First Steps in Multimodal Grammar Development in LKB & PET

Katya Alahverdzhieva
supervision: Alex Lascarides & Mark Steedman

School of Informatics
University of Edinburgh
K.Alahverdzhieva@sms.ed.ac.uk

July 6, 2010

Objectives

- analyse synchronous speech and co-verbal gesture
- ... so as to derive a single multimodal derivation tree
- use the Grammar Matrix as a basic grammar framework
- ... whereby extending it with types, features and construction rules
- use PET's token feature structures
- ... for parsing XML-based annotation data (ANVIL)
- ... containing gesture feature structures and prosodic information
- ... and for defining temporal constraints between speech and gesture

Outline

Outline

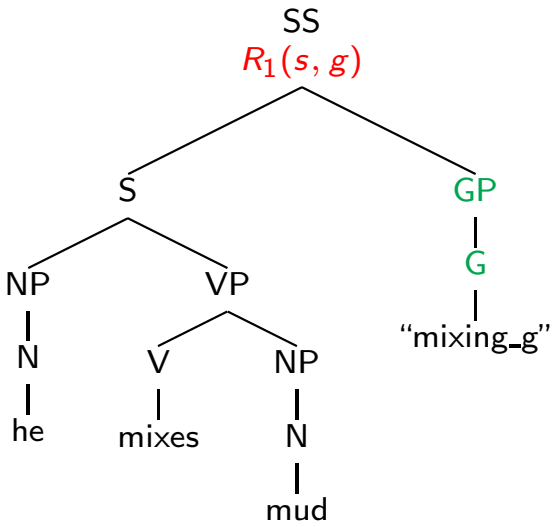
Speech-Gesture Synchrony

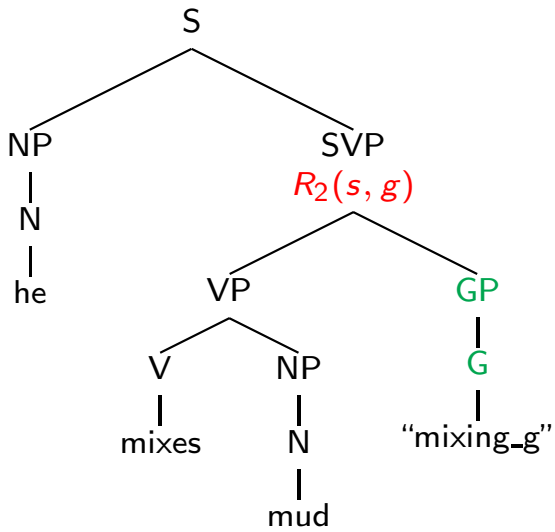
- Synchrony is prosodic, semantic and pragmatic [McNeill, 1992]
- But is it also syntactic?

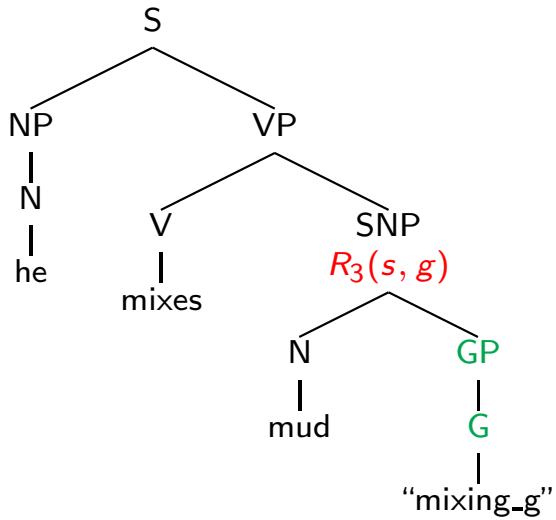
Example (Loehr, 2004)

So he MIXes MUD ...

