Underspecified quantification

A. Herbelot

Computer Laboratory
University of Cambridge

DELPH-IN, 2010

Outline

udef_q

- Cats sleep.

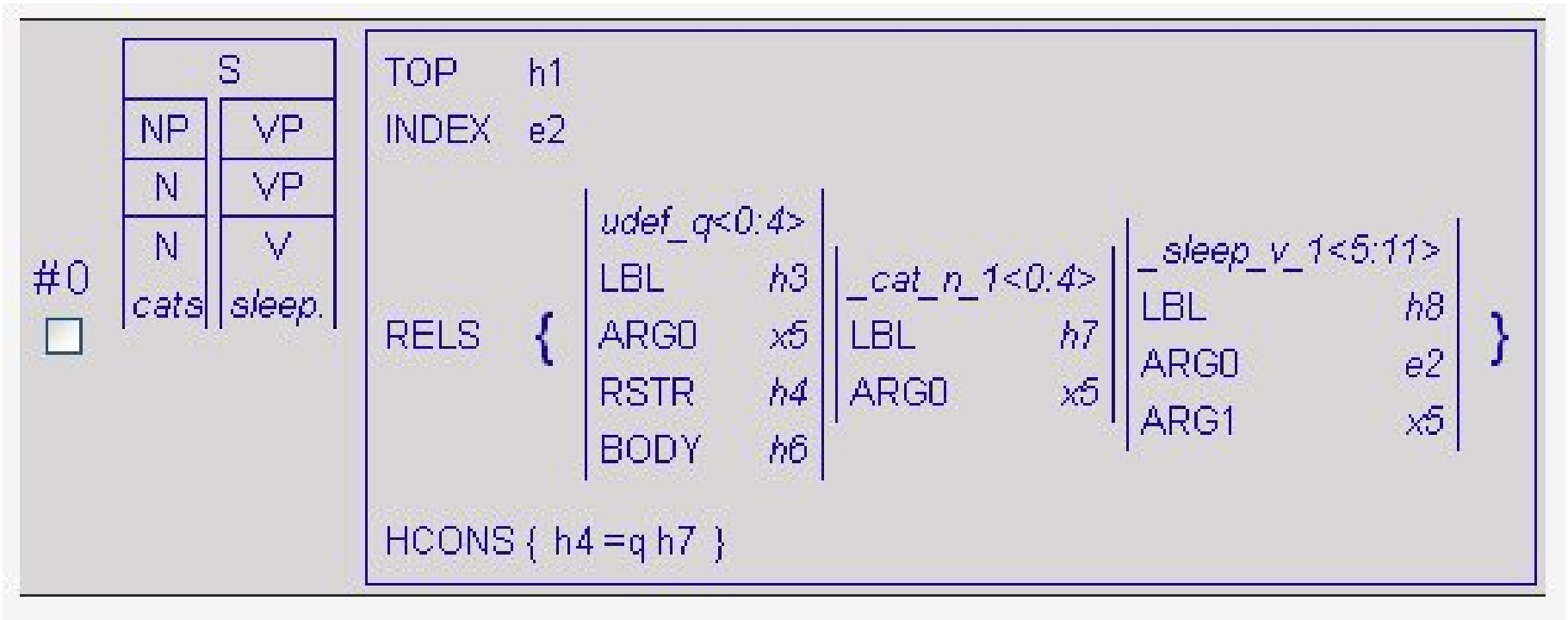

Figure: An example of udef_q

"We have underspecified quantifiers for bare plurals in the ERG, without ever having worked through what that would mean at all - but it's a lot prettier for the composition if that assumption can be maintained. I don't think the proper semanticists are enthused."

Ann (16.03.2009)

Wishes

- Answering the question ‘What does *udef_q* really mean?’
- Enthusing the formal semanticists.

