Jadwiga Linde-Usiekniewicz
Uniwersytet Warszawski

IV

The Encoding Grammar and Theme-Rheme Divisions

4.1. Preliminaries

The theme-rheme divisions discussed in this chapter corresponds to the notion of information structure invoked in the title of the book[1]. There are several reasons why the generally accepted term is not used for the phenomena in question within the general framework of the Encoding Grammar. The first reason is technical and terminological. One of the basic tenets of the Encoding Grammar presented here is the separation of representations, that present some features of sentences and utterances to be encoded, and structures, i.e. linguistic means of a specific nature available for encoding, cf. *The Architecture of the Encoding Grammar*. In contrast to the terms *semantics* and *syntax*, which can be invoked in such contrasting pairs as the *semantic representation of a sentence, the semantic structure of a language* and the *syntactic representation of a sentence,* the *syntactic structure of a language* respectively, the term *information structure* cannot be used in this way. Within the convention adopted here the respective terms would have to be the *information structure representation of a sentence* and *information structure structure of a language.* While the first is clumsy, the second looks ridiculously pleonastic.

The second reason is that the term *information structure* , although it could be applied to the theme-rheme division, is misleading. If used when talking about sentences, it suggests that besides their syntax, semantics, morphology and phonology, they also contain something called *information* which is somehow structured, or it may even be understood as implying that the information contained in the sentence (i.e. the semantic contents of the sentence) has a specific structure. Indeed, this is how the phenomenon was first conceived within the Prague School, and is still presented in some approaches, but would be inconsistent with the approach presented in this book, where the theme-rheme division is taken to be a part of semantics. When talking about language it would suggest that there is yet another level or subsystem of the language that has to be taken into account. This again is consistent with the classical Prague School approach, but would be incompatible with the general ideas presented here.

As already mentioned, the approach presented within the Encoding Grammar is that the division into themes and rhemes (which is the term adopted in this book for the phenomenon in question) is semantic in nature and consequently is a part of the semantic

---

[1] I am assuming here that reader is fairly familiar with the term *information structure* and with the terms *theme* and *rheme*. The reason for choosing this particular pair and not for example *topic* instead of *theme* and *comment* or *focus* instead of *rheme,* together with the definitions of the two terms contrasted here will be provided at later stage, .

representation of sentences , as already mentioned in Chapter II, and in particular represented in (2. 26). By the same token, it is devised by the speaker together as a part of meaning to be encoded. In consequence the division into themes and rhemes does not correspond to What is is another level imposed upon a semantic representation conceived in more traditional way, but that elements of the semantic representations carry in fact features of being either themes or rhemes.

In Chapter I *Why Encoding Perspective* it has been said that the existence of different and to a degree contradictory (or at least incompatible) accounts of the phenomena discussed here can be seen as the result of scholars adopting (explicitly or not) the speaker's perspective, the addressee's perspective, a mixed perspective (both speaker's and addressee's). The picture is further complicated by some approaches defining the terms in question fairly independently of discourse phenomena or speech act participants. On the other hand, many of the specific accounts, both within and without the speaker's perspective, provide interesting theoretical proposals, insights and analyses that the explicitly encoding perspective should somehow include, provided they can be reformulated within the proposed framework. These accounts also raise some important issues that any account of the division into themes and rhemes, including the one presented here, has to address. That is why the present chapter is divided into several sections. In the first two a brief and highly selective survey of approaches and their results is presented. The selective character is two-fold: firstly, only those approaches that are directly relevant to the formulation of my own proposal are mentioned; secondly, even within them only specific issues, reflected or reformulated within my proposal, are discussed. The third section discusses why even those approaches cannot be simply incorporated within my proposal. In all the three sections the original terms may be used, or the term *information structure*, instead of the pair *theme –rheme* adopted here. In the final section, my own proposal, similar to the one proposed in Linde-Usiekniewicz [2008a], albeit appended and elaborated, is presented.

4.2. A Brief Selective Typology of Approaches

One way in which the approaches can be classified is into functional and formal ones [Erteshik-Shir 2007: 72ff] . The distinction is not absolute and it is nowadays possible to find linguistic models that are at the same time formal and functional. Nevertheless with respect to the way information structure phenomena are presented the distinction can be applied to different ways in which the terms are defined and used to describe relevant phenomena.

