

TOPIC, FOCUS AND THE LOGIC OF LANGUAGE

Jaroslav Peregrin¹

[appeared in *Sprachtheoretische Grundlagen für die Computerlinguistik* (Proceedings of the Göttingen Focus Workshop, 17. DGfS, IBM Deutschland, Heidelberg, 1995)]

1 Topic as the "point of departure" of an utterance

The terms *topic* and *focus* are used by many theoreticians, but they often mean different things. In the most usual informal sense, topic is what an utterance is about (as contrasted to comment), and focus is what is emphasized in the utterance (as contrasted to background). Let us consider how this intuition can be sharpened.

Let us start with the notion of *topic*, of "that which an utterance (or a sentence) is about". Consider the sentence *John loves Mary*. What is it about? Clearly we can say that it is (says something) about John. However, it is hard to deny that it is - also - about Mary. And we can equally well say that it is about John loving Mary, or about John's love for Mary. "What a sentence is about" is thus an extremely vague notion; and to use it unregimented to define the concept of topic is hardly meaningful.

However, we can consider another, narrower sense of "being about" - a prototypical utterance of English (or indeed any other language) can be seen as consisting in picking out an item and saying something about it; thus the sentence *John walks*, in a "neutral" context and with a "neutral" articulation, picks out the person John and states that this person walks. In this sense, the sentence is about John, and *not* about, say, John's walking. Hence, in this sense, aboutness has to do with the subject-predicate structure; a sentence is "normally" taken to be about its subject.

However, it is important to realize that the subject-predicate structuring which is relevant from this semantic - or "informatoric" - viewpoint does not necessarily coincide with the syntactic structuring - what functions as a subject from the semantic viewpoint need not be identical with the syntactic subject. When one says *John walks* (answering the question *What does John do?*), he is addressing John and claiming that he is walking; whereas when he says *JOHN walks* (answering, e.g., the question *Who walks?*), he is rather addressing walking and is claiming that this activity is carried out by John. We can imagine that the separation of the syntactic from the semantic subject-predicate structure may represent a certain step in the genesis of language; a step by which language has gained another degree of flexibility - but this is of course nothing more than speculation.

¹The paper has been essentially improved thanks to the helpful criticism of Petr Sgall and Barbara Partee.

However, if we want to claim meaningfully that semantic subject and syntactic subject need not be one and the same thing, we must be more exact about what a semantic subject is (syntactic subject seems to be defined clearly enough). We have said that semantic subject is that part of the utterance which the utterance is about; but to take this as the desired definition would be begging the question, for we have explicated the concept of aboutness (in the narrower sense) by means of the concept of the semantic subject. To prevent this circle from being vicious we must find a way of defining semantic subject independently of aboutness.

A distinctive feature of the semantic subject is displayed by the fact that the failure of an utterance, seen as a communicative act, has different consequences when concerning the subject than when concerning the predicate. The former case of failure means the failure to point out something common to the speaker and the hearer, some shared basis which would anchor the following core of the utterance within the framework of knowledge and awareness shared by the speaker and the hearer; and by consequence generally making the whole utterance not quite intelligible. Thus this kind of failure leads to the utterance being either unintelligible or at least infelicitous. On the other hand, a later failure concerning that part of the utterance which aims at claiming something about the subject results in the utterance being simply false. If I say (as in the well-known example used by Strawson) *The exhibition was visited by the king* in the context where there is no exhibition to be referred to, my audience will simply fail to grasp what I am talking about and is likely to see my utterance as inappropriate; whereas if there is an exhibition which is understood to be referred to and if there is no king to visit it, the utterance is simply false.

Thus what we have called *semantic subject* is actually a basis, an information-anchoring point of departure for an utterance; whereas that which we have called *semantic predicate* is the information-conveying core of the utterance. Indeed it is precisely these two concepts - the *point of departure* and the *core* of an utterance - which were established as the basic means of analyzing a sentence by Mathesius (see esp. 1939) and his colleagues of the Prague Linguistics Circle (whose precursors were in France and Germany). In the more recent literature (see esp. Sgall et al., 1986), it has been proposed to identify this pair of concepts with the concepts of *topic* and *focus*; we have now indicated which considerations can substantiate such an identification in the case of *topic*.

