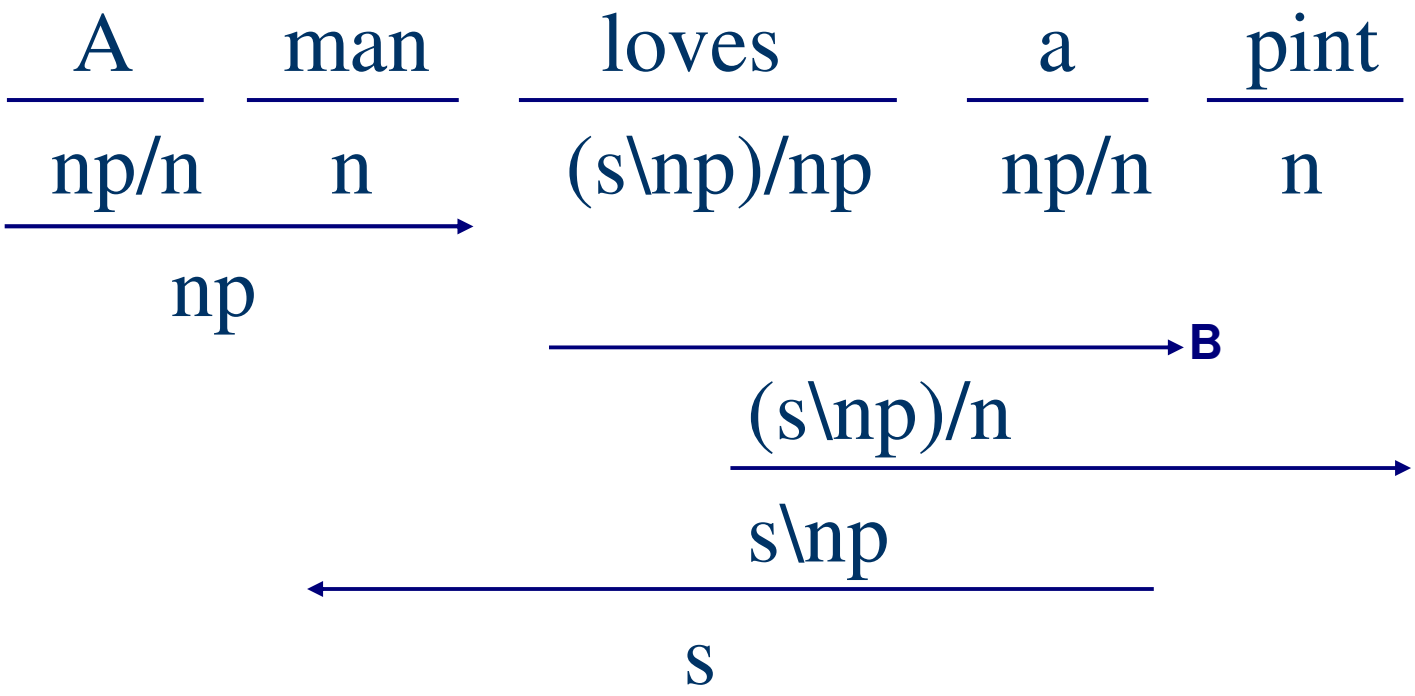
    a :- np/n
- Combinatory rules

$$\frac{a}{np/n} \quad \frac{man}{n} \Rightarrow \frac{a\ man}{np}$$

Combinatory Categorial Grammar (CCG)

- Categories
  - Basic categories: s, n, np
  - Complex categories: s\np, np/n, (n\n)/(s\np)
    - Directional slashes: **** and **/**
- Combinatory rules:
  - Forward application: X/Y Y → X
  - Backward application: Y X\Y → X
  - Forward composition: X/Y Y/Z → X/Z
  - Backward composition: Y\Z X\Y → X\Z
  - Etc.

