Bridging inferences in a dynamic, probabilistic frame theory

Ralf Naumann & Wiebke Petersen

University of Düsseldorf

Data and central issue. It is by now a well-known fact that that the semantic processing of an utterance usually involves different sources of information which are used in parallel to arrive at a coherent interpretation of this utterance in the given context. Three principle sources must be distinguished: (i) the (linguistic) meaning of the lexical items; (ii) (non-linguistic) world and situational knowledge and (iii) the prior linguistic context. A prime example of this interplay between different sources of information are bridging inferences. [AL98, 83p.] take bridging to be an inference that two objects or events that are introduced in a text are related in a particular way that isn't explicitly stated, and yet the relation is an essential part of the content of the text in the sense that without this information, the lack of connection between the sentences would make the text incoherent. Examples of bridging inferences are given in (1).

(1) a. Lizzy met a dog yesterday. The dog was very friendly. [AL98, 86p.]
  b. John unpacked the picnic. The beer was warm. [CH77]
  c. A car hit a truck. The windshield shattered. [DK09]
  d. I was at a wedding last week. [Geu11]
   (i) The bride was pregnant.
   (ii) The mock turtle soup was a dream.
  e. I've just arrived. The camel is outside and needs water. [AL98, 86p.]

Common to all bridging inferences is (i) a new discourse referent (dref) is introduced (see [Bur06] for neurophysiological evidence) and (ii) a dependency relation between this dref (corresponding to the bridged DP) and a dref that has already been introduced in the linguistic context (denoting an antecedent object). There are various forms of dependency relations, the most two prominent of which are: (a) identity (1-a) and (b) part-of (e.g. (1-b),(1-c)). Example (1-d), shows that no unique antecedent (car/truck) need be singled out so that one gets an ambiguous bridging inference. Finally, (1-d) and (1-e) show that the dependency can be indirect. E.g., the turtle soup is indirectly related as a part (starter) of the meal which was served at the wedding.

How the various sources of information are related to the factors playing a role in bridging inferences depends on the theory. For example, [AL98] analysis builds on Chierchia's [Chi95] analysis of definite descriptions as anaphoric expressions: 'The N' denotes an N which is related by some dependency relation to an antecedent y. On this account, lexical semantics provides an underspecified relation B which functions as the bridge or dependency relation to an object in the present discourse context. B is computed using world and/or situational knowledge. E.g., [AL98, 89p.] argue that in (1-e) one uses lexical semantics to infer that 'arrive', being a motion verb, defines a mode of transport. Next, one uses world knowledge to infer that camels can be used as such a mode of transport. When taken together, one gets a coherent interpretation of (1-e) because the two sentences are connected by the coherence relation *Result* and the bridging relation $B = \textit{Means-of-transport}$ with $B(x, y), x = \textit{camel}, y = \textit{arrival}$. Hence, one first assumes a form of lexical enrichment, providing B, and then uses both lexical semantic and world knowledge to compute B.

Informal outline of the analysis. We use a dynamic, probabilistic frame theory, which admits of a unified analysis. First, using frames the underspecified relation B is replaced by a (complex) attribute in a frame so that bridging inferences are not restricted to nominal DPs and can involve dependency relations of arbitrary depth. Second, world and pragmatic knowledge, lexical semantic information and discourse information can be combined seamlessly. E.g., semantic information at the lexical level can be combined with other sources of information by extending frames in the lexicon with additional attribute-value pairs representing such non-lexical information. Finally, it becomes possible to apply probabilistic reasoning at the level of frame properties which plays an essential role in determining a unique antecedent object. Frames are typed features structures in the sense that the value of each attribute must be of a particular sort (type). In frame theory the semantic information provided by lexical items like common nouns and verbs is not restricted to sortal information, e.g. it is of sort 'picnic', expressed by 'picnic(x)'. Rather it in addition includes information about attributes and possible values of those attributes in form of sortal restrictions on their target sorts. E.g., for 'picnic', attributes include FOOD and BEVERAGE. Processing a CN or a verb therefore activates a relational structure with sortal (type) restrictions but (mostly) unspecified values. One way of how such values can be specified is by bridged DPs. For example, in (1-b) the beer is a material part of the value of the attribute BEVERAGE of the frame associated with the picnic that was already introduced into the discourse. Hence, one has the path *picnic*? • BEVERAGE $=$ *beer*. In frame theory attributes can be chained so that the dependency relation can be more indirect. E.g., in (1-d) the mock turtle soup is the value of the STARTER-attribute which in turn is an attribute of the value of the MEAL-attribute of the wedding object already introduced into discourse, yielding the following path: *wedding?* • MEAL • STARTER $=$ *soup*. A limiting case is identity where the sort of the bridged DP matches that of the antecedent DP. Hence, in frame theory dependency relations used in bridging inferences can be of arbitrary length starting from 0, identity, to length 1, direct dependency relation, to length > 1, indirect dependency relation. So one gets (*):

