

REFERENCE AND THEME AND THE INTERPLAY OF THESE TWO TEXTUAL SYSTEMS

Eija Ventola

The University of Helsinki

The paper discusses how two textual systems, reference and theme, and their interplay, contribute to the cohesion of texts. The data used for exemplification are scientific texts produced by English academics and Finnish scholars writing in English. In contrast to native English writers, Finnish writers tend to have difficulties in coding the references to text participants appropriately in their texts. Furthermore, when organizing their texts thematically, Finnish writers appear to apply thematic patterns which are not typical to English texts. Moreover, the Finnish writers do not seem to utilize the possibilities of the interplay between the two systems at their textual optimum. The challenge that applied linguistics faces in the field of teaching academic writing in a foreign language is to develop learners' consciousness and linguistic skills of organizing information referentially and thematically appropriately in texts.

1. Introduction

Ever since grammars have been extended beyond sentence boundaries, linguists have focused their attention on questions of cohesive principles in text creation. For cohesion in English texts, two textual systems, **reference** and **theme**, are of vital importance. The functioning of both of these systems has been discussed extensively in literature (for 'reference', see Halliday & Hasan 1976; Martin 1983; Ventola 1987, 1992; for 'thematic progression', see Daneš 1974; Fries 1981). However, the **interplay** between these systems has not received much attention. This paper discusses how the two systems and their interplay contribute to cohesion in **English academic texts** written by both native and non-native writers.

Native writers of academic English, either consciously or unconsciously, make skilful use of the systems **reference** and **theme** and their **interplay** to build up clear argumentation, and good and experienced native writers do this more frequently and more naturally than inexperienced writers, but non-native writers may face a slightly different situation. They may be

fluent and competent writers in their native languages, but when writing in English they may have extreme difficulties in making these cohesive systems work, and since their texts are not cohesively appropriately coded they are also difficult to comprehend. The extent of the problems encountered in non-native texts depends on the linguistic background, training, experience in writing, etc. of the non native writers in question. The non-native writers studied here are Finnish researchers.

The paper is part of the work of a contrastive textlinguistic project carried out at Helsinki University Language Centre in 1989-1991¹. The ultimate goal of the project was to help Finnish writers in their task of reporting in English the results of their research and to raise the standard of academic writing of English generally in Finland through building up and organizing textlinguistically-oriented writing courses at tertiary level. Earlier work of the project has been reported, for example, in Ventola and Mauranen (1990), and several papers are in the process of being published (e.g. Mauranen 1992; forthcoming; Ventola 1991; 1992; Ventola & Mauranen 1991).

In short, then, the paper will highlight some problems discovered in the academic English of Finnish writers in three areas of cohesion: reference and theme, and the interplay of these two systems. The paper develops as follows: the first main section of the paper discusses the problems that Finnish writers seem to encounter in terms of consistency to their references to certain text participants; the second main section considers the difficulties Finnish writers seem to encounter when building up their clauses in terms of theme and rheme, when trying to construct various kinds of thematic patternings for their texts, and in establishing a certain method of development for their texts; the third main section focuses on the interplay between the systems of reference and theme, and on how the interplay might be used as a further tool to increase cohesion and clarity of argumentation in the texts of Finnish writers; the final section presents some considerations from an applied perspective.

¹ Directed by the author; Anna Mauranen as co-researcher.

2. Reference: Identity chains and participant tracking

Reference in this context will be understood in the way it has been described and analysed, for example, in Halliday and Hasan (1976:31-87), Halliday (1985:290-295), Martin (1983), Ventola (1987).

To summarize the basic principles of referential cohesion, reference items are taken to be those items which

instead of being interpreted semantically in their own right, ... make reference to something else for their interpretation.... The information to be retrieved is the referential meaning, the identity of the particular thing or class of things that is being referred to; and the cohesion lies in the continuity of reference, whereby the same thing enters into a discourse a second time. (Halliday & Hasan 1976:31)

Such items in English are of three types: 1) **personals** (*I, you, he*, etc.), 2) **demonstratives** (*this, those, the*, etc.), and 3) **comparatives** (*similar, other, -er*, etc.). Items from each category can function as heads of nominal groups (*he*) or they may carry other functions such as deictic (*the man*), possessive (*his hat*), or adjunct functions (*here*) (see e.g. Halliday 1985:295).

What is of specific interest to us here is the way the reference items are used to create **referential identity chains** in texts. Ultimately the identity of a referent in such chains can either be established in the context of situation or it can be non-existent (e.g. in fairy tales), that is, the identity is simply established in the verbal context without a real extralinguistic referent.

Whenever referential identity chains are established, lexical cohesion, too, plays a role in establishing these chains; thus, *the man* and *the chap* could be linked into the same chain begun by the presupposed item *a man*. The cohesive link would be two-fold, consisting of a referential link realised by *the* in both *the man* and *the chap* and a lexical cohesion link realised by repetition and synonymy in *man* and *chap*. Reference can be **specific** or **general** (*the men/men*), each type of linking being nevertheless cohesive.

Like Halliday and Hasan (1976), Martin (1983) sees referential items as a way of creating cohesion in texts. He speaks of **participant tracking** in texts, and defines it as a way in which various entities, such as people, places, things and events, are introduced into a text. The introduction of participants takes place by using **presenting reference** (realised by indefinite articles, indefinite pronouns, etc.). Once the participants are there, they are then kept track of throughout the text by **presuming reference** (by the personal and demonstrative reference items illustrated above) (Martin 1983:48). Some genres, narratives for example, have reference chains which track 'hero' participants.

Although English texts in academic genres do not have references to 'hero participants' in the same sense as they are found in narratives, cohesive referential identity chains and participant tracking can certainly be found in all kinds of academic texts. For example, almost any experimental research report contains a section similar to Example 1 which presents the studied subjects, items, etc. (the clauses in the example are numbered for easy reference).

- (1) (1) An 11-year-old Saudi Arabian boy was referred to the Eastman Dental Hospital in 1982, suffering from early loss of the permanent teeth, difficulty when eating, and with a history of premature shedding of deciduous teeth. (2) There was no history of other serious illness. (3) During the first 2 years of his life, the patient had had recurrent chest and face abscesses. (4) Since the age of 5 years, hyperkeratotic lesions had been present on his palms and soles. (5) The patient was diagnosed as having Papillon-Lefèvre Syndrome, (6) but unfortunately he had to return home immediately (7) and so treatment was not possible.... (Vrahopoulos et al. 1988:17)

Example 1 is from an article on dentistry, from a section where 'the studied case' is reported. An important participant is introduced to the text with a presenting reference item *an*, after which the writer refers to the participant by presuming reference items, and thus creates a cohesive referential identity chain consisting of items: *his*, *the* (patient), *his*, *the* (patient), *he*. (In addition to the reference items, the lexical relations between *boy* and *patient* also contribute to the cohesion of the extract, but are not considered here.)

Expressing referential relations and establishing identity chains in texts is not problematic for most native English writers of academic papers. Participants are introduced into the text and thereafter referred to systematically. Only very young native English writers might have some problems in realizing referential relations appropriately, especially when such writers are operating in a text genre not familiar to them. In Example 2 a seven-year-old child is writing a school report on snails and runs into problems in keeping track of the generic participant.

- (2) (1) Snails have a shell on their back to protect them selfs from enemy. (2) they like to hide behind a rock and it leavs a silvery track behide him. A snail has a strong foot. They stick to the-ground fermilee. (in Martin 1985:11)

As can be seen in Example 2, the child jumps from generic reference to specific reference (*snails* -> *their* -> *them selfs* -> *they* -> *it*), from the plural form of reference to the singular form (*they* -> *it*), from non-human reference to human reference, (*it* -> *him*). The function of the report is to be general (it is about 'snails'), but the child soon forgets this and emerges into an imagined context with an imagined real referent (*it*), which he treats almost as his pet (*him*). He then remembers the context of writing and returns back to the generic description, but instead of continuing with the already started presumed, general reference chain in its plural form, he begins a new reference chain by using a new singular presenting reference item (*a snail*)². The next clause begins with *they*, but the reader is no longer quite certain where to search for the identity of this participant. Due to these frequent reference changes, the reader, who most likely also evaluates the written performance of the child, thinks that the child has not yet mastered the consistent way of referring to the entities in the text and his report is evaluated as poor.

In adult native English writing, such shifts in reference chains rarely occur. However, referential problems which bear a resemblance to the young child's text in Example 2 can be found in many academic texts written in

² Singular possibly because he perceives a snail possessing only one 'foot', not 'feet'.

English by Finnish writers (see Example 3 below). Why referential problems in cohesive participant tracking can be expected in the academic texts of Finnish writers is explained, not by the unfamiliarity with the genre (as is the case with the child), but rather with linguistic differences between English and Finnish.

