CONTENTS

LETTER FROM THE EDITORS

Dear Readers,

After a pronounced diachronic focus in our latest VIEWS issue, this time synchronic issues are back – 'with a vengeance'. They address conceptual questions of delimiting and defining well-known notions such as formulaic sequences and parenthetical clauses as well as the more applied question of vocabulary teaching and learning.

Julia Hüttner explores the potential of linking formulaic sequences with genre analysis. This novel approach offers promising inroads into the

notoriously difficult question of how to pin down and classify formulaic language by linking it to the question of genre-specific functions.

Classifying and delimiting an elusive concept is also Gunther Kaltenböck's agenda, who investigates the class of spoken parenthetical clauses. Although widely used in linguistics, the term parenthetical often lacks a clear definition owing to its borderline status on the 'edge' of syntax. Gunther Kaltenböck systematises this diverse class of disjuncts and highlights its internal stratification.

Angelika Rieder's contribution deals with the difficult issue of linking linguistic research with teaching practice. More specifically, her focus is on vocabulary learning as a by-product of reading. Although widely researched and often taken for granted in foreign language teaching, the actual process of learning vocabulary through reading is by no means a straight-forward one. Most significantly, Angelika Rieder shows that in order to ensure successful vocabulary learning the reading process needs to be accompanied by relevant classroom practices. But if you want to 'learn' more about the topic: 'read' for yourselves...

We hope this issue makes an interesting read for your summer break. Who knows, you might even find the time to respond to one of the articles and give us your own VIEWS.

THE EDITORS

How to contact us:

VIEWz

c/o
Institut für Anglistik & Amerikanistik der Universität Wien
Universitätscampus AAKH, Spitalgasse 2, Hof 8
A – 1090 Vienna; Austria

fax	**(intern.) 43 1 4277 9424**
eMail	**julia.isabel.huettner @univie.ac.at**
W³	http://www.univie.ac.at/Anglistik/views.htm
	(all issues available online)

Formulaic language and genre analysis: the case of student academic papers

Julia Hüttner, Vienna*

1. Introduction

For most language learners, some areas of their target language remain highly problematic even up to advanced levels. University students with considerable experience as language learners often find that sounding appropriate and idiomatic in their foreign language, in addition to showing grammatical and lexical accuracy, is rather difficult.

This notion of learners sometimes 'not sounding quite right' in their foreign language despite a lack of errors has been linked previously to insufficient or incorrect use of formulaic language, i.e. those stretches of conventionalised language apparently so prominent in native speaker discourse. (cf. Pawley & Syder 1983) Also, this language problem of advanced learners has been seen in the connection of learners' difficulties in developing suitable familiarity with the institutionally demanded genres and their conventions. (cf. Johns 2002) Again, some language learners find that despite having attained high levels of lexical and grammatical accuracy, difficulties remain in producing appropriate texts in specific genres.

It is worth drawing attention to the fact that both of these areas of difficulty address aspects of conventionalised language use from macro-level genre structures to micro-level textualizations as instances of formulaic language. Given this apparent connection, it seems timely to investigate this link between various aspects of conventionalised language use further.

This article will present a new approach towards the study of conventionalised language, i.e. extended genre analysis, which combines elements of genre analysis and of research into formulaic language. This new approach also aims at addressing difficulties in previous methodologies of genre analysis following Swales (1990) and Bhatia (1993; 2004). The focus of this article will be on the investigation of formulaic sequences in particular

* Author's email for correspondence: julia.isabel.huettner@univie.ac.at.

genres and the ways in which extended genre analysis can shed more light on these patterns of use.

With regard to the situation of language learners described above, both factors of difficulty, i.e. the need to produce texts that are suitably idiomatic and also appropriate with regard to the genre conventions, combine in students' attempts at producing academic papers in a foreign language. For this reason, the study described later on in this paper will focus on Austrian students' academic papers in English. More specifically, I will address issues connected to students' attempts at finding appropriate and effective ways to support their communicative intentions within academic genres by taking recourse to formulaic sequences. Thus, while the impetus for this investigation was an issue of learning and teaching, what is explored in this article is the necessary first step of providing a more thorough description of the genre that appears to be so problematic to learn.

In order to set the background for a presentation of extended genre analysis, I will review some general questions of formulaic language and genre analysis and will then show the possibilities of this new approach towards the study of conventionalised language.

2. Conceptual background

The two research areas which inform extended genre analysis, i.e. research into formulaic language and genre analysis, have both received increasing interest over the past two decades. (cf. Bhatia 1993, 2004; Nattinger & DeCarrico 1992, Pawley & Syder 1983, Schmitt 2004, Sinclair 1991, Swales 1990, Wray 2002 among others) These areas have generally been considered as separate areas of research, although there are some clear overlaps in the issues they address: both deal with aspects of conventionalised language use, albeit at different levels, so that genre analysis addresses mostly issues of conventionalised structures of texts and their concurrent textualisations, and the study of formulaic language addresses mostly phraseological conventions. However, for a genre text to be considered conventionally appropriate, the expectations of listeners or readers have to be met, including the linguistic realizations of the genre. These can also incorporate specific instances of formulaic language, typical of the genre in question. It is this link which is to be specifically explored in extended genre analysis.

2.1 Formulaic language

In linguistics there is a long history focusing on the generative and creative aspects of language production, but a closer observation of our language patterns shows that a rather large proportion of everyday language does, in fact, consist of language which is used in (near-)identical form in similar situations. Estimates of just how much of the language produced is formulaic vary, ranging from 32.3% according to Foster (2001: 85) to 58.6% according to Erman and Warren (2000: 37). Despite the range of the estimates one can easily say that even the more conservative estimates justify the increased interest in this phenomenon. While linguistic observations on the formulaic nature of language go back as early as 1924 with Jespersen, this initial interest was not followed up systematically and the study of formulaic language has had quite a chequered history of varying degrees of attention from a wide range of (sub)-disciplines. Recently, however, the study of formulaic language has again attracted much attention, especially following on from the descriptive findings of corpus linguistics that clearly show the existence of an *idiom principle* (cf. Sinclair 1991) of favouring particular lexical combinations in language production.

In general, formulaic language can be considered as a counterpart to creative, generative language use. In fact, a wide range of language features can be considered formulaic, from proverbs and sayings, the more traditional idioms, like *Too many cooks spoil the broth* or *to kick the bucket*, to patterns of collocational choice in the idiom principle, based essentially on frequency of co-occurrence. Of great interest in this respect is the information obtainable on the differences in use of near-synonyms. An example would be the co-occurrence of the adjective *small* with *boys* and *little* with *girls*, although semantically there is no real difference between the two adjectives. (Stubbs 1996: 69-70)

This preference in native speakers for specific choices in collocational combinations is also shown through evidence from corpus linguistics and constitutes the type of formulaicity which is so frequent in language production. It is also this type of formulaic language use which is apparently responsibly for the fluency attained by native-speakers. (cf. Pawley & Syder 1983). This draws on a somewhat different notion of formulaicity than the one of proverbs. Here we are concerned with formulaic language or idiomaticity in the sense of sounding native-like or natural. In contrast to the idioms mentioned earlier, these patterns allow for some level of variation and many of the combinations thus established are semantically transparent.

The variety of language phenomena that are in some sense formulaic has led Schmitt and Carter (2004: 2) to observe that

> formulaic sequences seem to exist in so many forms that it is presently difficult to develop a comprehensive definition of the phenomenon. This [..] remains one of the foremost problems in the area.

Several attempts at defining formulaic language have focused on psycholinguistic factors that determine the status of any sequence of language as formulaic, such as whole-sale retrieval and storage. (cf. Wray 1999: 214, Wray 2002: 9, Sinclair 1991) There are, however, several other factors worth taking into consideration when attempting to define formulaic language. For the purposes of this article, the most relevant question relates to the functions fulfilled by these formulaic sequences. Thus, a distinction can be been made between recurring word-combinations that are pragmatically functional, such as *How are you?* as a typical greeting, and those that are not, such as *a large number*. Nattinger and DeCarrico (1992: 37) have indeed limited the definition of formulaic sequences to those that carry specific pragmatic or discourse meaning and distinguish collocations and lexical phrases, the latter being defined as collocations that have been assigned pragmatic function. Many of these functions relate to the importance of formulaic sequences in pragmatic uses, for instance for reiterative communicative tasks, such as greeting, apologizing, thanking or for aspects of discourse management.

While establishing a definition of formulaic sequences proves difficult enough, turning any such definition into operational criteria for identification of formulaic sequences proves even more so. Several possibilities have been formulated, such as frequency of occurrence, fixedness of form, or native speaker intuitions. Unfortunately, some problems remain with all of these operationalisations. While especially in the wake of corpus linguistics, frequency of occurrence has gained much support, it can fail to identify stretches of language as formulaic, which would be formulaic according to other criteria, notably native speaker intuition. Wray (1999: 214) gives the example of the phrase *The King is dead, long live the King* which most native speakers would classify as formulaic, but it does not appear in even the largest language corpora. However, relying solely on native speaker intuition is problematic in many ways. Importantly, native-speakers are familiar only with specific language registers and thus sometimes fail to identify formulaic language of special groups or discourse types they are unfamiliar with. Also,

native speakers seem to use different criteria for establishing what is in their view formulaic.[1]

One way in which extended genre analysis helps address these problems of operationalisation is by focusing only on one specific genre, and of attempting only to identify the formulaic sequences typical of that genre. Taking the example above, assuming it were possible to create a specialised corpus of spoken language recording at the time of death of male monarchs in English-speaking countries or of fictional texts dealing with such situations (quite a task!), the sequence *The King is dead, long live the King* would most probably occur. As the sequence is not likely to be referred to outside of this real or fictional situation, it comes as no surprise that it does not occur in the currently available language corpora.

2.2 Formulaic sequences in learner language

Despite these problems of identifying stretches of formulaic language unambiguously, its importance for language learners has long been noted. Pawley and Syder (1983) first pointed out that learners do not achieve native-like fluency because they do not use sufficient formulaic sequences in their language production. Yet, formulaic sequences have been observed in nearly all types of learner language. The amount of formulaic language observed in second language learners varies: typically more formulaic language use is found in naturalistic learners, especially those that acquired their second language as children, but to some extent also in adult learners. (Wong Fillmore 1979, Schmidt 1983). However, even in these cases there seems to be much inter-learner variation, so that, for instance Bohn (1986) and Hanania and Gradman (1977) report very low uses of formulaic language in their subjects.

In taught language learners, the use and further acquisition of formulaic sequences has been observed as well (cf. De Cock et al 1998; House 1996; Myles et al. 1999; Wiktorsson 2003). Following Wray's model (2002: 205-210) the assumption is strong that in this group of learners the usage of formulaic sequences is least high given the effects of formal instruction which result in learners favouring analytic approaches to language rather than making use of prefabricated blocks.

When looking at all the reports of second language learning of formulaic sequences, we can tentatively find a link between a readier acquisition of

---

[1] For a more extensive overview of methods and problems employed in the detection of formulaic sequences cf. Wray 2002, chapters 2 & 3 passim.

those formulaic sequences that have a high pragmatic relevance and those that have not. Thus, for instance in natural language acquisition, the phrases children use to enter games or to cope with classroom interactions were acquired comparatively soon (cf. Wong Fillmore 1979; Lynnäkyla 1980). Similarly, the formulaic sequences described by House (1996) are of high pragmatic function in that they help learners manage communication more effectively.

In sum, we can say that although formulaic language exists in all language production, including that of learners, the exact distribution of formulaic sequences might well vary between native and non-native speakers. There is some support for the notion that in all groups of learners those formulaic sequences most relevant to their pragmatic uses are acquired best.

2.3 Genre analysis

Genre analysis, the second area of research into conventionalised language constitutive of extended genre analysis, focuses on "situated linguistic behaviour in institutionalised academic or professional settings". (cf. Bhatia 1997: 181) The one followed here is referred to as the ESP (English for Specific Purposes) approach, and is most prominently represented by Swales (1990) and Bhatia (1993, 2004). The name of this approach derives from the focus of this group on contexts of professional and academic discourse, with a conscious exclusion of more general *pre-genres*, such as narratives or casual conversation. While the ESP approach developed from a commitment to teaching, its primary focus lies on a thorough description of professional writing. The precise definition of genre used in this approach is that of Swales (1990: 58):

> A genre is a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style. Communicative purpose is both a privileged criterion and one that operates to keep the scope of a genre as here conceived narrowly focused on comparable rhetorical action. In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience.

Although not clearly defined, an overriding *communicative purpose* can be seen as a superordinate concept, identifying the main purpose of the entire genre text. On a more specific level, a genre text can usually be subdivided into several sub-purposes by identifying the individual stages or so-called moves of a genre. Both within the overall genre and within individual moves typical linguistic realisations can be identified. This has also been

operationalised by comparing smaller genre-specific corpora with large reference corpora, and thus establishing statistically keywords and their patterns of use of the genre-specific corpus. (cf. Tribble 2001) The focus of any analysis of textualisations lies on the way in which these linguistic realisations support the overall communicative purpose of the entire genre or the more specific communicative intentions of individual genre moves. Thus, in contrast to more register based studies, the analysis of genre-specific textualisations always tries to give explanatory power to the descriptions of the patterns of language usage observed.

3. Extended genre analysis: a new approach to studying formulaic sequences[2]

As pointed out earlier in the example of *The King is dead. Long live the King*, we can see that some formulaic sequences are typical of specific situations in that they occur only or predominantly in these. This observation can also be made with regard to academic or professional genres, such as academic papers, sales letters, contracts etc. This focus on only particular genres acknowledges to some extent the reality of foreign language learners, who rarely aim at becoming proficient in all genres of their target language. Especially in professional settings, the target of many language learners is not the target speech community as a whole, but the target discourse communities, and their relevant patterns of language use, including the relevant formulaic sequences.

This typicality of formulaic sequences dependent on specific text-types or specific communicative situations has already been researched in some setting. While often not the overt focus of the investigations and never directly linked to professional genres, it could be established that there are, for instance, recurring formulaic sequences typical of children's verbal games (Wong Fillmore 1979), and typical formulaic sequences in the language use of supermarket cashiers. (Kuiper and Flindall 2000).

Also, specific written text-types have been studied regarding the use of formulaic language therein. If we consider the text-type of academic writing, we find several studies dealing with its formulaic language use, often in combination with focusing on the difference in patterns of use by learners and native speakers. (Howarth 1996; Jones & Haywood 2004) The focus of these

---

2 For a fuller discussion of the concept of extended genre analysis cf. Hüttner 2005: 94-113

studies lies on the accuracy of form of formulaic sequences achieved by learners rather than on addressing any functional relevance to the texts students produce.

Within the context of extended genre analysis, however, the focus is no longer only on the observation *that* specific formulaic sequences occur within specific genres, but on the *function* they fulfil in these genres. As a detailed discussion of this new approach would go beyond the scope of this paper (but cf. Hüttner 2005: 94-113 passim), a brief, diagrammatic representation of the complete methodology of extended genre analysis will be given in the following. Bold print indicates innovations of the approach when compared to Bhatia's (1993: 22ff) methodology. For the issue of linking formulaic sequences to genre, step five of this analysis is the most important one.

1: place genre or genre constituent in situational context

⇩

2: survey existing literature on the genre in question

⇩

3: **select a corpus of genre texts identified as belonging to one genre (or genre constituent) by the relevant discourse community, including information on authors' status within that community**

⇩

4: **Establish position of genre (constituent) within discourse community, addressing also question of hierarchical position of producers**

⇩

5: Levels of linguistic analysis: **a) if genre-constituent, define larger genre it is a constituent of** b) define genre structure**, including distinction between core and optional moves based on quantitative analysis and information from members of discourse community** c) analysis of lexico-grammatical features d) analysis of text-patterning **e) analysis of genre-functional formulaic sequences** **f) establish communicative purpose in bottom-up way**

⇩

6. elicit information on effects, acceptability and appropriacy of linguistic realisations as perceived by members of the discourse community

⇩

> **7: Establish sets of relationship of genre to other genres**
> a)establish 'genre-set'
> b)establish 'genre-(constituent)-colony'

⬇

> **8**: obtain feedback and further information on analysis by members of discourse community

Figure 1.: Step-by-step methodology of extended genre analysis

In sum, the link proposed here between formulaic sequences and genre is that the latter, which is essentially defined by communicative purposes, consists of individual moves, which are defined by their communicative intentions, and that these moves can be and frequently are realised by specific formulaic sequences. Thus, within extended genre analysis formulaic sequences are defined as prefabricated chunks that are pragmatically functional in a way that is apparent to both producers and recipients. In this sense, these formulaic sequences serve to further the communicative (sub-) purposes of specific genres. Depending on the specific genre, and possibly on individual moves, the dominance of formulaic sequences might vary; for instance, in contracts of sale, a preliminary study (Kastenberger 2005) found a considerable use of formulaic sequences, which highlighted the communicative intentions of nearly every move. One might argue that producers of especially such legal genres like contracts need to ensure that all important communicative intentions of the genre text are fulfilled and are textualized in completely unambiguous forms. In this context, the actual formulaic sequences might not even be readily understood by all users, but serve to indicate which pieces of information are about to come, and in a way act as a frame for the important content of the precise goods to be sold, prices arranged, and obligations entered into. Other genres, however, might make less use of formulaic sequences, or this use might be restricted to specific moves.

