Towards a Formal Characterization of Sentence Topics

Abstract for ESSLLI'98, Logic & Linguistics section
February 1998

In this paper I propose a model-theoretic approach intended to formally capture the notion of Sentence Topic (henceforth simply "Topic"). Such a formal characterization is still lacking in the linguistic literature to date (cf. the complaints to this effect by Chafe [5] [1]), and may be considered a necessary preliminary to further research on the behaviour of Topic as a Grammatical Relation.

The present proposal is based on the application of topology to predication, as presented by [17, 21, 18, 1], a framework congruent with a generalized form of situation semantics [2, p.75-8] (adopted as a framework for wider research of which this forms a part). The property of "topicalizability" for an NP [2] is given a formal characterization based on the notion of logicality in Generalized Quantifier Theory (GQT) [4, 22, 11, 16], namely as the class of non-logical NPs or those which fix the individuals in their denotation. This notion of logicality is then related topologically to the ability of the NP denotation to fix a model for the interpretation of the sentence. Actual topichood is distinguished from intrinsic topicalizability by the NP actually fixing such a model, whereas other topicalizable NPs (like all non-topicalizable ones) may be absorbed into the psoa. The structure resulting from this analysis is then used to characterize the dynamic process by which non-topics in one sentence may become topics in the next [9] in terms of the development of a computation.

Most (intensional) model-theoretic approaches assume at least the following apparatus: on the one hand a set W of worlds, times or situations, on which a structure such as an accessibility relation is defined and against which a proposition is evaluated; and on the other hand a domain of discourse D to which expressions in the object language are assigned by the interpretation function I. A dichotomy thus appears to exist between the objects inhabiting W and D. However there are good reasons for supposing that objects in W, at least times and situations [3, 7], may equally occur as arguments of predicates, in such a way as to have an extension in D. The motivation for my proposal is the converse idea that the referents of NPs show similar behaviour in potentially inhabiting W as well as D, and that this ambivalence can be used to capture the distinction between NPs as arguments and NPs as topics [3].

I assume a propositional geometric logic P (or "logic of finite affirmations" [21]), partially ordered by entailment ($p \preceq p'$ iff $p \Rightarrow p'$), and a family $\mathcal{E}$ of *partial* truth-assignment functions e from P to $\{0\ 1\}$. This induces a topological system (pt X, ΩX)[4] in which ΩX can be interpreted as an algebra of open sets or partial descriptions (or an algebra of infons [2]), while $\mathcal{E}$ forms the corresponding set of points, pt X. The latter is equivalent to a poset of models for the algebra, and may be thought of as points in a non-Hausdorff topological space. I also assume from GQT the standard semantics of NP denotations as sets of sets, making the explicit assumption that all sets referred to in their definitions are open. The proposed semantics for topic NPs will be driven by the interaction of the latter approach with the topological system described.

[1] I am referring to the provision of a principled denotational semantics, not the specification of algorithms for particular computational processes.

[2] In this paper generics are excluded from consideration

[3] cf. the work of Tsiplakou on clitic doubling [19, 20], in which Topic NPs, like Tense, are treated as database labels in a Labelled Deductive Systems framework.

[4] [21, Ch5]

In a standard intensional treatment, pt X might be interpreted as worlds, times, or in case of partial functions situations, these being generally treated as primitive rather than constructed (cf. [13]). In the approach being proposed here, the model itself and any internal structure it has are induced by relations between truth-value assignments to expressions in the object language, i.e. constructed out of the discourse itself[5]. Thus any single proposition p partitions pt X (the $\sqsubseteq$ relation in this instance being trivial) - so far as in standard approaches - and any subset of P induces a partially ordered model $\langle$pt X, $\sqsubseteq\rangle$.