Languages do not necessarily give us exactly the same means for realizing participant identification. For example, Martin (1983) reports that Tagalog, a Philippine language, codes participant identification both with the systems at group rank (with pronouns, proper names, and demonstratives) and with verbal affixes operating with the case system at clause rank. This coding contrasts with the participant coding in English, which is realised at the rank of group (by such items as personals, demonstratives, and comparatives). In participant coding, Finnish is presumed to be similar to English (see e.g. Hakulinen & Karlsson 1979:312-327), although there also are differences. One of the major causes of participant coding problems for Finns writing in English is that there are no articles in Finnish which can be used to identify the newly presented participants from those whose identities can be presumed (i.e. the distinction between *a man* - *the man* = *mies*). As in Tagalog, in Finnish the distinction between the presenting and the presuming of identities of participants can be made at the clausal level (e.g. presenting: *Huoneeseen astui mies* - 'Into the room stepped a man' vs. presuming: *Mies astui huoneeseen* - 'The man stepped into the room'). Different ordering is possible when identities are marked at group level, for example, by using indefinite pronouns like *eräs* 'a certain' (*Eräs mies astui sisään* - 'A certain man stepped in'). (A further difficulty for Finns is that Finnish does not differentiate third person pronouns according to the sex; *hän* can refer either to a male or a female third person; this causes many problems for Finns in spoken English discourse, though less in written academic discourse.) Thus, the major problem for Finnish writers, as far as reference is concerned, seems to be related to the use of articles. They are the most frequently corrected lexicogrammatical items in the articles of Finnish researchers (Ventola & Mauranen 1990:106; see also Ventola & Mauranen 1991; Ventola 1992).

Let us consider an example which demonstrates a typical case in the papers of Finnish writers where creating a consistent referential identity chain becomes a problem for the writer. The article from which Example 3 is taken discusses the teenage girls whom the researcher studied and divided into four groups according to certain variables. In the extract their characteristics are described group by group. The title and the beginning of the section in the extract, as well as the beginnings of the other paragraphs of the section, are given below (the numbering of sentences has been added for easy reference).

(3) (1) 4. Characteristics of the four groups.

(2) Below the groups are characterised by listing, for each group, those variables on which the group obtained extreme average values (i.e., was clearly differentiated from the other groups).

(3) GROUP A. (4) The girls in this group hoped more often than others did to obtain a job which had a high social status; (5) their parents also tended to have a higher social status than those of other girls. (6) They all lived with both of their parents ... [The paragraph continues.]

(7) GROUP B. (8) This group, the most numerous one (25% of the girls in the total sample), obtained an extreme mean value on one variable only: (9) they spent more time in hobby activities and handicrafts than other girls did ... [The paragraph continues.]

(10) GROUP C. (11) The reason given for smoking by girls in this group was usually "social" ("my/their friends smoke"); the reason for drinking mostly being "relaxation, to have fun". (12) The girls also valued sexual experiences more than other girls did (13) (though they on average were less "experienced" than the girls in group D). (14) When characterizing ideal ways of life, they mentioned conventional matters and "respectability" less often than the others did. (15) In their self-images, leadership & initiative, emotionality and impulsiveness were emphasised more than in those of other girls. (16) They were comparatively optimistic in respect of the future. (17) They valued intimate and secure friendship, and possibilities to meet new and interesting people, more than they believed that adults do. (18) (However, their own values did not differ from those they believed to be characteristic of adults in respect of e.g., health, use of alcohol, or family life.) (19) This group will be further discussed below.

(20) Group D. (21) The parents of girls in group D had on average a lower socioeconomic status than those of the other girls. (22) Their homes were more often "broken", (23) and their families spent comparatively seldom their free time together (though, in the opinion of the girls, often enough)... [The text continues.]

In the section concerning Group C, when the reference codings for the two most important participants, 'the C-group girls' and 'the girls in the other

groups, A, B, and D' in this text are examined, it will be noticed that the writer is inconsistent with the references to the participants. The references for 'the C-group girls' include: (11) *girls in this group* - (12) *The girls* - (13) *they* - (14) *they* - (15) *their* - (16) *They* - (17) *They* (17) *they* - (18) *their* - (18) *they* - (19) *this group*. At the first encounter of the text, native English readers will most likely pause after the nominal group *girls in this group* in the first clause dealing with group C, in clause (11). The reason for this is likely to be the question that arises in their mind as to whether the writer is still talking about 'the same Group C girls' that have been introduced earlier in the text or whether, in fact, the C group is subdivided and the writer now wants to introduce one of the subgroups. But, when reading on, the English readers realise that the whole of Group C is still being discussed and therefore a mental correction of the Finnish writer's error is made - an article is added to the nominal group: *the* -> *the girls in this group*. Adding *the* disambiguates the reference to specific and unifies the references in the writer's text.

Similar problems arise for readers as they read the references for the control groups B, C, and D: (12) *other girls* - (13) *the girls in Group D* - (14) *the others* - (15) *those of other girls*. Again in (12) and in (15) the writer is not talking of 'other girls in general', but of the specific groups of 'other girls' - the girls in groups A, B, and D. The Finnish writer, like the native child in Example 2, seems to make inappropriate choices for the realisation of reference, here generic instead of specific. The reason for inappropriate references is naturally not the same as it was in the case of the child. It appears that the child had difficulties in keeping his reference on the general level, instead of drifting to the imagined specific context where he observes an individual snail crawling on the ground. The problem with the Finnish writer is not that he does not realize that the reference should be specific instead of generic. Due to the linguistic differences between English and Finnish he simply has not coded the references correctly.

Naturally, in both Examples 2 and 3, readers not only have to rely on reference items and their coding in working out the intended meanings. Lexical

cohesion, too, will enable them to make appropriate cohesion links between the messages and to work out the intended identity participant relations, although the writers have coded them incorrectly. Halliday (1985:310) has noted that in the latter clause in (i) *Algy met a bear. The bear was bulgy*, there are two cohesive relations with the first clause (*the* and *bear*) whereas there is only one (*bear*) in (ii) *Algy met a bear. Bears are bulgy*. Following this, one could say that the Finnish writer in Example 3 does not seem to be able maximize the tools for 'cohesion', since his referential codings are inconsistent. (How such 'neglects' actually slow down the interpretative processes of English readers when they are reading texts written by non-native writers is an interesting psycholinguistic research question, which cannot, however, be pursued further in this context.)

In short, then, whereas the English adult native writer is likely to handle participant tracking effortlessly and make no errors in the realisations of referential items, the papers of Finnish researchers indicate inconsistencies in the use of articles for expression of referential participant tracking. As suggested elsewhere (see Ventola 1992), the solution to the problem might at least partly lie in making writers more aware of reference as a system, in explaining how it functions as a means of participant tracking, and what linguistic tools realize reference. In this way indefinite, definite articles, pronouns, etc. will be seen as realizations of one important textual function, reference, and not just as separate unrelated grammatical categories. When paying attention to referential chaining in his/her texts, the writer is partly taking responsibility for the texture of those texts: **prevailing referential consistency will improve the cohesiveness of texts.**

3. Theme-rheme, thematic progression, and method of development

The second system discussed here is theme, the system which constructs messages as themes and rhemes. From the point of view of cohesion, it is not only the actual individual choices of themes in clauses which are of interest here, but rather the sequences of choices which create certain kinds

of thematic patterning in texts, here called thematic **progression pattern** (Daneš 1974) and **method of development** (Fries 1981).

3. 1. Theme-rheme

Studies on **theme** and its counterpart, **rheme**, are numerous. For the present purposes it suffices to define theme as "the point of departure for the message" (Halliday 1985:38) and to refer to its "constructional role" in texts (see Daneš 1974:113). In English, a theme is considered as being realised by the first constituent in the clause, but clauses can also have multiple themes (textual themes and interpersonal themes preceding topical themes). A topical theme is either unmarked (i.e. the theme is realised by the subject) or marked (i.e. the theme is realised by clause constituent other than the subject; for details, see e.g. Halliday 1985). Rheme will here be taken to be "the part in which the theme is developed" (Halliday 1985:38), the part that "pushes the communication forward" (Daneš 1974: 113).

When theme and rheme structures of English clauses are under consideration, another construction, that of an information unit, must also be discussed. The structure of an information unit is typically that of **given**, that is, recoverable information, followed by **new**, that is, unrecoverable information (for details see, Halliday 1985:271-286). As far as the theme-rheme patterns in English clauses are concerned, what interests us in information structures in this context is the commonly proposed conflation between, on one hand, **theme** and **given** and, on the other, between **rheme** and **new**. In Halliday's (1985:278) words, under normal circumstances "a speaker will choose the theme from within what is given and locate the focus, the climax of the new [indicated by the tonic syllable], somewhere within the rheme". Theme-rheme and given-new are taken to represent two different ways in which the communicator relates to his/her own message, even in written discourse: **theme-rheme** is **speaker-oriented**, the emphasis being on theme, on "this-is-what-I-am-talking-about"; **given-new** is **listener-oriented**, the emphasis being on new, on "this-is-what-I-am-asking-you-to-attend-to" (Halliday 1985:278, 316). When discursive circumstances so

require, the common patterning given above can of course be also realised differently, that is, the functions of theme and new may conflate, for instance, when building up contrast.