Regardless of the actual frequencies of occurrence, investigating this link between formulaic sequences and genres systematically allows for a consideration of this aspect of conventionalised language as a logical extension of studying conventionalisation at a structural level.

It is also important to point out that in extended genre analysis, student genres can be considered separate genres when compared to their more expert counterparts, provided the analysis shows distinct genre structures. Thus student academic papers, , as will be shown later, have a partially different communicative purpose and their distinct genre structures and are considered a separate genre from expert research articles. However, as a part of the communicative purpose *does* overlap, student and expert academic genres are

considered to be both part of an overarching *genre-colony* of academic writing.

While the focus of formulaic sequences within extended genre analysis clearly lies on those sequences that fulfil the function of furthering the communicative intention of a specific move, it has to be noted that there are also other formulaic sequences that occur within genres. I thus suggest a clear distinction between two types of formulaic sequences within genres. Those discussed so far form the group of *genre-functional formulaic sequences*, and examples include chunks like *the aim of this paper is to*. These constitute the most important group from a perspective of genre analysis. The second group are the *non-functional, genre-specific formulaic sequences*. These include clusters formed through expected collocational patterns of genre-specific keywords, and incorporate sequences like *a large number* or *the development of*. Additionally, in all genres there might of course be formulaic sequences that are not genre-specific; thus, a phrase like *I would argue that* might occur in a wide variety of genres. In order to maintain a clear focus, the latter group will not be addressed in extended genre analysis.

4. Application: Genre-functional formulaic sequences in student academic papers

In the following section, I will present examples of genre-functional formulaic sequences established through extended genre analysis. The examples are taken from a study of student academic writing by Austrian learners of English. (cf. Hüttner 2005)

The participants of this study were 66 students of English at the University of Vienna. The L1 of the participants is German. At the time of the investigation, the students were all in their second or third semester, taking part in the introductory seminar in linguistics. This is the first class in the English Studies degree course where students have to write an academic paper, which ranges in length from 2,500 to 3,000 words.

The academic papers written by these 66 participants form the data base of this study. Altogether, due to the fact that some papers were jointly authored, there are in total 56 academic papers on various topics in linguistics. These non-native student data were compared with a corpus of 36 native speaker student academic texts written in language and linguistics departments in Great Britain and the USA, and with 56 published research articles in linguistics. These two sets of comparative data were chosen in order to address questions relating to the influence of factors both of language learning

and genre learning. Analysis of the latter two corpora followed the same procedure as for the non-native texts.

The student texts were analysed following the methodology of extended genre analysis as described earlier. Two genre constituents of the papers, namely introductions and conclusions, were analysed in greater detail, focusing on the textualisation patterns of their move structure. These particular sections were chosen for a variety of reasons; firstly, they are obligatory genre-constituents of all student academic papers, regardless of topic of the paper. Secondly, they constitute the passages that probably place the highest demands on the student writers as they have to produce dense texts with few possibilities of guidance from structures and quotations of published material.

Genre-functional formulaic sequences were classified as such if they recurred in the corpus in identical form in the same moves produced in texts by different authors, and served the function of supporting the communicative intention of the move in question.

5. Findings

5.1 Genre constituent structure

Before discussing the findings on the use of genre-functional formulaic sequences, I will very briefly present the genre structure established for introductions and conclusions in the student texts. In this context, it is important to note that an identical genre structure for both native and non-native student academic papers could be established, so that on the structural level the status of student authors as learners within academia seems to be more relevant than their status as language learners or native speakers.

For student introductions, the following genre structure of three obligatory moves was established:
- leading into the topic
- stating the purpose of paper
- previewing contents

Additionally, two optional moves occurred. These were:
- giving extra editorial information
- acknowledging gratitude

This is to some extent different from the genre structure established for experts, which noticeably includes attempts to position one's own research most favourably and thus vie for readership, especially in the introduction.

This is shown in the Create-a-Research-Space genre structure established by Swales. The latest version of this model is as follows (Swales 1990: 141):

Move 1: Establishing a territory
step 1: claiming centrality
and/or
step 2: making topic generalizations
and/or
step 3: reviewing items of previous research
Move 2: Establishing a niche
step 1A: Counter-claiming
or
step 1B: indicating a gap
or
step 1C: question-raising
or
step 1D: continuing a tradition
Move 3: Occupying the niche
step 1A: outlining purpose
or
step 1B: announcing present research
step 2: announcing principal findings
step 3: indicating research article structure

We can, however, also see some overlap in the structures described, so that the student move of *stating the purpose of the paper* corresponds to some extent to the expert move 3, step 1B *announcing present research*.

For student conclusions, the genre structure incorporates four obligatory and three optional moves, and is as follows:

- providing a summary statement or review
- qualifying and evaluating the paper
- providing a personal reflection
- providing a wider outlook/embedding the paper
  - acknowledging gratitude
  - providing new information
  - appeal to reader

In the expert conclusions analysed the following three moves were observed:

- providing a summary statement or review
- qualifying and evaluating the paper
- providing a wider outlook/embedding the paper

In the following presentation of findings, only results regarding the use of genre-functional formulaic sequences in comparable genre moves observed in both student and expert realisations are discussed.

5.2 Genre-functional formulaic sequences in the 'stating purpose' move in introductions

In both student and expert papers, the introductions include one move, either *stating the purpose of the paper* or *occupying the niche (announcing present research)*, where the author states what will be dealt with in the rest of the paper. The non-native students used a number of clearly identifiable genre-functional formulaic sequences in this move, which are listed below:

- This paper is about
- The (main) focus of this/my paper is
- The purpose of this paper is to
- This research paper deals with
- The basic emphasis of this paper is to
- This paper focuses on

As we can see here, the first sequence, *This paper is about,* is noticeably learner-like and does not seem entirely appropriate in an academic essay. Interestingly, also the L1 counterpart of this phrase, i.e. *Diese Arbeit ist über* would not correspond to expectations regarding academic style.

The native students used fewer formulaic sequences, the only two are listed below.

- This paper will (specifically) address
- my/the project focuses on

What is most interesting though, is that also in the 56 research papers by expert writers, fewer formulaic sequences were observed than in the non-native student papers. The sequences found were the following:

- this paper will explore
- the (prinicipal) purpose of this study is to
- the goal of this research project/the present study is to examine

There are no overlaps in the sequences used by the three groups of speakers, and on the face of it, the fact that the group with the highest proportion of formulaic sequences is the non-native learner group seems to contradict the notion that it is precisely the language learners who find formulaic language most difficult to cope with. However, the use of such formulaic sequences might provide especially the least secure writers in these groups with possibilities of formulating their communicative intentions in appropriate ways. As this move of stating the purpose of the paper is regarded as obligatory by instructors, it might also serve the writers well to signal its presence clearly.

5.3 Genre-functional formulaic sequences in 'presenting limitations' move in conclusions

In many of the conclusions, both by students and by experts, there was a move of qualifying and evaluating the paper/research presented. In this move, frequently the limitations of the present paper were acknowledged. With both non-native learners and native experts the phrase *X is/goes beyond the scope of this paper* occurred as a formulaic sequence. This was not, however, targeted by the native students, who realised this move less frequently and took recourse to non-formulaic realisations.

- X is/goes beyond the scope of this paper

In the case of the non-native writers, there was some evidence of learners aiming to produce this phrase and not quite achieving it. Examples are:

- beyond the reach of this paper
- beyond the boundaries of this paper

As these phrases have no L1 counterpart, it can reasonably be assumed that these are indeed instances of attempts at genre-functional formulaic learning.

Similarly to the formulaic sequences described above, this phrase might offer students the possibility of expressing in a suitable way the communicative purpose of acknowledging limitations. However, the same communicative purpose was also expressed through the use of non-formulaic expressions, notably focusing on a pattern involving a negation stating what had not been covered in the paper.

6. Conclusion

This paper addresses the possibilities of furthering research on formulaic sequences by extending the existing parameters of genre analysis to include a specific focus on the use made of formulaicity. These sequences have here been termed genre-functional formulaic sequences covering those recurring sequences that serve to support the communicative (sub-)purposes of a particular genre or genre move. From a theoretical point of view, I hope to have shown how extended genre analysis might be a helpful way towards a clearer identification of formulaic language in specific discourses. Firstly, limiting concordance searches to corpora made up of specific genres can show clusters typical of these genres, which might get lost in larger, more general language samples. Secondly, by focusing on formulaic sequences that support the communicative intentions of genre moves, a link between these two areas of conventionalised language is explicitly shown.

As an initial application of this extended genre analysis onto the investigation of formulaic language in student academic papers has shown,

genre-functional formulaic sequences appear slightly more frequently in non-native student writing than in expert writing, and also tentatively more frequently in non-native than in native student writing. This seems, on the face of it, to contradict previous research indicating that learners generally use fewer formulaic sequences than native speakers do. I would argue, however, that a closer look at the reasons for using genre-functional formulaic sequences and their status within these texts explains this apparent contradiction.

Firstly, a focus on genre-functional formulaic sequences excludes other formulaic sequences, notably those which have traditionally been covered under collocations, such as *to perform tasks* or *a large number*. These might, indeed, be especially hard to acquire for learners, seeing that in many cases their lack of accuracy will not seriously affect the intelligibility or communicative intentions of the text. Thus, even a student who, for instance, writes *a big number* will make his or her meaning clear to the reader. Thus, one might argue that the motivation for acquiring accuracy in these types of formulaic sequences might be comparatively low, even if we can conceive that the difficulty of learning these sequences can vary, and some, like many noun phrases with *of*-particle, such as *the development of,* might be comparatively easy to learn. (cf. Jones & Haywood 2004: 289) These are formed quite systematically, and German-speaking students who have realised that this construction corresponds to either compound nouns or genitive constructions in German should not find this too difficult.

While there is clearly some variation within the non-functional formulaic sequences as to their salience for the learners, they might be generally perceived as less salient, and therefore arguably less likely to be learnt. This could account for the fact that formulaic sequences can be found more regularly in non-native rather than native writing.

Coming back to the genre-functional sequences, these constitute ways of formulating the communicative intentions of the writers. This is an important task especially for inexperienced writers who are confronted with the need to satisfy genre-specific demands and at the same time struggle to find appropriate words. Using formulaic sequences deemed appropriate for this purpose by the writers can thus be seen as an aid towards formulating their own papers and also a means of sufficiently highlighting the presence of necessary information to the readers. The latter point might be especially relevant in combination with obligatory or near-obligatory pieces of information or genre moves. In this way, one might best view these genre-functional formulaic sequences as *building blocks* which help writers to create their texts. This view also helps explain the slightly higher frequency of such

formulaic sequences in the texts produced by the least proficient writers, i.e. the non-native students.

Interestingly, the formulaic sequences perceived by the writers as appropriate might not necessarily be so for native speakers and/or experts. The best example of this is probably the sequence *this paper is about,* which is particular to non-native learners, and neither follows the target language expectations nor the German L1 patterns. Tentatively, this points towards the need to create genre-functional formulaic sequences even if the most appropriate or most acceptable forms are unavailable to the writer.

With regard to the problems described initially of student learners producing academic papers, we can see firstly how an extended genre analysis shows differences in student papers from expert research articles that affect both genre structure and patterns of use of genre-functional formulaic sequences. Following on from this finding, a first step to improve the teaching situation and to aid students might be establishing clear and attainable goals, which take into account the specific communicative purposes of student papers and their possible textual realisations. The presence of formulaic sequences which appear to be used as *building blocks* to ease writing also points towards possibilities of exploiting this connection between communicative intention and particular forms by drawing students' attention to the function of such sequences and to the patterns used by more experienced writers. However, further research is needed to explore the effects of such genre-based teaching practice with regard to the use of formulaic language to achieve particular genre functions.

References

Bhatia, V. K. 1993. Analysing Genre: Language Use in Professional Settings. Harlow: Pearson Education Ltd.

Bhatia, V. K. 1997. "Genre-Mixing in Academic Introductions" English for Specific Purposes 16, 181-195.

Bhatia, V.K. 2004. Worlds of Discourse: A Genre-Based View. London, NY: Continuum International.

Biber, D., Johansson, S.; Leech, G. , Conrad, S. and Finegan, E. 1999. Longman Grammar of Spoken and Written English. London: Longman.

Bohn, O.-S. 1986. "Formulas, Frame Structures and Stereotypes in Early Syntactic Development: Some New Evidence from L2 Acquisition" Linguistics 24, 185-202.

De Cock, S. , Granger, S., Leech, G. and McEnery, T. 1998. "An Automated Approach to the Phrasicon of EFL Learners" in Learner English on Computer, S. Granger (ed), 67-79. London: Longman.

Foster, P. 1991. "Rules and Routines: A Consideration of their Role in the Task-Based Production of Native and Non-Native Speakers" In Researching Pedagogic Tasks: Second

Language Learning, Teaching and Testing. M. Bygate, P. Skehan and M. Swain (eds), 75-94. London, NY: Longman,

Hanania, E.A.S. and Gradman, H.L. "Acquisition of English Structures: A Case Study of an Adult Native Speaker of Arabic in an English-Speaking Environment" Language Learning 27 (1): 75-91.

Howarth, P. A. 1996. Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making. Tübingen: Niemeyer.

Howarth, P. A. 1998. "Phraseology and Second Language Proficiency." Applied Linguistics 19 (1): 24-44.

House, J. 1996. "Developing Pragmatic Fluency in English as a Foreign Language: Routines and Metapragmatic Awareness." Studies in Second Language Acquisition 18, 225-252.

Hüttner, J. I.. 2005. "Extended Genre Analysis: Exploring Student Academic Writing". Unpublished PhD dissertation, University of Vienna.

Jespersen, O. 1924/1976. "Living Grammar", reprinted in Readings in the Theory of Grammar, Bornstein, D.D. (ed), 82-93. Cambridge, MA: Winthrop Publishers.

Johns, A. (ed.) 2002. Genre in the Classroom: Multiple Perspectives. Mahwah, NJ: Lawrence Erlbaum Ass.

Jones, M. and Haywood, S. 2004. "Facilitating the Acquisition of Formulaic Sequences" In Formulaic Sequences, N. Schmitt (ed.), 269-300. Amsterdam/Philadelphia: John Benjamins.

Kastenberger, A. 2005. "Genre analysis: contracts of sale". Unpublished project paper, University of Vienna.

Kuiper, K. and Flindall, M. 2000. "Social Rituals, Formulaic Speech and Small Talk at the Supermarket Checkout" In Small Talk, J. Coupland (ed), 183-207. London: Longman.

Lynnäkyla, P. 1980. "Hi Superman: What is Most Functional English for a Finnish Five-Year-Old." Journal of Pragmatics 4, 367-392.

Myles, F., Mitchell, R. and Hopper, J. 1999. "Interrogative Chunks in French L2: A Basis for Creative Construction?" Studies in Second Language Acquisition 21, 49-80.

Nattinger, J. R. and DeCarrico, J. S. 1992. Lexical Phrases and Language Teaching. Oxford: OUP.

Pawley, A. and Syder, F.H. 1983. "Two Puzzles for Linguistic Theory: Nativelike selection and Nativelike fluency" in Language and Communication , J.C. Richards and R.W. Schmidt (eds), 191-226. London: Longman.

Schmidt, R.W. "Interaction, Acculturation, and the Acquisition of Communicative Competence: A Case Study of an Adult". In Sociolinguistics and Language Acquisition, N. Wolfson and E. Judd (eds), 137-174. Rowley, Mass.: Newbury House.