First of all, the very notion of theme and rheme (originally *starting point* and *nucleus* of the utterance [Mathesius 1939]) were introduced as functional phenomena, i.e. having to do with the way the message is organized by the speaker. In addition, the organization of the contents of the message was supposed to be fairly independent of either semantics or syntax. The original Prague School approach was indeed initially called *actual sentence division*, and only later the term was replaced by that of *functional sentence perspective* [Daneš 1994: 123]. Nevertheless, it was thought that the articulation into theme and rheme could simply be imposed on a given sentence by linearization alone in the so called *objective* order and by linearization and intonation in the so called *subjective order* [Mathesius 1939]). In addition it was stated that in the objective order the starting point preceded the nucleus and in the subjective order the nucleus came first. From there the notion that themes tend to be initial elements of sentences and utterances developed, but since (as it is widely known) not all languages allow re-alignments of sentence elements without altering the syntactic structure, formal phenomena, including passivization and specific topicalizing and focalizing constructions were studied. Thus even within the functional approaches the claim that the

information structure can be freely imposed on ready sentences had to be rejected, although the idea that to a degree it can, at least in the so-called free-order languages, persisted. Here I would like to argue that it can, but to a lesser degree than generally assumed. To illustrate my point I will use a pair of examples taken from Erteshik-Shir 2007: 77)[2]:

> (4.1a)   *The bank is next to the Post Office.*
> (4.1b)   *The Post Office is next to the bank.*

What is important here from the functional perspective is that the two sentences with the same prepositional contents differ in their syntactic structure because they differ in the way their contents is presented. In Erteschik-Shir terms, the subjects are topics, and the prepositional phrases are focused.

   In Polish, which is a free-order language, nevertheless the equivalent translations of (4.1a, b) have to maintain the original structure:

> (4.2a)   *Bank jest koło poczty.*
> (4.2b)   *Poczta jest koło banku.*

While the syntactic structure of (4.2a) can be maintained while imposing on it a different, actually almost a reverse information structure, i.e. the one in which the subject would be the focus, and the prepositional phrase 'next to the Post Office' would be the topic, as in:

> (4.3)   *Koło poczty jest bank.*

it would be wrong to assume that (4.3) has the same prepositional contents as (4.1a) and (4.2a). Translated back into English, (4.3) means something like:

> (4.4)   *Next to the Post Office [there] is a bank.*

While the presence of *there* in the English equivalent can be easily passed over when analyzing the meaning of (4.3), there is no way one can overlook the fact that (4.4), and by the same token (4.3), has not got the same prepositional contents as (4.2a) and (4.1a), since in (4.1a) the reference is made to a specific bank, since the appropriate nominal phrase is a definite one, and in (4.4) it is not, since the relevant nominal phrase is an indefinite one.

   Indeed, in Mathesius's original examples (taken from Czech) the division into theme and rheme could be imposed on a sentence without altering its syntactic structure and meaning, but only because in his examples also dealing with locations, both NPs were in fact proper names: *Václavské náměstí* 'Wenceslaus Square' and the National Museum. Similarly, Polish pair of utterances, discussed in Linde Usiekniewicz [2008a]

> (4.5a)   *Kolumb odkrył Amerykę*
>   'Columbus discovered America'
> (4.5b)   *Amerykę odkrył Kolumb*
>   'lit. America-acc. discovered Columbus; America was discovered
>   by Columbus'

represent the same phenomenon, because both NPs are in fact proper names.

---

[2] The example actually comes from one of the works discussed by her and serves to illustrate another point.

Nevertheless, fine-grained semantic distinctions of the kind presented above are not the main issue of functional approaches to information structure. Just the opposite, the meaning (or the message as such) tends to be taken for granted, and the main point of interest is the way the message is organized to fit its communicative function or communicative import. This preoccupation is very well illustrated by the very nature of the question test which is supposed to separate theme and rheme (or again whatever pair of terms is used.)

The very nature of question test consists in verifying if an utterance could be used as an answer to a wh-question. Thus (4.1a), and (4.1b) can be used as answers to different questions [Eretschik-Shir 2007: 77]:

(4.6a) *Where is the bank?*
The bank is next to the Post Office.// It is next to the Post Office.
#The Post Office is next to the bank,
(4.6b) *Where is the Post Office?*
The Post Office is next to the bank.// It is next to the bank.
 # The bank is next to the Post Office.

which is taken as evidence that they have different information structure even though their prepositional contents is the same.

The question test is also applied not just to show that the information structures differ, but to pick the theme and/or the rheme. It is assumed that the string corresponding to the question word represents the rheme, while the repeated string from the question repeated in the full answer corresponds to the theme. Moreover, it is said that in languages that allow ellipsis topical elements are those that can be absent in the answer, and in languages that do not allow ellipses, can be substituted by pronouns[3]. Thus for (4.1a) the question-answer test presented in (4.6a) would pick up *the bank* as obviously the theme by virtue of its being repeated or substituted by a pronoun, and *next to the Post Office* as the rheme, since it corresponds to the question word. Interestingly, the verb *is* appears to belong neither to theme nor to rheme; which shows that the question test does not work so good for verbs. (I will come back to it later on.)

