Vector Semantics: Lecture 7

András Kornai
SZTAKI Computer Science Research Institute

May 8 2024

The 'bound' schema

- Schema names are arbitrary (schemas are not linguistics signs)
- `bound_` has two participants (slots): `volume_` and `boundary_`
- These will play a role in definitions of things that are not in `4lang` but perhaps should be, such as *between*
- The reductivity claim amounts to the statement that given a sentence that uses the word, we can create a translation to `4lang` that uses only its defining vocabulary
- This is empirically testable! Take a random sample of English, grep for the word in question, and see how far you can go
- We divide the test in two parts: finding the words (lexical lookup), and putting them together (syntax)
- Lexical lookup may involve morphological analysis of word forms, and putting them together (syntax) – this is nothing to be ashamed of

LDOCE BETWEEN

- in or through the space that separates two things, people, or places
- in the time that separates two times or events
- within a range of amounts, numbers, distances etc
- which two places are joined or connected by something
- which people or things are involved in something together or are connected
- which people or things get, have, or are involved in something that is shared
- which two things or people you are comparing
- idioms: $\sim$ *you and me*, $\sim$ *the devil and the deep blue sea, . . .*

Reductivity homework task

- Actually take an English corpus and select 100 sentences
- Check if the use of *between* in these fits one or more of the above definitions
- If not, keep the example and let's discuss it in class
- This measures the coverage of LDOCE, which is not 100%, but quite decent (well above 90%)
- `=agt {bound} =pat, "by _" mark_ =pat`
- How do you express the fact that `=pat` is typically a coordinate phrase *A is between B and C*?
- Is this necessary? How about *Whatever took place between the sheets/the two of us was nobody's damn business but ours*
- How much of this carries over to idioms?
- Bonus task: look up *inbetweenie* on google n-grams, create analysis

Boolean picture not great for language

- BAs are a great theory! But we care about the prokaryotes . . .
- In standard logical semantics (model theory) positive and negative statements are perfectly symmetrical (Russell was very emphatic about this), but in any frequency-sensitive system they are not: it is much easier to say false things than true ones
- In BAs conjunction is a (commutative) binary operation, and negation is an (involutionary) unary operation, the NL story is different
- Double negation is very different, as is interaction with quantifiers
- The 'marked' status of negation in NL has long been noted

The main trick: NL negation is dyadic

- =agt `lack` =pat 'agent doesn't have patient, though normally it does' – negation of default
- blind 'lack sight' so *blind person* `person IsA blind` is fine but *#blind stone* is infelicitous (weird)
- `lack` appears in 58 definitions e.g. *water* `liquid, lack color, lack taste, lack smell, life need`
- *clean* `lack dirt` but not *dirt* 'lack clean', why? Because (i) things in general are somewhat dirty (by default); (ii) *dirt(y)* definable by mud, dust, etc, without reference to cleanliness; (iii) *tisztátlan, unclean, ?kosztalan, *undirty, *dirtless*
- We must have `gen` (generic quantifier) in the system (independently motivated)
- *no* `gen lack` (unary negation obtained from binary by quantifying over agent)
- Quantifiers (standard problem ever since Aristotle) read Resources/dyaneg.pdf

Force dynamics

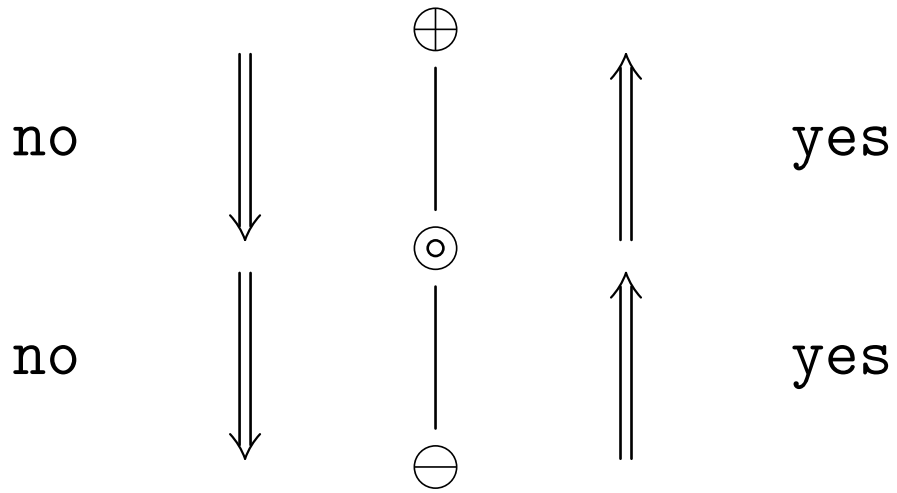

Figure: Forces in negation and affirmation

- `no yes yes no no`
- *Move up!* How to defy this command?
- Same as obeying *Don't move up!*
- There is more than one way
- Yes! We have no bananas
- *No No bananas