Schmitt, N. (ed.) 2004. Formulaic Sequences. Amsterdam/Philadelphia: John Benjamins.

Schmitt, N. and Carter, R. 2004."Formulaic Sequences in Action: An Introduction". In Formulaic Sequences, Schmitt, N. (ed), 1-22. Amsterdam/Philadelphia: John Benjamins.

Sinclair, J. 1991. Corpus, Concordance, Collocation. Oxford: OUP.

Stubbs, M. 1995. "Collocations and Cultural Connotations of Common Words". Linguistics and Education 7, 379-390.

Stubbs, M. 1996. Text and Corpus Analysis. Oxford; Cambridge, Mass.: Blackwell.

Swales, J. 1990. Genre Analysis: English in Academic and Research Settings. Cambridge: CUP

Tribble, C. 2001. "Small Corpora and Teaching Writing: Towards a Corpus-Informed Pedagogy of Writing" In Small Corpus Studies and ELT, A. Henry and R.L Roseberry (eds), 381-408. Amsterdam: John Benjamins.

Weinert, R. 1995. "The Role of Formulaic Language in Second Language Acquisition: A Review." Applied Linguistics 16 (2), 181 – 205.

Wiktorsson, M. 2003. Learning Idiomaticity: A Corpus-Based Study of Idiomatic Expressions in Learners' Written Production. Lund: Lund University Press.

Wong Fillmore, L. 1979. "Individual Differences in Second Language Acquisition" In Individual Differences in Language Ability and Language Behaviour, C. Fillmore, D. Kempler, and W.S.Y. Wang (eds), 203-228. NY, San Francisco, London: Academic Press.

Wray, A. 1999. "Formulaic Language in Learners and Native Speakers". Language Teaching 32 (4), 213-231.

Wray, A. 2000. "Formulaic Sequences in Second Language Teaching: Principle and Practice." Applied Linguistics 21 (4), 463-489.

Wray, A. 2002. Formulaic Language and the Lexicon. Cambridge: CUP.

Charting the boundaries of syntax: a taxonomy of spoken parenthetical clauses

Gunther Kaltenböck, Vienna[*]

1. Introduction

This paper investigates the class of spoken parenthetical clauses, typical examples of which are given in (1) in italics.[1]

> (1) a. You could *I suppose* commission some prints of you yourself (s1a-015-37)
>
> b. What I've done here *I hope you don't entirely disapprove* is try and limit the time taken on by this item by putting it in writing (s1b-075-180)

The term parenthetical is widely used in linguistics, often, however, without clear definition and surprising variation in the number of forms it subsumes. The present study tries to systematise this class of disjunct elements by providing a delimitation from related categories and highlighting its internal stratification.

Parenthetical clauses (PCs for short) are interesting especially because of their borderline status, crossing, as it were, the boundaries of syntax. On the one hand, PCs are part of syntax in terms of linear precedence: they intersect with other structures (their host structures) on the linear plane, sharing with them a terminal string. On the other hand, they fall outside the scope of syntax since this linear order is not controlled by independently motivated principles governing the linearisation of underlying structures (e.g. c-command). PCs, in other words, have no syntagmatic (i.e. paratactic, hypotactic) link to their host clauses. They are related to their host by linear adjacency but are not part of any larger syntactic unit, i.e. they do not form constituents. This 'peripheral' position of PCs, where the principle of linearity overrules that of hierarchical relations is, of course, difficult to account for in a grammar and has been a particular concern for generativists. On the one hand, it has led to proposals to

---

[1] All examples with a text code (e.g. s1a-075-37) are from the British component of the *International Corpus of English* (ICE-GB). Examples without such a code are invented ones.

extend the grammar to include such fringe phenomena by adding an extra level of representation (e.g. Emonds' 1979 extra E-node, Safir's 1986 level of LF-prime; cf. also Espinal's 1991 three-dimensional approach) or stipulating elaborate transformations (e.g. Ross' 1973 Slifting transformation, Jackendoff's 1972 sentential adverb source, McCawley's 1982 crossing branches in SS, Lakoff's 1974 amalgamation rules). On the other hand, it has led to analyses which exclude PCs from the domain of grammar altogether, treating them simply as utterance phenomena (e.g. Haegeman 1991, Burton-Roberts 1999, Peterson 1999).

The dilemma for grammatical analysis is clear: as units which lack an explicit (hierarchical) syntactic link with the host construction, relying merely on pragmatic bonds, they represent performance phenomena. As units which intervene on the linear plane, however, they interact with the level of competence. PCs thus straddle the distinction between (externalized) E-language, i.e. what speakers produce, and (internalized) I-language, i.e. what grammars generate (Chomsky 1986), or the distinction between speech and language (cf. Burton-Roberts 1999 for a discussion of the relation between the two).

Given their unclear status on the boundary between syntax and performance features, it is not really surprising that the term parenthetical has come to be used rather indiscriminately in a variety of different ways and lacks a clear definition. It is often used as a convenient cover term for a wide range of different, only vaguely related forms. The aim of the present paper is therefore to systematise the class of PCs with a view to establishing a clear definition which delimits them from related categories. Defining the class of PCs is, however, not an end in itself. It is part of, and motivated by, a larger research project which aims at investigating the functional properties of PCs based on corpus data provided by the British component of the *International Corpus of English* (ICE-GB). Corpus retrieval inevitably presupposes a clear and workable definition of the category in question.

The outline of the paper is the following. Section 2, first of all, establishes the need for action by providing a brief overview of pertinent uses and definitions of the term parenthetical. Section 3 discusses some methodological problems in defining the class of parentheticals, arguing for a clear separation of formal and functional criteria and for the adoption of a definition based on purely syntactic terms. Section 4, then, turns to the delimitation proper. It is based on three criteria, viz. syntactic form (4.1), lack of syntactic attachment (4.2), and positional flexibility (4.3), which successively narrow down the class of parentheticals. Section 5, finally, delimits PCs from related but nonetheless distinct categories, viz. question

tags (5.1), discourse markers (5.2) and anacolutha (5.3). Section 6 provides a brief summary.

2. *Status quo*: an overview

The concept of parenthesis is already part of the repertoire of traditional grammar and style manuals, where it is used primarily with reference to orthography (e.g. Fowler 1906: 247-248). Jespersen (1973: ch. 25) includes under "parenthetic clause" comment clauses of the type *This, I think, is modern*, reporting clauses, non-restrictive relative clauses, and so-called 'speaker's asides' (cf. Jespersen 1940, V, 252; 1924: 112). Kruisinga (1932, II § 2413) speaks of "parenthetic sentences" in terms of simple or compound sentences which interrupt another sentence and are used "to make a statement or an observation that may serve to make the situation…clearer, or to add a comment" (1932, II: 484).

The structuralist Bloomfield (1950: 187) uses the term parenthesis in a very wide sense, as "a variety of parataxis in which one form interrupts the other", with the specification that "in English the parenthetic form is ordinarily preceded and followed by a pause-pitch".

Parenthesis in a more restricted sense is discussed by the philosopher Urmson (1952), who exclusively refers to "parenthetical verbs", a label that was later adopted by numerous linguists. In this first in-depth investigation of parenthesis he makes the point that, contrary to the generally held belief in philosophy, there is a class of verbs which does not describe 'goings-on'. This group of verbs includes *suppose, know, believe, guess*, which in first person present tense simple can be used parenthetically. As such, they can take initial, medial, and final position in the clause and "orient the hearer aright towards the statement with which they are associated" (Urmson 1952: 491).

A predominantly functional definition is offered by Nosek (1973: 100), a member of the Prague School. For him parenthesis is "a dependent satellite part of the utterance, wedged into a non-compact primary (frame) utterance from which it differs. Parenthesis … expresses a secondary communication ... and a commentary". It is interesting to contrast this with the purely syntactic definition given more recently by a generativist such as Espinal (1991: 727). In this framework the term parenthesis is equated with disjunct constituents, i.e. those that are "identified as independent syntactic constituents or, more generally as independent syntactic structures within another syntactic structure".

Instead of parenthesis Quirk *et al*. (1972: 778-780) introduce the term 'comment clause'. Despite its functional label this clause is defined in terms

of its syntactic link as being "loosely related to the rest of the clause they belong to" (*op.cit*. 778). As such they are disjuncts or conjuncts and "may occur initially, finally, or medially, and have a separate tone unit" (*ibid*.). The following five subtypes are identified: main clause (*I believe*), adverbial clause introduced by *as* (*as you know*), nominal relative clause (*What's more*), *to*-infinitive clause (*to be honest*), *-ing* clause (*speaking as a layman*).

To this list *-ed* clauses (*Stated bluntly*) are added by Leech and Svartvik (1975: 216-217), who also give a brief functional description of comment clauses: they "are so called because they do not so much add to the information in a sentence as comment on its truth, the manner of saying it or the attitude of the speaker" (*op. cit*. 216).

The number of comment clause types is further expanded by Petola (1983) to a total of 15, mainly by a rather loose characterisation of the term 'clause'. His classification includes the following additional types: inserted main clause, sentence apposition (*But it won't happen – worse luck*), non-additive *and*-clause, non-alternative *or*-clause, non-conditional *if*-clause, elliptical predicative in front position (*More important,...*), interjections, adverbs, prepositional phrases. Petola's (1983: 103) definition of comment clauses is a combination of semantic and syntactic aspects: "1) their reference is metacommunicative, i.e. they comment on the truth value of a sentence ... on the organization of the text or on the attitude of the speaker; 2) they are parenthetic in relation to the 'head proposition'".

Quirk et al. (1985: 1112-1118) retain the term 'comment clause' and the previously identified five subtypes. Comment clauses are defined as "parenthetical disjuncts", i.e. adverbial clauses which may "occur initially, finally, or medially, and thus generally have a separate tone unit". This syntactic characterisation is complemented by the functional specification that they are "either content disjuncts that express the speakers' comments on the content of the matrix clause, or style disjuncts that convey the speakers' views on the way they are speaking" (*op. cit*. 1112).

Biber et al. (1999) operate with both the concept of parenthetical and that of comment clause, without however explicitly distinguishing the two. A parenthetical is defined as "an interpolated structure…a digressive structure...which is inserted in the middle of another structure, and which is unintegrated in the sense that it could be omitted without affecting the rest of that structure or its meaning" (*op.cit*. 1067). They are identified as features of written language (*op.cit*. 137) or "less conversational styles of speech" (*op.cit*. 1068) and take the form of noun/numeral phrases, complete independent clauses (*op.cit*. 137-138), or dependent clauses (*op. cit*. 1068, 201). 'Comment clauses', on the other hand, "are loosely connected to the main

clause, they normally lack an explicit link, and they are usually short and can appear in a variety of positions" (*op.cit.* 197), i.e. initially, medially and finally. They are characteristic of spoken language, typically occur in first or second person present tense and "comment on a thought rather than the delivery or wording" (*op.cit.* 197); e.g. *I think*.

The term used by Huddleston and Pullum (2002) is 'parenthetical'; the label 'comment clause' does not figure in their framework. By parentheticals they mean "expressions which can be appended parenthetically to an anchor clause but which also have a non-parenthetical use in which they take a declarative content clause as complement – expressions like *I think, don't you think?*" (*op.cit.* 895). Parentheticals thus resemble Biber et al.'s (1999) comment clauses, but are more restricted by their explicit 'transformational' relationship to non-parenthetical uses. On the other hand, they are also considerably wider since they are not accorded a specific commenting function, and as such include also, for instance, some of Biber *et al.*'s "reporting clauses", viz. those preceded by the reported speech (*op.cit.* 1024, 1027).

From this brief overview of definitions it becomes clear that there is no general agreement on the exact delimitation of parentheticals, let alone on the use of a uniform terminology. Table 1 summarises the different syntactic forms parentheticals have been associated with, with some representative references.

Table 1. Syntactic categories commonly included under parenthetical

A. Main clause / parenthetical parataxis
*He called John – **he is one of his best friends** – to find out what had happened*   e.g. Petola 1983: 103-104, Quirk et al. 1985: 977 note, Biber et al. 1999: 138
B. Coordinated main clause (with *and* / *or*) / parenthetical coordination
*For several years now – **and I don't mean to be cynical** – we have been trying to overcome this problem*   e.g. Petola 1983: 104-5, Aijmer 1980: 57, Quirk et al. 1985: 932, 977, Huddleston & Pullum 2002: 1361
C. Main clause-like 'comment clause'
*The solution**, it seems / I believe**, is an easy one*   e.g. Quirk et al. 1985: 1112, Biber et al. 1999: 197, Huddleston & Pullum 2002: 895-897, Stenström 1995, Petola 1983: 110-111, Aijmer 1972, Urmson 1952
D. Reporting clause
*She was very happy, **he said**. In the near future, **John announced**, I will move to Paris*   e.g. Huddleston & Pullum 2002: 1024, 1027, Schelfhout 2000, Wiechmann 2001: 179
E. Non-restrictive relative clause
- ad-nominal: *John**, who lives in London**, is travelling to France*

- nominal: ***What is more interesting***, *he finished his paper*
- sentential: *Mary went on holiday to Crete,* ***which is probably what you'd like to do***
        e.g. Quirk et al. 1985: 1112, Bolinger 1989: 191, 198, Espinal 1991: 726, Ziv 1985: 191, Petola 1983: 106-107, Hoffmann 1998: 307-309, Peterson 1999: 245, Burton-Roberts 1999: 34-40, Haegeman 1991

F. Content clause (appositive clause)
The excuse she gave **– *that there had been a traffic jam*** *– was ridiculous*
        e.g. Huddleston & Pullum 2002: 1358

G. Adverbial clause / clausal adjunct
- finite: ***As you probably know***, *I won't be here next week*
            That's a Ming vase, ***if I'm not mistaken***
- non-finite: *I'm a bit overwhelmed****, to be honest***
            I doubt, ***speaking as a layman***, *whether this will be the right solution*
            Stated briefly, *there is no quick solution to the problem*
        e.g. Petola 1983: 105-6, 111-3, Quirk et al. 1985: 1112-1113, Huddleston & Pullum 2002: 666, 1360, Espinal 1991: 726, Biber et al. 1999: 1068, Haegeman 1991

H. Question tag
Mary is coming tomorrow, ***isn't she?***
        e.g. Ziv 1985: 189-190, Knowles 1980, Huddleston & Pullum 2002: 891-892, 896-897, Mittwoch 1979

I. Right node raising / interpolated coordination / shared constituent coordination
He is, ***or at least he was***, *a great actor*
        e.g. Quirk et al. 1985: 976-977, Peterson 1999: 242-243

J. Amalgam(ation)
He gave this **I prefer not to know how awful** *paper*
        e.g. Lakoff 1974, Espinal 1991: 748, Plank 1981: 65-66, Aijmer 1997: 7

K. Verbless clause
The visitors, ***most of them students***, *were rather surprised*
        e.g. Huddleston & Pullum 2002: 1359

L. Adverbial phrase
Frankly, *I don't know what to say about this*
        e.g. Espinal 1991: 726, Petola 1983: 108-110, Nosek 1973: 108-109, Skrebnev 1959: 60, Corum 1975, Wiechmann 2001: 179, Asher 2000: 31

M. Adjective phrase
The chairman, ***angry at the delay***, *demanded a full report*
        e.g. Espinal 1991: 726, Petola 1983: 107, Huddleston & Pullum 2002: 1359

N. Prepositional phrase
In brief, *the film has been a great success*
        e.g. Petola 1983: 107, Espinal 1991: 727, Skrebnev 1959: 61

O. Noun phrase
- Apposition: *Annie Lennox,* ***my favourite pop singer***, *has a new song out*
        e.g. Huddleston & Pullum 2002: 1359, Petola 1983: 104, Peterson 1999: 243-248, Quirk et al. 1985: 1304, Burton-Roberts 1994: 186, Hoffmann 1998: 309-311.