What about themes and rhemes in the English texts of Finnish writers? Can Finnish writers produce thematically efficient texts in English, or are we to expect native language interference difficulties in theme-rheme choices, too?

Little in the linguistic literature on Finnish seems to suggest that Finnish writers will predictably encounter difficulties in the clausal theme-rheme structures of their texts. For example, Hakulinen and Karlsson (1979:300, 306) consider theme in declarative clauses in Finnish to be realised by the constituents occurring before the verbal element and rheme by the constituents after the verbal element. This viewpoint does not differ much from that of Halliday (1985). Furthermore, informational structure in unmarked cases in Finnish clauses seems to follow the same theme/given - rheme/new structure as in English. That is, in most clauses the new information will be mapped on to some element in the rheme part of the clause complement.

Thus, following the theme-rheme distinction in Finnish grammar, Finnish writers should have no interference problems when writing English clauses. Indeed, if, for example, we consider the theme-rheme choices which the writer of Example 3 has made for individual clauses, the choices seem appropriate, e.g. (1) *The reason given for smoking by [the] girls in this group* (<- unmarked theme) *was usually 'social'* (<- rheme), (2) *The girls* (<- unmarked theme) *also valued sexual experiences more than other girls did* (<- rheme), (5) *In their self-images, leadership & initiative* (<- marked theme), *emotionality and impulsiveness were emphasised more than in those of other girls* (<- rheme). For the most part in Example 3 the themes are unmarked, that is, the subject of the clause functions as the theme of the clause.

But marked themes also appear, for example, in (5) where a circumstantial adjunct functions as the theme. Then in rhemes, something is said about

'the departure of the message' and usually this is also informationally 'new', for example, in (1) the fact that the reason given by the girls was usually social, in (2) the fact that the girls seemed to value sexual experiences more than the other girls did, and in (5) the fact that what was emphasised in their self-images was leadership, initiative, emotionality and impulsiveness. The thematic and rhematic parts of the individual clauses seem fairly balanced and in that respect similar to English codings. Yet, as will be discussed later in sections 3.2, and 3.3, although the themes may seem to be appropriate when considered individually as thematic choices in the clauses, when they are analysed as text, they appear somewhat incoherent and unmotivated.

But considerably different thematic structures also appear in the clauses of Finnish writers, as shown in Example 4, the clauses which have been taken from the middle of a technical abstract. The themes in the clauses have been marked in bold.

- (4) ... (1) **The input devices (scanners and cameras) for different types of images** are reviewed. (2) **Image storage, especially optical disks, is studied.** (3) **Image computing including digital image processing, computer graphics, image pattern recognition and computer vision, visualization, and virtual realities** are shortly covered. (4) **Image retrieval methods together with hypertext and hypermedia systems** are presented.... (Heimbürger 1990:3)

The writer seems to establish a pattern of thematic choices which radically differs from the typical thematic choices in English clauses. As we can see, the themes in these clauses are all unmarked (that is, theme = subject), and they are fairly complex. Complex themes make the text difficult to read, but more surprising to English readers is that the rhemes that follow are extremely simple: *are reviewed, is studied, are shortly covered, are presented*. For English readers the clauses in Example 4 are peculiar because they systematically have an unbalanced theme/given - rheme/new structure. The reader expects something new to be said about the theme in the rheme part of the clause. But what appears to be said about the themes is trivially that they will be 'reviewed, covered, presented' (the tonic falls on the passive verbal elements in the clauses). The rhemes are practically empty - only the verbal process is there. From the English reader's point of view,

these matters are obviously **not** the most important things that could be said about the themes mentioned. Rather the 'news' would be what could be said about 'the input devices', etc. The Finnish writer's theme/new - rheme/empty' choices seem to be conflicting with the typical 'listener-orientation' in English.

It is interesting to note that we can frequently find Finnish texts where similar kinds of 'empty rheme' structures appear. An example is given in (5) (a paragraph taken from the middle of the materials-and-methods section of an article on dentistry). It begins with clauses where the rhemes in the theme-rheme structures consist only of verbal passive groups; here, too, the rhemes are practically empty (the English translation follows the Finnish structure as closely as possible).

- (5) (1) Potilaiden limakalvot tarkistettiin, (2) hammasstatukset kirjattiin (3) ja näistä laskettiin D- ja DMF-indeksit tai alle 6-vuotialta d- ja dmf-indeksit. (4) Kuusi vuotta täyttäneiden parodontium tutkittiin CPITN:n toteamiseksi. (5) Anamneesissa ilmenneet yleissairaudet, lääkitykset, allergiat tai raskaus huomioitiin (Knif 1989:697).

[a translation: (1) The patients' mucous tissue was examined, (2) the dental status was recorded (3) and from this were counted D- and DMF-indices or in case of under-6-year-olds d- and dmf-indices. (4) The six-year-olds' parodontium was examined in order to confirm CPITN. (5) The general illnesses, medication, allergies or pregnancy revealed in the anamnesis were noted.]

It is apparent that the theme/given - rheme/new constructions in both Example 4 and in the English translation of Example 5 are somewhat unbalanced in the clauses³. In Example 5 the clauses with empty rhemes are also 'empty' of new information - the rheme is realised by the passive verbal process only. The imbalance in such English clauses is not so apparent, when the clauses occur in isolation. Indeed, clauses of this kind of 'empty rheme' patterning can occasionally be found, for instance in the text

³ In Finnish the clauses do not appear as unbalanced; rather they are commonly accepted as 'academic style'. This raises a few queries, such as whether Finnish has adopted passive clauses which are almost empty thematically as its academic style more than English has, and, how this style has developed - independently or perhaps through influences from other languages. There will be no attempt to find answers to these queries here, but naturally it would be interesting to find the answers through detailed comparative studies of texts.

from which Example 1 was taken, e.g. *In this study, 11 permanent teeth and their associated soft tissues from an 11 year-old boy with PLS were examined.* (Vrahopoulus et al. 1988:17). But this clause was not followed by a clause of the same kind - there was no built-up sequence of 'empty rhemes'. English writers do not seem to use this kind of theme/given - rheme/new patterning often in clauses which follow one another immediately, whereas in Examples 4 and 5 rhematically empty clauses are used for creating a certain theme-rheme text pattern which to an English reader appears somewhat peculiar.

When texts such as Example 4 have been presented to some native English language teachers, they have suggested that Finnish writers writing in English are simply overusing the passive. It is a well known fact that scientific language favours the choice of passive in verbal groups (see e.g. Barnickel 1982:53-54). But this does not mean that in English scientific texts the most usual pattern is that in which only the verbal processes are used without any 'new' information in the rhemes of the clauses. Rather, the passive verbal groups are more frequently followed by something onto which new information is mapped, for instance, circumstantial material: *Plaque, cementum and periodontal tissues were examined by scanning (SEM) and transmission electron microscopy (TEM).* [circumstance: method]; *Few bacteria were found in any of the soft tissue layers.* [circumstance: location] (Vrahopoulus et al. 1988:17). The Finnish writer's peculiar sequential patterning of 'empty rhemes' cannot thus be explained by a mere over-use of the passive voice. Rather, there seems to be something in the theme-rheme realisation which cannot be explained by looking at the clause level only, but rather needs to be considered on the discourse level. The various principles of thematic construction have to be considered in larger sections of texts, not just in clauses.

To summarize the discussion so far, one major problem some Finnish writers seem to have when writing academic texts in English is that they construct thematically unbalanced clauses in their texts far more frequently and more sequentially than native speakers seem to. This tendency appears to go against the principles that linguists maintain also govern Finnish

unmarked theme-rheme structures in clauses. Thus the phenomenon cannot be explained by the mere notion of interference and the over-use of the passive voice. The situation becomes even more complex, because Finns seem to be accustomed to constructing the same kind of thematically unbalanced sequences in their texts even when writing in Finnish. If one wishes to help Finnish writers to improve their English thematic structures, it is not enough simply to draw their attention to the problem and to advise them to balance the information more appropriately from the English point of view. One should attempt to understand what Finnish writers are trying to do with these somewhat unexpected structures. Here the clausal analysis of themes and rhemes does not appear to be sufficient - a textual view is needed - an understanding of what motivates the Finnish writer to choose as his/her point of departure for the texts, both in Finnish and in English, certain aspects and to develop the texts round these aspects. In order to attempt an explanation, it is worth pursuing the investigation and examining the notion of thematic progression, as a possible motivating force for the Finnish writer's choices.

3. 2. Thematic progression

As the text unfolds, writers do not select randomly the items which work as points of departure (themes) and what will be said about those departure points (rhemes); rather, consciously or unconsciously, writers attempt to create in their texts various **thematic progression patterns**. Daneš (1974:118-120) proposes the following major patterns of thematic progression: 1) **simple linear progression**, where an item from the rheme of the first clause becomes the theme of the subsequent clause, 2) **constant progression**, where the item in the theme of the first clause is also selected as the theme of the following clause, 3) **hyperthematic progression**, where the item in the theme of the first clause functions as an element from which the themes of the subsequent clauses are derived, and 4) **splitting progression**, where the rheme of the first clause is split into two items,

each in turn being then taken as a theme element in the subsequent clauses. The motivation for these patternings is 'the inner connexity of texts' (Daneš 1974:114).