The model supporting a proposition is constrained by linguistic information. This is obvious for temporal logic, where the presence of tense or temporal adverbials will normally constrain the temporal parametrizations of the model which render the proposition "true". However, it is also true of certain NPs; essentially, I claim, those which function as Topics. To take a well-known example, a sentence containing "The King of France" (and evaluated at a time later than 1848) argubly cannot be assigned "true" or "false" if this NP is Topic ("*the King of France* is bald"), i.e. there is no model supporting either assignment precisely because the NP lacks a referent, though the same will not necessarily hold for a corresponding non-Topic NP ("If he scores against England, they will make him *the King of France*"). Thus a Topic NP can be considered as describing a point in the model, i.e. a function from P to {0 1}.

As the above arrangements have made $\langle$E, P$\rangle$ into a "topological system", the sets E and P, hitherto assumed to be disjoint sets, can be related as parts of the single system (pt X, ΩX), respectively the "points" and "frame" of a single domain related systematically by functions (hence mutually determining or constraining). From the present point of view it is useful to think of the frame as determining the points and the points as being functions from the frame to {0 1}, i.e. as a *locale* [21, p.61]. This operation raises the points of the original domain to "abstract points" which provide a denotational semantics or abstract meaning for the open sets in the frame. I invoke it to provide a denotational semantics for topics, analogous to the way in which it is used to provide a denotational semantics for a programming language (ibid p.63)[6].

Turning to the GQT perspective, an NP denotation can be represented as a subset of $\mathcal{P}$(D). As shown by [4, 22, 12], such NP denotations are confined to a highly restricted class of such subsets, the conditions for which themselves reflect the "logical topicality" of language [11, p.56] in a broader sense. However, even these conditions do not capture the notion of Topic in the grammatical sense (for example, they allow non-specific indefinites or cardinal NPs). In fact, the crucial characteristic of such non-topicalizable NPs is arguably their independence of particular models for their interpretation, i.e. their *logicality*. Topicalizable NPs, by contrast, have a semantics in which the identity of elements in the relevant set intersections (henceforth "pivot sets") is important. It is suggested in [10] that topicalizable NPs are precisely the class of non-logical NPs[7], and furthermore that the idea of a situation plays an important role in fixing their meaning. It is argued there that a common factor in the variety of topic constructions found cross-linguistically is their ability to be represented as operations on resource situations. However this notion of a situation can itself be interpreted formally as reflecting the partiality of interpretation functions in a suitable intensional logic, such as that outlined above.

Given the denotation of an NP as a set of sets whose membership is determined by particular set-theoretic relationships with a restriction set, I first distinguish between those NPs which fix the identity of the individuals in their denotation and those which do not [11, 14]. Prototypical examples of the former class include NPs with definite or possessive Dets, as well as proper names, while prototypical examples of the latter are NPs with cardinal Dets; in this paper I will confine my attention to these types. [8] In Situation Semantics it is well known that the former class fix their denotation by invoking a "resource situation". I propose that in fact these NPs (or

[5] More precisely, the two domains are mutually constraining (see below). However it is convenient here to concentrate on the way in which the discourse fixes the model.

[6] cf. [21, 17], with "Topic" corresponding to their notion of "Subject".

[7] Thus topicalizable NPs are the complement of those which can appear postverbally in existential constructions. I assume Lappin's [14] analysis of the latter as cardinal (and hebce logical) NPs.

[8] In particular, as already noted, I do not consider generics.

more precisely the function fixing their denotation) *constitute* a resource situation, and that it is in terms of operations on these that the natural-language notion of Topic should be understood (hence supporting my hypothesis that Topic NPs inhabit the world of situations W as well as the domain of individuals D).

To show this, I show that NPs in the first class, but not those in the second class, constitute frame homomorphisms.[9] Take for example the proper name "John", in the sentence "John drank some retsina". The sentence as such denotes a situation type / open set / infon, which is mapped onto a truth value by identifying the referent of "John". Holding that identification constant, however, will also assign *true* to any infon informationally contained in this, such as "John drank some wine", as also to any arbitrary disjunction or any finite conjunction of statements each of which is supported by the same identification of referent. It thus fulfills the requirements for a frame homomorphism. Frame homomorphisms, however, are the elements of E (or pt X).