- Vocative: *Today's topic, **ladies and gentlemen**, is astrophysics*
      e.g. Espinal 1991: 727, Nosek 1973: 101, Ziv 1985: 191

P. Interjection
Damn*, we've missed the train*
      e.g. Huddleston & Pullum 2002: 1360-1361, Ziv 1985: 190, Petola 1983: 107-108

Q. Discourse marker
*John**, you know**, is not going too come tonight*
      e.g. Biber et al 1999: 197, 140, 1075, Stenström 1995: 291, Quirk et al 1985: 1113-5

3. Methodological problems

The overview in Section 2 shows that most characterisations of parentheticals, either explicitly or implicitly, make use of different types of criteria, usually mixing formal (syntactic, prosodic) and functional ones without clear indication as to which is taken as primary. This may give the impression of comprehensive delimitation, but is often just an attempt to come to terms with an inherent problem: parentheticals cannot be defined by themselves, as "in isolation, there is nothing particular which identifies a parenthetical" (Espinal 1991: 728). They derive their existence, as it were, from their interaction with a host clause. This interaction, however, takes place purely on the linear plane, not on a relational (i.e. dependency) level. The relation between host and parenthetical is, in fact, a non-relation, at least syntactically. The extraneous nature of parentheticals has resulted in their being defined mainly negatively, in terms of what they lack, viz. a syntactic link to the host clause, phonological integration, a fixed position, meaning in truth-conditional terms. The only positive characterisation seems to be in terms of their communicative function (cf. Ziv 1985: 191), which explains the frequent combination of formal with functional criteria for definition. This blending of formal and functional properties is most conspicuous in the concept of 'comment clause' (e.g. Quirk et al. 1985, Petola 1983, Biber et al. 1999), where the term itself makes reference to a functional category. Despite the vagueness of the term, this suggests a one-to-one form-function relationship, which is not tenable.

The mixture of different types of criteria for definition also entails a serious problem for data analysis, that of circularity. As Wiechmann (2001: 181) points out with reference to prosodic definitions "[i]f parentheses are identified on the basis of prosody alone, the investigation of the prosody is a potentially circular business". The same applies to all other criteria, notably communicative function. If one of the aims is to explore the functional properties (cf. Section 1), an identification of parenthesis on the basis of

functional characteristics is highly problematic. The overall research goal therefore calls for a clear separation of functional and formal identification criteria and a definition in essentially non-functional terms.

Apart from the issue of circularity, we also need to consider reliability of the criteria and ease of identification. Because of their 'formal' nature, the syntactic criteria of form, structural independence, and positional flexibility are comparatively easy to identify. While syntactic form is often largely unconstrained in the literature and as such not exploited as a defining characteristic, the other two constitute core features of most definitions.

Prosodically, the most obvious feature identified for parentheticals is that of a separate tone unit (i.e. separate nucleus) often reinforced by pauses. It has been pointed out by various studies, however, that this is by no means a reliable criterion. A considerable number of parentheticals are not marked in this way. Bolinger (1989: 186), for instance, emphasises that any of the prosodic features typically found with parentheticals (pause, lowered pitch, terminal rise) may be suspended. Similarly, Espinal (1991: 734) notes that "having an independent intonational unit is neither a sufficient nor a strictly necessary property to identify parentheticals"; a view supported by Reinhart (1983: 178-179) and Mittwoch (1979: 407). The inconsistency of prosodic marking is also confirmed by corpus data, as pointed out, for instance, by Stenström (1995: 292) and Wiechmann (2001: 186). While it has to be conceded that prosodic realisation may be dependent on the definition of parentheticals in syntactic terms (i.e. the wider the range of permitted syntactic forms, the more inhomogeneous their prosodic realisation), it seems clear that there is no uniform prosody of syntactically disparate structures, which makes this criterion unsuitable for delimitation.

Functionally, parentheticals have been described as expressing some speaker comment (e.g. Quirk et al. 1985: 1112, Petola 1983: 103), which "in some way, qualifies that which is expressed by the sentence to which they are appended" (Ziv 1985: 182), with a wide range of possibilities: they may "strengthen or weaken the force, or specify the form, of the speaker's attitude to the content of the expression with which they occur" (*ibid.*). As such, "their reference is metacommunicative" (Petola 1983: 103). More generally still, they have been characterised as expressing "additional information" (Biber et al. 1999: 137), information of secondary importance (Hoffmann 1998: 304), which is syntactically "backgrounded" (Huddleston & Pullum 2002: 896).

Given the relative unreliability of prosodic features and the 'negative' character of syntactic criteria, which specify the relation between parenthetical and host in terms of what it is not, and the wide variety of syntactic forms involved, it is not really surprising that most definitions

incorporate a functional characterisation of some form. It provides a convenient uniform bracket for a syntactically diverse and rather 'intangible' phenomenon. The price for this smallest common denominator, however, lies in its relatively unconstrained and vague nature. Functional-pragmatic concepts, by their very nature, are often less clear-cut than formal ones, but an attempt to find a functional label able to accommodate a wide range of syntactic forms only adds to this. A term such as 'speaker comment' may be wide enough to subsume diverse syntactic subtypes, but it is also too unrestricted to provide a useful criterion for delimitation, especially when working with corpus data.

In view of the problems associated with prosodic and functional properties, the present study relies exclusively on syntactic criteria for the purpose of delimitation. They are discussed in detail in Section 4.

4. Delimiting the class: three criteria

This section investigates the three formal criteria of syntactic form (Section 4.1), lack of syntactic attachment (Section 4.2), and positional flexibility (Section 4.3), which will allow us to successively constrain the class of PCs and highlight its internal stratification.

4.1 Syntactic form

The overview in Section 2 has shown that the class of parentheticals may include a variety of different syntactic forms. As mentioned above, the focus of the present investigation is, however, exclusively on parenthetical clauses, i.e. the clausal category. The reason for this restriction is, first of all, a practical one, viz. to increase the formal homogeneity of the class. This is particularly important in view of the functional aim of the research project (cf. Section 1): in order to investigate the communicative functions it is useful to keep the class formally as uniform as possible, especially given the functional diversity of entities such as interjections, vocatives, discourse markers, and adverbials. Secondly, non-clausal items are arguably more prone to being integrated into the host clause than clausal ones. Unlike inserted clauses, which are always by definition syntactically external, non-clausal entities allow for varying degrees of integration, syntactically as well as prosodically. This potential for integration is most obvious with adverbs, such as in (2a), which can also be rendered non-parenthetically (i.e. without prosodic separation) as (2b). In contrast, a clausal constituent such as in (3) will always be parenthetical, irrespective of prosodic realisation.

(2) a. She described herself, *engagingly*, as an environmentalist

b. She described herself *engagingly* as an environmentalist

(3) She described herself, *I believe*, as an environmentalist

If we apply the criterion of clausal constituency a number of the parenthetical categories listed in Table 1 above are filtered out. Table 2, which distinguishes between clausal and non-clausal categories, illustrates this.

Table 2. Clausal vs. non-clausal parentheticals

clausal	non-clausal
A. Main clause	K. Verbless clause
B. Coordinated main clause	L. Adverbial phrase
C. Main clause-like comment clause	M. Adjective phrase
D. Reporting clause	N. Prepositional phrase
E. Non-restrictive relative clause	O. Noun phrase
F. Content clause	P. Interjection
G. Adverbial clause / clausal adjunct	Q. Discourse marker
H. Question tag	
I. Right node raising	
J. Amalgam(ation)	

There are of course various borderline cases. Verbless clauses, for instance, have been excluded here from the clausal category owing to their lack of a VP. Similarly, some discourse markers are by virtue of their form potential candidates for inclusion under 'clausal', e.g. *you know, I mean*. These cases have been disregarded here and will be given special attention in Section 5.2. The overwhelming majority of discourse markers are, however, non-clausal. On the other hand, instances of Right Node Raising and amalgamation have been grouped under clausal since they typically contain a verb, even though the VP is usually incomplete. These two categories will be discussed in Section 4.2.

Application of this clausal filter is a first step towards constraining the class of parentheticals, viz. to PCs. They will be further restricted in the following two sections.

4.2 Lack of syntactic attachment

It is generally agreed that PCs have no obvious syntactic link to the clause they seem to be attached to, i.e. their host. They are, in other words, syntactically disjunct, with non-syntagmatic relations, which lie outside the syntagmatic relations of parataxis (e.g. coordination) and hypotaxis (e.g. subordination and complementation). Huddleston and Pullum (2002: 1350)

speak of a "lack of integration into the syntactic structure", which distinguishes what they call "supplementation" from dependency constructions and coordination. An additional feature distinguishing supplements from dependency constructions (but one it shares with coordination) is their non-headedness, a logical consequence of their lack of syntactic integration. Parenthetical clauses are thus not only syntactically non-dependent on their host, but also syntactically not attached or integrated (i.e. do not form a syntactic unit with the host); they are merely adjacent to the host, i.e. they intersect with the host structure purely on the linear axis and are not linked by any syntactic nodes – a fact that has proved to be a particular problem for models of syntactic representation.

4.2.1 Reviewing the evidence

The syntactic independence of PCs can be demonstrated with a number of tests which show that they do not participate in syntactic processes and hence do not form a syntactic unit with their hosts. McCawley (1982: 96, 1998: 751), for instance, uses the VP-deletion test to show that VP ellipsis operates on VPs including a parenthetical as if the parenthetical were not there. Cf.

> (4)   John talked, *it seems*, about literature and Mary did too (= Mary talked about literature; ≠ Mary talked too; ≠ Mary talked, *it seems*, about literature)

Similarly, pronouns with a VP antecedent behave as if the parenthetical clause were not there; it does not count as part of the antecedent (McCawley 1982: 96-7; 1998; Peterson 1999: 234).

> (5)   John talked to us, *it seems*, about literature, but Mary would never do that (= Mary would never talk to us about literature; ≠ Mary would never talk to us, *it seems*, about literature)

Moreover, it has been demonstrated, for instance, that parentheticals cannot be the focus of a cleft sentence, cannot be questioned, are not under the scope of quantifiers, and are unaffected by negation in the host clause, i.e. they are outside its scope (e.g. Jackendoff 1977, Emonds 1979, Espinal 1991, Haegeman 1991, Fabb 1990, Burton-Roberts 1999).

These tests thus provide convincing evidence that PCs are unattached 'orphans' (Haegeman 1991) at syntactic structure. However, there are also some syntactic characteristics that seem to suggest control of the host clause over the PC. One such exception are **sequence-of-tense** facts. Ross (1973: 139-140), for instance, claims that a non-factive parenthetical verb in past tense is incompatible with present tense in the host clause. E.g.:

> (6)   *There is something funny about Venus, *it seemed to me* (Ross' ex. 19f)

This claim, however, is challenged by obvious counterexamples such as (7), which shows that Ross' example is less objectionable with other non-factive verbs.

(7)   There is something funny about Venus, *I thought/claimed/maintained*

In (7) the sentence can be read as a form of direct speech and as such is perfectly acceptable. The prototypical position of such a reporting PC is of course before the host and, interestingly, such a position also considerably improves Ross' example. Cf.:

(8)   *It seemed to me*, there is something funny about Venus

This seems to be an indication that tense agreement (back-shift of tense) in parentheticals is not so much a syntactic process restricted to the intra-sentential domain but rather a pragmatic one. It therefore does not invalidate a view which takes PCs to be syntactically independent from their host clauses.

Another characteristic of PCs which challenges the view of syntactic independence is discussed by Reinhard (1975, 1983), who distinguishes two types of PCs: parenthetical-subject oriented and speaker-oriented, exemplified in (9) and (10) respectively (Reinhart's 1983, ex. 16 and 19).

(9)   He$_i$ would be late, *John$_i$ said*   = parenthetical-subject oriented

(10)   John will be late, *he said*   = speaker-oriented

Parenthetical-subject oriented PCs are shown to differ from speaker-oriented PCs in that they have obligatory **'backward pronominalization'** (i.e. cataphoric pronominal reference; cf. **John would be late, he said*) and tense agreement (cf. **He will be late, John said*). While the back-shift of tense can be treated as a more general pragmatic principle (see above), obligatory cataphoric reference could be seen as a counterexample to the claim that PCs are not involved in any syntactic process. However, Reinhart (1983: 192) herself argues that the anaphora is the result of independently motivated performance constraints, viz. the convention that in our culture people do not normally refer to themselves with full nouns. Since the host clause in these cases is similar to direct quotes (*op.cit.* 178, 189, 192) and as such understood as having been uttered by John, pronominal reference is the only possibility here. As an alternative explanation it is also possible to argue along the following lines: with speaker-oriented PCs the focus (in terms of informational salience) is on the host clause, which *asserts* a particular state of affairs, while the PC only adds extra information about the speaker's source of information. Sentences with parenthetical-subject oriented PCs, on the other hand, have an entirely different communicative function, viz. that of *reporting* a speech act (John saying something). The focus is thus on both the

reporting frame plus 'quote'.[2] The prototypical position of such reporting frames is before the quote (*John said: 'He would be late'*), motivated by greater ease of processing, with the expected sequence of first the full NP (*John*) and then the (anaphoric) pronoun. Example (9) is simply a case of marked word order with a postponed (right-dislocated) reporting clause. Interestingly, the same issue of word order (but inversely) arises with example (10), the speaker-oriented PC. Fronting of the PC results in the unacceptable sentence **He said, John will be late*, owing to an inappropriate sequence of pronominal reference before full NP. The restrictions on pronominalization in Reinhart's examples therefore cannot be taken as sentence-internal syntactic constraints but are the result of functional requirements, viz. the question of communicative focus (asserting or reporting) together with word order preferences.

A third potential problem for the view that parentheticals do not partake in syntactic processes are **negative** and **interrogative** PCs. As pointed out by Ziv (1985: 186-188), the occurrence of negative and interrogative PCs is subject to considerable restrictions as regards the form of the respective host clause, which could indicate a syntactic constraint. A negative PC thus can only occur in combination with a host clause which is also negative, as illustrated in (11).

> (11) a.   The book is not very interesting, *I don't think*
>
>   b.   *The book is very interesting, *I don't think*

The question to be answered here is whether this restriction is the result of an intra-sentential syntactic process or whether it can be attributed to independently motivated pragmatic principles. Ziv (1985: 187) argues for the latter, pointing as evidence to the same restriction across sentence boundaries with non-parenthetical negative verbs (e.g. *The book is not very interesting. I don't think so* vs. **The book is very interesting. I don't think so*). Further evidence for a functional-pragmatic restriction comes from Knowles (1980: 388), who shows that syntactically negative PCs may actually be acceptable in combination with a positive host, provided the semantic 'sum' of the PC is positive (e.g. *The book is very interesting, I don't deny*). The same principle can be shown to be at work with a negative PC linked to a positive host, as in (11a). Although the PC is syntactically negative, its meaning is actually positive, viz. a positive claim to the truth of the proposition of the host. If this

---

2   The backshift of tense is a direct result of this reporting function enabling the speaker to create a distancing effect between him/her and the 'quote', indicating that s/he is not the original source.

positive meaning of the PC were changed to a negative one, i.e. one that reinforces the negative meaning of the host clause, the resulting sentence is also unacceptable (*_The book is not very interesting, the book/it is_), even though the PC is syntactically positive, a form typically allowed by negative hosts.

A similar functional explanation can be found for interrogative PCs, which, on the face of it, seem to be restricted to interrogative hosts (i.e. in syntactic terms), as illustrated in (12).

(12) a.   Is the book interesting, _do you think? / would you say?_

b.   *The book is interesting, _do you think? / would you say?_

However, it has been shown by Mittwoch (1979: 409) that a declarative host as in (12b) may in fact occur with an interrogative PC, provided the host has a rising intonation and thus functions as a question, cf. (13).

(13)   The book is ↗INteresting, _do you think?_

This is an indication that the link between interrogative PCs and their hosts is not a syntactic but a pragmatic one. The pragmatic function of interrogative PCs has been identified by Knowles (1980: 390) as "seeking confirmation of the truth of the main proposition". Consequently, an assertive (declarative) host results in a pragmatic contradiction: it only makes sense to ask for confirmation if the information of the host clause is not asserted.

4.2.2   Non-syntagmatic links

The evidence discussed so far thus provides strong support for the view that PCs are syntactically independent of their hosts. In other words, a PC is not a constituent of the host clause and hence there is no structural relationship between it and its host. There is, however, a semantic-pragmatic link with the host, which may take various forms with varying degrees of attachment. To analyse the different types of semantic links in more detail it is useful to systematise the class of PCs on the basis of formal criteria and distinguish very generally between syndetic (with a formal link) and asyndetic (without a formal link) PCs.

A) Syndetic PCs
Syndetic PCs are those that are introduced by an overt marker, which links the PC to the host clause. Typically, such markers are, what Huddleston and Pullum (2002: 1354) call "indicators", such as _namely, that is, that is to say, especially_ (cf. 14). Although limited by Huddleston and Pullum to precisely those items, it seems reasonable to extend the category of indicators to

include other elements, viz. coordinators, subordinators and even relative elements, as is illustrated in the following examples.

