A cognitive model of processing using LFG

Stephen Jones

stephen.jones@ling-phil.ox.ac.uk

Faculty of Linguistics, Philology and Phonetics, University of Oxford

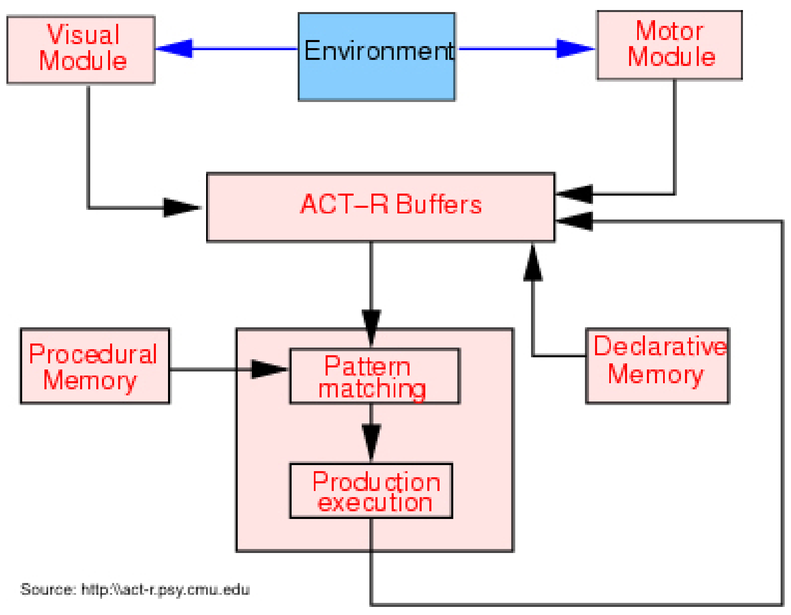
Aims

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).

ACT-R Cognitive Environment

- Represents a theory of mind (Anderson, 2007)
- Allows detailed models of cognitive processes
- Limited resources for cognition
- One buffer per cognitive system (e.g. visual)
- One chunk at a time per buffer
- Only buffered chunks available



Lewis & Vasishth (2005) parsing module:

- Modelled processing data for complex sentences
- Phrasal projection nodes (*e.g.* CP, V′) stored and recoverable as distinct memory chunks
- Structural chunks created in zero time
- Adds five new buffers (Engelmann, 2015)
- Not available in current ACT-R version

Model assumptions

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

Differences to L&V (2005):

- *No* additional cognitive resources for parsing
- *Functional* structure stored and recoverable
- C-structure not built, PSRs in productions
- *All* chunk creation has a time cost
- New chunks may be released *unattached*

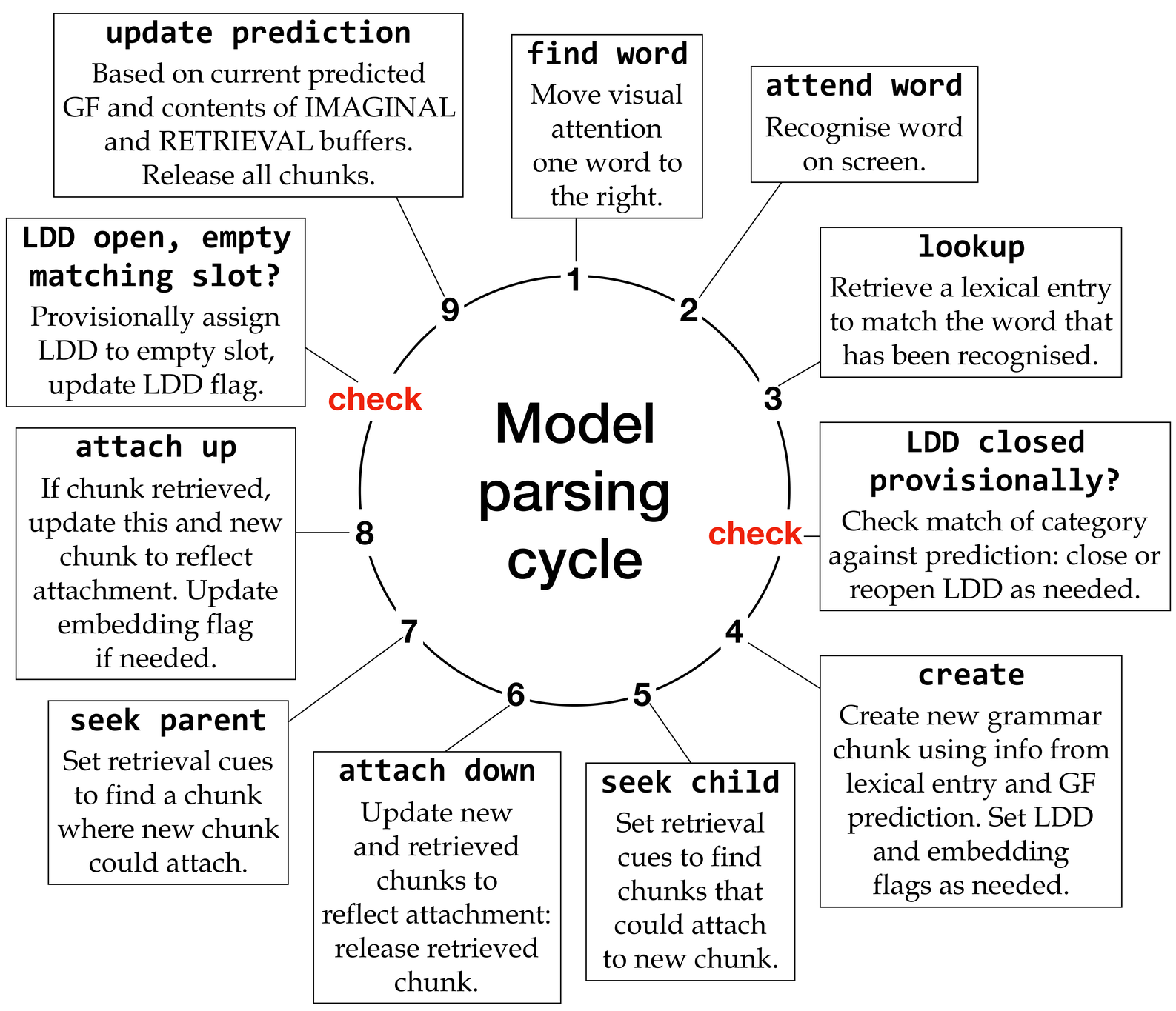
Grammatical memory		*Control memory*	
CHUNK ID	*e.g.* F7	GOALSTATE	*e.g.* attach-up
TYPE	*e.g.* N/P/V	TARGETGF	*e.g.* SUBJ/OBJ
PRED	*semantic info*	EMBED 1	n/y
FEATURES	*e.g.* DEF +	EMBED 2	n/y
LDD	n/poss/y/ID	LDDOPEN 1	n/poss/y
HOST ID	ID	LDDTYPE 1	*e.g.* N/P/V
HOST GF	*e.g.* OBJ	LDDOPEN 2	n/poss/y
child GFs,	nil/n/	LDDTYPE 2	*e.g.* N/P/V
e.g. SUBJ/	poss/reqd/		
OBJ/SPEC	child ID		