Speech-Gesture Synchrony, Cont'd



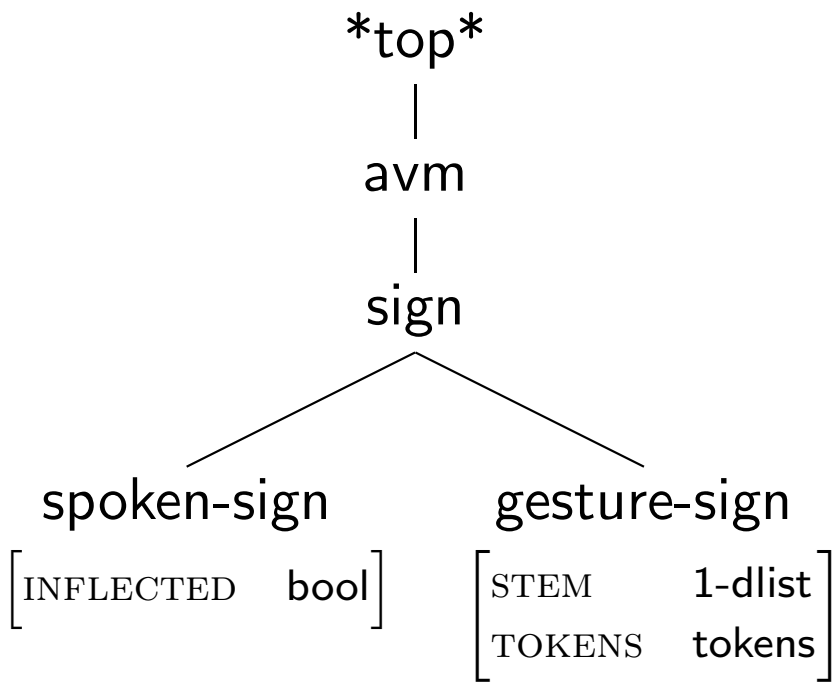




Outline

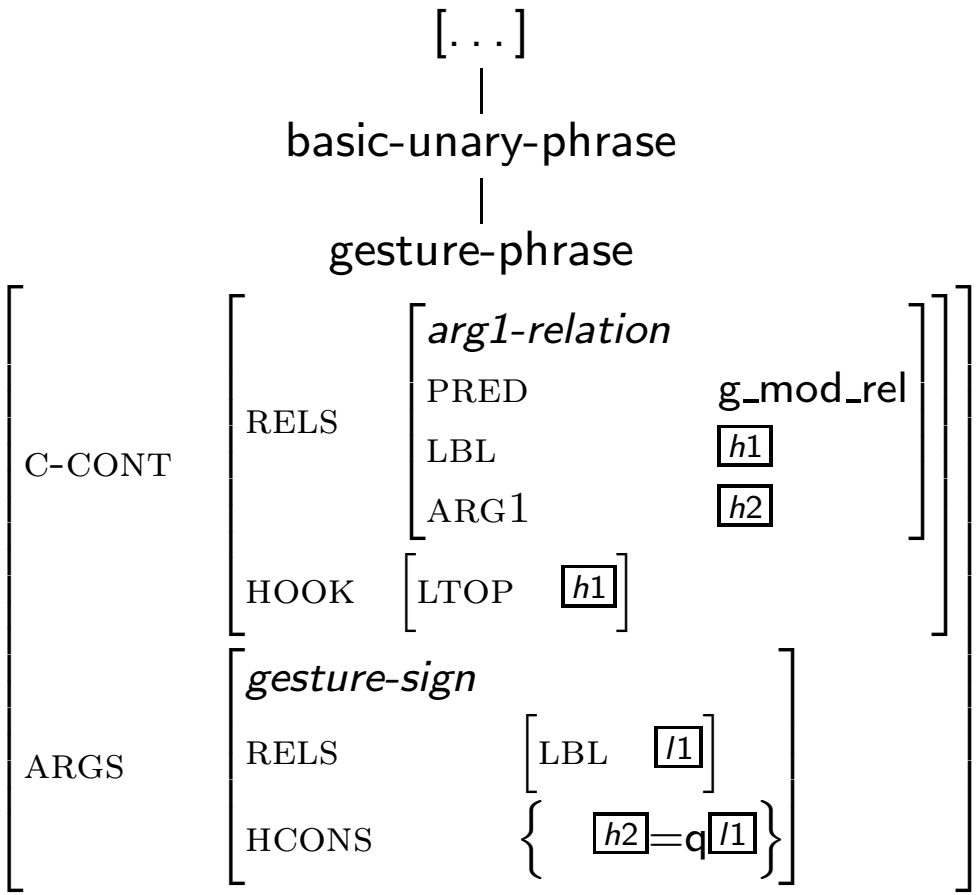
Type Hierarchy

Signs



Type Hierarchy, Cont'd

Phrases



... we'll review this rule later...