The question-answer test approach illustrates, albeit marginally, yet another important feature of functionalist approaches, i.e. reliance on context. While for the test only a minimal diagnostic context is created, it should be noted that many, if not all, attempts to capture the nature of information structure involve either discourse phenomena or the speech act situation and its participants. Nevertheless, an important contribution of the functionalist approach is establishing that information structure is of tantamount importance to studies of larger portions of texts, just to mention the seminal work by František Daneš [1974] and several papers and articles by Talmy Givón [1979, 1983 to mention just a few].

In very general terms functional analyses of information structure can be classified into binary and non-binary ones. Binary approaches prevail, while the non-binary nowadays seem to be of historic interest only. They include the Daneš's [1972] tripartite division into Topic, Transition and Comment, which is supposed to overcome the absence of parallelism between the syntactic and semantic perspective on the one hand and the communicative perspective on the other. It should be noted however that while Daneš posits the tripartite division within the syntactic perspective, i.e. Subject – Verb – Object and within the semantic perspective, i.e. Agent – Action – Patient, such division is but a construct, not shared by all syntactic and semantic theories. In fact, within the phrase structure theory the syntactic

---

[3] In my opinion the question test works fine enough as means of introducing the very notion of information structure, but as a diagnostic tool has a lot of shortcomings. I will discuss them in section 4.4. *Approaches and Their Results as Seen from the Encoding Perspective*

analysis into constituents is binary, and within dependency syntax there are no a priori restrictions on the number of elements depending on the verbal head. Similarly, semantic analysis of sentences is not necessarily tripartite, the propositional structure (i.e. the argument – predicate structure) is either binary or does not limit the number of arguments to just two. However, the notion of Transition seems to elegantly account for the fact that the question – answer tests and other theoretical devices for singling out either themes or rhemes do not always render a binary structure. Another well-known instance of non-binary approach, called Communicative Dynamism, was proposed within the Functional Sentence Perspective by Jan Firbas [1864; 1971].

Binary approaches can be divided into those that apply a single dichotomy and those that conceive the information structure as a result of two (or in rare instances more[4]) divisions. The dominant view within multiple dichotomy approaches singles out topic, as opposed to comment, and focus, as opposed to background. This type of analysis within the functionalist approach is relatively recent and seems to be influenced at least partly by formal criteria of topichood and focushood, initially established within formal approaches. While the notion of topic seems fairly consistent within differing models (the definitions vary but try to capture broadly equivalent phenomena), the notion of focus seems to be far more diverse (see below).

Within functionalist approaches a lot of emphasis has been put on the definitions of the terms in question. However, this has resulted not in a definition generally accepted within the linguistic community, but in a whole array of partly matching and partly contradictory concepts. Some of them have been presented and critically discussed in such works as Bogusławski [1977], Huszcza [1991a], Mel'čuk [2001] and Erteshik-Shir [2007]. Again a bird's view typology allows to single out approaches that limit themselves to the sentence or utterance itself (the sentence-internal approaches) and those that somehow include elements that are external to them. Among the latter one can find approaches that rely on context, but still remain within the realm of linguistic phenomena. This approach can be called sentence-external but text-internal, since it does not involve extralinguistic phenomena. The two language-internal types of approach can be further contrasted with the approaches that rely heavily on extralinguistic phenomena, principally on the so called *knowledge* of the participants, attention of the participants being brought to something etc. Interestingly, some of the approaches that can be considered extralinguistic tend to confine their analyses to a single sentence.

A consistently sentence-internal approach is the one that applies the concept of *aboutness*: theme or topic is what the sentence or message is about, while comment is the message itself. Proponents of this approach seem to assume that it is easy to recognize what each sentence is about. The problem with this approach is that a sentence is about everything its contains. Thus our (4.1a, b) are both about the bank, the Post Office and the spatial relation between the two, and our (4.5a, a) are both about Columbus, America, and the discovery of America. This is rightly pointed out by Bogusławski [1977: 142-147], although he argues that the aboutness criterion is valid, provided it is used in an appropriate way[5].

Sentence-external approaches rely on context or discourse and define topic or theme as something already present in the context. The main problem they face is that it either means that discourse initial utterances cannot have topics, and have to be considered thetic, unless

---

[4] For example [Wlodarczyk 2004a] and [Wlodarczyk 2004b]. In their account simple SVO sentences are divided just into topic and comment, while complex sentences, with multiple complements, have the focus singled out within the comment. Both topic and focus qualify as centers of interest. However, they also consider the grammatical subject as another center of interest, and as yet another, they postulate theme as a center of interest for larger texts.