How elements are organised as themes and rhemes in thematic progression in a text serve textual cohesion purposes. Themes and rhemes are used to organize an argument in terms of the logical relations of cause, consequence, concession, etc., achieving an effect, maximizing clearness of presentation, and so on (Daneš 1974:126). Daneš (1974:122) says all texts are "interwoven with expressions of signalling significant points of thematic progression of the text" and of texts having "networks of orientation" captured in thematic progression patterns. Furthermore, he asserts that all these patterns are observable in many languages and in many types of text, in Czech as well as in German and English, in 'scientific and other professional texts' (Daneš 1974:118). There seems to be no reason for assuming that such thematic patterns are not also available for text construction in Finnish, since theme - rheme structures are recognised in Finnish (see e.g. Hakulinen & Karlsson 1979). Finnish writers should therefore be able to utilize these patterns for the same functions as English writers do. Learning to use various thematic progression patterns should thus cause no problems for Finnish writers writing in English.

Tentative results of the contrastive English-Finnish textlinguistic study on academic writing mentioned above seem to confirm that English writers, as well as Finnish writers writing in Finnish, use all the Danešian thematic progression patterns for creating their text (for details, see Ventola & Mauranen 1990, 1991). When writing in their native languages, the writers use patterns intermittently for introducing and developing topics. But when Finnish writers write in English, they seem to develop their texts predominantly by using linear and, to a lesser degree, constant thematic patterns. The effect of the frequent choice of the linear pattern is as follows: the academic texts in English by Finnish writers give an impression of the texts' moving rapidly from topic to topic and of a considerable amount of material being discussed in the texts (for examples, see Ventola & Mauranen 1990, 1991).

In addition to the tendency to move very rapidly from topic to topic, the texts of Finnish writers also appear to include unexplained jumps in the thematic patternings. These jumps are usually considered to be indications of 'breaks in the argument' (see Fries 1981). In writing 'stream of consciousness' novels, they may be a convenient method of awakening the readers, but they are inappropriate in academic writing. In argumentative texts "each sentence should follow logically from what has gone before" and "the point of departure of each sentence should relate in some way to what has preceded." (Fries 1981:6). But in the texts of Finnish writers, unless the themes are developed from the previous rhemes, new themes are often introduced quite abruptly. There are also sudden jumps which confuse the readers in constant theme patterns developed by Finnish writers, as Example 3 shows. Example 6 lists the themes of the discussion on 'Group C girls' in Example 3.

- (6) Group C.
1. The reason given for smoking by girls in this group...
 2. The girls,...
 3. though they...
 4. they...
 5. In their self-images, leadership & initiative...
 6. They...
 7. They...
 8. (However, their own values...
 9. This group...

In listing the themes used by the writer, the thematic pattern that will most prominently emerge is that of a constant theme of 'Group C girls'. But the coding of the individual thematic items has been a problem for the Finnish writer. The theme of the paragraph, 'the girls', should be introduced to the readers in the most obvious position of the paragraph, that is, as the head of the first nominal group in an unmarked theme. However, here, at the beginning of the paragraph, the writer introduces 'the girls' indirectly in a qualifier, *The reason given for smoking by girls in this group....* Had the writer begun the text with *The group C girls gave as their reason for smoking....*, for example, he would have tuned his readers immediately to the constant theme pattern to be developed. In clauses (5) and (8) the constant theme pattern is also coded less significantly as a referential posses-

sive in a nominal group (*their*) and thus the focus of the reader is turned to 'self-images, leadership & initiative' and to 'values'. Such unmotivated breaks and jumps in the thematic development make the text less focused and reduce the effectiveness of the thematic pattern developed in the paragraph.

In Example 6, in spite of the jumps, the continuity of the constant theme pattern is still largely maintained referentially. Sometimes, however, Finnish writers begin the development of constant theme appropriately, but then seem to get confused in the structures they have created and not to know how to continue. This is illustrated by Example 7, the beginning of the same text as Example 4. The thematic part of the text in Example 7 is in bold, with topical themes underlined.

(7)

Electronic Images

- (1) **This report** is the state-of-the-art study of electronic images and image database systems.
- (2) It is based on the international literature.
- (3) First of all, the basic concepts of electronic images are given.
- (4) The input devices (scanners and cameras) for different types of images are reviewed.
- (5) Image storage, especially optical disks, is studied.
- (6) Image computing including digital image processing, computer graphics, image pattern recognition and computer vision, visualization, and virtual realities are shortly covered.
- (7) Image retrieval methods together with hypertext and hypermedia systems are presented.... (Heimbürger 1990:3)

Here in Example 7, the writer's focus at the beginning of the text is *the report* in clauses (1) and (2). The text begins with a constant theme pattern. But the rhematic information in (2) is so superfluous that the writer cannot continue text creation with this constant theme. The writer has painted herself into a corner, from an English reader's point of view. The writer is forced to turn back to the rheme of (1), in order to find a way of developing the text further. The topical theme of (3), *the basic concepts of electronic images*, seems to follow in a linear 'skip' from the rheme of (1), *the state-of-the-art study of electronic images and image database systems*. This seems to suggest that the writer might now adopt a linear progression. However, the linear patterning cannot be followed because the theme of

(4), *The input devices (scanners and cameras) for different types of images*, cannot be perceived as developing from items appearing in the rheme of (3), since the rheme contains only the passive verbal group, *are given*. Clause (5) has as little potential for development of the linear theme pattern as (4), because its rheme, too, consists only of the passive verbal group, *are reviewed*. In both cases the writer has started with something 'new' in the thematic position. Thus, in (4) and (5) neither a constant nor a hypertheme pattern are developed in the text, since neither do *the basic concepts of electronic images* of (3), *the input devices (scanners and cameras) for different types of images* of (4), and *image storage, especially optical disks* of (5) refer to the same referent, nor are they in any way even related as parts of a referent, which Daneš (1974) implies are the necessary conditions for establishment of constant and hypertheme patterns.

None of the thematic progression patterns discussed above seem to explain the way in which the text in Example 7 is organised (since the rhemes practically consist of verbal groups only, no clear thematic progression pattern emerges). The whole construction pattern with which this text opens seems to follow principles quite different from those suggested to form thematic progression patterns. The question arises as to whether the nature of this text production principle can be explained textlinguistically and whether textlinguistic tools can be developed for tuning it into a more native-like text.

In the following section, section 3.3., a further kind of explicative tool, a 'method of development' (Fries 1981), will be used to capture these less English-like patterns in texts. Through this notion an attempt will also be made to capture some further differences of cultural, orientation in academic texts written by Finns.

3. 3. Method of development

In the two previous subsections on the discussion on theme, it has been pointed out that neither a mere clausal view of theme-rheme structures in

texts nor the thematic progression patterns found operating in English discourse, seem sufficient to explain some of the features in the academic texts which Finnish writers produce in English. Both are fairly inoperative mainly for the same reason: we cannot look at text merely by considering one clause, or even one cohesive thematic pattern, at the time. It seems that a more global explanation, one related to situational and cultural differences, is necessary. A possible explanation has been introduced by Fries (1981:20) who suggests that "the information contained within the themes of all the sentences of a paragraph creates the **method of development of that paragraph.**" [my emphasis]. He continues: "If the themes of most of the sentences of a paragraph refer to one semantic field, then that semantic field will be perceived as the method of development of the paragraph." (Fries 1981:20). Some examples of choices of method of development are patterns of contrast, features of some object, locations, etc. Perhaps here it may be useful to take one of Fries' own examples to illustrate the notion in Example 8 (the themes which Fries considers as realizing the method of development have been set in bold):

- (8) (1) **The room they entered** took up most of the cottage, and served for dining room, parlour and kitchen combined.
 (2) **Two roughly made sofas or bunks with feather cushions and some three-legged stools, with a table in the centre,** comprised most of the furniture.
 (3) **There was, however, a large cupboard** occupying the recess between the fireplace and the end wall, which contained many of Auntie's household goods.
 (4) **Iron or tin utensils** hung from numberless pegs, or were ranged on shelves.
 (5) **Of earthenware** there was very little.
 (6) **Everything** was scrupulously clean. (Fries 1981:13)

The method of development for this description is the "existence of the various items" in the room mentioned, not the room itself, nor the location of the utensils. Most of the themes capture the semantic field of 'kitchen utensils'. It is this semantic field aspect that makes 'method of development' different from the Danešian notion of a hypertheme. For example, the 'kitchen utensils' cannot be related to 'the room' in the kind of a part/whole (meronymy) relation that is implied in the hypertheme development. There is no common referent to be referred to in the themes of Example 8. This example is, in a way, very similar to the technical abstract example already

discussed in Examples 4 and 7. Let us now consider the whole of the abstract in Example 9. (The thematic section is in bold and the topical themes are underlined.)