Type Hierarchy, Cont'd

Construction Rules

[. . .]

basic-binary-headed-phrase

situated-phrase

$$\begin{bmatrix}
\text{SS} \mid \text{LOC} & \begin{bmatrix} \text{CAT} & \boxed{1} \\ \text{CONT} \mid \text{HOOK} & \begin{bmatrix} \text{IND} & \boxed{2} \\ \text{LTOP} & \boxed{3} \end{bmatrix} \end{bmatrix} \\
\text{S-DTR} & \begin{bmatrix} \text{SS} \mid \text{LOC} & \begin{bmatrix} \text{CAT} & \boxed{1} \\ \text{CONT} \mid \text{HOOK} \mid \text{LTOP} & \boxed{h1} \end{bmatrix} \end{bmatrix} \\
\text{G-DTR} & \begin{bmatrix} \text{SS} \mid \text{LOC} \mid \text{CONT} \mid \text{HOOK} \mid \text{LTOP} & \boxed{h2} \end{bmatrix} \\
\text{C-CONT} \mid \text{RELS} & \begin{bmatrix} \textit{vis-relation} & \\ \text{PRED} & \text{vis_rel} \\ \text{LBL} & \boxed{3} \\ \text{M-ARG} & \boxed{2} \\ \text{S-HNDL} & \boxed{h1} \\ \text{G-HNDL} & \boxed{h2} \end{bmatrix}
\end{bmatrix}$$

Outline

Multimodal Input

[Loehr, 2004]

- text annotated with timing

```
<el index="117" start="33.155910491" end="33.261188507">
  <attribute name="text">he</attribute>
</el>
<el index="118" start="33.261188507" end="33.667694091">
  <attribute name="text">mixes</attribute>
</el>
<el index="119" start="33.667694091" end="34.289043426">
  <attribute name="text">mud</attribute>
</el>
```

- .. and prosody

```
<el index="35" start="33.19949" end="33.32139">
   <attribute name="text"> H* </attribute>
</el>
<el index="36" start="33.71449" end="33.84776">
   <attribute name="text"> X* </attribute>
</el>
```

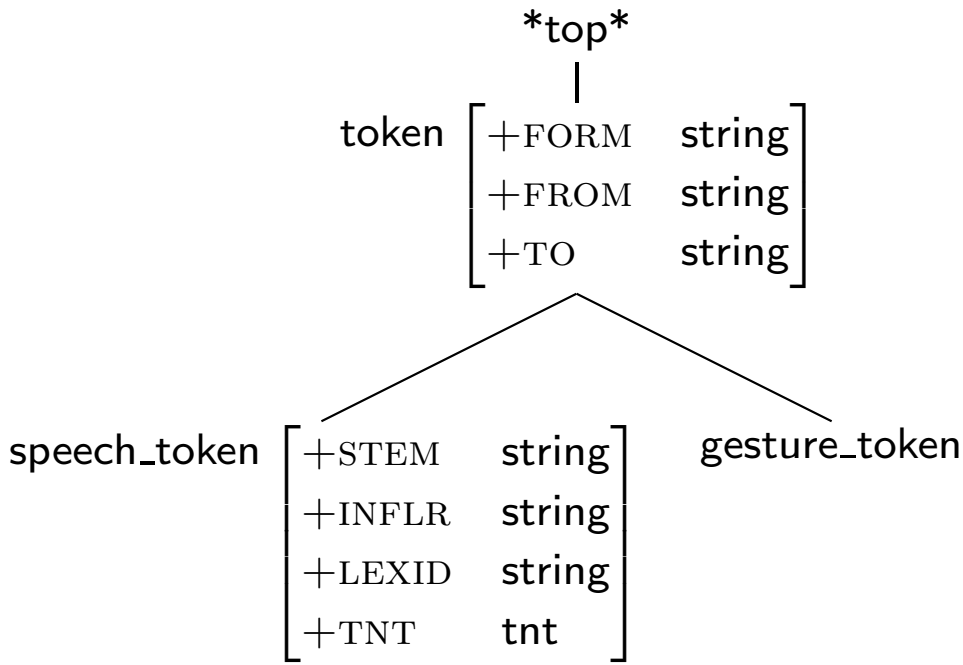
- ... and gestural performance

```
<el index="24" start="59" end="59">
  <attribute name="Hand Shape"> Open-Flat </attribute>
  <attribute name="Palm Orientation"> TC (Towards Centre) </attribute>
  <attribute name="Finger Orientation"> Towards Down </attribute>
  <attribute name="Location"> Centre-Low </attribute>
  <attribute name="Move"> Circular </attribute>
  <attribute name="Gloss"> mixing drywall mud </attribute>
</el>
```