Outline

Quantification resolution

- The task: translation of ambiguously quantified NPs into unambiguous ones:
  - Cats are mammals = All cats are mammals
  - Cats were sleeping by the fire = Some cats were sleeping by the fire
  - Water was dripping through the ceiling = Some water was dripping throught the ceiling
  - The beans spilt out of the bag = Most/all beans spilt out of the bag
- Why? For information extraction, inference, entailment, etc.

Quantification resolution

- More formally, quantification resolution is the process of annotating an ambiguously quantified noun phrase with a fully specified quantifier. A fully specified quantifier is a quantifier for which we have a quantification semantics with a unique, unambiguous set relation.
- *Cats sleep*
  - Some cats sleep: $0 < |\phi \cap \psi|$
  - All cats are known to sleep: $|\phi \cap \psi| = |\phi|$

  where ϕ is the set of all cats and ψ the set of all things sleeping.

Quantifiers

- We assume a continuous quantificational space divided into three basic quantifiers: *some*, *most* and *all* (also *one*).
  - *some*(ϕ, ψ) is true iff $0 <| \phi \cap \psi |<| \phi - \psi |$
  - *most*(ϕ, ψ) is true iff $| \phi - \psi |\leq| \phi \cap \psi |<| \phi |$
  - *all*(ϕ, ψ) is true iff $| \phi \cap \psi |=| \phi |$

Reference

- Following Link (1983), we take a lattice view of plurals where any point of the lattice under the supremum refers to a proper subset of the supremum.

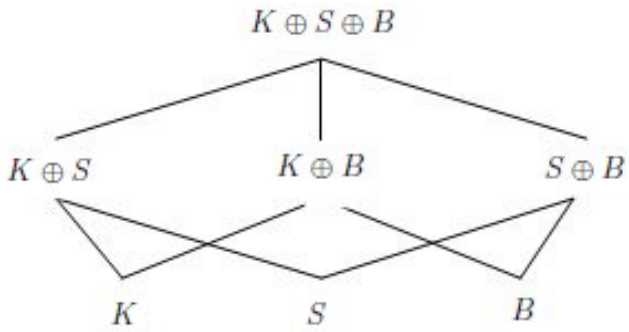

Figure: The join-semilattice of all cats (Kitty, Sylvester and Bagpuss) in world *W*

Scoping the study

- We assume three forms give rise to quantification ambiguity in NPs: the definite form, the indefinite singular *a* and the bare form.
  - A cat can sleep rolled up, with its head on its hind legs. (Most cats – those without arthritis).
  - The Galapagos turtle lives over 150 years. (Some lucky Galapagos turtles).
  - Dodos are extinct. (The kind *dodo*?).
  - At the end of the lecture, the/her/his students asked questions about the dodo. (Some/Some of her/Some of his students).
  - Water was dripping through the ceiling. (Some water).
  - Furniture has a practical purpose. (Most furniture – except contemporary art tables and chairs).

Outline

Genericity: some definitions

- Genericity is ambiguous when it comes to quantification!
- Krifka et al (1995): two phenomena
  - Habituality: *John smokes after dinner.*
  - Reference to kind: *The potato was first cultivated in South America.*

What is a kind?

- Species? Specimen?
- Well-established kinds
- Carlson (1977): basic ontology of kinds, objects and stages
- NOT definite plurals

The GEN operator

- Generalised quantifier form

  GEN $x_1 \ldots x_n; y_1 \ldots y_n[\mathit{Restrictor}(x \ldots x_n); \mathit{Matrix}(x_1 \ldots x_n, y_1 \ldots y_n)]$
- No explicit, pronounced form in any known language (Dayal, 1999)
- No unique quantifier corresponds to GEN. Do generics quantify at all?
  - Yes. (At least in some cases) inference is possible at instance-level. No inference possible with $\phi(\psi)$.

Outline

Underspecified quantification

- Replace ambiguous quantification with **underspecified quantification**.
- No silent *GEN* quantifier but an empty slot for the appropriate quantifier.
- We can always paraphrase ‘X does Y’ as ‘There is a set of things X, *a certain number of which* do Y’ (note the partitive construction).