A definite NP such as "the men", or a possessive NP such as "Napoleon's men" will similarly fix the individuals in the relevant set of men, either anaphorically or via the reference already fixed for the proper name "Napoleon". The denotation of the NP as a whole will be the set of those sets which stand in the specified relation to the fixed individuals (the "pivot set").

A cardinal Det, by contrast, does not fix its reference in the same way.[10] Thus in the sentence "John had four wives", the denotation of "four wives" is not fixed relative to a particular set of individuals but relative to an equivalence class (of sets of the specified cardinality, subsets of the restriction set of wives). Although the NP contributes descriptively to the situation *type*, it cannot be used to anchor it in a situation and thus give it a truth value. Topologically it is easy to see why, especially thinking of the topological system spatially. Members of the set of sets denoted by the GQ may be completely disjoint, and cannot therefore serve as descriptions of any particular point, hence they do not provide any model.

Thus there is a formal distinction between "topicalizable" NPs which provide a model for the sentence and "non-topicalizable" NPs which do not. However, not all topicalizable NPs in a sentence need be actual Topics. In the following example (based on [5]), the NP "your wife" is definite, old information, with its reference fixed by a resource situation, but is not a Topic.

(1) I went to a party last night, and guess what, I saw *your wife* there.

The characterization of actual Topichood, as opposed to Topicalizability, therefore requires an account of the process by which situations are built up in the discourse.

I take the Topic of the second conjunct as being the party situation. (The speaker is also a Topic, and can perhaps be regarded as an extension of the party situation, but for the present I ignore this detail). In the first conjunct the Topic is "last night", and the party is introduced as (presentational) focus. The Topics in both conjuncts are examples of what Erteschik-Shir [9] calls "stage topics", realized by explicit or implicit space-time adverbials rather than an NP argument. (I prefer the traditional term "frame adverbials", a term which unrelated of course to the term "frame" as used in the topological theory invoked in this paper).

Erteschik-Shir (op.cit.) describes the changing informational status of "the party" in this kind of discourse (where it acts as presentational focus and then topic) by the metaphor of first adding a file-card to the top of the stack (focus) and then having it available as the most salient background information (topic). Although I do not make any formal use of the popular image of file-cards here, it is essentially this process that has to be captured. In terms of the account given so far, this would seem to require the following elements:

(2) 1. In the first conjunct, the open set representing the predicate should be representable as a relation, one term of which is a Generalized Quantifier. This Generalized

[9] More precisely the interpretation function for the NP constitutes a frame homomorphism.

[10] I refer here to "pure" cardinal Dets, as opposed to the use of cardinal Dets as what Enc calls "covert partitives" [8]; in the latter, whether or not Enc's exact formulation is correct, there is some anaphoric reference to a definite referent already established in the discourse.

Quantifier will not in general be anchored to the situation which forms the Topic of the sentence, nor necessarily to any situation.

2. For the second, the topic situation includes this Generalized Quantifier, and should also bear a specifiable relationship to the Topic situation of the first conjunct.

I assume the standard GQT account of the denotation of (extensional) Direct Object NPs given by Cooper [6][11]. (I use the subscript 2 to refer to Direct Objects). Again the NP denotation for the Direct Object, notated GQ_2, is the set of all sets Y satisfying a given relation with an N′ set N_2 or individual n_2. The relation R of the transitive verb denotes a set of ordered pairs $\subseteq A \times B$ where A and B are the sets of (descriptively speaking) potential "actor" and "goal" arguments respectively. The denotation V of a transitive verb is then the set of ordered pairs $<a,GQ_x>$ such that $B \in GQ_x$:

(3) $V = \{<a,\mathbf{GQ}_x>: \{b: <a,b> \in R\} \in \mathbf{GQ}_x\}$

The transitive VP then denotes the set X ($\subseteq$ A) of all x such that $<x,\mathbf{GQ}_2>$ is an element of V, and this then combines with the Subject NP denotation in the normal way.