Outline

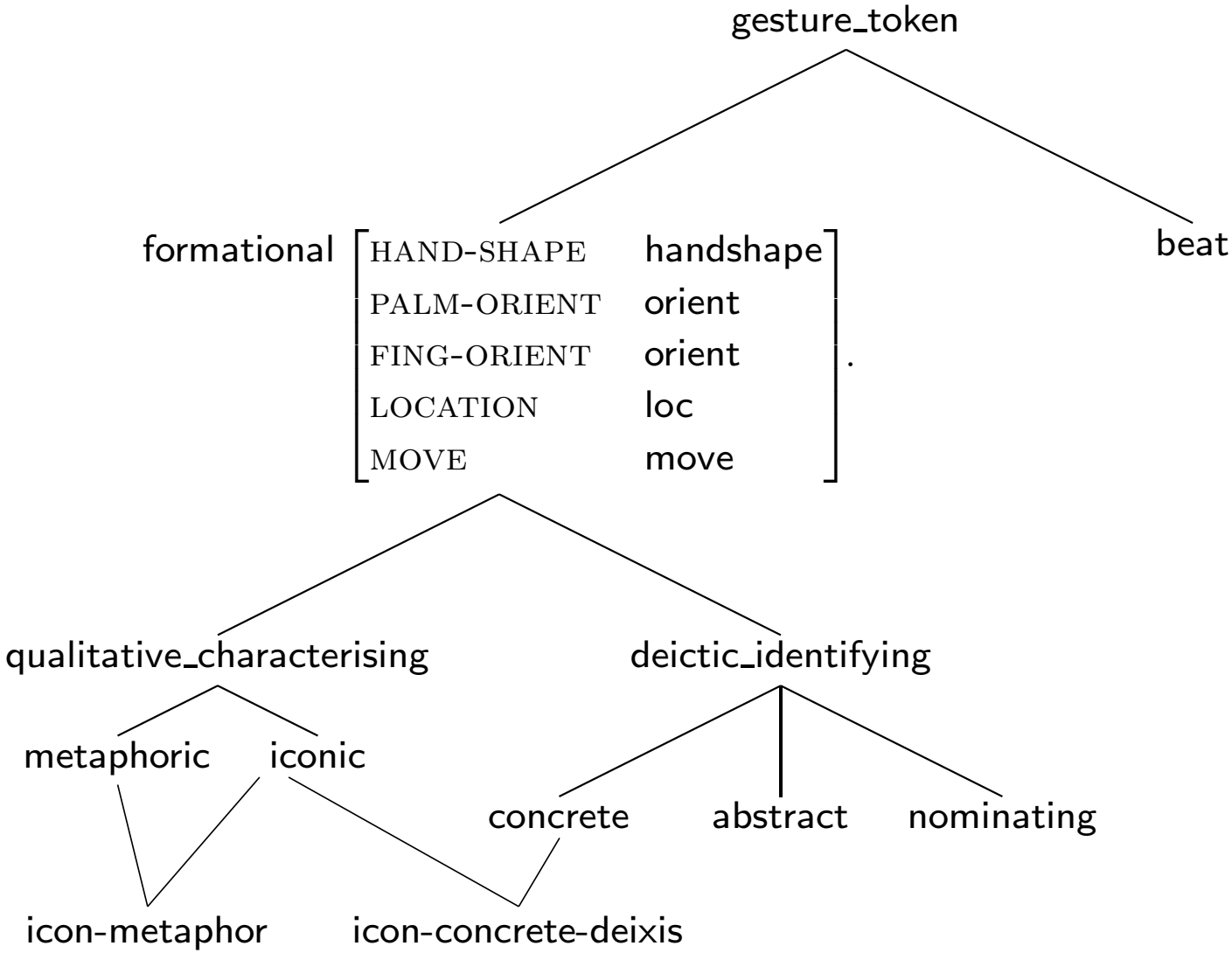
Extending Token Type Hierarchy

Token Types

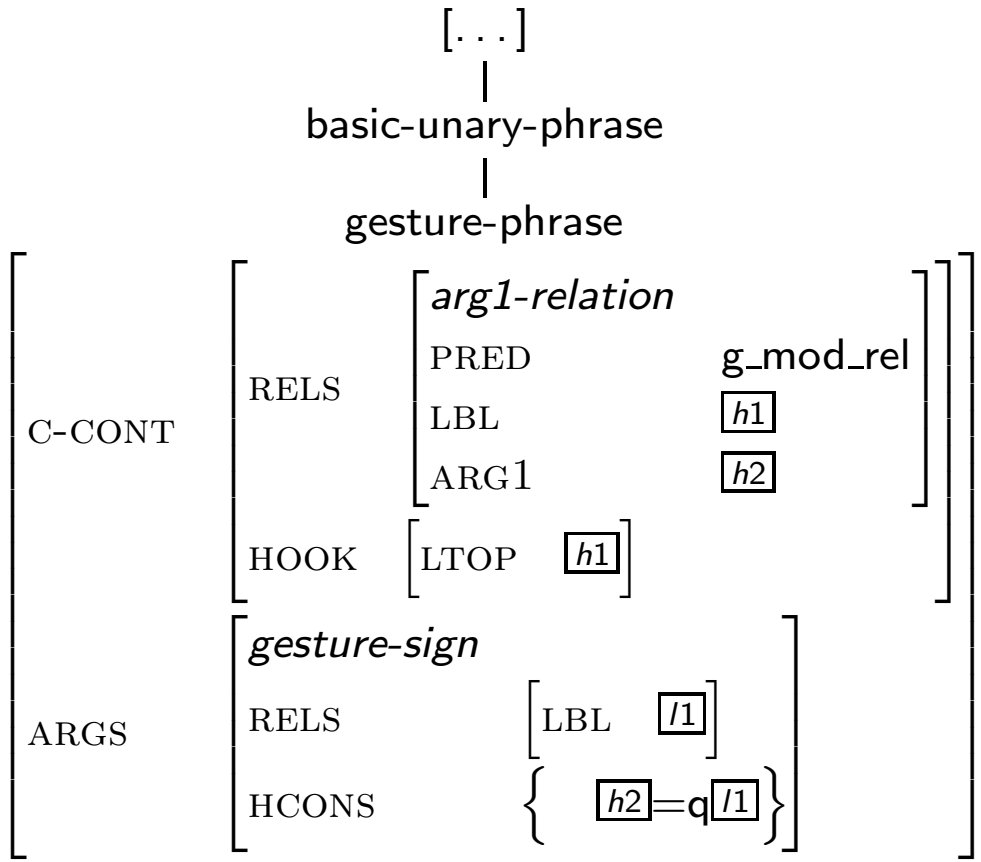


