

# Revisiting compositionality

Ann Copestake

Natural Language and Information Processing Group  
Computer Laboratory  
University of Cambridge

June 2016

## Revisiting compositionality: aims

- Revisit \*MRS representations by looking at them as graphs (TFS encoded or otherwise).
- Suggest a route for revising the algebra.
- Demonstrate DMRS composition.
- If time, look at incremental semantic construction.

## Some preliminaries: use of events and intrinsic variables

- No e semantics:  
every(x, big(x) & dog(x), quick(past(run))(x))
- Moderate e semantics:  
every(x, big(x) & dog(x), quick(e) & past(e) & run(e))
- Full e semantics (intrinsic arguments):  
every(x, big(e',x) & dog(x), quick(e'',e) & past(e) & run(e))

Note: even more e:

every(e''',x, big(e',x) & dog(x), quick(e'',e) & past(e) & run(e))

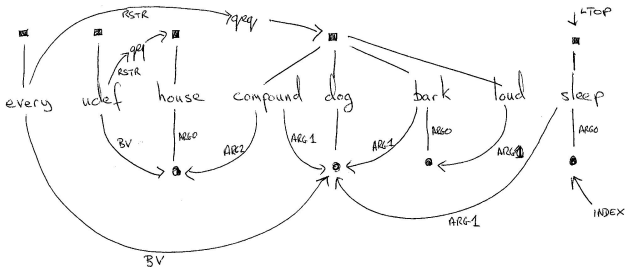
## Some preliminaries: use of events and intrinsic variables

- No e semantics:  
every(x, big(x) & dog(x), quick(past(run))(x))
- Moderate e semantics:  
every(x, big(x) & dog(x), quick(e) & past(e) & run(e))
- Full e semantics (intrinsic arguments):  
every(x, big(e',x) & dog(x), quick(e'',e) & past(e) & run(e))

Note: even more e:

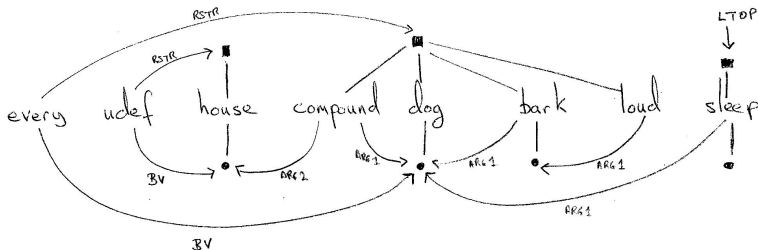
every(e''',x, big(e',x) & dog(x), quick(e'',e) & past(e) & run(e))

# every house dog which barked loudly slept



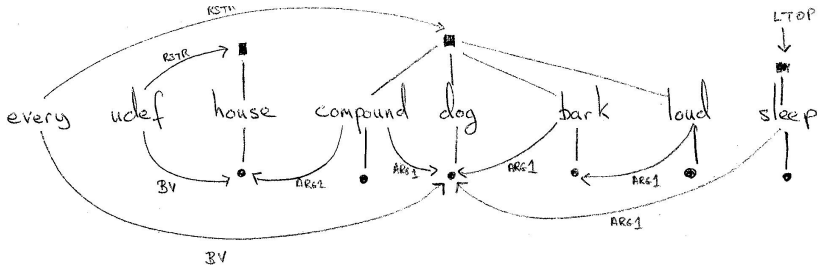
- mod e version: squares are labels, circles are individuals
- simplified! omitting list that links all EPs; using predicate names directly; not writing out qeqs in full
- conversion to MRS involves generating identifiers:  
 $I1:every(x2,h1,h2)$ ,  $I2:udef(x1,h3,h4)$ ,  $I3:house(x1)$ ,  
 $I4:compound(x2.x1)$ ,  $I4:dog(x2)$ ,  $I4:bark(e3,x2)$ ,  $I4:loud(e3)$ ,  
 $I5:sleep(e4,x2)$ ,  $h1 \text{ qeq } I4$ ,  $h3 \text{ qeq } I3$

# Redrawing for readability



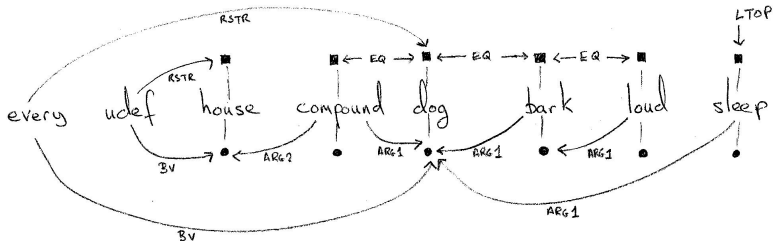
- qeqs entirely implicit
- labels for quantifiers left off (these are anyway inaccessible)

