Modelling the role of grammatical functions in language processing

Stephen Jones
stephen.jones@ling-phil.ox.ac.uk
Faculty of Linguistics, Philology and Phonetics, University of Oxford

Aims

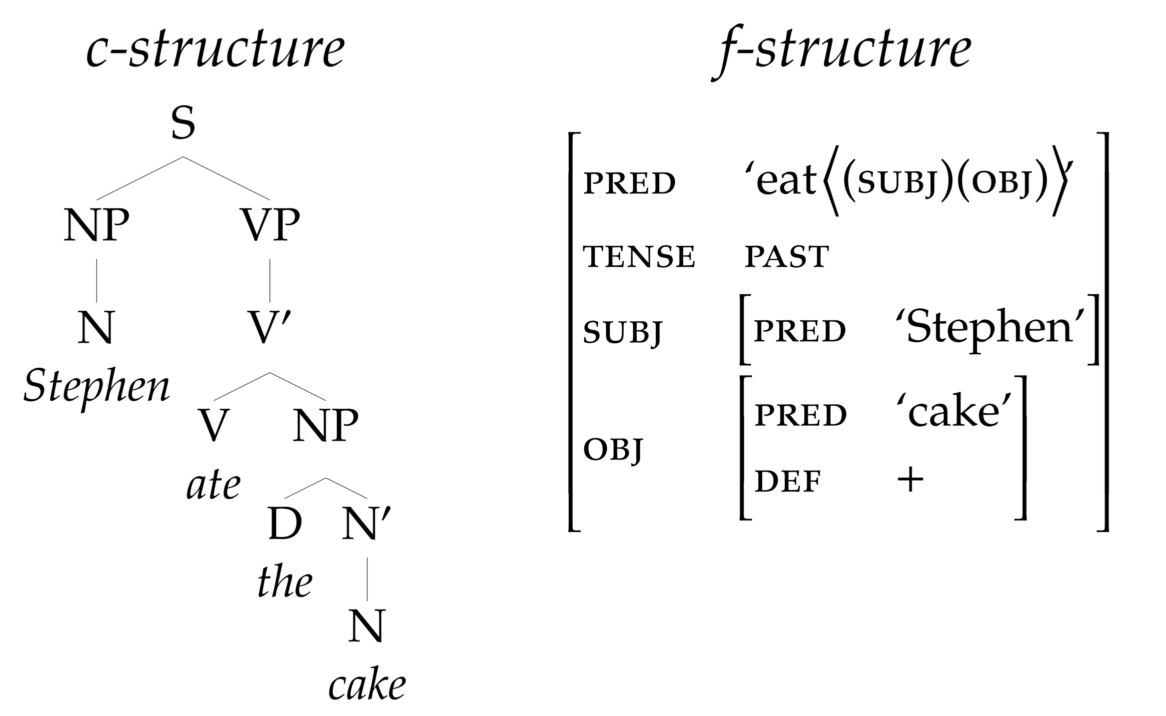
Overall: Develop a cognitive parsing model based on LFG using grammatical functions (GFs) for memory structure and retrieval cues.

This project: Test a model that uses only base ACT-R 7 resources against experimental data from Grodner and Gibson (2005).

Lexical Functional Grammar

In LFG (Dalrymple, 2001) syntax distinguishes
- c-structure: word class, phrase structure
- f-structure: semantic content, GF

GFs are seen as language universals; meaning is derived from f-structure.



For English, word order determines GF. In the model:
- c-structure rules are encoded in productions
- c-structure as read does not persist
- grammar chunks record f-structure
- word order is not recoverable from DM

References

Dalrymple, M. (2001). *Lexical Functional Grammar.* Academic Press, San Diego. **Grodner, D. and Gibson, E. (2005).** Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, 29(2):261-290. **Lewis, R. L. and Vasishth, S. (2005).** An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29:375-419.

Assumptions

- GF prediction order:
  SUBJ ≺ PRED ≺ OBJ ≺ OBL ≺ ADJ
- Lexical info trumps predicted GF
- Try to close long-distance dependencies (LDDs) whenever possible
- Repair and reopen LDDs if trial fails

Grammar chunk info:
- type
- parent attachment
- f-structure attribute-value pairs
- coreference of LDDs