(9)

Electronic Images

- (1) **This report** is the state-of-the-art study of electronic images and image database systems.
- (2) **It** is based on the international literature.
- (3) **First of all, the basic concepts of electronic images** are given.
- (4) **The input devices (scanners and cameras) for different types of images** are reviewed.
- (5) **Image storage, especially optical disks,** is studied.
- (6) **Image computing including digital image processing, computer graphics, image pattern recognition and computer vision, visualization, and virtual realities** are shortly covered.
- (7) **Image retrieval methods together with hypertext and hypermedia systems** are presented.
- (8) **Finally, some examples of image database systems** are given.
- (9) **In Appendixes** online databases referring to audiovisual materials, meteorological online databases containing radar and satellite images, online databases of crimes and criminals containing pictorial information, some examples of databases with statistics, tables, and charts as well as online databases in chemistry and in physics containing graphical information are listed.
- (10) **Furthermore, a list of CD-ROM databases containing graphics or/and images** is given together with some pictures showing examples of image database systems. (Heimbürger 1990:3)

In the discussion of the thematic choices made in this example earlier, it is noted that as the text unfolds, the writer seems to have developed none of the thematic progression patterns. One could suggest a hypertheme pattern, but it seems that the Danešian hypertheme pattern is based on a clear meronymy relation (a part/whole relation which is lexically taxonomic, compositive and constitutive, see Ventola 1987:132), for example a machine and its parts functioning as themes. The relations between the themes in clauses (3)-(7) are more general in nature. *The basic concepts of electronic images* is followed by items which are not in meronymy relation to *concepts* in (3), but rather the lexical relation is hyponymic in nature (a lexical relation which is taxonomic, superordinating and inclusive, see Ventola 1987:132). Thus *The input devices (scanners and cameras)* of (4), *Image storage* of (5), *Image computing* of (6), and *Image retrieval methods* of (7) are related to *the basic concepts of electronic images* of (3) hyper-

nymically, rather than seen to form a hypertheme pattern in the text. Even though we would interpret the pattern which the writer develops in clauses (3)-(7) as a hypertheme pattern, from clause (8) onwards, it becomes harder to relate the topical themes comprising of *some examples* (8), *In appendixes*, (9) and *a list* (10) in hypernymic fashion to *the basic concepts of electronic images* of (3). There seems to be an abrupt change in the pattern the writer develops.

Taking into account these reservations, it is hard to see any linear constant or any clear hypertheme development in this text. Yet the writer has attempted to develop a pattern for the text in Example 9 - a pattern which is very similar to the 'method of development' in Example 8. As has been said, the text in Example 9 is an abstract - an abstract to a monograph. When we analyze the themes in the example from the point of view of 'method of development', it appears that the writer has taken 'the contents of the monograph' as the method of development. The topical themes in most clauses capture the contents of the chapters in the monograph. If one compares the thematic organisation of clauses in the abstract to the table of contents in the monograph, this organisation becomes very clear. For instance, the theme of clause (6) *Image computing including digital image processing, computer graphics, image pattern recognition and computer vision, visualization, and virtual realities* is in the table of contents listed as:

5 IMAGE COMPUTING

- 5.1. *Digital image processing*
- 5.2. *Computer graphics*
- 5.3. *Image pattern recognition and computer vision*
- 5.4. *Visualization*
- 5.5. *Virtual realities* (Heimbürger 1990:5)

So the method of the development for the abstract up to clause (8) seems to simply list the contents of major chapters in the monograph. Why the text seems awkward from the point of view of an English reader is not just because of the use of the passive voice in the clauses, but because the whole 'method of development' is different. The writer has taken each

chapter and its contents as her method of development for writing the abstract, but in the rhemes the readers are told nothing about the items selected for method of development in the themes. Furthermore, the writer is not even consistent in developing this method of development. The reader begins to get a different kind of orientation to the 'semantic field' in the method of development in the clause (8), where *some examples of image database systems* is selected as the theme. The reader would have expected something like *Finally, image database systems are illustrated by some examples*, if the previously selected method of development had been followed. In the table of contents the equivalent chapter is coded as:

9 EXAMPLES OF IMAGE DATABASE SYSTEMS

(Heimbürger 1990:6)

Even a more radical change occurs in (9) of Example 9 when *In appendixes* appears as a topical theme, whereas the reader would have expected the selection of *online databases referring to audiovisual materials, meteorological online databases containing radar and satellite images, online databases of crimes and criminals containing pictorial information, some examples of databases with statistics, tables, and charts as well as online databases in chemistry and in physics containing graphical information* as the topical theme for the particular method of development selected previously by the writer. In the table of contents of the monograph the same informational content appears as:

- | | |
|-------------|--|
| Appendix 1. | <i>Online databases referring to audiovisual materials</i> |
| Appendix 2. | <i>Meteorological online databases containing radar and satellite images, and databases of crimes and criminals containing pictorial information</i> |
| Appendix 3. | <i>Some chemical and physical online databases supporting graphical input and display of data</i> |
| Appendix 4. | <i>Some online databases of economics and business with statistical data, tables, charts etc.</i> |
| Appendix 5. | <i>CD-ROM databases containing graphics and images</i> |
| Appendix 6. | <i>Examples of image database systems</i> |

(Heimbürger 1990:6)

In the abstract (Example 9) the writer has conflated the information in Appendixes (1) - (4) into the subject of clause (9), which is not, however,

the topical theme of the clause and thus not part of the method of development selected for the text. In clause (10), the topical theme is *a list of CD-ROM databases containing graphics orland images*, which captures the information in Appendix (5). The information of Appendix (6) has been coded in the topical theme in clause (8). The comparison between the abstract (Example 9) and the Appendix (above) seems to indicate that clause (9) breaks the method of development and that clause (8) is actually misplaced and should in fact be placed at the end of the text in the abstract.

It appears that the writer is not very consistent in her selection of method of development for the text. As the text begins, the focus is on 'the report', that is, the monograph. The constant theme and the method of development is *This report*. Then, the constant theme pattern breaks up and the method of development is 'the contents of chapters in the monograph'. However, at the end of the text the reader's attention is directed again to the report, but not to the whole of the report, but to its 'parts', *examples, appendices, a list*, which are in turn taken as the method of the development. It is obvious that the Finnish writer has attempted 'the report and its component chapters' as her method of development. One can almost perceive the structure of the monograph chapters and their contents in the organisation of the abstract, followed by the appendices and a list of CD-ROM databases. But the writer has not managed to portray this method of development linguistically in the most efficient manner. She not only confuses the reader by using passive clauses and coding the theme-rheme structures in the individual clauses in an unbalanced way, but also by shifting the method of development from 'the report' to 'the content of the major chapters' and further to 'the examples and appendices'.

In addition, it appears that the method of development is partly obscured by the use of such internal clause initial conjunctions as *First of all, Finally, and Furthermore* (for a discussion on internal conjunctions, see Halliday & Hasan 1976; Halliday 1985; Ventola 1987). The analysis of the text, when it is carried out from the point of view of its method of development, seems to indicate that the writer has used the internal conjunctions inappropriately in the text. The conjunctions *First of all, Finally, and Furthermore* corre-

spond closely to the textual information given *In Appendixes* of the clause (9). It seems justified to claim that the writer actually intended these conjunctions as topical themes equivalent to *the first part of the report, the final part of the report, the latter part of the appendix* or some such reader-orientation expressions. These codings correspond more closely to the method of development 'the report and its component chapters' which the writer seems to have attempted in the abstract. When Example 9 is perceived from this point of view, it will be seen similar to native texts belonging the same or agnate genre, for example those presented in Examples 10 and 11 (not written out in full).

- (10) What this book contains
Macwork MacPlay has four parts. **The first part** contains two chapters that introduce the Mac novice to the parts of the machine: the screen... (the rest of the paragraph).
The second part of the book describes three projects in complete detail... (the rest of the paragraph).
The third part of the book contains 24 additional projects... (the rest of the paragraph).
The fourth part of the book contains reference material ... (the rest of the paragraph). (Pool 1984:x-xi)
- (11) ... Let us take a brief tour through the book. **Chapter 1, "What is Think-Tank?"**, introduces the program, ... (the rest of the paragraph).
Chapter 2, "Getting Started", is designed to get the package up and running.... (the rest of the paragraph).
In Chapter 3, "The Grand tour", you are guided through all of Think-Tank's major functions.... (the rest of the paragraph) (Kamin 1986:xviii)

Functionally, Examples 9, 10, and 11 are, if not of the same genre, at least very closely related agnate genres, Example 9 being an abstract to a monograph and Examples 10 and 11 being prefaces to a book, but all having the function of informing the reader of the contents of the particular document. From the point of view of 'method of development' they are constructed by following totally different kinds of principles. In the clauses of Examples 10 and 11, after the initial theme, the themes capture the parts of the book as the method of development, and what will be said about the contents of chapters in that book, that is, new information, is portrayed in the rhemes. As pointed out above, Example 9 is not organised according to the same method of development. In the clauses of Example 9, the actual new infor-

mation (the contents of the chapters in the monograph) is, for the most part, taken as the method of development in the themes. This procedure leaves the rhemes practically empty. In Examples 10 and 11, the clearly recognizable thematic progression pattern is a hypertheme pattern - the book and its parts - and the contents of the parts are described in the rhemes. The development could be described as advancement 'from global contents to local contents'. In Examples 10 and 11, the method of development seems to be the parts of the book, and the contents of those parts are presented in the rhemes. In Example 9, the major method of development seems to be the content of the chapters in the monograph, but then nothing is said about this content, except that it 'will be discussed', and so on. The text is minimally informative.