Some bare plurals

- Dogs are in my garden = some dogs are in my garden.
- Frenchmen eat horsemeat = some/relatively-many Frenchmen eat horsemeat. (For the *relatively many* reading, see Cohen, 2001.)
- Cars have four wheels = most cars have four wheels.
- Typhoons arise in this part of the Pacific = some typhoons arise in this part of the Pacific OR most/all typhoons arise in this part of the Pacific.

Other constructs

- Bare plurals are quantifiable. How about other constructs?
- (In)definite singulars: *The cat is a mammal* / *A gentleman opens doors for ladies*. [(Nearly) trivial]
- Bare singulars: *Furniture has a practical purpose*.
- Definite singulars: *The reporters asked questions...*

Are bare singulars singular?

- Following Chierchia (1998) we treat bare singulars as regular bare plurals:
  - Water was dripping through the ceiling. (Some water).
  - Furniture has a practical purpose. (Most furniture).
  - Water consists of H_2O. (All water).

Definite plurals included

- Traditionally, definite plurals have been considered as universals (Lyons, 1999).
- But: *The reporters asked questions after the press conference* (Dowty, 1987).
- The best paraphrase is: 'There is a large set of reporters – all those present at the press conference – and some of them asked a question'.
- So definite plurals can also be said to be underspecified:
  - Your employees are dedicated. (True, even if one out of 50 likes a lie-in.)
  - Those apples have turned bad. (True, even if 10% are still okay.)
  - The rice spilt out of the bag. (True, even if three grains are still in the bag.)

The partitive construct

- We can always paraphrase ‘X does Y’ as ‘There is a set of things X, *a certain number of which* do Y’.
- Brogaard (2007) gives an account of definite plurals as partitive constructions. We follow this reading and expand it to all underquantified constructs.

Outline

Formalising the partitive construction

- Distributive predicate:
  The reporters asked questions at the press conference.

  $X = \sigma^* x\ \mathit{ReporterAtPressConference}'(x) \wedge$
  $\exists Y[Y \prod X \wedge \forall z[z \cdot \prod Y \rightarrow \mathit{askQuestions}'(z)]]$

- Collective predicate:
  Americans elect a new president every five years.

  $X = \sigma^* x\ \mathit{American}'(x) \wedge \exists Y[Y \prod X \wedge \mathit{electPresident}(Y)]$

Adding the quantifier

- *The reporters asked questions at the press conference.*

$$X = \sigma^* x\ \mathit{ReporterAtPressConference}'(x) \wedge \exists Y[Y \prod X \wedge \forall z[z \cdot \prod Y \rightarrow \mathit{askQuestions}'(z)] \wedge 0 < |Y| < |X - Y|]$$

- *Americans elect a new president every five years.*

$$X = \sigma^* x\ \mathit{American}'(x) \wedge \exists Y[Y \prod X \wedge \mathit{electPresident}'(Y) \wedge |X - Y| = |X|]$$

The underspecified quantifier

- $X = \sigma^* x\ P'(x) \wedge \exists Y[Y \prod X \wedge Q(Y)] \wedge \mathit{quantConstraint}(X, Y)]$

  where the quantConstraint ensures the correct cardinality of Y for various quantifiers and the predicate Q applies distributively or collectively depending on the semantics of the sentence. X denotes the N-bar referent while Y denotes the NP referent.

Formalising kinds

- A kind reading for *The dodo is extinct*: $\phi(\psi)$?
- But: true kinds can be expressed as bare plurals: *Dodos are extinct* (14,700 Google matches)
- But: how to cater for anaphora?
  The dodo is extinct but Mary says she's seen one.
- A kind is the supremum of all instances with property
  $Kind(X) = \sigma^* x\ X'(x)$ (Chierchia, 1998)
- $X = \sigma^* x\ dodo'(x) \wedge \exists Y[Y \prod X \wedge extinct'(Y)$
  $\wedge(|Y - X| = 0) \wedge \exists Z[Z \prod Y \wedge see'(Mary, Z)] \wedge (|Z| = 1)]$
- Tick: Enthusing formal semanticists.

udef_q explained

- Tick: Defining udef_q.
  $X = \sigma^* x\ P'(x) \wedge \exists Y[Y \prod X \wedge Q(Y)] \wedge \mathit{quantConstraint}(X, Y)]$

In practical terms...