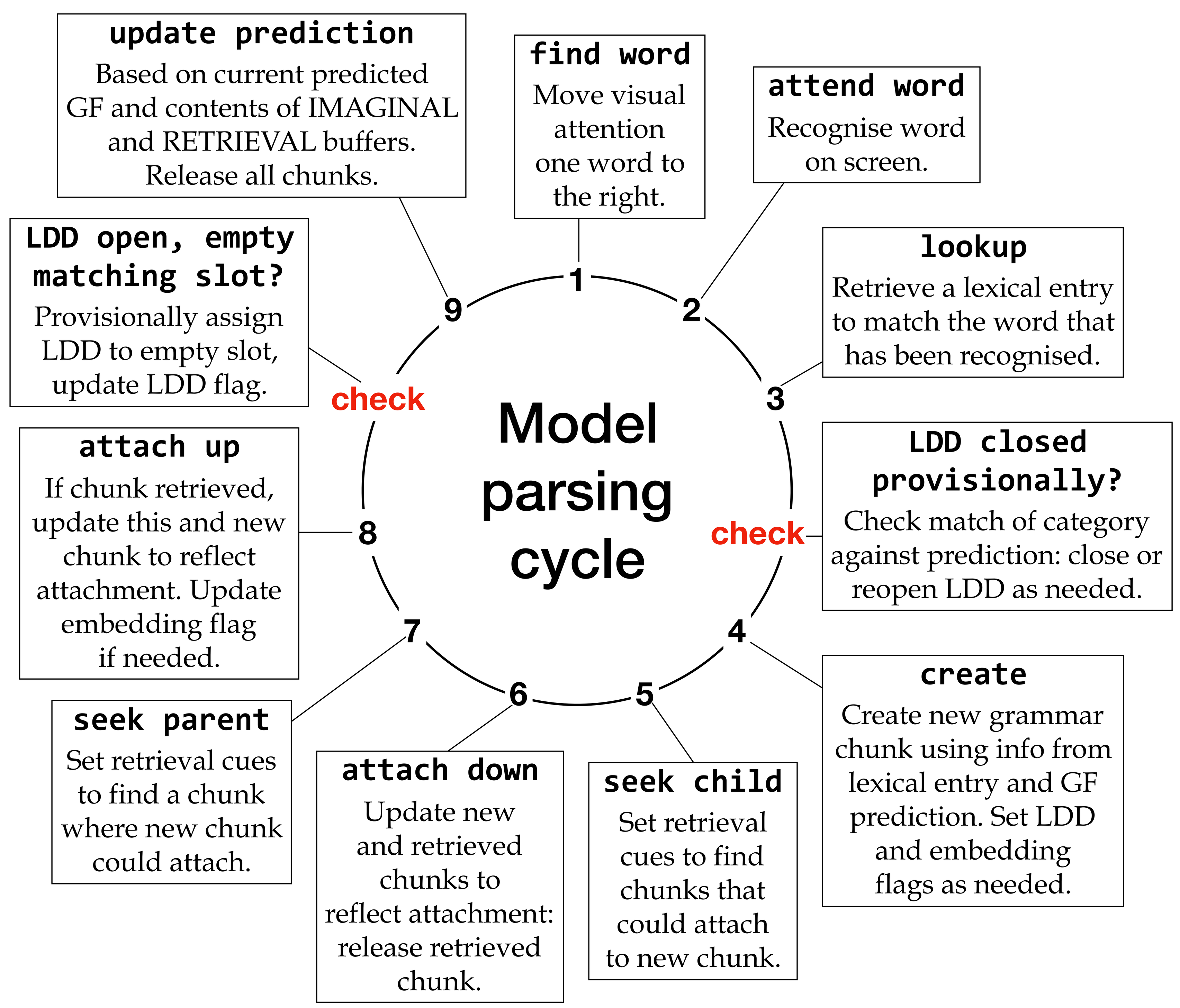
CHUNK ID	*e.g.* F7
TYPE	*e.g.* N/P/V
PRED	*semantic info*
FEATURES	*e.g.* DEF +
LDD	n/poss/y/ID
HOST ID	ID
HOST GF	*e.g.* OBJ
child GFs, e.g. SUBJ/OBJ/SPEC	nil/n/poss/reqd/child ID

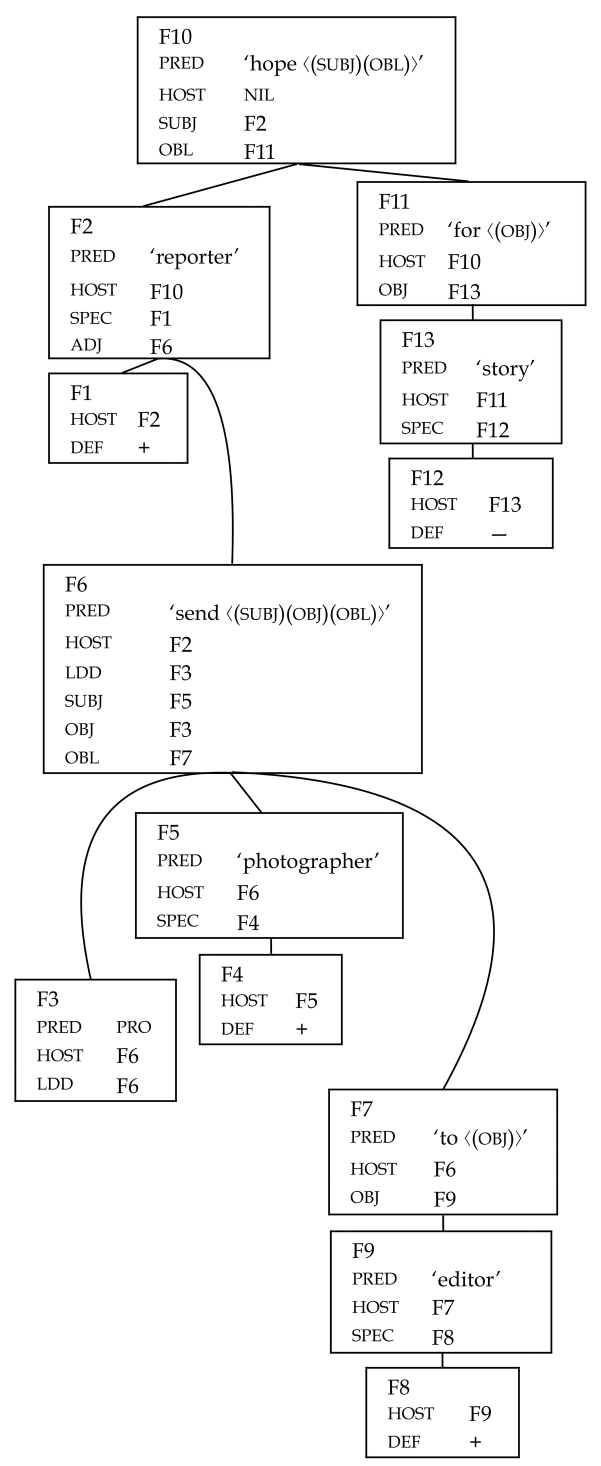
Differences to L&V (2005):
- *No* extra cognitive resources for parsing
- *Functional* structure recoverable
- Constituent structure *not* built
- *All* chunk creation has a time cost
- New chunks may be released *unattached*