2 ... and focus as its "core"

Having presented some evidence that *topic*, as the term is usually used today, is comparable to what the fathers of the Prague Linguistic School called the *point of departure* of an utterance, we would now like to indicate that it is the Prague *core* of the utterance that can serve as a

plausible generalization of the notion of *focus*.

To be sure, focus is often conceived today as something occurring only within some utterances, and which becomes especially relevant in combination with focalizers (i.e. particles like *only*, *even*, *also* etc.). Focus is an emphasized part of the utterance commonly distinguished by way of intonation and stress; it is, as Krifka (ms.) put it, the "intonationally highlighted" part of the utterance. However, the notion stemming from the Prague tradition is different: focus, taken as the core of the utterance, is not some accidental surplus of certain utterances, but an essential constituent of every meaningful and "pointful" utterance; and intonational highlighting can be seen as a mere way of marking focus in cases where the topic-focus articulation cannot be read off the syntax (or sometimes possibly as a means also of marking a strengthened exhaustiveness or contrastiveness claim of the utterance).

The plausibility of such a view grows if we do not restrict our attention to English but consider also languages with relatively free word-order like Czech. Such languages allow for organizing the utterance according to the 'functional sentence perspective' - going from the topic, the point of departure, to the focus, the core of the utterance. The topic picks up an entity (in a very broad sense) familiar to the participants of the communicative act (thus presupposing its existence, or, more precisely, its referential availability) and thereby anchors the utterance within the informational pool shared by the communicants (thereby making the utterance intelligible). Focus then says something about the entity thus specified by the topic (usually without presupposing anything) and develops a new theme (extending the common pool).

It might seem that topic (what the utterance is about) is something to a large extent independent from focus (conceived as that which the utterance stresses). However, the view advocated here is based on the conviction that all utterances have a "logic" which prevents these two things from being totally independent. The point of an utterance is to get from something known and agreed to something new and informative, and it is the new information that is primarily stressed; so that the topic and the focus emerge as two aspects of a single articulation pattern. This is to say that any "intonational highlighting", characteristic of the focus, makes *eo ipso* that which is highlighted into that which is the "point" of the utterance, and hence into that which is the core of the utterance.

3 Three perspectives

The intuitions discussed so far allow for various forms of explication of the topic-focus articulation; these in turn lend themselves to various kinds of formal articulation. Let us present at least three "metaphors". The three viewpoints lead to three ways of formalization; but these are not intended to represent three different ways of topic-focus structuring - they are simply

three ways of presentation of a single ("triune") pattern.

(1) Topic is a subject (picking up a piece of information "as an object", thereby triggering an existential presupposition) and focus is a predicate (presenting some further specification of the object).

(2) Topic and focus are arguments of an implicit generalized quantifier, or they are - in terms of Partee (1991) - the restrictor and the nuclear scope of a "tripartite structure". In certain cases, the implicit topic-focus-binding quantifier can be overridden by an explicit focalizer, such as *always* or *only*, but also by negation (cf. Haji_ová, 1984; 1994).

(3) Topic and focus are two phases of an information-conveying act (and they can be pictured as two segments of a dynamically viewed proposition). Topic corresponds to the phase where the information gets anchored to the existing "informational structures", and focus to that where the genuine new information is being added. Therefore, the failure of the act during the topic-phase (i.e. the falsity of presupposition) means the failure of the whole act (which may precipitate a - possibly temporary - breakdown of communication), whereas that during the focus-phase (i.e. the falsity of assertion) engenders the failure to add new information.

4 General Questions of Formalization

The three perspectives outlined in the preceding section lead to three different ways in which we can develop a logical formalism to account for the topic and focus.