Combinatory Categorial Grammar



Unification Categorial Grammar

- Unification
- Feature structures called signs
- Forward and backward application

LOVE
sent[fin]/W_1:np:x:pre/W_2:np:y:post
[e]love(e,x,y)
O

Unification-based Combinatory Categorial Grammar

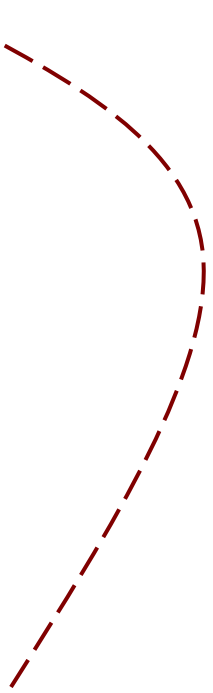
- Combinatory rules
- Directional slashes

CCG

- Unification
- Signs
- Compositional semantics

UCG

- Basic categories: s, n, vp
- DRT semantics (1st order)
- Information structure in semantics



Signs

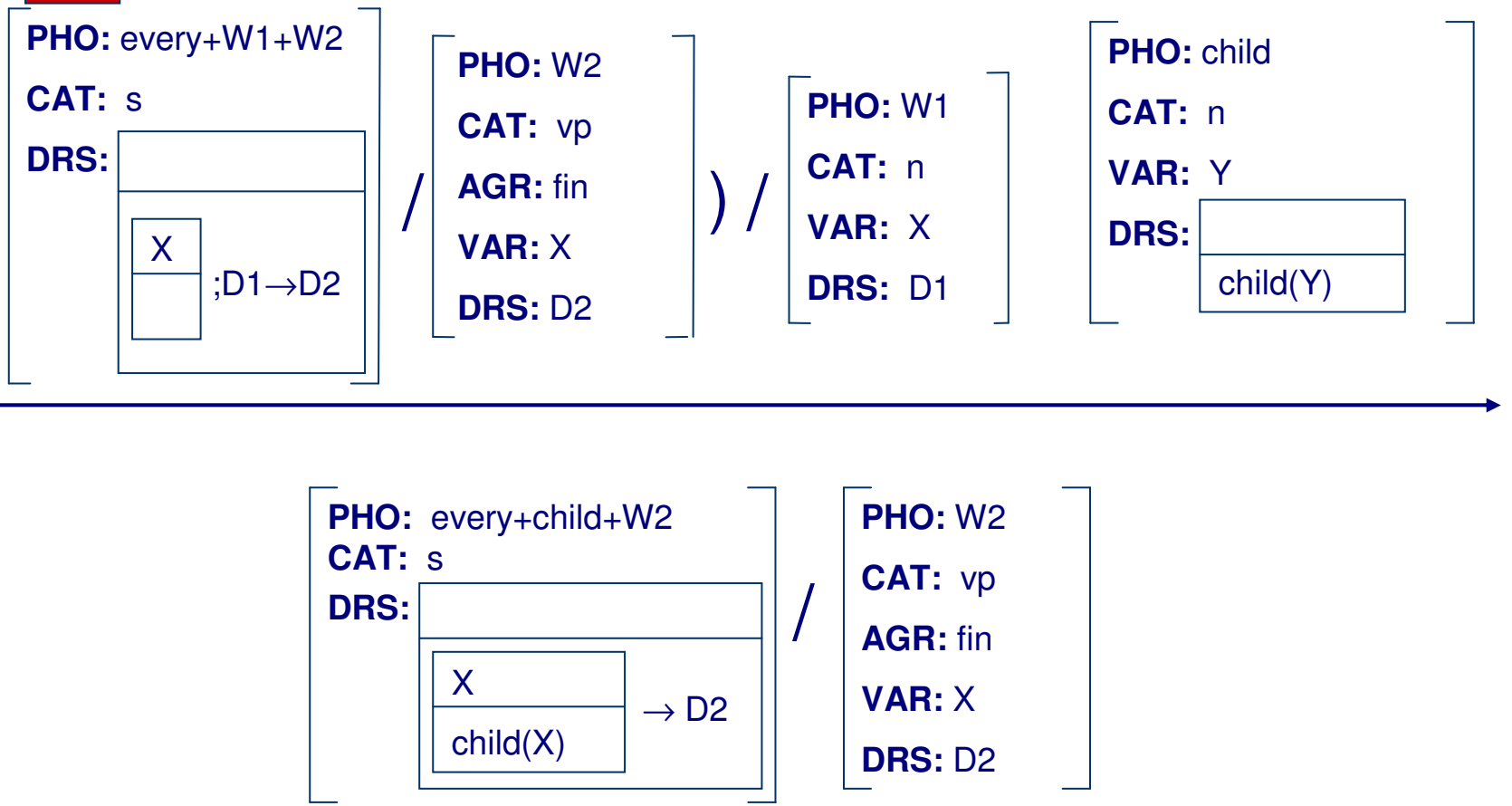
- Basic signs

PHO: child
CAT: n
VAR: X
DRS: | | child(X) |

- Complex sign

(**PHO:** every+W1+W2
CAT: s
DRS: | X | ;D1→D2 |) / **PHO:** W2
CAT: vp
AGR: fin
VAR: X
DRS: D2) / **PHO:** W1
CAT: n
VAR: X
DRS: D1