(*) Bridging inferences are done at the level of frames. The bridged DP provides a value for some attribute in the frame associated with an object that was already introduced in discourse. Therefore, the sort of the frame associated with the object denoted by the bridged DP must be an admissible sort for an attribute chain of the antecedent object. The relation between this attribute chain and the sort is called the *dependency relation.*

Using (*) only requires lexical information together with access to the information provided by antecedent objects. These sources do in general not yield a unique antecedent object as shown by example (1-c). Let the set of admissible antecedents computed by applying (*) be S. This uncertainty for a comprehender can be (partly) resolved by applying world and situational knowledge involving probabilities. For example, in the case of (1-c) the probability that the windshield of a car shatters in an accident will in general be higher than the corresponding probability for the windshield of a truck. Hence, a decision rule applies according to which one chooses that admissible antecedent, i.e. element from S, for which the probability is highest. Hence, a choice between different accessible antecedent objects can often only be made *after* the bridged DP has been parsed and therefore information which is not directly related to the dependency relation can play a role in singling out a unique antecedent. Furthermore, this example shows that bridging inferences are non-monotonic: bottom-up information encountered after the bridged DP can influence the choice of the antecedent object.

Rough outline of the formalization. We define a probabilistic dynamic update semantics with frames. Therefore, we assume a finite set of possible worlds W and a domain D. D is structured by a (material) part relation $\sqsubseteq$. The extended domain D^* is $D \cup \{\perp\}$ with the additional minimal element $\perp \sqsubseteq d$ for all $d \in D$. The elements in D are assigned sorts of the partially ordered sort hierarchy $\langle \Sigma, \sqsubseteq_\Sigma \rangle$ with basic sorts like 'event' or 'individual'. $F = \{f_{d,\sigma,w}\}_{d \in D, \sigma \in \Sigma, w \in W}$ is the domain of frames. Each frame is of a sort σ and is a (generated) submodel of a possible world w, namely, the information associated with a particular object d in that world which is the root of the frame. F_w is the set of frames in world w. A frame corresponds to a set of relations on $D \times D$. Each relation R corresponds to a finite path (chain of attributes) ≥ 0 starting at the root d. The domain of R is given by the source-sort of the first attribute in the path and the range of R by the target-sort of the last attribute in the path.

The probabilistic component. Following [Gär88], we define two probability measures Pr_w and B, one for the probability of properties and the other for the probability of worlds. In order to define (*) it must be possible to speak for a world w about the probability that among the frames of sort σ those having an attribute (chain) with value of sort σ' is p with $0 \leq p \leq 1$. This is captured by a probability distribution Pr_w. If $S \subseteq D$, then $Pr_w(S)$ represents the probability that an individual belongs to the set S in w. Pr_w reflects the fact that for a comprehender two states of the world can differ in their frequency of a property among the objects. For example, a comprehender may not know the exact probabilities of a windshield shattering in a car accident but he knows that it is between 0.5 and 0.7. This uncertainty is reflected in different elements $w \in W$, representing different possible states of the world for the comprehender, having different proportions of shattered windshields in car accidents. Pr_w is local in the sense that it gives the probability of a property in a particular world w. However, in order to determine a unique antecedent object it is necessary to look at *all* worlds which are epistemically possible for the comprehender and to determine the probability of a property in the whole epistemic state. To this end, we define a probability distribution B over all subsets of W. If V is a subset of W, $B(V)$ is a measure of the probability that the actual world is among those in V. For example, B is used to prefer worlds in which during an accident of a car with a truck the windshield of the former and not the windshield of the latter shattered. Using B and Pr_w, it becomes possible to define the probability that an object in the epistemic state of a comprehender has a certain property w.r.t. to a subset V of W (see [Gär88]):

(2) $Pr_V(Q) = \sum_{w \in W} \frac{Pr_w(Q) \cdot B(\{w\})}{B(V)}$, provided that $B(V) \neq 0$.

Information states in a frame theory. An information state s must model both (i) the local discourse information component as well as (ii) the global world and situational knowledge component. (i) Local discourse level: Information in frames is of two kinds: sortal (non-relational) one and relational information about paths of length ≥ 1. Sortal information by itself does not relate two objects. By contrast information given by paths of length ≥ 1 always relate two or more objects: $R(d_1, d_2)$ where d_2 is the value of an attribute in the frame of d_1. Hence, d_1 and d_2 play different roles. This double perspective is captured by defining the discourse component as a stack c (following Incremental dynamics, [vE01]). Each position on the stack is assigned a pair $\langle d, F_{w,d} \rangle$ of an object $d \in D^*$ and a set of frames $F_{w,d}$, which is the set of frames in w with root d. (ii) The global component is given by a set of worlds W_s and the two probability distributions Pr_w and B. An *information state* s is a triple $\langle \alpha, Pr_w, B \rangle$ s.t. α is a set of possibilities i which are pairs $\langle w, c \rangle$ consisting of a world $w \in W_s$ and a stack c.

