

typically indefinite NPs or focal sentence intonation or later word order. Such introductions can be thought of as moving a referent from off stage onto the stage, or from some long term memory store into the current conceptual representation.

Speakers use other linguistic forms, most typically anaphoric forms, pronominalization or definite NPs, to signal referents which are already available to the listener. Such referents can be thought of as activated or enplaced within the conceptual representation. Speakers may also reintroduce a referent after a long hiatus or some other interruption.

Keeping track of referents over time involves important interplay between the activation status of a referent and the granularity of discourse. Of numerous linguistic and psycholinguistic approaches to reference management in discourse, the episode model (van Dijk and Kintsch, 1983; Fox, 1987; Marslen-Wilson et al., 1982) has been the most influential. It considers the use of *anaphora* to be a function of a particular discourse structure – the paragraph or *episode*. The basic assumption underlying the model is that while texts may be produced in a linear fashion, they are nevertheless hierarchically organized and processed as episodes – semantic units dominated by higher level macropropositions (van Dijk and Kintsch, 1983). This episode organization has dramatic consequences for reference tracking in texts.

The notion of episode as a semantic unit dominated by a macroproposition has been found to have psychological relevance in several studies. Black and Bower (1979), for example, demonstrated in a psychological study of story processing the existence of episodes as chunks in narrative memory. Similarly, Guindon and Kintsch (1984) in their experiment studying the macrostructure of texts found that macrostructure formation appears to be a virtually automatic process. That is, people appear to form macrostructures during reading and derive relevant macropropositions of a passage as soon as possible. Their findings provided evidence for the episode and the *macrostructure* theories of Kintsch and van Dijk (1978) and Schank and Abelson (1977).

Gernsbacher (1990) supports the episodic organization of stories, reporting that comprehenders capture the episode structure of narratives in their mental representation by building separate substructures to represent each episode. Readers build new mental substructures for new episodes, where information on the previous episode is less accessible to them. It is therefore harder for readers to draw coherence inferences across two episodes than within the same episode.

The cognitive basis of discourse organization helps us further understand the relationship between discourse structure and anaphora. An episode, as a semantic unit subsumed under a macroproposition, is the textual manifestation of a memory chunk which represents sustained attentional effort and endures until an episode boundary is reached. Attention shifts when the processing of the episode is completed. In other words, 'the macroproposition remains in Short Term Memory for the rest of the

interpretation of the same episode. As soon as propositions are interpreted that no longer fit that macroproposition, a new macroproposition is set up' (van Dijk, 1982: 191). At an episode boundary, where a change of macroproposition occurs (that is, new agents, places, times, objects or possible worlds are expected to be introduced), the encoding load is much heavier, the reference under concern is less accessible, and hence a more explicit anaphoric form (such as an NP) is required to code the referent. Within an episode, when the macroproposition is maintained, the referent under consideration is more accessible and hence a less explicit anaphoric form (such as a pronominal) is sufficient to code the reference.

Indeed many studies on anaphora have reported the alternation between nominal and pronominals to be a function of the paragraph or episodic structure. Hinds (1977), for example, discusses how paragraph structure controls the choice of NPs and pronouns. He finds that noun phrases are used to convey 'semantically prominent' information in peak sentences of a paragraph while pronouns are used to indicate 'semantically subordinate' information in non-peak sentences. Fox (1987) demonstrated that structural factors of discourse establish the basic patterns of anaphora: NPs are generally used at the beginning of a 'development structure' to demarcate new narrative units, whereas pronominals are used within that structure. Marslen-Wilson et al. (1982) also argue that a speaker's use of referential devices is governed by discourse structure and the context of speaking. The general pattern of anaphora is that NPs and proper names are used to establish initial reference at an episode when a particular referent is in a state of low focus, whereas pronouns are used to maintain reference within an action sequence when a particular referent is in a state of high focus.