Extending Token Type Hierarchy, Cont'd

Gesture features, a fragment

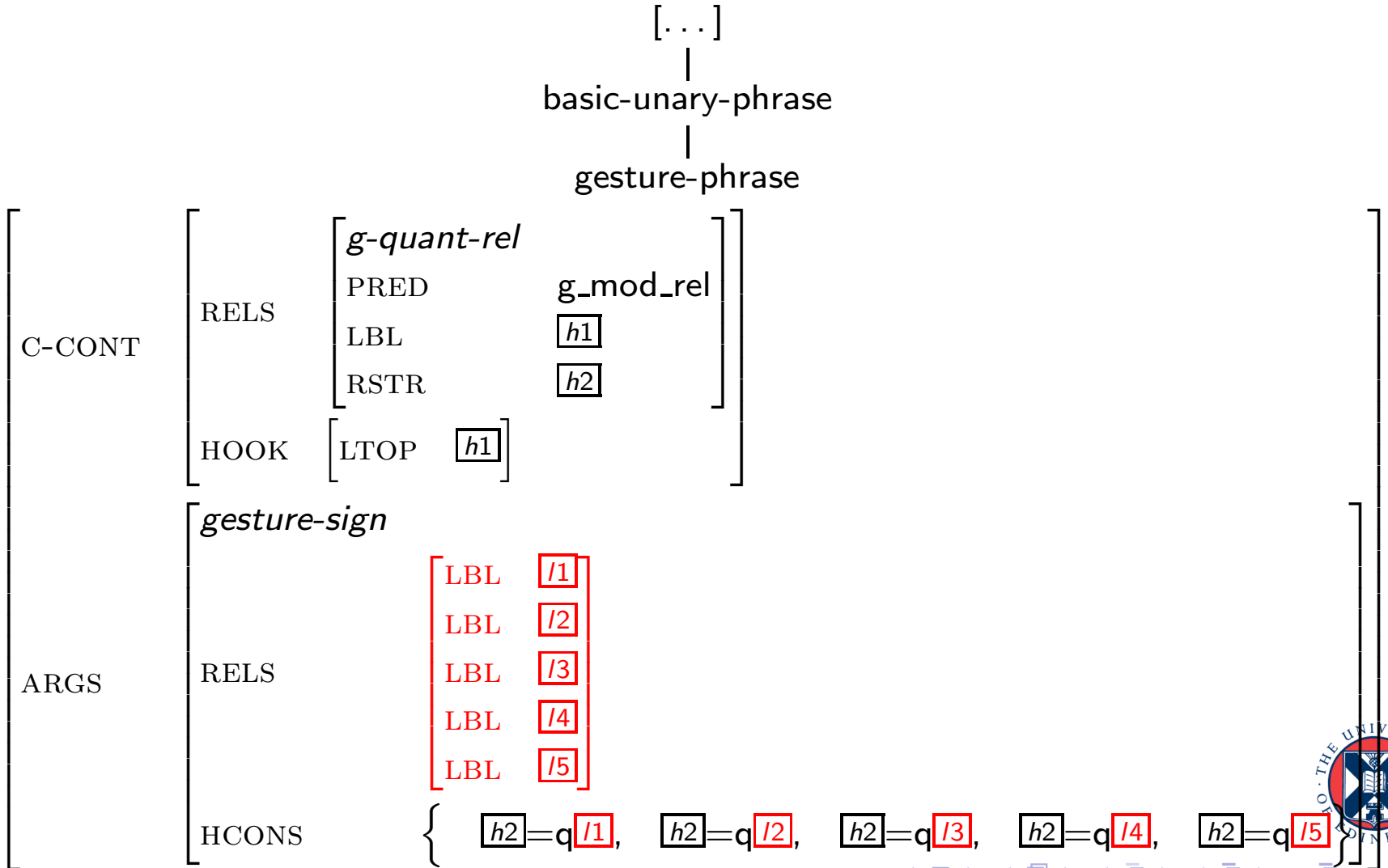


How do the gesture features reflect our gesture rule?



How do the gesture features reflect our gesture rule?