Indicator:

(14)   Many clauses are asyndetic, *that is they do not have an overt marker*

Coordinator:

(15)   For several years now – *and I don't mean to be cynical* – we have been trying without success to overcome these problems

Subordinator:

(16) a.   The warning – *that prices should be lowered* – was ignored

b.   He is a real bastard – *if you don't mind the expression*

Relative element:

(17)   Mary is away on business, *which is convenient*

Although all of these PCs have some formal link to the host clause, we need to remember that these 'links' do not have any syntactic value, as the PCs are not constituents of the host clause. Hence the coordinator *and* is not a real coordinator and neither are the subordinators real subordinators (i.e. are not syntactically dominated by another node, S/CP or NP). The same is true for the relative element in (17). The terms coordinator, subordinator, relative element may therefore be misleading and are probably best subsumed under 'indicator'.

Although all the formal syndetic links are non-syntagmatic, they are, nonetheless, indicative of a semantic-pragmatic link to the host clause, or what I refer to as the anchor.[3] In example (14), for instance, *that* indicates that the PC provides an 'explanation' of its anchor (*asyndetic*). In example (15) the 'coordinating indicator' *and* has 'additive function'. For the examples in (16) and (17) the type of semantic-pragmatic link is more difficult to specify; nonetheless it is clearly there.

For **content clauses** (or complement-like clauses) such as (16a) Huddleston and Pullum (2002: 1315) point out that, although these clauses are not syntactically licensed by a head (like proper complement clauses), they must still be semantically compatible with their anchor, as illustrated by the adapted versions (18a) and (18b).

---

[3] Unlike Huddleston and Pullum (2002), I use the term anchor here for the element in the host clause to which the PC (or elements of it) is semantically related. It is either a constituent of the host clause (as in examples 14 and 16) or the entire host clause (as in example 17).

(18) a. *The question – *that prices should be lowered* – was ignored

    b. * The warning – *whether prices should be lowered* – was ignored

For **peripheral adverbial clauses** (adverbial-like clauses) such as (16b) Haegeman (1991: 246-248) suggests that, although syntactically unattached orphans, they are semantically linked to the host, with the head of the PC (the subordinator, e.g. *if*) being coindexed with the host clause. The open variable in the form of an index in the propositional form of the PC imposes certain conceptual constraints on the host at the level of discourse rather than syntax. The host (or modifee of the PC), in other words, is only established at the level of utterance processing, which requires that the index of the PC be identified with the index of the host clause. The full interpretation of the PC is therefore achieved only pragmatically in a given context, when it becomes integrated into the representation schemas associated with other propositions, following the general principles of utterance interpretation, as developed in Relevance Theory (Sperber & Wilson 1995).

For **non-restrictive relative clauses**, such as (17), a similar non-syntagmatic link has been identified (e.g. Fabb 1990, Burton-Roberts 1999). Although differing in details of their analyses, modern accounts of non-restrictive relative clauses assume a link between relative pronoun and its antecedent only on a semantic-pragmatic level. Burton-Roberts (1999: 38-40), for instance, argues that the pronoun-antecedent relationship between the non-restrictive relative clause and its host is one that is only contextually interpretable and therefore does not invoke (syntactic) coindexing but (semantic-pragmatic) coreference. This syntactic independence is exemplified by B's answer in (19) below, where the only link to the preceding clause is one of contextual/pragmatic coreference.

(19)   A: Mary is away on business
          B: Which is convenient (cf. That's convenient)

There is one further candidate for the category of syndetic PCs, viz. instances of **Right Node Raising**, which will be dealt with below, together with asyndetic amalgamation.

B) Asyndetic PCs

Let us now turn to the category of asyndetic PCs. Unlike syndetic PCs, whose syntactic form in some way indicates that they are dependent utterances which need to be pragmatically linked to a host, asyndetic PCs have no overt marker indicating such a link. In terms of syntactic form, it is possible to distinguish three different subtypes, which also differ in their relationship to

the host. They are referred to here as 'self-contained PCs', 'reduced PCs' and 'Right Node Raising' and are briefly discussed in the following.

Self-contained PCs are main clauses, whose syntactic independence is already signalled by their form, i.e. that of independent and self-contained clauses. They are exemplified in (20).

    (20) a.   Mary – *I hate to tell you this* – is coming over to visit

        b.   Mary – *is that her real name?* – is coming over to visit

        c.   Mary – *don't forget* – is coming over to visit

        d.   Mary – *is she persistent!* – is coming over to visit

As a clear sign of their independence, there are no constraints on the clause type of the PC. The examples in (20) show that the PC may take the form of a declarative, an interrogative, an imperative, or an exclamative. The relationship between PC and host is thus a non-syntagmatic one, sometimes reinforced by a pronominal link (as in 20a, b, d) but not necessarily so (20c). The connection between host and PC is established by the interruption and hence structural incompleteness of the host clause, which forces the listener to establish some pragmatic link between the two, often aided by contextual features, as in (21).

    (21)   And what we found was uhm – *could you turn the slide projector off please* – uhm very substantial mortality differences within this population (s2a-047-110)

Limitations on permitted combinations of host and PC only apply on a pragmatic level, as in the case of clashing illocutions illustrated in (22).

    (22) a.   *Mary – *I hate to ask you this* – is coming over to visit

        b.   *When is Mary – *don't forget* – coming over to visit?

Owing to their syntactic form, self-contained PCs represent a particular problem for identification. While they are unmistakably identifiable in interpolated position (as in 23a and the examples above), their exact status is less clear in peripheral position (as in 23b, c).

    (23) a.   Mary, *she has left her boyfriend* , is coming over to visit us

        b.   Mary is coming over to visit us; *she's left her boyfriend*

        c.   *Mary has left her boyfriend*; she's coming over to visit us

According to Peterson (1999: 241-242), such juxtaposed clauses are best treated as non-syntagmatic relations rather than coordination based on their different syntactic behaviour: unlike coordination they cannot serve as complements and do not allow 'across-the-board' extraction. They are therefore seen as analogous to disjunct dependent clauses. When working

with corpus data, however, there is no unambiguous way of distinguishing peripheral cases of juxtaposition (such as 23b, c) from coordination. They are therefore excluded from our class of PCs.

A special type of self-contained PCs are ***semantic gap-filling clauses*** (Biber et al. 1999: 1066), as illustrated in (24).

> (24) a.   Isn't it where all the last century upper middle classes used to drive around on Sundays in their uhm *what are they called* (s1a-006-218)
>
> b.   Well most most database packages if you're producing a package most most database manufacturers let you buy um *I can't recall the word for it* but they let you buy it so that is you buy it once for a bit more the runtime version of it basically (s1a-029-88)

These PCs differ from other self-contained PCs as they occur in a position within the host clause where normally a single word or phrase would be expected. Without them the host clause would remain syntactically incomplete. Such gap-filling PCs need to be distinguished from cases of anacoluthon (cf. Section 5.3) since they follow a relatively predictable pattern and fulfil a specific communicative function, viz. indicating that the speaker – for whatever reason – is unable to supply a more exact lexical item.

The second category of asyndetic PCs is that of **reduced PCs** (RPCs) (Schneider 2005), or gap-containing PCs. They are illustrated in (25).

> (25) a.   You could *I suppose* commission some prints of you yourself (s1a-015-37)
>
> b.   Britons *he said* could compete and win (s2b-005-129)

As can be seen from the examples, these PCs are linked to the host in that they contain a syntactic gap (typically the complement of the verb) which is filled conceptually by the host clause. This gap-antecedent relationship holding between the PC and host is one that is contextually interpretable and, as argued by Burton-Roberts (1999: 40) and Peterson (1999: 237), does not involve (syntactic) coindexing but (semantic/pragmatic) coreference. The lack of syntactic dependence of the PCs is particularly interesting given the existence of a non-parenthetical counterpart construction, illustrated in (26), where the verb in question functions as matrix predicate governing a complement clause (cf. also Section 4.3).

> (26) a.   *I suppose* that you could commission some prints of you yourself
>
> b.   *He said* that Britons could compete and win

It is possible to distinguish two subtypes of reduced parenthetical clauses (RPCs) based on the semantic category of the verb used: ***comment RPC*** and ***reporting RPCs***. While both types contain assertive predicates (Hooper 1975), comment RPCs typically make use of some verbs of thinking, as illustrated in example (25a) above, and reporting RPCs make use of message

conveying verbs (reporting verbs, *verba dicendi*), as illustrated in example (25b).

Comment RPCs usually involve transitive verbs (e.g. *I believe, I guess*) without the object but may also consist of an adjective which elsewhere requires a *that*-clause object (e.g. *I'm afraid, I'm sorry to say*). They are typically in the present rather than the past tense and often do not have a corresponding progressive form (cf. Urmson 1952: 481). Unlike reporting RPCs, their subject is usually in the first or second person rather than the third, although impersonal third person subjects are possible, too (e.g. *it's true, there's no doubt, one would have thought*).

Reporting RPCs, on the other hand, are not limited to present tense but typically take a third person subject, as a result of their reporting function. More specifically, their function consists in identifying the speaker's source of information. As such, they represent a particular type of 'evidential', particularly in the narrow definition of evidentials as "markers of the speaker's information source" (Traugott 1989: 32). Reporting RPCs also differ from commenting RPCs in allowing a certain amount of flexibility in their word order, provided the subject is non-pronominal:[4] cf. *The flight will be delayed, John says/says John, by two hours.*

Despite the formal and semantic differences between the two categories there is also considerable room for overlap. The typical (cognition) verbs of commenting RPCs, *think* and *suggest*, for instance, also commonly function as reporting verbs (in the third person). Conversely, message conveying verbs can also occur with a first person subject, e.g. *I said, I wondered, I was told (by John)*. Moreover, pigeonholing expressions with impersonal *it*, such as *it is said/reported/claimed/rumoured* and *it seems/appears/transpires* may also prove difficult. Is the latter set best classified alongside the former as reporting RPCs, or best singled out as commenting RPCs, despite the formal and semantic similarity with the former? The present study therefore adopts a restrictive view of reporting RPCs. It includes only cases of explicit third person source identification of the type 'source = X' (X ≠ 1st or 2nd person), allocating all references to some unspecific source, such as the hearsay evidentials *they say* or *it is reported*, to the class of commenting RPCs. Accordingly, the expressions *John/He said, I was told by John, it is reported/pointed out by John* are classified as reporting RPCs, while *I was told, it is reported/pointed out* together with *it appears/transpires/seems* and

---

4 Postposing of a personal pronoun subject (*said he*) is considered archaic or non-standard. Compare also the jocular *says I* (cf. Huddleston & Pullum 2002: 1027).

I/you said are taken to be comment RPCs, owing to their lack of a specific source of information.

This leaves us with one further problem: second person subjects. Although Quirk et al. (1985: 1115) and Biber et al. (1999: 197) include them under comment clauses, this class is far less homogenous than might appear at first glance. While expressions such as *you may know*, *you can see* may qualify for the category commenting RPCs, there is also the potential of overlap with reporting RPCs, as illustrated by example (27).

> (27) So you're quite good at it *you thought* (s1a-089-186)

Most importantly, however, in the majority of cases second person subjects directly address the hearer with the aim of claiming his/her attention or inviting agreement. This 'interpersonal' function is particularly obvious in interrogative forms such as in (28).

> (28) a. Uh you do though *don't don't you think* (s1a-084-76)
>
> b. I mean again Sir Simon Gourlay you talk about the uh countryside as a as the farmer's workshop and and workplace and actually <,,> nature *would you not agree Rodney Legge* is in fact the creation of man (s1b-037-36)

It is tempting therefore to establish a third category alongside commenting and reporting RPCs, thus creating a three-fold division which mirrors the first-second-third person trinity or, in functional systemic terms, the distinction of ideational-interpersonal-textual (cf. Widdowson 1997: 158 for a correlation of the two systems). However, given the amount of overlap with the other categories (cf. the unclear status of *you think, you said, you warned*, which would warrant classification as commenting, reporting, and interpersonal), introducing a third subclass would only create unnecessary profusion of categories at this stage of classification. I therefore preserve the traditional two-fold distinction and commenting and reporting, with all second person subjects being subsumed under commenting RPCs.

The third type of asyndetic PCs are instances of **amalgamation** (Lakoff 1974), where one constituent is shared by the host sentence and the inserted clause, as in example (19a). Syntactic amalgams can also take syndetic form (with a coordinator) typically referred to as **Right Node Raising**, "interpolated coordination**"** **(**Quirk et al. 1985: 976-977), or "elliptical parenthetical clause" (Peterson 1999: 232), as in (29b).

> (29) a. And I uh used to get *maybe it was* five five pounds as a birthday present (s1a-076-78)
>
> b. He is, *or at least he was*, a great actor

As with RPCs, the link between host and PC in these cases is established by a conceptual 'gap' which is, however, not filled by the entire host clause but by

a complement in the host clause. This link is arguably much stronger than with RPCs, which can be seen from the fact that the PC does not have its usual positional flexibility but is tied to one specific position in the host clause. The inserted clause can also be questioned with a tag question (*He is, or at least he was, a great actor, wasn't he?*), which is normally impossible with PCs. Cases of amalgamation and Right Node Raising are therefore excluded from the present taxonomy of PCs.

4.3   Positional Flexibility

The third syntactic criterion for delimiting the class of PCs is that of their positional flexibility, a direct result of their lack of syntactic attachment. It has variously been pointed out (e.g. Ziv 1985: 182, Corum 1975: 137, Peterson 1999: 237, Urmson 1952) that PCs are not tied to a specific position in the host clause but may occur in several locations, i.e. in initial, medial or final position.[5]

(30) a. *It must be admitted*, Mary is very clever

 b.   John, *I believe*, is coming to the party

 c.   Paul is having a good time, *it seems*

The general flexibility notwithstanding, there seem to be certain constraints on what position within the host can serve as a 'niche' for the insertion of a PC (e.g. Emonds 1973: 335-336, McCawley 1998: 751, Peterson 1999: 238-240). Compare, for instance, the following examples (from Emonds 1973: 335, ex. 12).

(31) a. *He likes every, *I believe*, friend of John

 b.   He likes, *I believe*, every friend of John

(32) a. *John pushed, *they claimed*, a child into the street

 b.   John pushed a child, *they claimed*, into the street

 c.   John, *they claimed*, pushed a child into the street

(33) a. *F. Mitterand was in front of – *there was no other place to be* – the Parliament

---

5   Some scholars use the term 'parenthetical' exclusively for interpolated juxtaposition, while end-to-end sequencing is referred to, for instance, as "peripheral juxtaposition" (Peterson 1999) – a practice that is not adopted in the present study. Cf. also the distinction between 'interpolation' and 'appendage' by Huddleston & Pullum 2002: 1355.

b. F. Mitterand was in front – *there was no other place to be* – of the Parliament   (Espinal 1991: 735, ex. 54a)

Emonds (1973: 335) tries to capture these restrictions within his 'Parenthetical Formation' transformational rule by specifying that what follows the parenthetical in the host clause must be a phrasal constituent. This approach, however, takes the PC to be a constituent of the host clause, in contrast to the non-syntagmatic view outlined above (in Section 4.2).

What the examples (31) – (33) indicate is that there are certain 'weak spots' in the syntax of the host clause which admit insertion of PCs more readily than others. Restrictions on interpolation, although correlating with syntactic constituent structure, however, seem to be mainly a matter of performance and processing constraints (as suggested for instance by Espinal 1991: 753 and Peterson 1999: 239). Such a view is supported by the fact that the asterisked examples in (31) – (33) can be improved by either changing the semantic type of PC into a metalinguistic comment or by increasing end-weight of the host clause or both, as can be seen from the following acceptable versions.

(34)   He likes every, *I hope this is the right word / if I may say so*, friend of John

(35) a.   John pushed, *I still find it hard to believe / you probably won't believe me,* a child into the street

b.   John pushed, *I believe*, a child with a red anorak and a woolly hat into the street

(36)   F. Mitterand was in front of, *you probably won't be surprised to hear this*, the French Parliament in Paris

The question of preferred positions of PCs in the host clause is also closely connected with the scope of a parenthetical. What is meant by scope here is of course not a syntactic relationship (defined in terms of c-command), but a semantic-pragmatic association between the PC and some 'anchor' in the host clause. Depending on the position of the PC its scope may change as in (37a), where it covers the entire host clause, and (37b), where it may also be over the NP *some prints* (depending on context and delivery).