The semantic orientation in the two texts, as far as method of development is concerned, is somewhat different. Both native texts start from the 'whole' and their component parts the books - which are clearly captured in the hypertheme progression as well as in the semantic field of 'book'. The Finnish writer, although she begins with a global view to the monograph, very soon runs into difficulties in creating her text, as indicated in Section 3.2; the writer sways between various thematic progression patterns and keeps shifting her method of development, and the consequence is that none of the patterns nor the method of development function well in the text. In the initial part of the abstract, the method of development created by the writer captures the semantic field of the content in the monograph, but towards the end the method of development changes to one similar to that of the English texts, that is, of capturing the component parts of the monograph in the themes (examples, appendixes, a list). In order to make the Finn's text more like a native text one would have to transfer all the monograph content information in the themes to the rhematic parts of the clauses and develop the text systematically from the global structure to its parts.

Why the Finnish writer has selected this method of development (but does not continue it) is somewhat more difficult to explain. Are the English and Finnish methods of developments in this text type different? We have seen

in the Finnish Example 5 (where 'what has been studied in the experiment' has been taken as the method of development in themes) that in Finnish a method of development can be realised similarly to that realised in Example 9. (The question as to whether patterns of method of development of this sort can be found more generally in Finnish texts needs to be clarified in further studies.) The Finnish writer has obviously not learnt (or has not even been taught) the typical English way of realizing methods of development in the kind of genre discussed, and she simply uses a different method of development for producing English texts.

Certainly differences such as these have to be explored further and to be explained somehow in terms of cultural rhetorics. That cultural preferences for different rhetorical patterns exist is a well-established fact. For example, Eggington (1987) reports that Korean students writing academic essays follow a different text construction pattern from their American counterparts. The American way is to develop an argument linearly, advancing from the general to the specific. The Korean way is first to introduce an argument and develop it; but then the direction changes abruptly and a sub- or an unrelated argument is introduced, a return back to the main argument not appearing until the last section. Korean and American writers are socialised into two different rhetorical patterns and the Korean readers are reported as being able to understand the Korean 'sidetracking' way of argumentation better than the American linear way, although the Korean way seems totally illogical and incoherent to an American reader.

Finnish rhetorics as a whole may not necessarily be very different from the Anglo-American rhetorics. But certainly more research is necessary in such linguistic areas as method of development and cultural rhetorics. In conducting such studies, it is extremely important to link the investigations with the study of the relationship between the text type and the situation in which it has been produced as well as with the study of culture. For instance, it is well known that, culturally, Finns prefer implicit coding of information (see e.g. Widen 1985, 1988). Consequently, the writer in Example 9 may have perceived the whole writing situation and the coding of this type of text from a totally different point of view than an English

writer. For instance, Example 9 is intended to appear in the monograph, not separately from the main text, and this may have had its consequence to the construction of text. If one tried to follow the implicit logic that the Finnish writer might have used when writing, it would involve an understanding of certain situational and cultural principles of information retrieval between the writer and the reader. The Finnish writer codes 'the book and its contents' in method of development as given and known, because the reader actually holds onto the monograph at the same time as reading the abstract and because the writer can expect the reader who presumably comes from culturally the same academic background to understand the abstract. This kind of 'assuming the shared context' with the peer reader would at least partly explain the somewhat awkward thematic coding of clauses in Example 9⁴. In other words, a partial explanation suggested is that Finnish writers when reporting their research might make bolder assumptions of the sharedness of information and the familiarity and closeness of the scientific community, whereas English writers have been trained to approach the task of reporting research results from the point of view of presenting something totally new to the reader.

4. Interplay between reference and theme

Since both the reference choices and the various thematic choices discussed create cohesion in texts, to say that these discourse systems at least partly overlap from a functional point of view comes as no surprise. As Halliday & Hasan (1976:296) point out, cohesion in texts is a matter of 'more or less', but frequently we find sections in texts where the meanings are strongly interrelated and consequently the cohesive texture of the text at those points is tightened. Here it will be suggested that in native English academic writing one of the means by which such tightening takes place, is the interplay of the systems of reference and theme. The smoother the interplay, the easier the text is to read. Halliday & Hasan (1976:310-312) point out that this interplay between reference and theme plays an impor-

⁴ It would also explain the use of an awkward presuming item in clause (2) in Example 9, First of all, the basic concepts of electronic images are given, where an English writer would most likely have used a presenting reference, e.g. some/a few concepts.

tant role in the unravelling of ambiguities in texts. They ask, for example, how, in Examples like 12, the readers know who beat or caught or eluded whom, or who are staying with the grandparents?

- (12) (1) **Spurs (Actor) played Liverpool. They (Actor) beat them (Goal).**
 (2) **The cops (Actor) chased the robbers. They (Actor) caught them (Goal).**
 (3) **The cops (Actor) chased the robbers. They (Actor) eluded them (Goal).**
 (4) **These ponies (Theme) the children (Subject) were given by their grandparents (Actor). They're (Actor/Subject/Theme) staying here, now.**
 (5) **These ponies the children were given by their grandparents. Have you seen them? (from Halliday & Hasan 1976:310-312)**

Halliday & Hasan suggest that the interpretation of the referential relations can take place through grammatical criteria, the assumption being that in subsequent clauses Actor/Subject and Goal/Complement roles match. Thus, in (1) and (2) *they* would refer to *Spurs* and *the cops* and *them* would refer to *Liverpool* and *the robbers*. But in (3) the only interpretation is that *they* refers to *the robbers*. The lexical meaning of *eluded* enables the disambiguation in (3). Halliday & Hasan further note that, when lexical meaning does not resolve the interpretation of reference, then Theme is a more likely candidate for clearing the ambiguity than the transitivity roles of Actor and Goal, or the mood functions of Subject and Complement. Thus, in (4) the preferred interpretation of *they* is *these ponies*, because *these ponies* also functions as the Theme of the clause, although generally English prefers human Subjects. It seems common then that, if other means of interpretation fail, the ambiguous reference relation is usually interpreted as referring to the head of the nominal group functioning as the Theme in the preceding text. It also seems to be the case that, when the ambiguous reference item itself is not in the thematic position (e.g. *them* in (5)), it is still more likely to be linked with the thematic element (i.e. *them* is linked with *these ponies* rather than with *the children*).

If one considers what has been said in the previous section about reference and matters concerning theme, one can predict problems for non-native writers. It has been shown that Finnish writers have difficulties in coding referential relations correctly and that they do not necessarily seem to follow the same lines of thematic choices as their English-speaking colleagues. It appears that in their texts they do not maximize the inter-

play between the systems of reference and theme for the sake of cohesion. The discussion on the interplay problems which Finnish writers encounter can be initiated by considering the question: what does this presumption of thematic and referential interplay mean to writers in general? Writers must disambiguate referential problems for readers in advance, not leave the disambiguation to the reader to do. This is the only way in which a writer can ensure that references will be interpreted in the way the writer intended. Writers should bear in mind the principle that the themes in texts are common sources of interpretation for referential relations. Naturally non-native writers are prone to make more mistakes in this area, as will be illustrated in Example 13 in which certain extracts that have already appeared in Example 3 are analysed.

- (13) GROUP A. (1) **The girls in this group** hoped more often than others did to obtain a job which had a high social status; (2) **their parents** also tended to have a higher social status than those of other girls. (3) **They** all lived with both of their parents, and (4) **the family** often spent weekends together; (5) **the girls** also tended to have very good relations with their fathers.

The constant thematic progression pattern and the method of development intended seems to be 'the Group A girls'. The constant theme pattern requires that the reference items should refer to the same referent, and the method of development should also necessitate the appearance of 'the Group A girls' as head of the thematic nominal groups. However, the Finnish writer has problems in establishing these thematic patterns in the text so that they are coded referentially appropriately. 'The girls in Group A' in clause (2) are coded only by the use of a possessive reference *their*. Since the head of the nominal group in the Subject/Theme position is *parents*, the reader's focus shifts to 'the parents', who should not be involved in the constant thematic progression pattern or in the method of development. When the reader reads *they* in (3), there is a danger that he/she might process *they* linearly as referring to the theme of the previous clause, that is, to *parents*. Furthermore, when he/she reaches *both of their parents* in (3), the reader will realize that the assumption that *they* referred to *parents* is not correct and will have to make a mental backtrack, making a referential link from *they* in (3) to *their* in (2) and even further back to *the girls in this group* appearing in (1). The theme of (4), *the family*, naturally includes

'the girls', but the appearance of this lexical item in the thematic position breaks up the constant thematic progression pattern which the writer seems to cultivate in the rest of the section. Minor linguistic changes would improve the extract considerably from the point of view of the constant thematic progression and the method of development.