How does prediction work in the model?

- Prediction and lexical information interact to determine which GF is assigned.
- Where predicted and lexical category match, the new chunk receives the predicted GF.
- Where there is a mismatch, next actions depend on the overall buffer contents.
- Trial closure (★) of open LDDs if LDD type matches predicted category and predicted GF ≠ SUBJ

Parsing *The girl who Mary sent the book to ate the cake.*

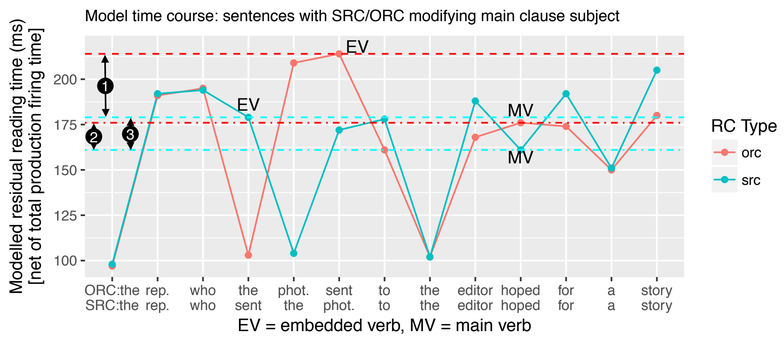
Open LDD?	–	–	–	+	+	+	?	?	+	+	?	–
LDD type	–	–	–	N	N	N	(N)	(N)	N	N	(N)	–
Predicted GF	SUBJ	SUBJ	PRED	SUBJ	PRED	OBJ	OBL	OBL	OBL	OBJ	ADJ	OBJ
Predicted category	N	N	V	N	V	N	P	P	P	N	A/P	N
Lexical item	*the*	*girl*	*who*	*Mary*	*sent*	★	*the*	*book*	*to*	★	*ate*	...
Category	D	N	N	N	V		D	N	P		V	
Match?												
Assigned GF	SPEC	SUBJ	DF	SUBJ	PRED	OBJ	SPEC	OBJ	OBL	OBJ	PRED	



References

Anderson, J.R. (2007). *How can the human mind occur in the physical universe?* OUP. **Engelmann, F. (2015)**. *Act-R-Parsing-Module*. https://github.com/felixengelmann/ACT-R-Parsing-Module **Grodner, D. and Gibson, E. (2005)**. Consequences of the serial nature of linguistic input for sentential complexity. *Cogn Sci*, 29:261-290. **Lewis, R. L. and Vasishth, S. (2005)**. An activation-based model of sentence processing as skilled memory retrieval. *Cogn Sci*, 29:375-419.

Results



Three processing asymmetries are relevant. Comparison experimental data is 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 is qualitatively against data.

Discussion

- Model successfully parses SRCs and ORCs.
- Parsing production path varies between words.
- Main determiner of time variation is number of attachment productions required (0-6).
- Reducing path variation by only attaching upward needs more IMAGINAL capacity.
- Simultaneous attachment to parent and children needs more RETRIEVAL capacity.

Conclusions

The LFG-based model transparently parses complex sentences using only base ACT-R 7. Replicating experimental time courses will require additional buffer capacity, more complex productions, or both.