(37) a.   You could *I suppose* commission some prints of you yourself (s1a-015-37)

b.   You could commission *I suppose* some prints of you yourself

The change in scope, depending on the position in the host clause, could be taken as an indication that PCs are in fact less independent of syntax than was suggested in Section 4.2. However, the scope of the PC can be shown to depend on the semantic value of the parenthetical and potential anchors in the host. Compare, for instance, examples (38a) and (38b), where the scope

changes with the different (semantic) type of PC from being over the entire host clause in the former to only the preceding NP *his strange and unexpected behaviour* in the latter.

(38) a. His strange behaviour – *it must be admitted* – caused a terrible outrage

b. His strange behaviour – *what else would you call turning up for dinner dressed as a smurf* – caused a terrible outrage

Similarly, in the different versions of (39) the scope of the PC (indicated by underlining) is manipulated by the choice of lexical items.

(39) a. His strange behaviour caused a – *it must be admitted* – <u>terrible outrage</u>

b. His strange behaviour caused a – *I'm sorry to say* – a slight <u>outrage</u>

c. His strange behaviour caused a – *I'm relieved to say* – a <u>slight</u> outrage

While syntactic flexibility is not a necessary characteristic of PCs (syntactic independence alone is already sufficient), it identifies the core/prototypical members of the class. These are RPCs, self-contained PCs, and 'adverbial' PCs. With RPCs, however, there is a particular problem for delimitation if they occur in clause-initial position. I will briefly discuss this in the following. Unlike medial and final RPCs, as in (40b) and (40c), initial RPCs (40a) are difficult to distinguish from matrix clauses, especially if the *that*-complementizer has been omitted.

(40) a. *I suppose* (that) John has come back from London

b. John has come back, *I suppose*, from London

c. John has come back from London, *I suppose*

Various different views have been expressed on the status of such initial clauses which allow shifting to other locations in the sentence. For constructions with a *that*-complementizer the following three views can be found. They are either taken to be parenthetical (e.g. Ross 1973, Thompson 2002, Kärkkäinen 2003), ambiguous, i.e. allowing interpretation as both matrix clause and PC depending on context and type of 'matrix' predicate (e.g. Urmson 1952: 481, Recanati 1984: 326, 346, 348, Aijmer 1972: 46, Quirk et al. 1985: 1113, Biber et al. 1999: 197, Huddleston & Pullum 2002: 896), or matrix clauses (e.g. Stenström 1995: e.g. 293, 296).

For constructions without a *that*-complementizer exactly the same three views can be found. According to some studies complementizer omission is a clear indication of a parenthetical use of the first clause (e.g. Kruisinga 1932: 486, Thompson & Mulac 1991). In another view, cases of *that*-omission are treated as ambiguous, allowing a double reading as parenthetical and matrix clause (e.g. Biber et al. 1999: 197). A third view takes them as matrix clauses only (e.g. Svensson 1976: 375, Peterson 1999: 236).

Most studies, however, seem to agree that a parenthetical use of initial clauses without *that* can unambiguously be assumed in cases where they are separated by a pause (or comma) from the following clause (e.g. Peterson 1999: 236, Biber et al. 1999: 197). Other prosodic marking, for instance a separate tone unit, also seems to strongly suggest a parenthetic reading but are no guarantee (cf. e.g. Bolinger 1989: 186, Stenström 1995: 292).

The reason for the unclear status of these initial clauses is that we are dealing here with varying degrees of grammaticalization (cf. Thompson 2002, Thompson & Mulac 1991) or rather, as Aijmer (1997) puts it, "pragmaticalization", depending on the lexical item involved. The difference between initial matrix clauses and initial PCs is therefore not necessarily a question of either-or but is best represented on a scale of gradience, as illustrated in Figure 1.

Figure 1. Scale of gradience distinguishing sentence-initial matrix clauses from PCs

Matrix clause	*I suppose* that John has come back from London
	I suppose John has come back from London
	I suppose, John has come back from London
Parenthetical clause	John has come back from London, *I suppose*

This of course raises the question of where to draw the line between the two categories. To answer this question let us briefly consider the evidence used in support of the different views.

One way to show that there is a difference between initial clauses with a *that*-complementizer and those without is by way of the **tag-question test** (e.g. Aijmer 1972: 52, 1997: 8, Hand 1993: 501, Knowles 1980: 405). The argument is that the 'subordinate' clause in (41) has lost some of its subordinate status since it allows various 'main clause phenomena' (Hooper-Thompson 1973, Green 1976), such as the tag question.

(41)   *I think* ∅ John is in London, isn't he

However, the same seems to be true also for sentences with *that*, as in (42) (Aijmer 1997: 8), although Hand (1993: 501) marks it as questionable.

(42)   *I think* that John is in London, isn't he

Conversely, the 'matrix clause' in a sentence without *that*, as in (43a) does not seem to allow questioning in this way (Aijmer 1997: 8, Knowles 1980: 405). This is equivalent to the behaviour of 'real' PCs, as in (43b) and can be taken as an indication that the clause lacks illocutionary force.

(43) a.   *I think* ∅ John is in London, *don't I

       b.   John is in London, *I think*, *don't I

The validity of this test, however, is somewhat questionable, as the unacceptability of the tag in (43) could also be attributed to a pragmatic restriction, viz. the inappropriateness of a speaker questioning (doubting) his/her own expression of belief or 'cognition'. Indeed, if we substitute a pragmatically more likely tag, as in (44), the result is a different one in both cases.

(44) a. *I think* John is in London, don't you

    b. John is in London, *I think*, don't you

Further tests are the **question test** (Asher 2000: 33, Huddleston & Pullum 2002: 896) and the **negation test** (Erteschik-Shir & Lappin 1979: 46). All in all, however, acceptability of the examples in these tests varies and much may depend on prosodic delivery. They can therefore not be taken as providing conclusive evidence.

Corpus evidence for a difference in status between 'matrix' clauses with *that* and without is provided by Thomson and Mulac (1991), who show a correlation between *that*-omission and grammaticalization of the 'matrix' clause into an epistemic (parenthetical) phrase, which typically takes the form of high frequency predicates such as *think* and *guess* in the first or second person singular. In other words, much depends on the type of lexical verb. Moreover, while suggesting that clauses without *that* function as parentheticals, this does not suggest that clauses with *that* automatically don't. In fact, Thompson (2002) expands the view expressed by Thompson and Mulac (1991) that matrix clauses without *that* function as formulaic stance markers (i.e. parentheticals) to all complement-taking predicates, no matter whether they take a *that* complementizer or not.

Finally, evidence for the syntactic equivalence of initial clauses (with and without *that*) and non-initial ones comes from their alleged **equal function** (Kärkkäinen 2003: 41). However, since no systematic functional comparison has been carried out, this assumption has to be taken with some caution, too.

In the absence of hard and fast evidence for any of the positions the present study takes a cautious approach and excludes all instances of initial clauses with a *that*-complementizer from the class of PCs. Initial clauses without *that* are only taken into account if they are clearly separated from the complement/host clause by means of a pause or some intervening material such as hesitation sounds (*uh, uhm*) or other fillers (*you know, I mean*). This restrictive policy, on the one hand, avoids the inclusion of possible ambiguous examples and, on the other, allows for a more systematic investigation of potential differences in use between instances with and without *that*.

5. Related categories

In the process of delimiting the class of PCs we have narrowed down the category of potential candidates by excluding all non-clausal parentheticals as well as syntactically dependent clauses and finally by positing the criterion of positional flexibility for core members of the class. As a final step in the delimitation process I will now distinguish PCs from three closely related categories, viz. interrogative tags, discourse markers, and anacolutha.

5.1 Interrogative tags

As clausal constituents interrogative tags look superficially similar to the class of PCs. There are indeed a number of similarities. Ziv (1985: 189), for instance, detects a functional similarity between interrogative PCs and tags. Compare the following two examples.

> (47) a.   The project should be finished by next week, *wouldn't you agree?*
>
>   b.   You are not proposing to finish the project by next week, *are you?*

Knowles (1980: 381-384), on the other hand, suggests that tags are a type of truth parenthetical (e.g. *it is true*), with which they share a number of syntactic and semantic properties. In the present framework, however, tag questions are not included in the class of PCs, as some properties of question tags seem to indicate a syntactic link to their host clause:

(a) Anaphoric reference: the pronoun of a tag question has obligatory anaphoric reference. However, this property has to be taken with caution since anaphoric processes can be relegated to the area of discourse rather than sentence grammar (cf. Ziv 1985: 189-190, Knowles 1980: 383) and therefore do not represent clear evidence for a syntactic link.

(b) Reversed polarity of question tags: a negative tag attaches to a positive host clause and vice versa. As argued by Ziv (1985: 190), however, this property seems to be pragmatically motivated rather than conditioned by sentence grammar. This view is also supported by the existence of so-called "constant polarity tags" (Huddleston & Pullum 2002: 892), e.g. *He is ill, is he; He isn't ill, isn't he.* [6]

(c) Tense constraints: the tag preserves tense and aspect of the host. Thus, a progressive in the host clause requires a progressive in the tag, e.g.

---

[6]   The latter example is rejected by many native speakers. However, it cannot be ruled out completely; cf. Huddleston (1970), Knowles (1980: 380), Huddleston & Pullum (2002: 892).

(48) a   *You are working for IBM, *don't you?*

    b   *You work for IBM, *aren't you?*

(d) Elliptical form dependent on the syntactic form of host clause: The syntactic link between tag and host clause is particularly obvious in the way in which the exact (elliptical) form of the tag is shaped by the form of the host VP, more precisely the type of main verb and possible auxiliaries (e.g. *John is ill, isn't he/*doesn't he/*hasn't he*). This contrasts with the relative freedom in the choice of (lexical) verbs in PCs (e.g. *John is ill, I think/suppose/reckon*).

(e) Sentence-final position: owing to their close dependence on the form of the host VP, tags typically occur in sentence-final position (hence the term): unlike PCs, a tag can only be added once the form of the host VP has been established, as illustrated in (49).

(49) a.   **Hasn't he*, John has worked hard

      b.   *John, *hasn't he*, has worked hard

      c.   ???John has, *hasn't he*, worked hard

      d.   ??John has worked, *hasn't he*, hard

      e.   John has worked hard, *hasn't he*

Although Knowles (1980: 383) argues that the reduced movability of tags can be explained pragmatically in terms of anaphoric constraints on the pronominal subject and the auxiliary of the tag, claiming that "these anaphoric elements can only occur to the right of their anaphors", this is questionable given the possibility of cataphoric reference for personal pronouns.

5.2   Discourse markers

The category of discourse markers has been discussed under a wide range of different headings with discourse marker (e.g. Schiffrin 1987, Blakemore 2002) only being one of them. Other terms are discourse particle, pragmatic particle, pragmatic marker, pragmatic expression, discourse connective (cf. Jucker & Ziv 1998: 1-5 for an overview). Not surprisingly, there is no uniform, generally agreed definition either, and considerable disagreement on which items are to be included in the category of discourse markers. Typical examples, as discussed in the literature (cf. Brinton 1996: 32, Aijmer 2002: 2 for inventories), are the following: *actually, ah, (all) right, I mean, like, ok, really, sort of, you know well,* etc.

From these selected examples it is clear that overlap with the class of PCs is minimal: most discourse markers are single non-verbal items and as such are formally distinct from PCs, which are a clausal category, i.e. contain a

verb. Possible overlap between the two categories is restricted to a small set of elements, more precisely: *I mean, I see, I think, you know, (you) see, mind you, look, listen*. All of these have previously been discussed under the heading of discourse marker (e.g. Östman 1981, Erman 1987, Schiffrin 1987, Schourup 1985, Palander-Collin 1999) as well as under the heading of comment clause (cf. Table 1 in Section 2). The reason for regarding these expressions as discourse markers lies in their formulaic nature, which is a direct result of the process of grammaticalization that has given rise to them in the first place. Grammaticalization is of course equally at work with RPCs such as *I guess, I think, I suppose* (e.g. Thompson & Mulac 1991), but unlike discourse markers, which have already completed the process, RPCs are not yet fully grammaticalized. While it is difficult to decide where exactly to draw the dividing line between fully grammaticalized clauses (i.e. discourse markers) and those that may just be on their way to full grammaticalization (i.e. RPCs), the practice adopted in the present study is that of excluding all of the above listed expressions from the class of PCs, with the exception of *I think*. *I think* is not counted as a discourse marker for the following reasons: First, parenthetical *I think* – despite its relatively fixed character as independent epistemic fragment – still permits considerable variation in form, as is evidenced by the following attested variants: *I don't think, I certainly/just think, I would/should think, I thought, we think*. Such variations are excluded from typical discourse markers such as *I mean*, *you know*; only in other uses (e.g. as matrix clause) is some variation possible (albeit less so than with *I think*; cf. *I don't/didn't mean, I meant, ?I certainly/just mean, ?we mean, *I would/should mean*). Second, *I think* differs from typical discourse markers in terms of distribution and possible syntactic functions. As pointed out by Stenström (1995: 293, 296), *I think* occurs substantially more often in interpolated position than *I mean, you know, you see* and is also exceptional in its greater likelihood to act as a matrix clause.

5.3   Anacoluthon

Another area of potential overlap is between PCs and anacolutha (or syntactic blends), which are also a typical feature of spoken language. In both cases the speaker leaves a construction incomplete: in the case of PCs only temporarily so, with the construction being resumed and completed after the parenthetical insertion; in the case of anacoluthon it stays incomplete. An example is given in (50), which is in fact a case of multiple blending with two incomplete constructions (indicated by underlining).

> (50)   So when you graduated then <u>when did</u> because you said I think coming up
>   the stairs that that you've been applying for lots of different jobs (s1a-034-
>   144)

The difference between PCs and anacolutha is thus that the former follow a clear and relatively predictable pattern, whereas the latter fall into the category of performance error, caused by working memory limitations (Biber et al. 1999: 1065). Despite the clear structural difference between the two constructions, there may be cases where identification of the corpus material as one or the other is difficult.

The problem for delimitation may be two-fold. First, there are cases where the first clause (host clause) remains incomplete but where the intervening second clause can be seen as part of the first, as in example (51).

> (51)   Now Dickens setting his uh Tale of Two Cities has the meeting of the <u>I</u>
>   <u>forget the name of the girl or the man</u>   indeed so uh probably boring both of
>   them uhm characters (s1a-020-203)

What might look like an anacoluthon, owing to the incompleteness of the first clause, is, however, best classified as an instance of a semantic gap filling PC (cf. Section 4.2), as here the second clause completes a semantic gap in the first.

Second, there are cases where a clause simply interrupts the first, without filling a semantic gap, and the resumption of the first is not clear, as illustrated in (4).

> (52)   The only problem with that Estelle Road place <u>I mean I'd still be happy to</u>
>   <u>have it at the price that I offered</u> <short PAUSE> uhm is the other thing that
>   we'll need <long PAUSE> about all those places (s1a-023-207)
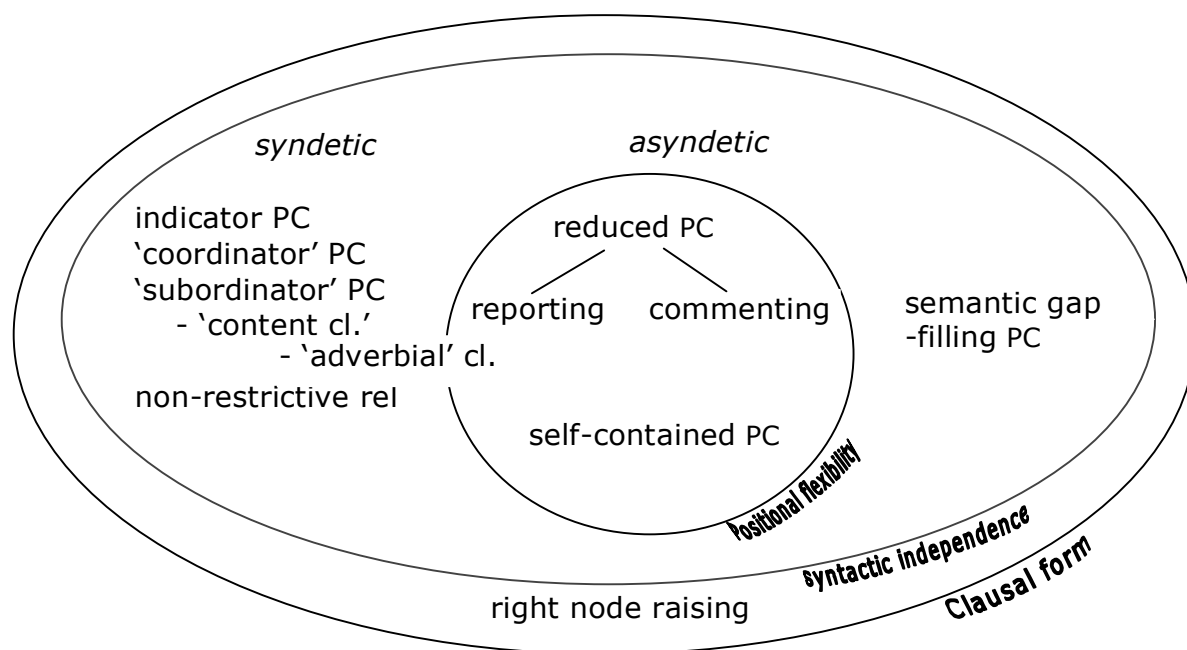
Example (52) is a typical case where the speaker loses track of the original construction after a lengthy insertion (of 16 words), as is indicated by the hesitation (cf. pause and hesitation sound). However, the speaker still manages to pick up the original (reversed pseudo-cleft) construction, which is shown by the use of *is* after the insertion (cf. *The only problem is...*), only to abandon it after that. Since there is a clear sign of resumption (even if abandoned), examples such as these are classified as PCs rather than anacolutha.

