A Semantic Explication of *Information Status* and the Underspecification of the Recipients’ Knowledge

Starting with Prince (1981), various classifications have been proposed for annotating the *information status*, i.e. degree or mode of *givenness* of referential categories such as NPs, DPs or PPs in ordinary text. Finding the explanatorily most adequate classification has proved a difficult task. More recent approaches such as Nissim et al. (2004) differ considerably from Prince’s original taxonomy. In Prince (1992), we find a basic categorization along two dimensions, hearer’s knowledge and discourse occurrence, for which the following terminology has been proposed by various researchers:

<table>
<thead>
<tr>
<th>...</th>
<th>mentioned</th>
<th>unmentioned</th>
</tr>
</thead>
<tbody>
<tr>
<td>hearer-known</td>
<td>old/given</td>
<td>accessible/mediated/inferrable</td>
</tr>
<tr>
<td>hearer-unknown</td>
<td>-</td>
<td>new</td>
</tr>
</tbody>
</table>

There are, however, a couple of open questions.

I. What to do with definite entities such as proper names or definites like “the Pope” upon their first occurrence in a text: are they *new* (due to being “unused”) or *accessible/inferrable/mediated*?

II. What about bridging entities: should they be considered *given/old* because they have some sort of discourse antecedent or merely *accessible* because their antecedent is not coreferential?

III. How can annotations of *text* possibly account for the fact that different addressees have different knowledge?

There are basically two ways to improve our understanding of what might be the linguistically most appropriate taxonomy. The first one is to conduct prosodic investigations in order to detect significant correlations between certain categories of information status and intonational parameters, like e.g. in Baumann (2006). The second one, which I am going to pursue in this talk, is to investigate the most commonly used categories in formal semantic terms and reduce them to their presuppositional and anaphoric properties. This will provide semantic arguments that help finding answers to questions (I) and (II).

The third question has surprisingly been completely neglected in discussions on information status, although it should be clear that a notion which is explicitly said to depend on the recipient’s knowledge simply cannot be annotated as a static property of the textual item itself, a problem which shows up with many kinds of definites. The individual denoted by “Stuttgart mayor Wolfgang Schuster” certainly won’t be familiar to a global audience. Still, it might get classified as *mediated-general* (or generally *accessible*) on a par with “the Pope”, according to Nissim et al. (2004), because it doesn’t seem inappropriate despite its “unknown” status. This is certainly not convincing. On the other hand, classifying the item as *new* wouldn’t do justice to those addressees who do know the item’s referent. How can we escape this dilemma? I would like to argue in favour of two things: (i) the introduction of a category *accessible by description*, which is supposed to do justice to the fact that certain definites – despite denoting a previously unfamiliar entity – allow for a unique mental categorization, other than e.g. the phrase “the man”. (ii) To finally acknowledge the fact that, quite often, text might be suitable for different types of recipients at the same time and, hence, classification should be kept underspecified or ambiguous.

The point here is that text of this kind is explicitly meant to be suitable for all such kinds of readers – not leaving the less knowledgeable out in the rain, which would be the case if the description simply had been “Wolfgang Schuster”, but still not sounding annoying to those already informed. It is furthermore clear, that presuppositional accommodation plays a crucial role, functioning as the necessary “shock-absorber”. Let us look at the way how such a definite description is processed in the minds of different recipients of the same message. We assume, moreover, a discourse hierarchy reminiscent of Kamp (ms.), according to which anaphoric resolution is attempted stepwise, first, in the actual discourse context, second in a situative context, and third in the so-called encyclopaedic context comprising the entire “world knowledge” of the respective individual.

According to Kamp (2001) proper names, on a par with other definites, are in some sense anaphoric. Their anaphoric nature consists in the requirement they place on the addressee to uniquely identify the bearer of the name. A sentence like “Fred is ill.” would therefore be represented as in (1), using the compact DRT-notation of Geurts and van der Sandt, as well as the $\partial$ presupposition operator from Beaver (1995).

\[
\begin{align*}
(1) & \quad [ : \text{ill}(x), \partial [x : \text{Fred}(x)] ]
\end{align*}
\]
Proper names don’t seem ideal discourse anaphors, at first glance. This impression might derive from the fact that they are typically used to introduce an entity into the discourse for the first time. This, however, does not make them equal to *indefinites* like “a man”, which – unless used specifically – creates an “entirely” new referent. Proper names, on the other hand, are *definites*. Their definiteness, however, does not necessarily manifest itself in the picking up of an already available discourse referent (which would amount to their being *discourse given/old* and which only seems to occur in longer texts), but in their ability to identify with an individual present in the recipient’s world knowledge. Therefore, although names introduce a discourse referent we should not say that their information status was *new*. Resolution of (1) – as long as a unique person called Fred can be singled out – results in (2).

\[(2) \ [x: \text{ill}(x), \text{Fred}(x)]\]

Unfortunately, we cannot anticipate in which context the proper name is going to find its antecedent. Thus, the behaviour of the item can’t be marked in the representation of its presupposition itself but has to be handled in a separate module of anaphoric retrieval which is sensitive to the context hierarchy.

Definites like in the German sentence (3a) will likewise just get one representation, shown in (3c), which, however, is going to function in different ways with different addressees. (Note, that the most commonly used English translation seems to differ from German with respect to the missing definite article.)

\[(3) \begin{align*}
\text{a.} & \quad \text{Der spanische Formel-1-Weltmeister Fernando Alonso ist krank.} \\
\text{b.} & \quad '\text{Spanish Formula 1 world champion Fernando Alonso is ill.}' \\
\text{c.} & \quad [\text{krank}(x), \partial[x: \text{spanisch}(x), \text{Formel}_1\text{-Weltmeister}(x), \text{Fernando}_\text{Alonso}(x)]]
\end{align*}\]

Suppose now, that the message is received by three addressees. A is a sports fan and has followed Alonso’s career very closely, B is not particularly interested in Formula 1 but has heard of Alonso, while C doesn’t know anything about what’s going on in motor sports despite of being familiar with the term “Formula 1”.

For A, (3c) will function just like the name in (1), i.e. she is going to identify the referent of the definite description in her knowledge space and verify that the presuppositional content is compatible with what she knows about the described individual. B will do the same, although for her the information that Alonso is Spanish will be informative, so she is going to accommodate it. C, finally, who is able to cope with the meaning of the word “Formel-1-Weltmeister” (and deciding to interpret “Weltmeister” as the incumbent champion) has enough information to open a unique “mental drawer” for the referent of the anaphor in order to introduce a discourse referent *based on the referent in the drawer*.

In all three cases the final representation is going to look like in (4).

\[(4) \ [x: \text{krank}(x), \text{spanisch}(x), \text{Formel}_1\text{-Weltmeister}(x), \text{Fernando}_\text{Alonso}(x)]\]

Anaphors that are resolved via bridging may be represented just like the other definites above. Consider the mini-discourse in (5a) and its preliminary representation in (5b).

\[(5) \begin{align*}
\text{a.} & \quad \text{The UN held a meeting. The secretary-general gave a speech.} \\
\text{b.} & \quad [x: \text{UN}(x), \text{held\_meeting}(x), \text{gave\_speech}(y), \partial[y: \text{secretary\_general\_of}(y, z), \partial[z:]]]
\end{align*}\]

In this case, the anaphoric variable \(y\) can neither be resolved in the discourse nor in the encyclopaedic context for the reason that the former doesn’t contain a suitable referent and the latter contains too many (in other words, its description is insufficient). However, after identifying \(z\) as the UN, the descriptive content of the presupposition suffices to identify and therefore introduce \(y\) into the main discourse representation.

**Short References**