# Full e TFS graph



- additional events, giving intrinsic argument property (Oepen and Lønning, 2006)
- as current ERS

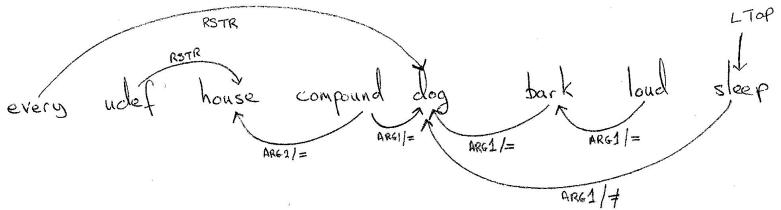
# Full e TFS graph, distinct labels plus label eq links



- Expository purposes only!
- labels are 1:1 with EPs, label equalities rather than coindexed
- cf formalisation of the algebra, early RMRS in-group etc

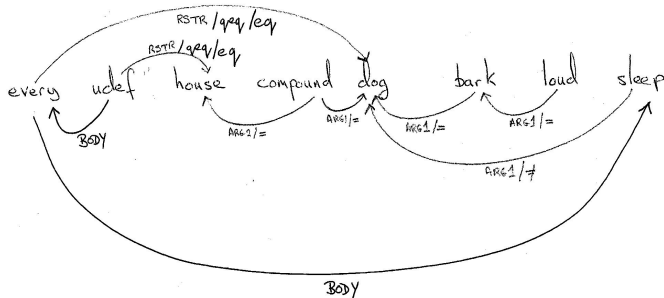


# DMRS graph



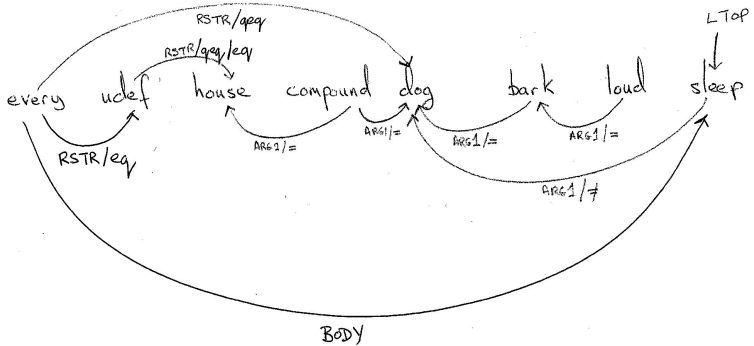
- no labels or individuals
- take advantage of the regularity of composition to combine arcs (cf algebra again)

# DMRS scoping



- Add /heq links consistent with /qeq and noun position. New BODY /heq links for quantifiers.
- Interpret heq to give a scope tree, with = links giving conjunction.

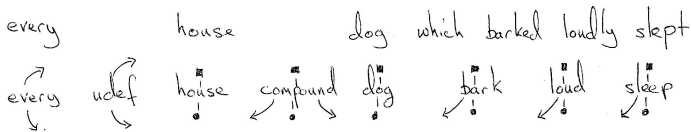
# DMRS scoping



## Compositionality: general ideas

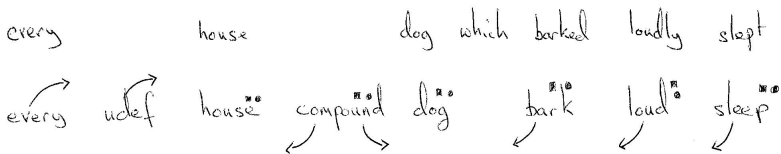
- Underlying rationale: learnability of language (by humans).
- Assumption: syntactically-driven compositional semantic representation possible for sentences out of context.
- Then: regularity in production of this representation is a condition for learnability, given infinite nature of language.
- Note: assumption is compatible with context-specific predicate interpretation.
- Composition has two components (not simply ordered):
  1. partial map between words and elementary predications (plus syntactically-motivated construction EPs)
  2. combination of EPs according to syntax and (reasonably) straightforward principles