Update semantics. We provide simplified update operations for 'a', 'the' and atomic formulas. Update operations directly change the value of the stack component and indirectly the probability component consisting of Pr_w and B. The general idea is to first leave the object component of a stack element relatively unconstrained by imposing only sortal information and to then use the frame component to determine the sort of possible objects for the corresponding stack element.

Update related to domain extension: The update operations for 'a' and 'the' are defined in terms of two atomic update operations $[c^1]$ and $[c^2]$ that extend the domain. They both increment the stack by one element s.t. the new element is $\perp$, i.e. the bottom element of D^*. This captures the intuition that incrementing the stack and getting information about this object are two distinct operations: the first introduces a new topic whereas the second provides factual information about this object. In contrast to other update semantics, 'branching' is introduced at the level of the frame component associated with this object. For $[c^1]$ (corresponding to the indefinite determiner 'a'), this is the set $F_{w,d}$ of all frames in w for the frame component (no factual information is known), thus the new incremented stack is $c\hat{\ }\langle \perp, F_{w,d} \rangle$. By contrast, in the case of a bridged DP (operation $[c^2]$) the frame component is restricted by the set of frames that have already been introduced in relation to a previous stack element. The new incremented stack is $c\hat{\ }\langle \perp, F_{w,d,c} \rangle$ where $F_{w,d,c}$ is the set of all frames which are the value of a chain of attributes in a frame $F_{w,d}$. Both update operations do not change the two probability distribution Pr_w and B. This reflects the fact that an update only introduces a new topic and (possibly) imposes a constraint on its frame components. $[c^2]$ can be strengthened by imposing further restrictions on the dependent frame component, e.g. by requiring that the last attribute in R be an instance of the part-of or of the identity relation.

Update related to atomic information. The effect of atomic information is threefold. First, possibilities which do not satisfy this information are eliminated. Hence, atomic information must be satisfied in the world of a possibility. Second, for surviving possibilities the frame components of the arguments are updated. Locally this is again eliminative: only those frames in a component survive which satisfy this information. Finally, the probability distributions are updated.

Singling out a unique antecedent. Update by atomic bottom-up information can yield a unique frame so that a unique antecedent object is determined. However, by itself, the compositional process does in general only determine a set of frames, each of a particular sort which are possible antecedents (the set S from above). In order to determine a unique antecedent the comprehender uses the global probability distribution Pr_V from (2). She calculates the global probability $Pr_{W_s}(Q_\sigma)$ for properties Q_σ s.t. f is of sort σ and an element of $\pi^2(\pi^2(c[i]))$ for $c[i]$ corresponding to the interpretation of a bridged DP and s.t. W_s is the set of worlds underlying her current epistemic state. So world knowledge is used to single out a unique antecedent because this operation is a global one since it depends on the information as a whole and not on a single possibility. In fact, other factors will come into play as well, like accessibility e.g., which have not been considered in this abstract.

[AL98] Nicholas Asher and Alex Lascarides. Bridging. *Journal of Semantics*, 15:83–113, 1998.

[Bur06] Petra Burkhardt. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98(2):159 – 168, 2006.

[CH77] H. H. Clark and S. E. Haviland. Comprehension and the Given-New Contract. In R. O. Freedle, editor, *Discourse Production and Comprehension*, pages 1–40. Ablex Publishing, Hillsdale, N. J., 1977.

[Chi95] Gennaro Chierchia. *Dynamics of Meaning: Anaphora, Presupposition and the Theory of Grammar*. University of Chicago Press, Chicago, 1995.

[DK09] Alex Djalali and Stefan Kaufmann. Probabilistic inference in dynamic semantics. In *Proceedings of the 10th Symposium on Logic and Language*, pages 99–107. Research Institute for Linguistics, Hungarian Academy of Sciences (HAS), and Theoretical Linguistics Program, Eötvös Lorand University (ELTE), 2009.

[Gär88] Peter Gärdenfors. *Knowledge in Flux*. MIT Press, Cambridge, MA, 1988.

[Geu11] Bart Geurts. Accessibility and anaphora. In Klaus von Heusinger, Claudia Maienborn, and Paul Portner, editors, *Semantics*, volume 2 of *Handbooks of Linguistics and Communication Science*, chapter 75, pages 1988–2011. DeGruyter, 2011.

[vE01] Jan van Eijck. Incremental dynamics. *Journal of Logic, Language and Information*, 10(3):319–351, 2001.