A similar problem occurs later on in Example 3, when the writer discusses Group D, as reproduced in Example 14:

- (14) Group D. (1) **The parents of girls in Group D** had on average a lower socio-economic status than those of the other girls. (2) **Their homes** were more often "broken" (3) **and their families** spent comparatively seldom their free time together (though in the opinion of the girls, often enough).

The heads of the thematic nominal groups indicate a shift from *parents* to *homes* to *families*. It is obvious that 'the Group D girls' is also intended as the method of development here - yet, the Finnish writer reduces the power of that thematic patterning by inconsistencies with reference (*girls in Group D* -> *the girls in Group D*) and by not making 'the girls' the heads of the thematic nominal groups (*their, their*). When readers begin to interpret the *their* in (2) and *their* in (3), especially if they are reading rapidly or carelessly, they are likely to link up these references with the head of the nominal group functioning as the theme in (1), namely *parents*. Thus, the train of interpretative thought for referential relations becomes: *the parents* <- *their homes* <- *their families*, which naturally is not at all the writer's intention. The head noun in the theme of (1), *parents*, is not the appropriate source of interpretation coded in the referential chain of *their - their*, since the whole text, as already discussed, is about four different groups of girls. Here the main focus should be on the girls, 'the girls in Group D'. This the Finnish writer has been unable to portray, since he does not maximize the interplay between reference, constant thematic progression pattern, and the method of development. *The girls*, not *the parents*, should begin the constant theme pattern and appear as the head of the thematic group. 'The girls in Group D' do appear in the thematic part of clause (1) in the qualifier part of *girls in Group D*, but this position does not maximize the constant thematic progression and the method of development which the writer is attempting to establish in the rest of the section. In other words, the Finnish

writer confuses his English readers by making *the girls* qualify an informationally less important item, *the parents*. Furthermore, not only the 'wrong' position of the girls, but also the erroneous reference - 'general' instead of 'specific' - causes interpretation problems for the reader. To maximize the interplay between reference and theme, the text could be improved simply by moving *the girls in Group D* to the head position of the element which realizes the theme and opens the discussion in the paragraph and reformulates the other references.

These referential and thematic changes would pattern the discussion of Group D in a similar way to Groups A and B, in which either the group as a whole or the girls of each group become the head of the first thematic group which begins the paragraph. The paragraphs about Groups A, B, and D would be more consistently developed referentially and thematically. However, the paragraph dealing with Group C still remains problematic. The problem is basically the same as that in the Group D paragraph. The themes in the first two clauses are: *The reason given for smoking by girls in this group* and *the girls*. Again, naturally, informationally the most important item, 'the girls', should occupy the head position of the thematic nominal group which, as in the other paragraphs dealing with the groups of girls, initiates the whole paragraph. So changing the beginning of the paragraph to something like *The girls in this group usually gave social reasons...* would improve the internal cohesion of the paragraph by creating a constant thematic pattern at the beginning of the paragraph. Further consistent references to this group would also improve the cohesiveness of the rest of the paragraph. This change would bring the Group C paragraph in line with the beginnings of the other paragraphs, now each having 'the girls' as the head of the first thematic nominal group, and the further clauses would include appropriate referential links to these heads, thus creating an expectation of 'the girls' as the constant theme pattern and the method of development for each paragraph.

Not only would the internal cohesion of each individual paragraph improve, but the changes would also increase cohesion across the four

paragraphs by creating a 'macromethod of development' pattern under the title of "4. Characteristics of the four groups.". What now stands as

Group A. The girls in this group hoped more often than others...

Group B. This group, the most numerous one (26% of the total...

Group C. The reason given for smoking by girls in this group was...

Group D. The parents of girls in this group D had on average...

could for instance, be coded as

Group A. The girls in this group hoped, more often than the others...

Group B. The group B girls, (the biggest group, 26% of the total...

Group C. The girls in this group usually gave social reasons for...

Group D. The girls in Group D were born of parents who, on average, had a...

Thus, the macromethod of development for the whole of the Section 4 would also be 'the girls', but not referentially identical girls, but girls belonging to different groups.

The suggested linguistic referential and thematic changes would maximize the referential and thematic patterning intentions of the Finnish writer and clarify these to native English readers. Organizing the text around the method of development of 'the girls', as individual groups in separate paragraphs and as a whole group for the whole section, is just one organisational option that is possible. Had the writer's purpose been different, he might have selected a totally different method of development. He could, for example, have alternated the group C girls and the other girls in the thematic part, thus creating a clear contrastive method of development pattern in the form of *Group C was.... - The other groups were ...*, or the variables studied could have been taken as the method of development: *Smoking seemed to have social causes in Group C, whereas drinking was indulged in by the Group C girls for relaxation. Sexual experiences were valued by Group C more than by the girls in the other groups....*

Writers have obviously many ways of organizing the content of their texts, depending on the choices they make as to thematic progressions and methods of development. As has already been emphasised, from the point of view of cohesion, the systems of reference and theme seem in coherent texts to work harmoniously together. But when non-native writers create

text, the difficulty of conveying meanings in a foreign language demands so much of the non-native writer's attention and energy on the lexicogrammatical level that matters of cohesion may easily be neglected (see Ventola & Mauranen 1991). And even when one of the cohesive systems seems to work in a non-native text and the intended cohesive pattern has been achieved, even so the interplay between the systems is often forgotten. This appears to be the case in Example 9.

The Finnish writer of the abstract in Example 9 has clearly aimed at a method of development pattern that will hold the text together, beginning with the introduction of 'content' of each of the chapters in the monograph where the abstract appears. However, as already indicated, from the English reader's point of view, the writer has chosen a very unusual pattern which makes the text non-native like, presenting in thematic positions what English readers would consider to be new, non-retrievable information in the text and context and leaving the rhemes empty. In Example 9 there is also very little interplay between the systems of reference and theme. The referential chains are minimal: the anaphoric relationship between *this report* and *it* in the initial clauses seems to be the only relation of that kind in the text. Most of the participant information in the text is coded with presenting reference, and no endophoric (neither anaphoric nor cataphoric) references are made to the participants thus introduced into the text. Furthermore, none of the elements in either the themes or the rhemes seem to be tracked down referentially. It is therefore not surprising that such a text will appear to be incoherent to an English reader. The Finnish writer of Example 9, although in many ways extremely competent in English, does not seem to be able to maximize the interplay between the systems of reference and theme (the text can be contrasted with Example 11 where referential, thematic, and lexical cohesion all contribute to the cohesion of the text).

5. An applied concern

The focus in this paper has been on Finnish writers writing in English. It seems appropriate to make some final observations about training in writing provided for these groups of writers (for a more detailed discussion, see Ventola & Mauranen 1991; Ventola 1991). Both English and Finnish academic writers naturally receive some training in writing in their native languages at some time during their education. However, the nature of the training may be 'marked' culturally. A quick look at the writing manuals published in Finland gives one an impression that the teachers of writing in Finland are more occupied with lexicogrammatical matters (making sure that the matters of form and punctuation are grammatically correct) than are their English colleagues in equivalent works. Even the fairly recently published writing manuals, which are otherwise very advanced and useful books for researchers, seem to neglect the cohesion side of text creation (see e.g. Hirsijärvi et al. 1986); many (but not, of course, all) English language writing manuals include, in addition to such matters as the structure and advancement of argumentation, some discussion on cohesion, although often cohesion matters are not dealt with very effectively from the point of view of textlinguistics. Ellis and Hopkins (1985), for example, give writers a useful example which captures the kind of referential and thematic cohesion problem on which also this paper has focused. Their example is reproduced as Example 15 below.

- (15) A period of probation on the city streets is a wonderful education for any young police recruit. The city is full of tests of nerve and resolve. You cannot imagine how difficult it is to remember all that you were told about the law when you are suddenly presented with an accident on the street. Accidents will suddenly happen. Criminals will not wait until you have consulted your police handbook. Previous to recent reform, police on probation were expected to learn as they went along. Now they will have to act on their own initiative. Police work is becoming more and more difficult, and the need is great for talented and intelligent police recruits. (Ellis & Hopkins 1985:85; their underlining).

The discussion which follows Example 15 points out a few of the cohesive problems that readers should encounter:

...lacks coherence. The underlined topic sentence sets a course which the subsequent sentences do not follow. The subjects of the sentences change from

'period of probation' to 'the city' to 'you' to 'accidents' to 'criminals' to 'police work'. The tenses change, too, (Ellis & Hopkins 1985:85)

Then a rewritten version is introduced to the readers to help them out in their writing problems; here Example 16.