# Compositional component 1: MRS



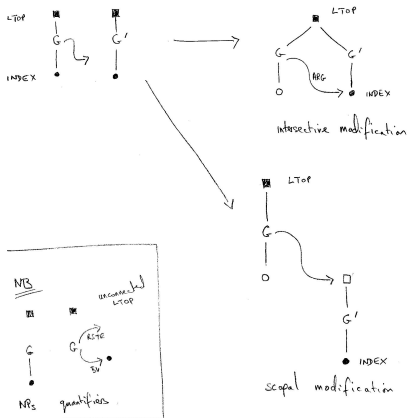
- unconnected EPs, mostly lexical
- filled squares are LTOPs, filled circles INDEXs (xarg omitted for simplicity)
- open links lexically specified (not a necessary assumption)
- assumptions required by learnability
  - ▶ the lexical items which have irregular relationships with EPs are in a closed class
  - ▶ contribution from constructions is systematic

# Compositional component 1: DMRS



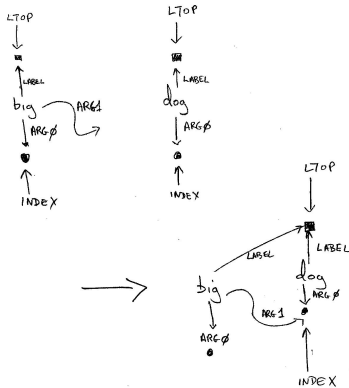
- very similar to MRS
- LTOP, INDEX are nodes of the DMRS graph

## Compositional component 2: MRS



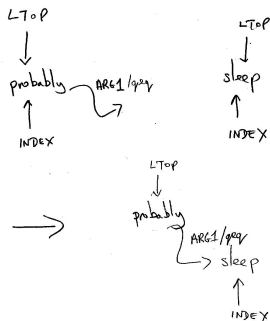
- small number of general patterns for semantic combination operations (to be confirmed on current ERG)
- cf the algebra (which unfortunately doesn't work as specified with full e MRS)

# Compositional example: MRS





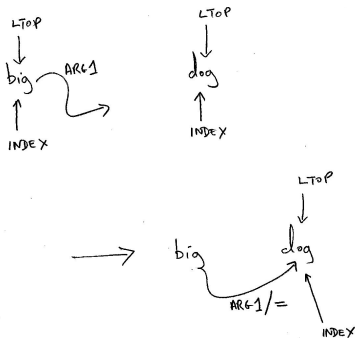
# Scopal modification: DMRS



Very similar to MRS:

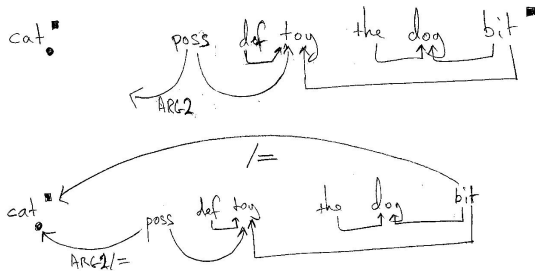
1. LTOP of result is the modifier's LTOP
2. INDEX is the head's INDEX
3. slot (i.e., open link) on modifier has head's LTOP as target

# Intersective modification: DMRS



1. LTOP of result is the head's LTOP
2. INDEX is the head's INDEX
3. the open link's target is the head's index
4. an eq link exists from the modifier LTOP to the head LTOP: usually this is the same link as the open link

## cat whose toy the dog bit



- if the modifier's LTOP is not the source of the open link, an additional EQ link is needed to satisfy condition 4.
- link label is /= in current DMRS, but possible alternative is MOD/=, where MOD is general (and semantically vacuous)

## DMRS composition principles

- General principles can be formalised in terms of graphs.
- Syntactic head always provides INDEX of result (except for null semantic heads), but LTOP varies with rule class.
- Open links (i.e., slots in algebra terms) typed to target either LTOP or INDEX (but these are often the same node).
- There are  $\neq$  links to all NP targets, h (i.e., qeq) links to all scopal targets (plus some heq links in ERS), and = links to everything else.
- Principle of intersective modification allows for (directed) EQ link in 'whose toy' type examples.
- Null semantic items have no INDEX: auxiliaries have to specify tense etc as a property of the main verb.