Combining of Signs



Theme and Rheme in UCCG

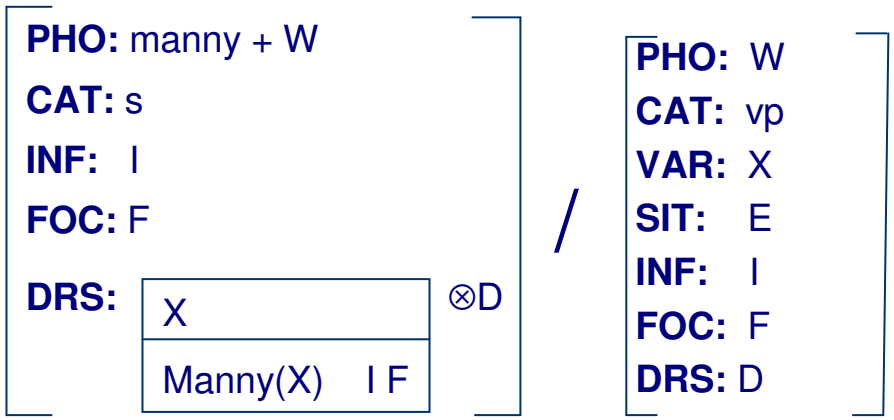
- Themeness or rhemeness of a lexical item depends on the type of pitch accent that occurs on it
- If no pitch accent on the item, its information structure value is a variable → gets its value from a marked item with which the current item is combined
  - theme marked item can only combine with
    - another theme marked item
    - an unmarked item
  - rheme marked item can only combine with
    - another rheme marked item
    - an unmarked item

Information Structure in UCCG Feature Structures

- INF feature

  theme, rheme, phrase

- FOC feature
  - + the word carries a pitch accent
  - – the word does not have a pitch accent on it



Pitch Accent

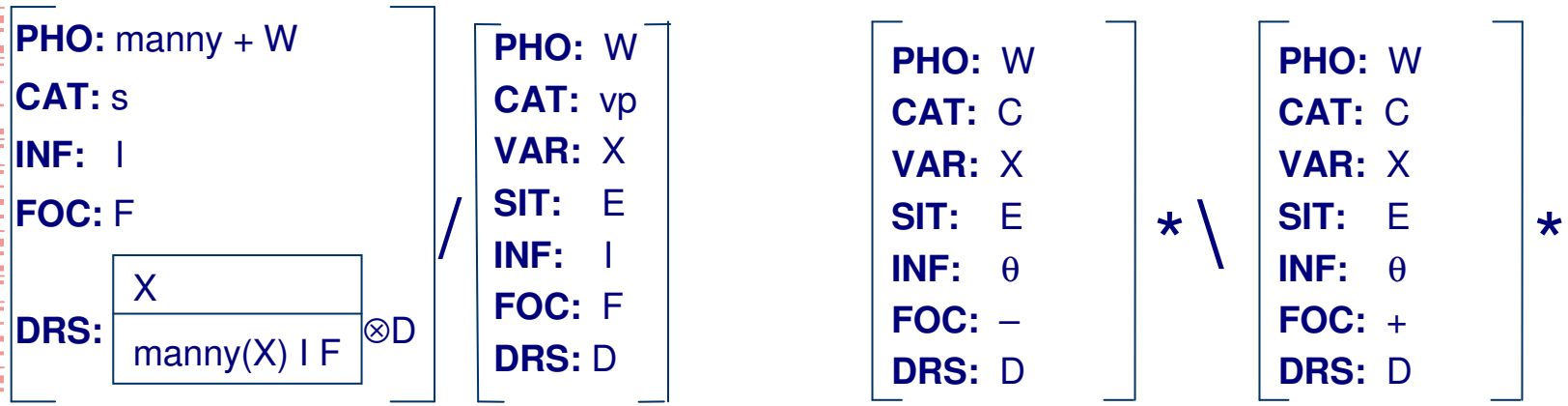
- Appears as an individual lexical item
- Copies the sign on its left, and instantiates the information structure feature value to theme (θ) or rheme (ρ)
- Via unification marks the semantics of the lexical item as being focused (+) (different values in active & result parts)