6. Summary

The aim of this article has been to delimit and systematise the class of parenthetical clauses, which lacks a clear definition in the literature (cf. Section 2). For methodological reasons (cf. Section 3) this has been done with the help of purely syntactic criteria: clausal form (4.1) and lack of syntactic

attachment (4.2) as two necessary characteristics and positional flexibility (4.3) as an additional feature of core members of the class. This procedure has enabled us not only to delimit the class of PCs from other, related categories (cf. e.g. Section 5), but also to establish the internal stratification of this class in the form of a taxonomy of subtypes, which is summarised in Figure 2.

Figure 2. Syntactic types of parenthetical clauses

This taxonomy, it is hoped, will provide an operational basis for the identification and corpus extraction of PCs. It will subsequently have to be tested against the corpus data and, if necessary, further adapted and refined.

References

Aijmer, Karin. 1972. *Some aspects of psychological predicates in English*. Stockholm: Almqvist & Wiksell.

Aijmer, Karin. 1980. *Evidence and the declarative sentence*. Stockholm: Almqvist & Wiksell.

Aijmer, Karin. 1997. "*I think* – an English modal particle", in : Swan, T.; Westvik, O. J. (eds.). *Modality in Germanic languages. Historical and comparative perspectives*. Berlin: Mouton de Gruyter, 1-47.

Aijmer, Karin. 2002. *English discourse particles. Evidence from a corpus*. Amsterdam: John Benjamins.

Asher, Nicholas. 2000. "Truth conditional discourse semantics for parentheticals", *Journal of Semantics* 17 (1): 31-50.

Biber, Douglas et al. 1999. *Longman grammar of spoken and written English*. Harlow: Longman.

Blakemore, Diane. 2002. *Relevance and linguistic meaning. The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1950. *Language*. New York: Holt.

Bolinger, Dwight. 1989. *Intonation and its uses. Melody in grammar and discourse*. London: Edward Arnold.

Brinton, Laurel J. 1996. *Pragmatic markers in English. Grammaticalization and discourse functions*. Berlin: Mouton de Gruyter.

Burton-Roberts, Noel. 1994. "Apposition", in: Asher, R. E. (ed.). *The encyclopaedia of language and linguistics*. Oxford: Pergamon Press, 184-187.

Burton-Roberts, Noel. 1999. "Language, linear precedence and parentheticals", in: Collins, Peter; Lee, David (eds.). *The clause in English*. Amsterdam and Philadelphia: Benjamins, 33-52.

Chomsky, Noam A. 1986. *Knowledge in language. Its Nature, Origin and Use.* New York: Praeger.

Corum, Claudia. 1975. "A pragmatic analysis of parenthetic adjuncts", *Chicago Linguistic Society Papers* 11: 133-141.

Emonds, Joseph. 1973. "Parenthetical clauses", in Corum, C. et al. (eds.). *You take the high node and I'll take the low node*. Chicago: Linguistic Society, 333-347.

Emonds, Joseph. 1979. "Appositive relatives have no properties", *Linguistic Inquiry* 10: 211-243.

Erman, Britt. 1987. *Pragmatic expressions in English. A study of* you know, you see *and* I mean *in face-to-face conversation*. Stockholm: Almqvist & Wiksell.

Erteschik-Shir, N., Lappin, S. 1979. "Dominance and the functional explanation of island phenomena", *Theoretical Linguistics* 6 (1): 41-86.

Espinal, Teresa M. 1991. "The representation of disjunct constituents", *Language* 67 (4): 726-762.

Fabb, Nigel. 1990. "The difference between English restrictive and nonrestrictive relative clauses", *Journal of Linguistics* 26: 57-78.

Fowler, H. W. 1906. *The King's English*. Oxford: Clarendon Press.

Green, Georgina M. 1976. "Main clause phenomena in subordinate clauses", *Language* 52: 382-397.

Haegeman, Liliane. 1991. "Parenthetical adverbials: the radical orphanage approach", in: Chiba, Shukin et al. (eds.). *Aspects of modern English linguistics. Papers presented to Masamoto Ukaji on his 60th birthday*. Tokyo: Kaitakushi, 232-254.

Hand, Michael. 1993. "Parataxis and parentheticals", *Linguistics and Philosophy* 16: 495-507.

Hoffmann, Ludger. 1998. "Parenthesen", *Linguistische Berichte* 175: 299-328.

Hooper, Joan B. 1975. "On assertive predicates", in: Kimball, J. P. (ed.). *Syntax and Semantics*, Vol. 4. New York: Academic Press, 91-124.

Hooper, Joan B.; Thomson, Sandra A. 1973. "On the applicability of root transformations", *Linguistic Inquiry* 4: 465-498.

Huddleston, Rodney; Pullum, Geoffrey K. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, Mass. MIT Press.

Jackendoff, Ray. 1977. *X' syntax: a study of phrase structure*. Cambridge, MA: MIT Press.

Jespersen, Otto. 1924. *The philosophy of grammar*. London: Allen & Unwin.

Jespersen, Otto. 1940. *A modern English grammar on historical principles*. Part V. Copenhagen: Munksgaard.

Jespersen, Otto. 1973. *Analytic syntax*. London: Allen & Unwin.

Jucker, Andreas H., Ziv, Yael. 1998. "Discourse markers: introduction", in: Jucker, A. H.; Ziv, Y. (eds.). *Discourse markers: descriptions and theory*. Amsterdam: John Benjamins, 1-12.

Kärkkäinen, Elise. 2003. *Epistemic stance in English conversation*. Amsterdam: John Benjamins.

Knowles, John. 1980. "The tag as a parenthetical", *Studies in Language* 4: 370-409.

Kruisinga, E. 1932. *A handbook of present-day English*. Part II. Groningen: Noordhoff.

Lakoff, George. 1974. "Syntactic amalgams", *Papers from the tenth regional meeting*, Chicago: Linguistic Society, 321-343.

Leech, Geoffrey; Svartvik, Jan. 1975. *A communicative grammar of English*. Harlow: Longman.

McCawley, James D. 1982. "Parentheticals and discontinuous constituent structure", *Linguistic Inquiry* 13: 91-106.

McCawley, James D. 1998. *The syntactic phenomena of English* (2nd ed.). Chicago: University of Chicago Press.

Mittwoch, Anita. 1979. "Final parentheticals with English questions: their illocutionary function and grammar", *Journal of Pragmatics* 3: 401-412.

Nosek, Jirí. 1973. "Parenthesis in modern colloquial English"; *Prague Studies in English* 15: 99-116.

Östman, Jan-Ola. 1981. You know: *a discourse functional approach*. Amsterdam. John Benjamins.

Palander-Collin, Minna. 1999. *Grammaticalization and social embedding. I think and methinks in Middle English and Early Modern English*. Helsinki: Société Néophilologique.

Peterson, Peter. 1999. "On the boundaries of syntax: non-syntagmatic relations", in: Collins, Peter; Lee, David (eds.). *The clause in English*. Amsterdam and Philadelphia: Benjamins, 229-250.

Petola, Niilo. 1983. "Comment clauses in present-day English", in: Kajanto, Iiro et al. (eds.). *Studies in Classical and Modern Philology*. Helsinki: Suomalainen, Tiedeakatemia, 101-113.

Plank, Frans. 1981. "Modalitätsausdruck zwischen Autonomie und Auxiliarität", in: Rosengren, I. (ed.). *Sprache und Pragmatik. Lunder Symposium 1980*. Lund: CWK Gleerup, 57-71.

Quirk, Randolph et al. 1972. *A grammar of contemporary English*. London: Longman.

Quirk, Randolph et al. 1985. *A comprehensive grammar of the English language*. Harlow: Longman.

Recanati, François. 1984. "Remarques sur les verbes parenthetiques", in: Attal, P.; Muller, C. (eds.). *De la syntaxe à la pragmatique*. Amsterdam: Benjamins., 319-352.

Reinhart, Tanya. 1975. "Whose main-clause-point of view in sentences with parentheticals", in: Kuno, S. (ed.) *Harvard studies in syntax and semantics I*, 127-171.

Reinhart, Tanya. 1983. "Point of view in language: the use of parentheticals", in: Rauh, G. (ed.). *Essays on Deixis*. Tübingen: Gunter Narr, 169.194.

Ross, John R. 1973. "Slifting", in: Gross, Maurice, Halle, Morris, Schützenberger, Marcel-Paul (eds.). *The formal analysis of natural languages*. The Hague: Mouton, 133-169.

Safir, Kenneth. 1986. "Relative clauses in a theory of binding and levels", *Linguistic Inquiry* 17: 663-690.

Schelfhout, Carla. 2000. "Corpus-based analysis of parenthetical reporting clauses", in: Eynde, F. von et al. (eds*.). Computational Linguistics in the Netherlands 1998*. Amsterdam: Rodopi, 147-159.

Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: Cambridge University Press.

Schneider, Stefan. 2005. *Reduced parenthetical clauses. A corpus study of spoken French, Italian and Spanish*. [unpublished postdoctoral thesis, University of Graz].

Schourup, Laurence C. 1985. *Common discourse particles in English conversation*. New York: Garland

Skrebnev, J. M. 1959. "Parenthese und Absonderung", *Zeitschrift für Anglistik und Amerikanistik* 7: 58-63.

Sperber, Dan; Wilson, Deirdre. 1995. *Relevance: communication and cognition* (2nd ed.). Oxford: Blackwell.

Stenström, Anna-Brita. 1995. "Some remarks on comment clauses", in: Aarts, B.; Meyer, Ch. F. (eds.). *The verb in contemporary English*. Cambridge: Cambridge University Press, 290-299.

Svensson, Jan. 1976. "Reportindicators and other parentheticals", in: Karlsson, F. (ed.). *Papers from the 3rd Scandinavian Conference of Linguistics*. Turku: Textlinguistics Research Group, Academy of Finland, 369-380.

Thompson, Sandra A. 2002. "'Object complements' and conversation. Towards a realistic account", *Studies in Language* 26 (1): 125-164.

Thompson, Sandra A.; Mulac, Anthony. 1991. "The discourse conditions for the use of the complementizer *that* in conversational English", *Journal of Pragmatics* 15: 237-251.

Traugott, Elizabeth Closs. 1989. "On the rise of epistemic meanings in English: an example of subjectification in semantic change", *Language* 65 (1): 31-55.

Urmson, J. O. 1952. "Parenthetical verbs", *Mind* 61: 480-496.

Widdowson, Henry G. 1997. "The use of grammar, the grammar of use", *Functions of Language* 4 (2): 145-168.

Wiechmann, Anne. 2001. "Spoken parentheticals", in: Aijmer, K. (ed.). *A wealth of English. Studies in honour of Göran Kjellmer*. Göteborg: Acta Universitatis Gothoburgiensis, 177-193.

Ziv, Yael. 1985. "Parentheticals and Functional Grammar", in: Bolkestein, A.M. et al. (eds.) *Syntax and Pragmatics in Functional Grammar*. Dordrecht: Foris, 181-199.

Vocabulary learning through reading: influences and teaching implications

Angelika Rieder, Vienna*

1. Introduction

In second language acquisition, part of a learner's vocabulary is usually acquired through intentional teaching and learning. Intentional learning, however, cannot account for all the words second language learners know; this implies that a considerable number of words are acquired incidentally, e.g. as a by-product of reading second language texts, by guessing their meaning from context[1] (cf. Nation & Coady 1988, Nation 2001). The assumption that vocabulary is acquired through reading is also reflected in teaching methodology (cf. Krashen 1989), where particularly for more advanced learners, the tacit assumption seems to be that vocabulary acquisition happens primarily as a side-effect of reading activities. As we shall see, however, this established view appears to be oversimplifying the complexity of the processes and influences involved.

Naturally, most second language texts learners read will contain words they do not know. While research usually focuses on the fact that some of these words <u>are</u> learned incidentally (Hulstijn, Hollander & Greidanus 1996; Knight 1994), it is just as interesting to note that the majority of the unknown words encountered in a text are in fact <u>not</u> acquired – because they are not attended to, because the learners cannot guess their meanings, or because they simply do not remember them after reading. The present paper investigates the various variables involved in vocabulary learning through reading and attempts to translate the insights gained in this analysis into suggestions for relevant teaching practices which can enhance the conditions for incidental vocabulary acquisition.

---

[1] In this article, the term *context* is used to cover the more general 'surrounding' of a word, including linguistic, situational as well as world knowledge aspects. The term *co-text*, in turn, is used when specifically referring to narrow, purely linguistic context.

2. How learners deal with unknown words in texts

When reading a second language text, the meanings of the words in the text will generally form one of the bases on which a learner constructs the text meaning. Thus, if he/she encounters an unfamiliar word, no meaning will be available for this word, and consequently the learner's mental representation of the text meaning will remain unspecified with regard to this particular word meaning gap.

What possible learner reactions can we imagine in this case? If the word only contributes to the text meaning peripherally, it is likely to be skipped (or perhaps not even noticed) as it does not impede the learner's text comprehension (Stein 1993). Only if the word meaning is considered sufficiently important for determining the text meaning, will it attract the learner's attention and he/she will invest some effort in inferring the unknown meaning through contextual clues. Thus, from the text comprehension perspective, whether a learner actually decides to focus on an unknown word and attempts to figure out its meaning at all will primarily depend on the word's (estimated) importance for the text meaning (cf. Rieder 2002: 56).

Another factor to be taken into account is that in text comprehension, learners will generally concentrate on new word meanings only as a means for reaching their ultimate goal of comprehending the text (cf. Beaugrande & Dressler 1981: 108) unless they are personally interested in acquiring this new vocabulary item. This focus on the text meaning level has two implications: on the one hand, it implies that learners will stop trying to specify unknown meanings as soon as they feel that the specification is sufficient for comprehending the text, and on the other hand they will not automatically form a connection between the meaning contribution they have specified and the unknown word. Thus, the fact that learners have inferred an unknown word meaning from the context during the reading process does not automatically imply that they have also acquired this new word; moreover, they will frequently not even remember the word after reading unless they have taken this active step from the text meaning to the word meaning level and focused on the word itself (cf. Rieder 2002: 65, 2004: 66).

This phenomenon can be illustrated by text passage (1), in which the word *dungeon* is assumed to be unknown to the learner:

(1)   Finally he was arrested and punished for his crimes. He was kept in a dungeon for 25 years.

When trying to comprehend this passage, clues from the co-text (e.g. arrested, punished for crimes, kept in, for 25 years) and the learner's world knowledge will enable him/her to specify the overall meaning of the passage as describing that a male criminal was captured and locked up in some sort of prison for 25 years. But this does not automatically imply that the learner has acquired any knowledge about the meaning of the word *dungeon.* In order to learn this new vocabulary item, he/she has to focus on the word meaning level, isolate the meaning contribution of the word *dungeon* from its context and form a mental connection between the word form and the specified meaning.

These characteristics of text comprehension have two important implications for incidental vocabulary acquisition: firstly, that an unknown word has to attract the learners' attention, otherwise they will not attempt to infer its meaning at all; and secondly, that for meaning inference to lead to vocabulary acquisition, a link has to be formed between the word form and the meaning guess, which is a step not automatically taken in text comprehension.