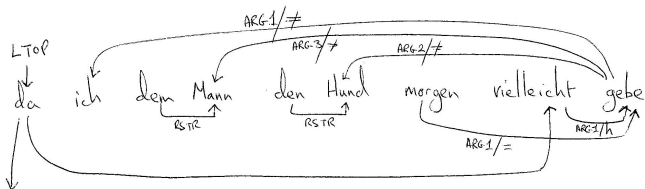
# Left-to-right semantic construction and human memory

- Composition learnability principles are about acquisition of language.
- But processing and semantic construction has to be incremental (though how strictly left-to-right disputed).
- Limits on short term memory.
- Rabagliati (2016) on experiments with adjective processing.

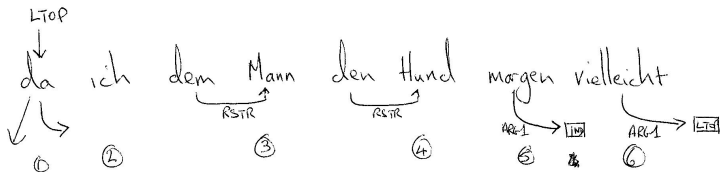
## Left-to-right semantic construction

da ich dem Mann den Hund morgen vielleicht gebe,  
as I the man the dog tomorrow perhaps give,

as I'll perhaps give the dog to the man tomorrow, I'm coming

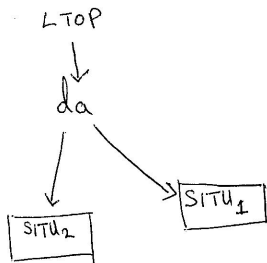


# Left-to-right construction



- At *vielleicht*, 6 distinct components ...
- Not sure of upper limit in German examples, but 6 already stretches plausibility.
- But this is a perfectly normal sentence, which speakers report no difficulty in processing.

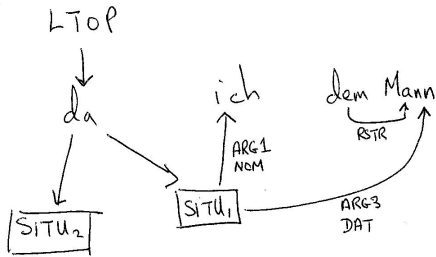
# Left-to-right construction



- Intuition: *da* gives speakers a notion of two situations, with first active.
- Subsequent information attached to the active situation.

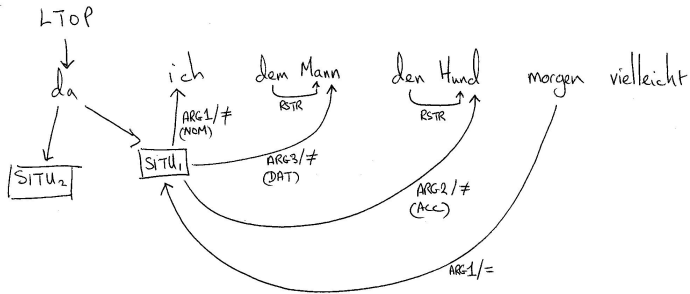


# Left-to-right construction

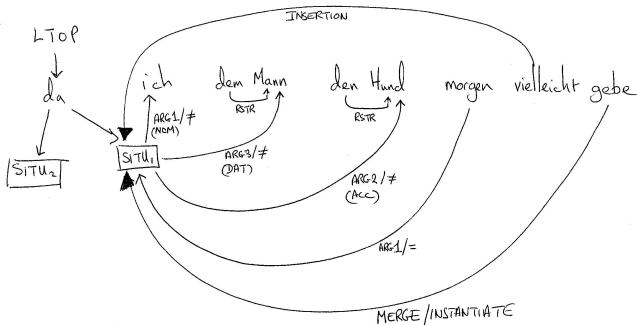


- Some sort of generic role labels may well be required: here NOM, ACC, DAT etc

# Left-to-right construction



# Left-to-right construction



- If we're constructing something like \*MRS, *vielleicht* has to be inserted and *gebe* instantiates the underspecified situation.

# Conclusions

- Normal DMRS composition works so far . . .  
see dmrscomp grammar
- practical advantages: substantial reduction in size of TFS,  
semantic packing (yet to implement)
- theoretical issues:
  - ▶ direct DMRS scoping, translation to models (for simple  
extensional fragment)
  - ▶ DMRS can be converted into no e LF and moderate e LF  
as well as full e LF (no implementation yet)
  - ▶ there's useful 'wiggle room' in working with DMRS itself
    - ▶ adjectives: *stone frog, probable winner* etc
    - ▶ conjoined scopal verbs etc? *we could and should stop*