While the episode model presupposes the importance of cognitive constraints and hierarchical organization of discourse, it faces difficulties to the extent that structural units such as paragraph, episode, event and theme are not well defined theoretically. Many structural units are hard to identify in spoken and written texts, and are prone to misinterpretation.

Tomlin (1987a) attempts to solve the problem of prior studies by introducing the attention model, where he ties the use of anaphora directly to the cognitive activities of attention and memory. He argues that an episode represents sustained attentional effort and endures until attention is diverted (that is, an episode boundary is reached). In his study, each episode was represented by a slide picture; the shutter release cycle of the slide projector, which imposed a sufficiently strong perceptual disruption for the subject, served as the episode boundary. He demonstrated experimentally that NPs are used at the boundary of episodes when attention shifts, while pronominals are used within episodes when attention is sustained (see Tomlin and Pu, 1991).

Another important model of referential management is the distance model proposed by Givón (1983), which argues for a correlation between anaphora and referential distance in discourse, such as number of clauses between a given anaphor and its antecedent. The distance model may be a

manifestation of a psychological factor, such as short term memory decay. The 'iconicity principle' underlying the model is that the longer the distance, the harder it is for the listener to identify the referent, and so a more explicit referential form (such as a full noun phrase) is required. The shorter the distance, the easier it is for the listener to identify the referent, and hence a less explicit referential form (a lexical pronoun or a zero anaphor – ellipsis) is required.

The Thematic Management of Discourse

The key insight within thematic management is that certain concepts and propositions seem more central or important to the development of the discourse than others. Such central concepts and propositions provide the framework or scaffolding around which the details of the discourse are emplaced. Such central concepts and propositions seem to be what is better remembered when a discourse is interpreted. The key questions are: (1) What makes a concept or proposition more central? (2) How is that centrality tied to the developing discourse? (3) What does the speaker do to convey centrality to the listener? and (4) How does the listener know when to construe a concept or proposition as central?

Thematic management in discourse concerns four interrelated problem areas. First, at least historically first, is the problem of clause level *theme* or *topic*. In general, it appears that each clause in a discourse contains one key element – the *theme* – which is somehow the *central* referent at that moment, a *point of departure* for the clause, the referent *about which* the remainder of the clause predicates something. Second, there is the problem of higher level paragraph or *discourse theme*. In this area, one is concerned with either the centrality or the significance of a referent globally or with the aggregate propositional goal of a discourse or some major component unit. Third, there is the problem of *foregrounding* in discourse. Propositions in a discourse can be sorted into at least two coarse sets. Some propositions relate key concepts or events in discourse; others in one way or another elaborate or provide supporting information. The former are called *foregrounded* propositions; the latter *background*. Finally, there is growing interest in the cognitive underpinnings of thematic management, in particular the relationship of theme to cognitive processes of memory, attention, and consciousness.

Theme or Topic: Conceptual Starting Point

There are several quite excellent reviews of ideas and issues of clause level theme or topic (Goodenough, 1983; Gundel, 1988b; Jones, 1977). Each clause or utterance is theorized to contain one element which is more important or central to the discourse; or, more technically, which serves as the *starting point* of the utterance; or which serves as the element *about which* the predication is asserted.

To understand the idea of theme or topic requires that one deal with a number of interrelated matters. One, there is the theoretical definition of theme or topic, which is generally articulated in terms of *starting point* or *aboutness*. Two, there is its manifestation through syntactic form, most generally discussed in terms of constituent order (initial or preposed), or syntactic subject, or other morpho-syntactic cues (such as *wa* in Japanese: Hinds et al., 1987). Three, there is the interplay between these two areas, the extent to which the definitions of theme or topic include information about its syntactic manifestation.

In the paragraph fragment in (5), the bold-faced NPs are ostensibly clause level themes. In each case, the referent seems introspectively to satisfy requirements of importance, centrality, starting point, and aboutness.