All in all, the points made so far show that the chances for successful vocabulary learning through reading are affected by various conditions and forces. In a next step, we will therefore investigate the specific factors which may enhance or hinder the chances of learning unknown words from reading.

3. Factors influencing the chances that a word is acquired

In an attempt to enumerate the factors influencing incidental vocabulary acquisition, two basic dimensions have to be distinguished. One dimension concerns the quantity and quality of the meaning specification and learning process: the *quantity* of learning that will take place will depend on the focus or attention a learner decides to devote to an unknown word, whereas the *quality* of meaning specification will be determined by the degree of possible specification of the word meaning through meaning clues in the text etc. Both focus and elaboration factors will in turn be determined by interacting conditions on the text level, the learner level and the situation level (cf. Rieder 2002: 69).

On the textual level, the amount of attention a learner devotes to a word will be influenced by factors such as a word's (estimated) centrality for the text meaning (Hulstijn 1993) or its number of occurrences in the text (or in previous texts) (Sternberg 1987). The degree to which a learner can infer an unknown word meaning, on the other hand, will depend on the availability and nature of the available meaning clues.

Generally, meaning clues for unknown words can come from three different sources (cf. Haastrup 1991). They may either come from the word form itself or the learner's knowledge of English (intralingual clues); e.g. the meaning of *operative* can probably be deduced by a learner who is familiar with the word *operation* and the suffix *-ive*. Alternatively, clues may result from similarities between the word and cognates in other languages known to the learner (interlingual clues) – in the case of so-called 'false friends' (compare English *sensible* and German *sensibel* (sensitive)), however, these clues can also be misleading. Thirdly, surrounding words in the text or supplementary world knowledge of the learner may offer hints to the meaning of unknown words (contextual clues).

With regard to this third category, a point of criticism often levelled at the conditions for guessing the meaning of unknown word in texts is that the word meaning clues contained in a text are frequently insufficient or even misleading (Schatz & Baldwin 1986). Comparing, for instance, the contextual constellations for the word *dusky* in sentences (2) and (3), the different nature of these clues can be illustrated:

(2)   They enjoyed the late **dusky** summer evening.
(3)   He returned to the **dusky** room.

In sentence (2), the key clues 'summer evening' and 'late' together with the learner's world knowledge would be valuable contextual clues to narrow down the meaning of *dusky*, whereas in sentence (3) no specific clues are available which are characteristic for the word meaning (here, properties of a room could just as well be 'untidy' or 'small'), so that in the latter case the co-text of 'dusky' does not offer any help for narrowing down the meaning. However, even in example (2) it is unclear which quality of a 'late summer evening' the word *dusky* refers to (e.g. the quality 'mild' would also be appropriate), so that the final meaning guess of a learner who does not know

the word will probably remain vague or even be erroneous.[2]

Apart from the meaning clues in the surrounding co-text, another major influencing factor to be taken into account is the overall percentage of unknown words in the text. The amount of unknown words will influence both text comprehension and the learner's ability to infer unknown word meanings. The optimal ratio of unknown words in a text is generally held to be 2% (this means that 98% of the words should be known to the learner, cf. Nation & Coady 1988: 97). This ratio, however, can vary with regard to the particular reading goal (see below).

With regard to learner-specific influencing variables, the focus on a word will primarily be determined by individual factors like the learners' general motivation to extend their vocabulary or their particular interest in a specific word. As to the successful specification of unknown word meanings, in turn, the learners' background knowledge and language knowledge are vital for exploiting meaning clues in the text. Finally, the strategies learners employ for figuring out unknown word meanings have been shown to vary considerably from learner to learner (cf. Haynes 1993). Consequently, even if the surrounding co-text offers ample clues to the meaning of a word, not every learner will be equally skilled in exploiting their full potential.

Besides textual and learner-specific influences, a third variable to be taken into account is the reading situation: here, factors such as, for example, the specific reading goal will determine the amount of attention a learner pays to individual words. Activities which can be distinguished in terms of reading situation are e.g. intensive reading (involving the close deliberate study of short texts with special attention paid to language features, and thus to vocabulary) and extensive reading (i.e. reading long texts with the aim of globally understanding the text, cf. Nation 2001: 149). The particular reading goal perceived by the learner will thus influence the degree of attention he/she devotes to specific words in the text. With regard to language teaching, factors related to the teaching/learning context created for a text in a particular learning situation will be further influences.

---

2 An important distinction in this connection is the difference between the abstract, denotational meaning of a word on the one hand, and the specific, contextual meaning it takes on in a text on the other hand. Different terms are used in the literature to characterize these phenomena (e.g. *symbolic* vs. *indexical* meaning, cf. Widdowson 1983, or *instantiation* of contextual meaning, cf. Anderson & Shifrin 1980). In Widdowson's terms, we could say that the learner firstly forms hypotheses about the indexical meaning of 'dusky' in the texts and then attempt to draw conclusions about the symbolic meaning based on these hypotheses.

Finally, we always have to keep in mind that acquiring vocabulary should be regarded as a gradual process, and that a single encounter with a new word – if at all – is only sufficient for building up a first, vague knowledge basis. Thus, reading and inferring an unknown word meaning once is not sufficient for vocabulary acquisition, but repeated exposure to the same word in different contexts is vital for meaning elaboration and knowledge rehearsal (cf. Hulstijn 2001: 284). Furthermore, we need to keep in mind that mere exposure to words in texts can generally only be expected to lead to a receptive level of vocabulary knowledge

In the light of all the complex moderating variables mentioned, it becomes clear that incidental vocabulary acquisition is not at all a straightforward process that merely 'happens'. Since various influencing factors appear to hinder this process, I will suggest some relevant guidelines for teacher actions or classroom activities which aim at optimizing the conditions for acquiring vocabulary through reading.

4. Teaching implications

4.1 Choosing appropriate texts

Choosing appropriate reading material for the level and interests of the targeted learner group will form a major prerequisite for improving both reading comprehension and reading motivation, both of which will positively affect incidental vocabulary acquisition. In line with the optimal ratio mentioned before, texts chosen for reading activities should only contain 2% of unknown words. Nation (2001: 261), however, states that learners can work with texts containing a percentage of unknown words of up to 5% when practising guessing from context, and that intensive reading tasks can even be planned with a coverage of less than 95% (p.150). One possibility of ensuring that the percentage of unknown words does not exceed this threshold would be the use of graded readers or, in individual cases, the simplification of texts by substituting low frequency words with known words in order to increase the coverage.

4.2 Integrating meaning information for new vocabulary

As has been mentioned above, most meaning clues to unknown words encountered in texts will only offer partial insights into the meaning of unknown words. In order to overcome these limitations, some researchers have suggested elaborative modification of texts through e.g. the addition of redundancy through paraphrase or synonyms. Elaboration is best seen as a

complementary alternative to simplification which is preferred by some researchers, as it is said to retain more of the original text (Yano, Long & Ross 1994). However, it should be kept in mind that elaboration will always result in a text which is longer than the original.

Another way of integrating meaning clues would be the integration of glosses (brief descriptions or synonyms in the L1 or L2) for unknown words in the text. Glossing is certainly an attractive mechanism for annotating unknown vocabulary items, as it allows for a condensed and precise provision of meaning information which can be retrieved quickly. But while glossaries can contribute to vocabulary learning and constitute a useful tool for enlarging learners' knowledge of word meanings, they will not be helpful for learners to develop their abilities to cope with unknown words in non-annotated texts.

4.3 Directing learner attention to unknown words in the text

Selecting appropriate texts and giving sufficient meaning information for unknown words are certainly prerequisites for successful guessing. But in order to make sure that words which are deemed important by the teacher are not skipped in the reading process, reading goals should be chosen which ensure reader focus on these specific words. Attention can be drawn to words on different levels (cf. Nation 2001: 252): either on the level of the word form by presenting a vocabulary list before reading, or on the level of word meaning by preteaching or glossing the respective words. Thirdly, learner focus can be directed to specific words through content related activities such as questions about specific features of the text presented before reading.

4.4 Including complementary vocabulary activities

Complementing reading tasks with vocabulary exercises can improve vocabulary learning chances on two levels: on the one hand it directs learner attention towards the target words, and on the other hand it ensures the elaboration and rehearsal of acquired knowledge. Depending on the knowledge level addressed, text-related follow-up exercises can range from matching words with definitions or synonyms to morphological analyses or collocation matching activities (cf. Paribakht & Wesche 1996). Various suggestions for pre-text activities and vocabulary related activities while working with texts (including sample texts) are given in Morgan & Rinvolucri (1986).

4.5 Ensuring repeated exposure to 'old' words in 'new' texts

As stated in the previous section, encountering a word only once will rarely result in lasting vocabulary knowledge. Even if an initial stage of acquisition is reached, repeated encounters are necessary for extending the word knowledge and for ensuring that the acquired information is retained. Reading new texts with old words will not only ensure that the receptive knowledge of these words is consolidated, but will also improve learner motivation through the experience of grasping the contents (cf. Hulstijn 2001: 284f.). If productive vocabulary knowledge, that is the learners' ability to use these words actively, is aimed at, the reading activities should eventually be supplemented by complementary vocabulary activities focusing on use.

4.6 Training strategies and raising learner awareness

Learners appear to differ widely in the skill with which they use strategies for guessing unknown word meanings from context. Therefore, training learners in their strategy choice and use should be part of any planned vocabulary development programme.

Most books on vocabulary teaching offer exercises for guessing from context in one or the other format, but do not contain any instructions about specific guessing strategies (cf. Gairns & Redman 1986, Taylor 1990, Wallace 1982). Various valuable methods for integrating strategy training into language classes, however, can be found in Nation (2001: 222ff.). He suggests a sequence of strategy modelling by the teacher, followed by strategy practice in steps, by application and report through the learner, and by final testing and feedback through the teacher. Planning such a mini-syllabus will of course involve deciding which strategies to focus on (e.g. analysing word parts, using parallels to other languages, using the co-text, etc.) and how much time to spend on them. For guessing word meanings from context, Nation (1990: 162) proposes a deductive 5 step approach which starts with the word itself (step 1) and moves from analysing its immediate co-text (step 2) and its operation in the wider co-text (step 3) to a meaning guess (step 4) and finally to checking the appropriateness of the guess by analysing the word structure or replacing the word (step 5). [3]

---

[3] As stated in section 3, the guessing procedure will start from the indexical (context-specific) word meaning level, and aim at forming hypotheses about the symbolic (denotational) word meaning on the basis of these context-specific interpretations.

Ideally, strategy training should not only include strategies for guessing meanings but also strategies for recording, learning and recycling words. Various guidelines for strategy training can be found in Thornbury (2000); for a useful list of vocabulary learning strategies grouped into different classes and categories (e.g. strategies for discovering meaning vs. for consolidating vocabulary knowledge) see Schmitt (2000: 132ff.).

Apart from the importance of strategy teaching, Sternberg (1987: 96) also stresses that learners have to be made aware of the role of the processes, clues, and moderating variables involved in incidental vocabulary acquisition, in order to apply these strategies effectively. In his eyes, this theoretical awareness will lead to gains in vocabulary acquisition from reading, as it will enable learners to consciously engage in inferring unknown meanings, while at the same time taking into account the limitations involved. The importance of learner awareness is also stressed by constructivist approaches to language learning, in which comprehension and learning are seen as an active and subjective construction process on the part of the learner (cf. Wolff 1994).

5. Conclusion

As the analysis of the factors influencing incidental vocabulary acquisition shows, the process of learning vocabulary through reading is by no means as straightforward as some researchers would make us believe. However, even though the initial probability that a learner acquires an unknown word through reading – i.e. that he/she firstly focuses on this word, secondly succeeds in specifying its meaning and thirdly commits form and meaning to memory – seems minute, teachers can significantly enhance the chances that new words in texts are acquired and remembered by the learners if they integrate relevant classroom practices into their teaching. On the one hand, teacher-directed actions such as choosing appropriate reading materials and reading goals, ensuring repeated exposure, or integrating complementary activities (fostering receptive knowledge and/or productive use) serve to provide favourable conditions in connection with class-based reading. On the other hand, training strategy choice and use and increasing learner awareness aim at fostering autonomous learner behaviour, and will eventually have positive effects on learners' future vocabulary gains from any type of reading.

References

Anderson, R.C. and Z. Shifrin. 1980. 'The meaning of words in context'. In: Spiro, R.J., B.C. Bruce and W.F. Brewer. (eds.). 1980. *Theoretical Issues in Reading Comprehension*. Hillsdale, N.J.: Lawrence Erlbaum, 330-348.

Beaugrande, R.A. de and W.U. Dressler. 1981. *Einführung in die Textlinguistik*. Tübingen: Niemeyer.

Gairns, R. and S. Redman. 1986. *Working with Words*. Cambridge: Cambridge University Press.

Haastrup, K. 1991. *Lexical Inferencing Procedures or Talking about Words*. Tübingen: Narr.

Haynes, M. 1993. 'Patterns and perils of guessing in second language reading' in T. Huckin, M. Haynes and J. Coady (eds.). *Second Language Reading and Vocabulary Learning*. Norwood, N.J.: Ablex Publishing Corporation.

Hulstijn, J.H. 1993. 'When do foreign language readers look up the meaning of unfamiliar words? The influence of task and learner variables'. *Modern Language Journal* 77: 139-147.

Hulstijn, J.H. 2001. 'Intentional and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity.' in P. Robinson (ed.). *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.

Hulstijn, J.H., M. Hollander and T. Greidanus. 1996. 'Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use and reoccurrence of unknown words'. *Modern Language Journal* 80: 327-339.

Knight, S. 1994. 'Dictionary use while reading: The effect on comprehension and vocabulary acquisition for students of different verbal abilities'. *Modern Language Journal* 78: 285-299.

Krashen, S. 1989. 'We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis', *Modern Language Journal* 73: 440-464.

Nation, I.S.P. 1990. *Teaching and Learning Vocabulary*. Boston, MA : Heinle and Heinle.

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I.S.P. and J. Coady. 1988. 'Vocabulary and reading' in R. Carter, and M. McCarthy (eds.). *Vocabulary and Language Teaching*. London: Longman.

Paribakht, T.S. and M. Wesche. 1996. 'Enhancing vocabulary acquisition through reading: a hierarchy of text-related exercise-types.' *The Canadian Modern Language Review* 52/2: 155-178.

Rieder, A. 2002. 'A cognitive view of incidental vocabulary acquisition: from text meaning to word meaning.' *VIEWS* 11/1&2: 53-71.
http://www.univie.ac.at/Anglistik/views/02_1&2/AR.PDF.

Rieder, A. 2004. 'Der Aufbau von Wortbedeutungswissen beim Lesen fremdsprachiger Texte: ausgewählte Fallstudienergebnisse.' *Fremdsprachen Lehren und Lernen* 33: 52-71.

Morgan, J. and M. Rinvolucri. 1986. *Vocabulary*. Oxford: Oxford University Press.

Schmitt, N. 2000. *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Schatz, E.K. and R.S. Baldwin. 1986. 'Context clues are unreliable predictors of word meanings'. *Reading Research Quarterly* 21: 439-453.

Stein, M.J. 1993. 'The healthy inadequacy of contextual definition' in: Huckin, T., M. Haynes and J. Coady (eds.). *Second Language Reading and Vocabulary Learning*. Norwood, N.J.: Ablex Publishing Corporation, 203-214.

Sternberg, R.J. 1987. 'Most vocabulary is learned from context' in M.G. McKeown and M.E. Curtis (eds.). *The Nature of Vocabulary Acquisition*. Hillsdale, N.J.: Lawrence Erlbaum.

Taylor, L. 1990. *Teaching and Learning Vocabulary*. New York: Prentice-Hall.

Thornbury, S. 2002. *How to Teach Vocabulary*. Harlow: Longman.

Wallace, M. 1982. *Teaching Vocabulary*. London: Heinemann.

Widdowson, H.G. 1983. *Learning Purpose and Language Use*. Oxford: Oxford University Press.

Wolff, D. 1994. 'Der Konstruktivismus: Ein neues Paradigma in der Fremdsprachendidaktik?'. *Die Neueren Sprachen* 93/5: 407-429.

Yano, Y., M.H. Long and S. Ross. 1994. 'The effects of simplified and elaborated texts on foreign language reading comprehension'. *Language and Learning* 44:189-219.