- (16) A period of probation on the city streets is a wonderful education for young police recruits. They will be fully tested by the demands of law enforcement in the city and by having to apply lectures on law to the handling of accidents in the streets. Nor will recruits have time to consult their police handbooks while criminals obligingly wait. They will not have time to learn so much on the job either. They will be expected to act on their own initiative. The demand is for talented and intelligent recruits to measure up to the demands of a police force facing more and more difficult challenges. (Ellis & Hopkins 1985:86; their underlining).

Finally, an explanatory commentary is given to the learner of the revised version:

"the sentences do not jump from idea to idea ... The idea in the topic sentence, referring to the education of young police recruits, is developed consistently, and the tenses are more consistent, too." (Ellis & Hopkins 1985:86)

In the light of the matters presented in this article, the explanation given to the learner is not very explicit. The learner is not explained how the increased consistency is achieved in the text. Ellis and Hopkins have, to a certain extent, built up a constant theme progression pattern in the rewritten version and also used referential chains reasonably efficiently. The changes naturally increase the internal cohesion of the text. But the question as to whether the result is a good text, and to what degree is it better than the first version, will remain open. Reader-studies or psycholinguistic research would have to determine this, but so far there seems to be relatively little theoretical or applied information available on questions such as these.

Although one can criticise the examples of cohesion which appear in writing instruction books published in Anglo-American context, these instruction books, seem, to a certain degree, to draw the attention of their readers to problems of cohesion, although clearly discussions in such writing manuals would benefit from a thorough textlinguistic approach, such as explicated, for example, by reference and thematic analyses in this paper. In the Finnish context there is an urgent need for writing manuals which

concentrate on cohesion and other writing problems specific to Finnish writers when writing in English. So far no such contrastively oriented writing manual exists, and research in this area has only fairly recently begun. (Ventola & Mauranen 1990 and Ventola & Mauranen 1991 are the first publications to give a comprehensive account of the problems Finns have when writing scientific English.) But every day many Finnish scholars have to struggle with writing in English, since practically all their academic writing is in that language. For them, this is a difficult task since they have had no training for it, no instruction as to how, for instance, to manipulate efficiently cohesive principles such as reference and thematic patterning in texts written in English. Yet they are expected to publish in English internationally.

To summarize, it seems obvious that Finnish writers who have had no linguistic training are not always aware of the importance of interplay between such cohesive systems as reference and thematic patterning. Naturally, if training were given in writing in English, this would help Finnish writers to help themselves and to improve the cohesiveness of their texts. But, before linguists can help the writers, more textlinguistic information on interplay between the various cohesive systems is needed. For example, at present, we seem to know, comparatively, a great deal about the choices that the writer can make in such matters as, reference, theme, and the interplay of these two systems, but we know considerably less about the patterns which are effective from the readers point of view. We do not know, for instance, when a certain thematic pattern or a method of development which is also successful in its reference codings of relevant participants becomes too repetitive and monotonous. Here reading studies and psycholinguistic research may perhaps offer some help. A strong textlinguistic and interdisciplinary orientation would be invaluable in all the applied work intended to help writers, both native and non-native, in developing their writing skills.

6. Conclusion

This paper has approached the question of cohesion in the articles of non-native writers. The paper has discussed the problems Finnish writers encounter when reporting their work in English. In this paper it has been possible to analyse only a few text extracts as representative examples of the kind of problems typically occurring in such academic texts in English. What has been achieved in the paper is summarised below.

The paper first focused on the system of reference and the problems which Finnish writers face when they have to code text participant information in English. It pointed out that in this area many problems occur, the majority of which are caused by the textual coding differences between the two languages, one being the absence of the article system in Finnish as a means for coding participant information in terms of presenting and presuming reference. The paper also dealt with the construction of texts in terms of theme-rheme structures of clauses, of thematic progression patterns, of methods of development, as well as with the interplay of the two systems of reference and theme and the effect of this interplay in texts. No clear reason for the difficulties encountered by the Finnish writers were detected in the language differences, but it became obvious that the Finnish writers tended not to maximize the potential offered by the cohesive systems for the benefit of their text organisation. The reference and thematic problems which typically occur in the texts of Finnish writers must not only be attributed to the linguistic differences between the two languages, but also to contextual and cultural factors, but these factors are in turn very hard to study and verify, and their effects are often extremely difficult to remedy in non-native writing, even through conscious teaching.

It is hoped that studies such as the one exemplified here will provide some immediate help for non-native writers struggling with creating meanings in a foreign code. Here it has been possible to discuss only one relatively small area that has proved to be problematic. Further help can be provided when textlinguists conduct more explorations into the relationships which exist between non-native writing problems and the various text conven-

tions, specifically generic qualities in texts, and situational and cultural factors, all of which play a part in creating meanings in a foreign code.

References

- Barnickel, K.-L. 1982. *Sprachliche Varianten des englischen. Register und Stile*. (Hueber Hochschulreihe 45/II). München: Max Hueber.
- Connor, U. & R. Kaplan (eds.) 1987. *Writing Across Languages: Analysis of L2 Texts*. Reading: Addison-Wesley.
- Daneš, F. (ed.) 1974. *Papers on Functional Sentence Perspective* (Janua Linguarum, Ser. Minor, 147). The Hague: Mouton.
- Daneš, F. 1974. Functional sentence perspective and the organization of the text. In Daneš, F. (ed.), 106-128.
- Eggington, W. G. 1987. Written academic discourse in Korean: implications for effective communication. In Connor, U. & R. Kaplan (eds.), 153-168.
- Ellis, R. & K. Hopkins 1985. *How to Succeed in Written Work and Study*. London & Glasgow: Collins.
- Fries, P. 1981. On the status of theme in English: Arguments from discourse. *Forum Linguisticum* 6:1, 1-38.
- Hakulinen, A. & F. Karlsson 1979. *Nykysuomen lauseoppia* (Modern Finnish Syntax). Helsinki: SKS.
- Halliday, M. A. K. 1985. *An Introduction to Functional Grammar*. London: Arnold.
- Halliday, M. A. K. & R. Hasan 1976. *Cohesion in English*. London: Longman.
- Hirsjärvi, S. & P. Liikanen & P. Remes & P. Sajavaara 1986. *Tutkimus ja sen raportointi* (Research and Reporting It). Helsinki: Kirjayhtymä.
- Lindeberg, A.-C. & N. E. Enkvist & K. Wikberg (eds.) 1992. *Nordic Research on Text and Discourse. NORDTEXT Symposium 1990*. Åbo: Åbo Akademi Press.
- Martin, J. R. 1983. Participant identification in English, Tagalog and Kâte. *Australian Journal of Linguistics* 3:1, 45-74.
- Martin, J. R. 1985. *Factual writing: exploring and challenging social reality*. Geelong, Vic.: Deakin University Press. (Republished by Oxford University Press, 1989.)
- Mauranen, A. 1992. Reference in academic rhetoric: a contrastive study of Finnish and English writing. In Lindeberg, A.-C. et al. (eds.), 237-250.

- Mauranen, A. forthcoming. Cultural Differences in Academic Rhetoric. A Textlinguistic Study. A Ph.D. thesis. Department of English, The University of Birmingham.
- Ventola, E. 1987. *The Structure of Service Encounters*. London: Pinter.
- Ventola, E. 1991. Research on English academic writing in a non-English context: implications to teaching. A paper presented at the International Conference on Teacher Education in Second Language Teaching, Hong Kong, 17.4.1991.
- Ventola, E. 1992. Text and reference. In Lindeberg, A.-C. et al. (eds.), 223-235.
- Ventola, E. (ed.) 1991. *Functional and Systemic Linguistics: Approaches and Uses*. Berlin: Mouton.
- Ventola, E. & A. Mauranen 1990. *Tutkijat ja englanniksi kirjoittaminen* (Researchers and Writing in English). Helsinki: Helsinki University Press.
- Ventola, E. & A. Mauranen 1991. Non-native writing and native revising of scientific articles. In Ventola, E. (ed.) 1991, 457-492.
- Widén, P. 1985. Interkulturelle Verständigung am Beispiel der finnisch-deutschen Kommunikation, *Lebende Sprachen* 4, 167-170.
- Widén, P. 1988. Suomalainen kommunikointi eurooppalaisen kulttuurin silmin (Finnish communication viewed from the European cultural perspective), Kielikoulutuksen ideapäivä. Vientikoulutussäätiö. 26.10.1988. (Unpublished manuscript.)

Secondary references:

- Heimbürger, A. 1990. *Electronic Images* (Research Notes 1983). Helsinki: Technical Research Centre of Finland.
- Kamin, J. 1986. *Mastering Think Tank on the 512 Macintosh*. Berkeley: SYBEX.
- Knif, J. 1989. Karttakielen esiintyminen kajaanilaisilla terveystieteiden opiskelijoilla, *Suomen hammaslääkärilehti* 36, 12: 696-700.
- Pool, L. 1984. *MacWork MacPlay*. Washington: Microsoft Press.
- Vrahoupoulos, T. P., P. Barber, H. Liakoni & H. N. Newman 1988. Ultrastructure of the periodontal lesion in a case of Papillon-Lefèvre syndrome (PLS). *J Clin Periodontol* 15: 17-26.