(1) The elaboration of the idea of topic and focus being semantic subject and semantic predicate, respectively, calls for a formalization allowing for the notion of presupposition, i.e. for one based on partial or three-valued logic. (To say that A is a presupposition of B is to say that if B is not true, A cannot be but truth-valueless, which makes a nontrivial sense only if we allow for a nontrivial truth-valuelessness.) The basic subject-predicate nexus has to be analyzed in such a way that subject, in contrast to predicate, triggers a kind of existential presupposition. The subject-predicate pattern of the formal language we use to analyze its natural counterpart thus loses its role as a reflection of the overt, syntactic structure and is intended instead to reflect the topic-focus articulation. This means that we analyze an utterance consisting of the topic T and focus F as a subject-predicate statement $F(T)$ which has a truth value only if any presupposition associated with T is true. The sentence *John walks* (with the "neutral" intonation) may be thus analyzed as $walk\{John\}$ (the curly braces indicating non-classical, "presuppositional" predication) with the presupposition $\exists x.x=John$; whereas the sentence *JOHN walks* as $\lambda p.p(John)\{walk\}$ presupposing something like $\exists x.walk(x)$. Moreover, the fact that the focus is usually in a sense exhaustive (*JOHN walks* is usually understood as not only claiming that walking is carried out by John, but also as indicating that John is the only, or at least the

most significant, walker) can be accounted for by a further modification of the apparatus of predication yielding the formula $\lambda p.p(\text{John})!\{\text{walk}\}$ - where $P!\{T\}$, roughly speaking, presupposes the existence of T, claims that P applies to T and that there is no alternative P' to P that would apply to T. For details see Peregrin (1994 and 1995).

(2) From the second viewpoint we see an utterance as essentially consisting of three parts (not all of which have to be overt). Each utterance is seen to consist of a topic (which may be void), a focus, and an operator binding them together (which may be implicit). In the prototypical case the operator can be seen as nothing more than a "higher-level" realization of predication, so that $PRED(\text{John},\text{walk})$ yields $\text{walks}!\{\text{John}\}$ and $PRED(\text{walk},\text{John})$ gives $\lambda p.p(\text{John})!\{\text{walk}\}$. (For details see Peregrin, 1994). The operator $PRED$ is in fact nothing else than Jacobs' (1984) operator $ASSERT$; the place of this operator can be assumed by overt focalizers such as *ONLY*, but also by negation.

(3) If we stick to a dynamic view of language, as articulated by various dynamic semantic theories, we have to see an utterance as something that alters the context, or the information state, in which it is produced. The utterance uses the input information state and works towards an articulation of a message yielding a new, output information state. (This perspective has been suitably formalized by Groenendijk and Stokhof's, 1989a and 1989b, *dynamic logic*; see also Peregrin & von Heusinger, 1995). Topic and focus now present themselves as two different phases in this process; thus, an utterance consisting of a topic T and a focus F has to be analyzed as $T \} \&! F$, where $\} \&!$ is a new kind of concatenation operator signalling the switch between two modes of evaluation (see Peregrin, 1995). In contrast to the Groenendijk-Stokhofian logic, it is vital here that the logic needed to accommodate such a proposal can distinguish between two kinds of failure of an utterance: between falsity and infelicity: the formula $A \} \&! B$ is then (1) *true* w.r.t. an input evaluation f iff there exists an evaluation f' such that $\langle f, f' \rangle \in A$ and an evaluation g such that $\langle f, g \rangle \in B$; it is (2) *false* w.r.t. f iff there exists a f' such that $\langle f, f' \rangle \in A$, but there exists no f'' such that $\langle f, f'' \rangle \in B$; and it is (3) *infelicitous* w.r.t. f iff there exists no f' such that $\langle f, f' \rangle \in A$.

5 A Case Study: Bill Meets the Dallas Clan

To illustrate the foregoing of this, let us take the sentence (1) borrowed from Krifka (1991).