- Annotation performed on 600 random subject NPs from Wikipedia. Inter-annotator agreement on 300 (Kappa=0.72).
- Two classifiers:
  - Tree-based classifier with simple syntactic features (78% precision)
  - Similarity-based classifier (can increase precision but recall is poor)

Further work

- Annotating the Wikipedia data in the Redwoods treebank (?)
- Add deeper features to the tree-based classifier.
- Re-run everything with more data...

Questions

- Where is the udef_q in *The cat sleeps*?

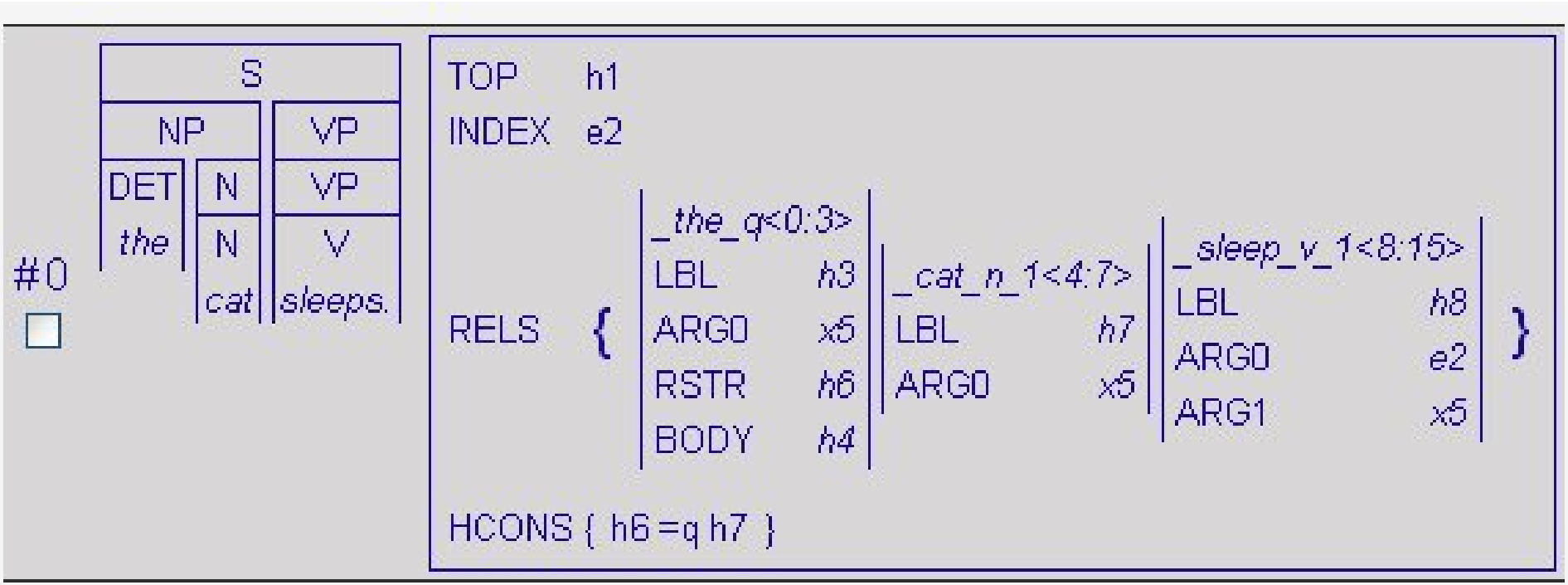

Figure: Something missing?