Returning to the situation-theoretic perspective, the psoa or open set comprises a relation with one or more argument parameters (I ignore for the moment the location parameter) by which it can be anchored. It is clear from the semantics just given that an Object NP cannot by itself anchor the psoa, since its combination with the transitive verb denotation does not fix a unique pivot set in the sense required (its set denotation still contains the unanchored variable x). This is true whether the Object is definite or indefinite. The information contained in the Object NP is "absorbed" into the predicate, providing an additional constraint on the open set characterizing the final proposition. This set however will also partially fix the denotation of the Object NP, restricting it to one element of the set of sets comprising its GQ. Thus *after* the processing of the entire proposition, the denotation of the Object NP is available as a partially fixed point which may contribute a model for the next sentence.

In conclusion, it is argued that the topological system outlined is an appropriate formal model of topichood. Combined with assumptions taken from Generalized Quantifier Theory it not only respects the standard account of NP denotations but interacts with it to predict the difference between topicalizable and non-topicalizable NPs, and to suggest a mechanism by which a non-topic NP in one sentence can serve as a topic for the next. In other research in progress I consider the marked cases involving non-topic Subjects and/or topic Objects, and also apply the theory presented here to outstanding problems in the preservation of topical information through discourse.

References

[1] S. Abramsky. *Domain Theory and the Logic of Observable Properties.* PhD thesis, QMC, University of London, 1987.

[2] J. Barwise and J.Etchemendy. Information, infons and inference. In *R.Cooper K.Mukai and J.Perry eds. 1990*, pages 33–78.

[3] J. Barwise and J.Perry. *Situations and Attitudes.* MIT Press, Cambridge MA, 1983.

[4] J. Barwise and R.Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, pages 159–219, 1981.

[5] W. Chafe. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. Li, editor, *Subject and Topic*, chapter 2, pages 25–55. 1976.

[6] R. Cooper. *Quantification and Syntactic Theory.* Reidel, Dordrecht, 1983.

[11] cf. also [15, p.273-4] for a convenient summary.

[7] R. Cooper. The role of situations in generalized quantifiers. In S. Lappin, editor, *The Handbook of Contemporary Semantic Theory*, pages 65–86. Blackwell, Oxford, 1996.

[8] M. Enç. The semantics of specificity. *LI*, 22:1–26, 1991.

[9] N. Erteschik-Shir. *The Dynamics of Focus Structure*. forthcoming.

[10] H. Gregory. Topic noun phrases, specificity and resource situations. Term Paper, SOAS, 1996.

[11] E. Keenan. The semantics of determiners. In S. Lappin, editor, *The Handbook of Contemporary Semantic Theory*, pages 41–63. Blackwell, Oxford, 1996.

[12] E. Keenan and L.Faltz. *Boolean Semantics for Natural Language*. Reidel, Dordrecht, 1985.

[13] F. Landman. *Towards a Theory of Information: the status of partial objects in semantics*. Foris, Dordrecht, 1986.

[14] S. Lappin. The semantics of 'many' as a weak determiner. *Linguistics*, 26:977–98, 1988.

[15] S. Lappin. Donkey pronouns unbound. *Theoretical Linguistics*, 15(3):263–285, 1988-9.

[16] S. Lappin. Semantic types for natural language. Inaugural Lecture, SOAS, February 1997.

[17] W. Rounds. Domain theory and its applications in language, logic and computation. Course given at the European Summer School in Language, Logic and Computation, Aix-en-Provence, 1997.

[18] M. Smyth. Topology. In S.Abramsky D.Gabbay and T.Maibaum, editors, *Handbook of Logic in Computer Science*, volume 1 of *Oxford Science Publications*, pages 641–761. OUP, 1992.

[19] S. Tsiplakou. On the representation of focus in greek - an lds perspective. *SOAS Working Papers in Linguistics and Phonetics*, pages 233–250, 1995.

[20] S. Tsiplakou. *Focus in Greek: its structure and interpretation*. PhD thesis, SOAS, University of London, 1998.

[21] S. Vickers. *Topology via Logic*. CUP, 1989.

[22] D. Westerstahl. Quantifiers in formal and natural languages. In D. Gabbay and F.Guenthner, editors, *Handbook of Philosophical Logic*, pages 1–131. 1989.