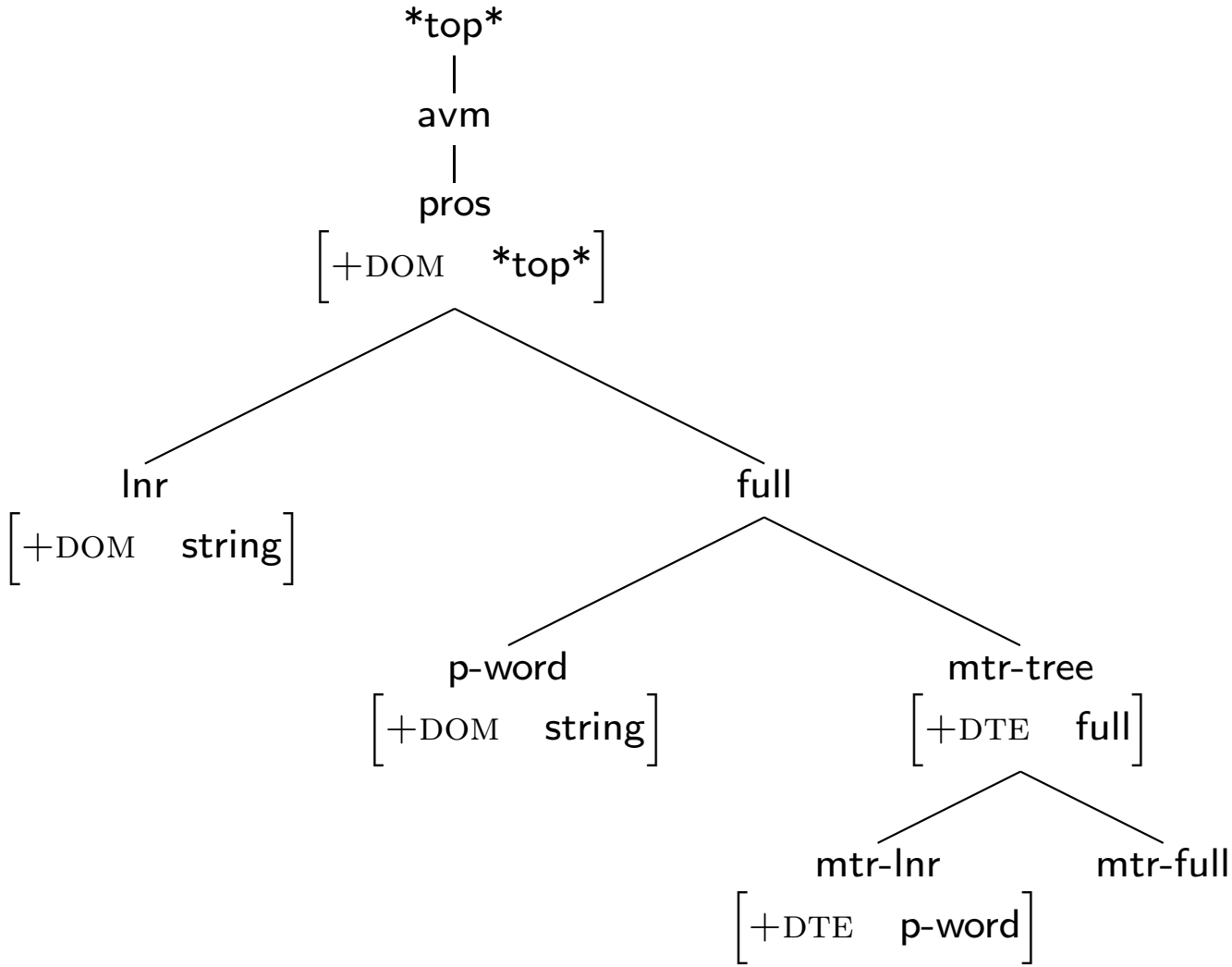
- every feature-value maps to an underspecified predicate, [Lascarides & Stone, 2009]:



Outline

Extending Type Hierarchy with Prosody

Prosodic types and features, [Klein 2000]



Outline

Future Work and Discussion

- derive structured phonology

$$\begin{bmatrix} \text{PHON} & \langle \text{HE, MIXES, MUD} \rangle \\ \text{SYNSEM} & \text{synsem} \end{bmatrix}$$

Figure: Unstructured Phonology

vs.

$$\begin{bmatrix} \textit{mtr(full)} & & \\ & \begin{bmatrix} \textit{mtr(lnr)} & & \\ & & \\ \text{PHON} & \text{DOM} & \left\langle \begin{bmatrix} \textit{mtr(lnr)} & \\ \text{DOM} & \langle \textit{he}, \boxed{1}\ \textit{mixes} \rangle \\ \text{DTE} & \boxed{1} \end{bmatrix}, \boxed{2}\textit{mud} \right\rangle \\ & & \\ & \text{DTE} & \boxed{2} \end{bmatrix} \\ \text{SYNSEM} & \text{synsem} & \end{bmatrix}$$

Figure: Structured Phonology

Future Work, Cont'd

- integrate a gesture and a prosodic structure (which is not a syntactic constituent)
- add semantics to a prosodic structure (which is not a syntactic constituent)
- add temporal constraints

References I


Ewan Klein.
Prosodic Constituency in HPSG,
Grammatical Interfaces in HPSG, Studies in Constraint-Based Lexicalism, 2000.


Alex Lascarides and Matthew Stone,
A Formal Semantic Analysis of Gesture,
Journal of Semantics, 2009.


Daniel Loehr.
Gesture and Intonation,
Doctoral Dissertation, Georgetown University, 2004.


David McNeill.
Hand and Mind. What Gestures Reveal about Thought,
University of Chicago Press, 1992.