John only introduced Bill to Sue (1)

This sentence, devoid of topic-focus articulation, can be schematized as (1') (for simplicity's sake we remain on the level of extensions)

$$\mathbf{introduce(J,B,S)} \tag{1'}$$

Notice that the same sentence (still without any topic-focus articulation) could be equivalently analyzed in many other ways; for (1) is equivalent to many other formulas, and the particular one we employ is a purely technical matter. Thus, instead of (1) we could alternatively use any of, say, (1'a) through (1'c) (where x is a variable ranging over individuals, p a variable ranging over properties of individuals, r a variable ranging over binary relations among individuals, and q a variable ranging over properties of binary relations among individuals):

$$\lambda x \mathbf{introduce(J,x,S)(B)} \tag{1'a}$$

$$\lambda p.p(\mathbf{B})(\lambda x.\mathbf{introduce(J,x,S)}) \tag{1'b}$$

$$\lambda q.q(\lambda xy.\mathbf{introduce(x,B,y)})\{\lambda r.r(\mathbf{J,S})\} \tag{1'c}$$

All these formulas are provably equivalent (by lambda-conversion) to (1'), and so they furnish the very same semantic analysis of (1').

Now let us take (1) with *Bill* stressed, i.e. (2)

$$\mathbf{John\ introduced\ BILL\ to\ Sue} \tag{2}$$

Speaking informally, this sentence expresses the claim that the person introduced by John to Sue is Bill, hence it is about the property analyzable as $\lambda x.\mathbf{introduce(J,x,S)}$ and it claims that this property is instantiated by Bill (and moreover that Bill is in some sense the only significant - in the simplest case the unique - instantiant of the property). To account for this intuition, we cannot simply take (1'b) (for this would in effect be to license no semantic difference between (1) and (2)); we have to make use of the modified predication

$$\lambda p.p(\mathbf{B})!\{\lambda x.\mathbf{introduce(J,x,S)}\} \tag{2'}$$

(2'), in contrast to (1'), has a truth value only if there is someone whom we introduced to Sue, and it is true only if Bill is - in the present context - the only "relevant" person we introduced to her. In fact, (2) is compatible also with another topic-focus articulation, with not only *Bill*, but *introduced Bill* constituting the focus, which yields the following analysis:

$$\lambda q.q(\lambda xy.\mathbf{introduce(x,B,y)})\{\lambda r.r(\mathbf{J,S})\} \tag{2''}$$

An alternative way to express (2') and (2'') would be

$$\mathbf{PRED}(\lambda x.\mathbf{introduce}(J,x,S),\{\mathbf{B}\}) \quad (2'a)$$

$$\mathbf{PRED}(\lambda r.r(J,S),\{\lambda xy.\mathbf{introduce}(x,B,y)\}) \quad (2''a)$$

These analyses amount to comparing two classes and claiming that the latter exhausts the "significant" part of the former; in case of (2'a) the two classes are the class of persons introduced by John to Sue and the class consisting of Bill, in case of (2''a) the class of all that John does to Sue and the class consisting of introducing Bill. (That these new analyses are equivalent to the old ones is ensured because **PRED** is defined so that $\mathbf{PRED}(A,\{B\}) = \lambda p.p(B)!\{A\}$).

This perspective is plausible if we consider vocalizers, as in the sentence (3)

$$\textit{John only introduced BILL to Sue} \quad (3)$$

In this case we may assume the focalizer to simply assume the place of the implicit general quantifier **PRED**; the two resulting analyses are

$$\mathbf{ONLY}(\lambda x.\mathbf{introduce}(J,x,S),\{\mathbf{B}\}) \quad (3')$$

$$\mathbf{ONLY}(\lambda r.r(J,S),\{\lambda xy.\mathbf{introduce}(x,B,y)\}) \quad (3'')$$

claiming now that the two compared classes coincide. This perspective can be also used to account for what Krifka (ms.) calls "second occurrence focus", like in (4), which can be analyzed as (4') or (4''):

$$\textit{John also only introduced Bill to Pamela} \quad (4)$$

$$\mathbf{ALSO}(\lambda y.\mathbf{ONLY}(\lambda x.\mathbf{introduce}(J,x,y),\{\mathbf{B}\}),\{\mathbf{P}\}) \quad (4')$$

$$\mathbf{ALSO}(\lambda y.\mathbf{ONLY}(\lambda r.r(J,y),\{\lambda xy.\mathbf{introduce}(x,B,y)\}),\{\mathbf{P}\}) \quad (4'')$$