H*

$$\begin{bmatrix} \textbf{PHO:} & W \\ \textbf{CAT:} & C \\ \textbf{VAR:} & X \\ \textbf{SIT:} & E \\ \textbf{INF:} & \rho \\ \textbf{FOC:} & - \\ \textbf{DRS:} & D \end{bmatrix}^{*} \backslash \begin{bmatrix} \textbf{PHO:} & W \\ \textbf{CAT:} & C \\ \textbf{VAR:} & X \\ \textbf{SIT:} & E \\ \textbf{INF:} & \rho \\ \textbf{FOC:} & + \\ \textbf{DRS:} & D \end{bmatrix}^{*}$$

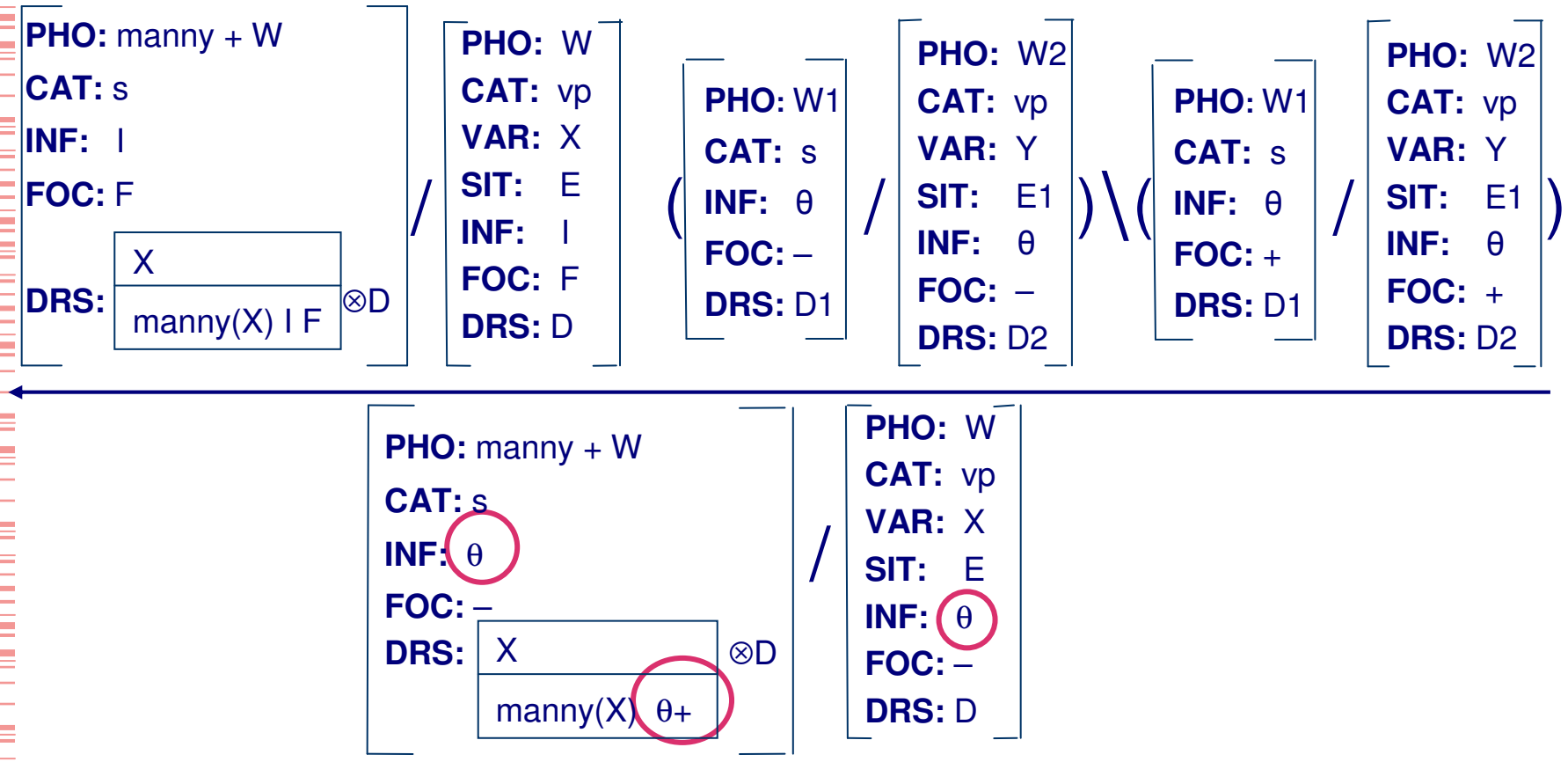
Combining with a Pitch Accent

- Manny L+H*



Combining with a Pitch Accent

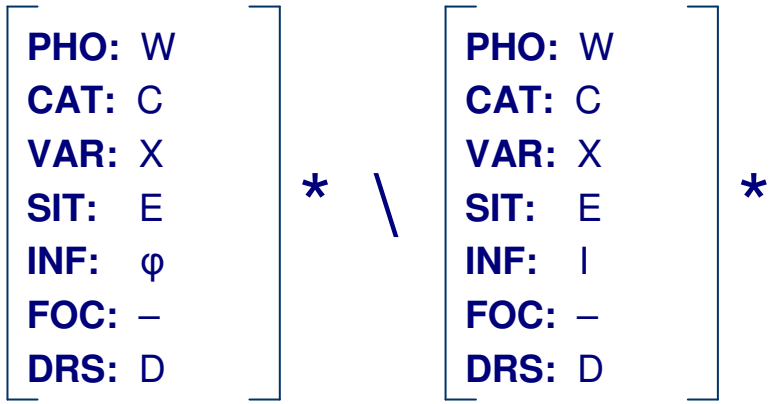
- Manny L+H*



Boundary Tones

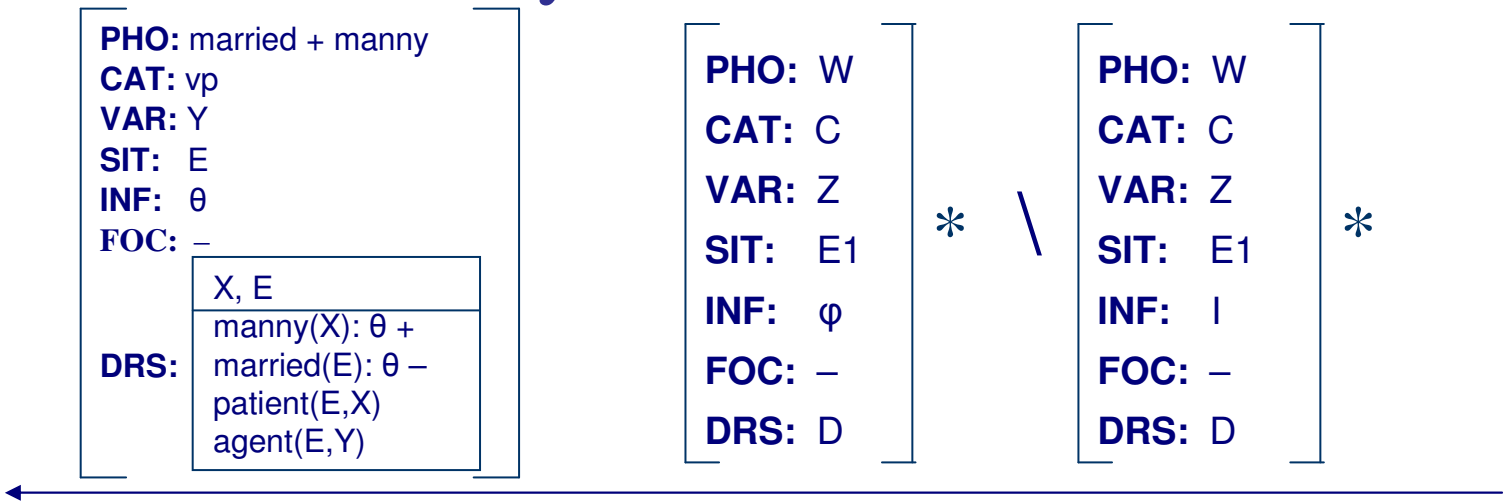
- Boundary tones "freeze" intonational phrases:

cannot combine with anything else but a complete phrase



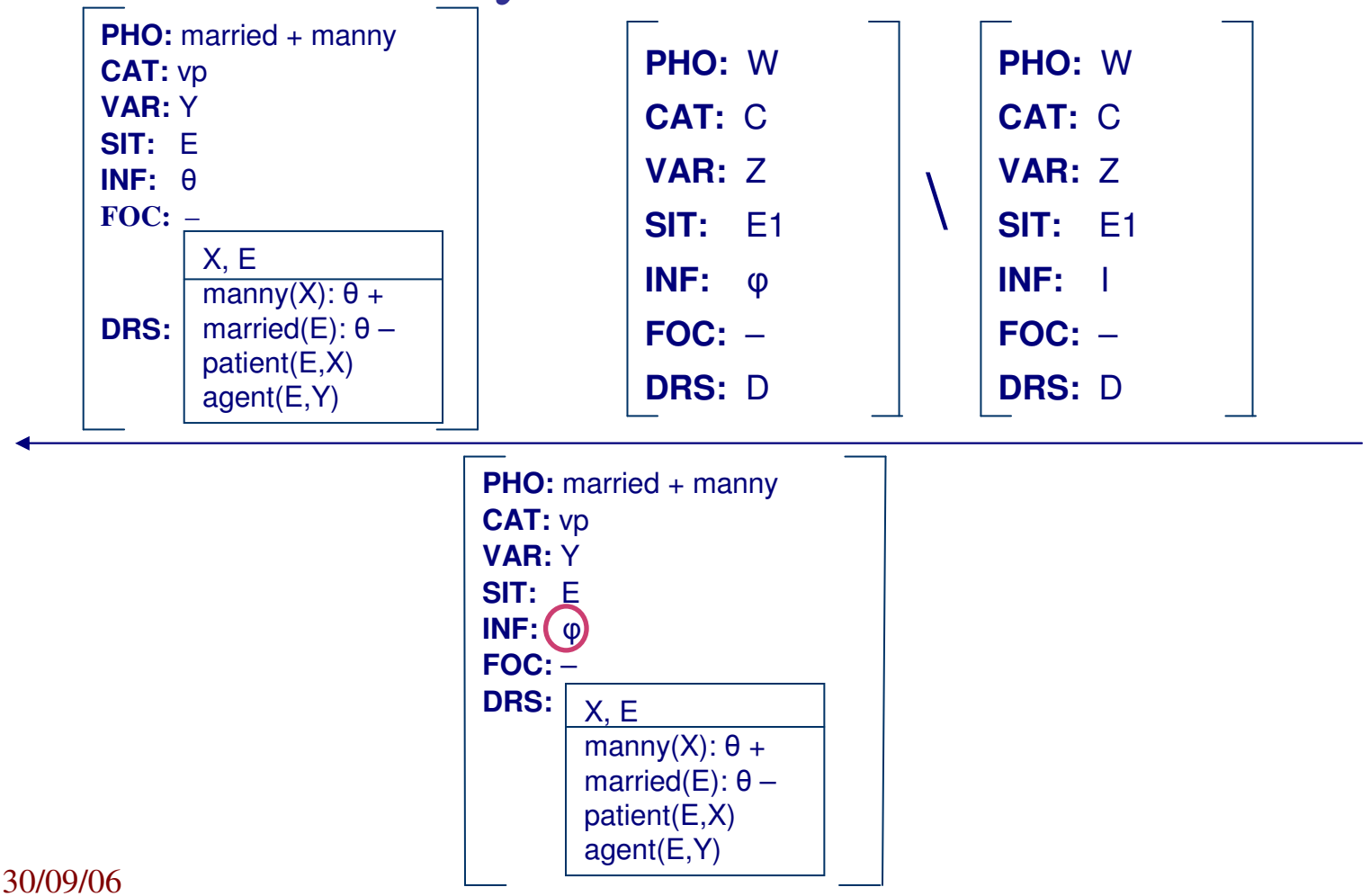
Combining with a Boundary Tone

- married Manny L+H* LH%



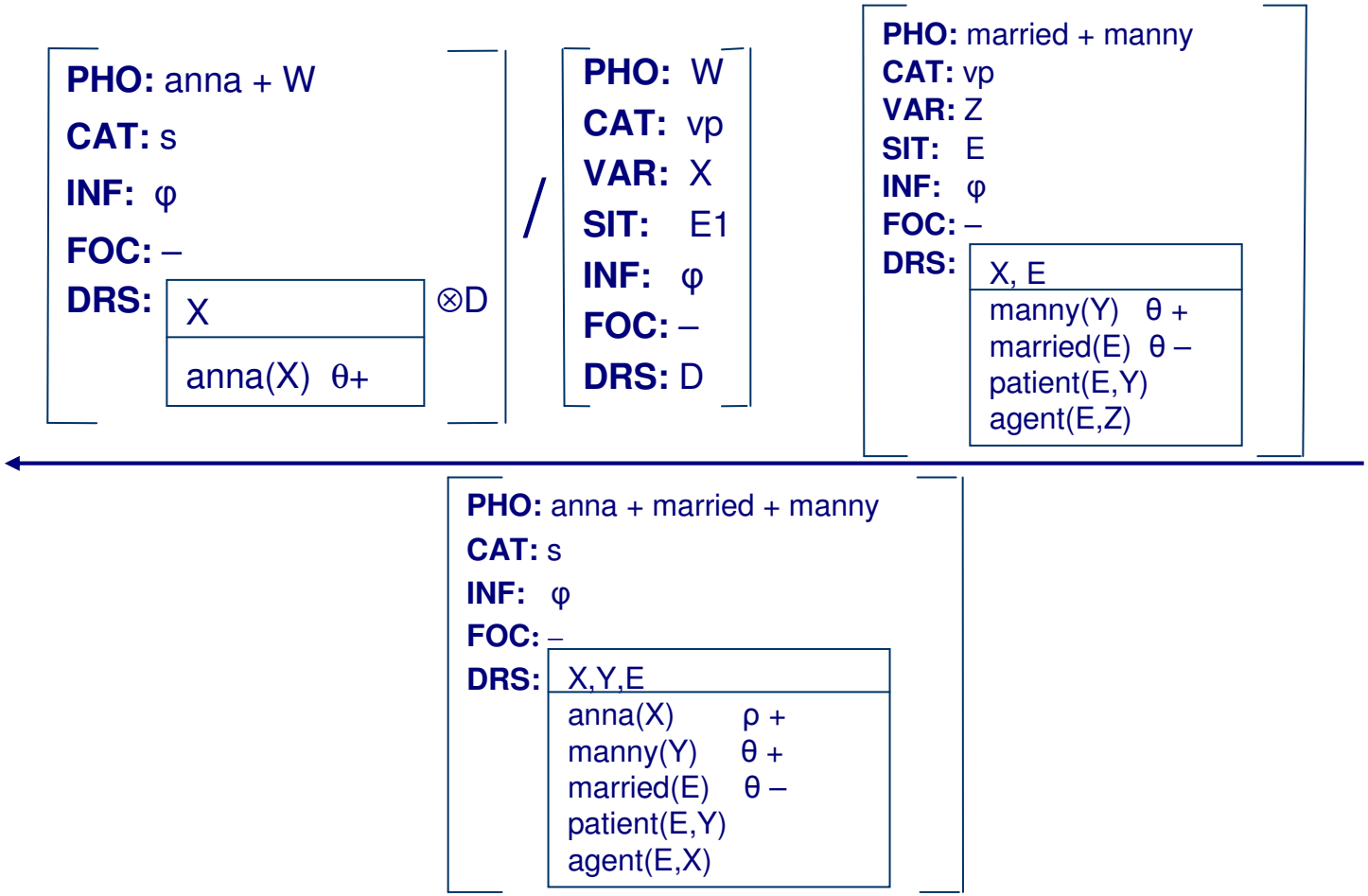
Combining with a Boundary Tone

- married Manny L+H* LH%



Combining of Full Intonational Phrases

- Anna H* LL% married Manny L+H* LH%



Conclusion

- Integrated DRT, IS and CCG into a single framework
  - IS-DRS
    - DRT combined with a theory of information structure
  - Unification-based Combinatory Categorial Grammar
    - parsing and generating prosodically annotated text
    - compositional analysis of information structure

Future Work

- Enhance the existing UCCG parser
- Implement a generation component based on UCCG
- Use the formalism in existing spoken dialogue systems

For a more complete account see:

http://www.iccs.inf.ed.ac.uk/~s0129610/thesis.ps.gz