(5) Text fragment from popular novel *Sarum* by E. Rutherford (1988: 206)

- 1 Late in the winter, while **the snow** was still on the ground,
a new figure of great significance arrived on the island.
 He was tall, middle aged, with a thin, kindly face and receding
 hair.
 He had two peculiarities that Porteus observed:
- 5 **he** stooped when he spoke to people,
 as though \emptyset concentrating intently on what they said;
 but when \emptyset not involved in conversation
his eyes often seemed to grow distant
 as though **he** were dreaming of some far off place.
- 10 **He** was Julius Classicianus, the new procurator and replacement
 for the disgraced Decianus Catus.
His responsibility included all the island's finances.
 Under the Roman system of divided authority, **he** reported direct
 to the emperor.

From a global point of view, this paragraph introduces a new character to the novel, one Julius Classicianus, and provides his initial description.⁴ This paragraph illustrates well the basic ideas and issues surrounding clause level theme or topic. First, several of the clauses exhibit prototypical cases of clause level theme (3, 4, 5, 10, 12). In each of these cases, the referent of the bold-faced NP is the central or most important character in this paragraph. It is also the referent about whom the predication is asserted. And, it is also the referent from which the description proceeds. Though it cannot be a defining characteristic of theme, it is also not an accident that the relevant NP in each case is the subject of the clause. It is on such cases that the central theoretical ideas of theme and topic have been constructed.

Second, several of the clauses illustrate cases which remain problematic in discussions of theme and topic. One, it is not clear how to treat the subject of clause 2, *a new figure of great significance*. Some argue that this NP is *not* a theme because the information is referentially new, unlike the NPs in the prototypical cases. Others argue that this NP *is* thematic precisely because its referential status is in principle independent of its

thematic status, and in this clause it is this NP which is the starting point of the utterance and about which a predication is asserted.

Two, it is not clear how to treat the subjects of clauses 8 and 11. Neither denotes the principal character of the paragraph, so it is difficult to sustain a view that the referent is somehow important. Yet for each clause itself, the bold-faced NP seems to be the starting point of the utterance and the NP about which the utterance is predicated.

Three, there are a number of subordinate clauses, some embedded, which have either explicit or elliptic subjects. It remains unclear exactly how the thematic status of key NPs in these clauses is to be treated.

With these observations in mind, we can turn to explore more carefully major ideas about clause level theme and topic.

Conceptual Foundations for Theme and Topic

Despite individual variation in the formulation of definitions and the specific terms defined, there are essentially three basic ideas of what constitutes a clause level theme or topic: (1) the theme is what the sentence is *about*, (2) the theme is the *starting point* of the sentence, and (3) the theme is the *center of attention* for the sentence.

Theme as Aboutness Classical scholarship on language and logic distinguished those portions of a sentence about which something is predicated and that which is predicated of it. In classical terms the difference between *subject* and *predicate* was exactly this sort of difference. Modern and contemporary research developed this idea, leading to many formulations of clause level theme or topic in terms of aboutness. A particular referent counts as theme when it is this referent that the remainder of the sentence is about. One classic 'test' for aboutness comes from Gundel (1974; 1988b): in a given context, a particular NP will count as a topic (for Gundel) if it can be included in a preposed adverbial with *As for*. Thus, the bold-faced NP in example (5), clause 8, counts as a topic because this referent felicitously fits into the following sentence:

(6) As for **his eyes**, they seemed to grow distant . . .

Theme or topic as *aboutness* dominates current research in this area. It begins with classical research from the Prague School, proceeds through Halliday (1967b; 1973; 1976) and Halliday and Fawcett (1987), and features prominently in much current research (Gundel, 1974; 1988b; Hajicová, 1984; Lambrecht, 1994; Reinhart, 1981; Vallduví, 1992).

Theme as Starting Point Another way of conceptualizing theme is as the starting point of the utterance as a message. In this view, the speaker plans his message to proceed from a point of view held in common with the listener out to a message novel to the listener. The starting point helps to frame the utterance, tying the predication to something already known