Back to (2) - another way to analyze it, by means of a suitably modified version of dynamic logic, would lead to the analyses

$$\mathbf{introduce}(J,d,S) \}!\& d=\mathbf{B} \quad (2'a)$$

$$r(\mathbf{J},\mathbf{S}) \}!\&! r=\lambda xy.\mathbf{introduce}(x,B,y) \quad (2''a)$$

6 Conclusion

In this paper we have tried to bring forward some arguments for the following theses:

1. Topic and focus are two aspects of a single articulation pattern which is basic for every sentence.
2. It is just this pattern that was pointed out by the linguists of the Prague School under such names as the 'topical structuring' ('aktuální _len_ní') or 'functional sentence perspective'.
3. This pattern is relevant semantically, namely in that it triggers an existential presupposition connected with the topic, and that it gives the focus a certain claim of exhaustivity of the significant.
4. The pattern can be viewed from three different perspectives which lead to formalizations in the spirit of three different formal semantic theories (predicate logic, theory of generalized quantifiers, dynamic logic).

References.

- Groenendijk, J., M. Stokhof. 1989a. "Dynamic Predicate Logic". *ITLI Prepublication Series LP-89-02*, University of Amsterdam. Printed in *Linguistics and Philosophy* 14.39-101. 1991.
- . 1989b. "Dynamic Montague Grammar, a first sketch". *ITLI Prepublication Series X-89-04*. University of Amsterdam.
- Haji_ová, E. 1984. "On Presupposition and Allegation". Sgall 1984. 99-122.
- . 1994. "Topic, Focus and Negation." *Focus & Natural Language Processing* (ed. P. Bosch & R. van der Sandt), IBM Deutschland, Heidelberg.
- Jacobs, J. 1984. "Funktionale Satzperspektive und Illokutionssemantik". *Linguistische Berichte* 91.25-28.
- Krifka, M. 1991. "A Compositional Semantics for Multiple Focus Constructions". Proceedings of SALT I. *Cornell Working Papers* 11.
- . Ms.: *Focus and/or Context: A Second Look at Second Occurrence Expressions*. To appear.
- Mathesius, V. 1939: "O takzvaném aktuálním _len_ní v_tném". *Slovo a Slovesnost* 5.171-174. Translated as "On Information-bearing Structure of the Sentence". *Harvard Studies in Syntax and Semantics* ed. by S. Kuno. Harvard: Harvard University 1975.467-480.
- Partee, B. 1991. "Topic, Focus and Quantification". Proceedings of SALT I. *Cornell Working Papers* 10.
- . 1994. "Focus, Quantification, and Semantic-Pragmatic Issues." *Focus & Natural Language Processing* (ed. P. Bosch & R. van der Sandt), IBM Deutschland, Heidelberg.

- Peregrin, J. 1994. "Topic-focus articulation as Generalized Quantification." *Focus & Natural Language Processing* (ed. P.Bosch & R.van der Sandt), IBM Deutschland, Heidelberg.
- . 1995. "Topic and Focus in a formal Framework". To appear in *Discourse and Meaning* (ed. by B.Partee & P.Sgall).
- Peregrin, J. and von Heusinger, K. 1995. "Dynamic Semantics with Choice Functions". In: U. Egli & K. von Heusinger (eds.). *Choice Functions in Natural Language Semantics*, Konstanz: Universität Konstanz, 43-67.
- Sgall, P. ed. 1984. *Contributions to Functional Syntax, Semantics and Language Comprehension*. Amsterdam: Bejnamins - Praha: Academia.
- Sgall, P., E. Haji_ová & J.Panevová. 1986. *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Praha: Academia.