Goal chunk info:
- parse state
- predicted GF
- embedding and LDD states (max. depth 2)
- LDD type

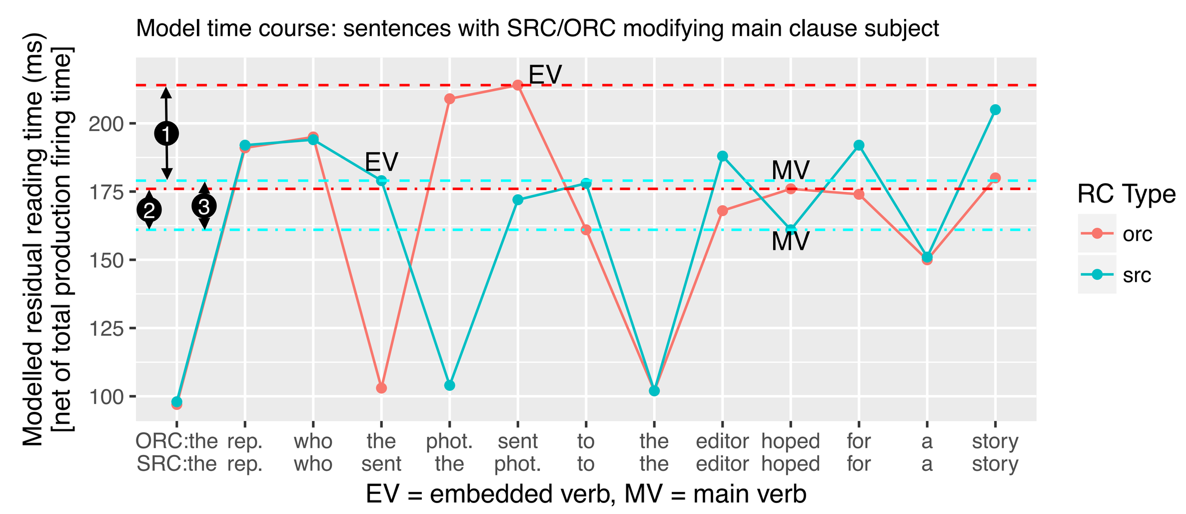
GOALSTATE	*e.g.* attach-up
TARGETGF	*e.g.* SUBJ/OBJ
EMBED 1	n/y
EMBED 2	n/y
LDDOPEN 1	n/poss/y
LDDTYPE 1	*e.g.* N/P/V
LDDOPEN 2	n/poss/y
LDDTYPE 2	*e.g.* N/P/V

Model parsing cycle and recoverable syntactic structure





Results



Three processing asymmetries are relevant. Experimental reference data from Grodner and Gibson (2005) Experiment 1.

❶ SRC-ORC at the embedded verb: model qualitatively matches data.

❷ SRC-ORC at the matrix verb: model qualitatively matches data.

❸ Matrix-embedded verb in SRC sentence: model asymmetry qualitatively *against* data.

Discussion

- Model successfully parses SRCs and ORCs.
- Production path length varies by word.
- Main determiner of time variation is no. of attachment productions required (0-6).
- Productions are smallest possible steps, could combine to streamline.
- Reducing path variation by only attaching upward needs more IMAGINAL capacity.
- Simultaneous attachment to parent and children needs more RETRIEVAL capacity.

Conclusions

GFs can parse sentences with embedded verbs using only the resources of base ACT-R 7. Improving fit with experimental timing data requires additional buffer capacity, more complex productions, or both.