[5] I will discuss his proposal later on.

stage topics [Erteschik 2003: 16] and permanently available topics [Erteschik 2003: 17] are allowed. However both types of topics are extralinguistic in nature, as they cover not only the discourse participants and the deictic situation but also general world that surrounds the discourse participants. (Erteschik-Shir actually gives the moon and the president as examples of permanently available topics).

Extralinguistic approaches rely either on such concept as the distinction between given and new or pertaining to the shared knowledge, or not, with all the pitfalls the very concept of shared knowledge entails, cf. [Bogusławski 1977: 162-168; Sperber, Wilson 1986: 15-21, 38-46], or, directly or indirectly, upon human cognitive abilities. Under this heading I include both those proposals that invoke differences in prominence (saliency, the figure-ground distinction, etc.) and those that invoke some reference, somehow metaphorical, to the way information is processed (file-card definitions). Some of the definitions are couched in cognitive terms, while other are more pragmatic in their outlook.

The extralinguistic approaches appear to be particularly fraught with problems, since prominence or being the figure is something that can concern both topics and comments. Actually it is the property of both contrastive topics and contrastive foci [Huszcza 2001; Erteschik-Shir 2007: 76-77]. It may lead to inability of distinguishing between topics and foci themselves. One instance of such situation appears in Frajzyngier, Shay [2003: 164]. Of a Tagalog preposition *ang* the authors write: "This preposition has a pragmatic function whose pragmatic nature is still somewhat controversial but which may be conceived of as topic or perhaps clausal focus."

The shared knowledge approaches and the file-card approaches are not mutually exclusive, as can be seen in Erteschik-Shir [2007: 43-44]. According to her, in the work she presents[6]:

> "The common ground or context set is (metaphorically) represented by as set of file cards. Each file card represents a discourse referent. Entries on each card correspond to what is presupposed about the discourse referent in question. The cards are organized so that the most recently activated cards are to be found on top of the stack of cards. […] These are the discourse referents with provide potential topics in the discourse. How do cards get to be on top of the file? This follows implicitly from the definition of the focus. If the attention of the hearer is drawn to (the referent) of X, then the hearer (metaphorically) selects the card for X and puts it in a place of prominence, namely on top of his stack. […]A topic remains on top of the file and can therefore be continued. A focused element can become a topic in the next sentence since focusing a discourse referent requires the positioning of its card on top of the stack.
> The file card system thus involves locating the cards on top of the stack (topics) or positioning them there (foci). In addition, each card manipulated through the processing of an utterance is also updated with the information contained in the utterance."

However, what is most noteworthy in this account is the fact that the same referents (or expressions referring to them) can be topics and foci, since each has a specific file card devoted to them; the difference being that topics have to have appeared in previous discourse, while foci are newly introduced. By the same token, sentences can be either about topics or about focus. Moreover, the file-card system described does not limit itself to a single sentence, but applies to what happens (metaphorically) in conversation containing many utterances. Interestingly again, both information presupposed and information newly obtained is recorded as the entry on the file card. Finally, this model may metaphorically reflect the way hearers process long texts, but it says nothing about what happens in a particular sentence.

---

[6] The original work discussed there is [Reinhart 1981].

This way of applying the file-card metaphor to defining elements of information structure contrasts dramatically with the filing-system definition proposed by Romuald Huszcza [1983: 57]:

> „Przez temat zdania będziemy rozumieć człon składniowy lub grupę składniową […] wskazane w zdaniu przez nadawcę jako pewnego rodzaju hasło wywoławcze, pod którym odbiorca ma umieścić treść tego zdania w swoim umyśle"[7].

Other relevant features of Huszcza's approach include the existence not only of ordinary themes and rhemes but also of prominent[8] themes and rhemes which appear in sentences with so called prominent theme-rheme division. Prominent themes correspond to some kind of warning to the addressee that the speaker does not want them to understand that the contents of the sentence may apply to something or somebody else than the element marked as theme, while prominent rhemes correspond to warning the addressee that the contents of the sentence is supposed to rectify some misconceptions, and resemble contrastive topics or foci of other accounts. In this account sentences with prominent division contain either prominent themes or prominent rhemes, as in the following Polish sentences [Huszcza, 2000].

(4.7a)   *SAMOLOT ↑ to leci !*
             'The plane *is* flying / As for the plane, it is flying (but the pilot is not there). [9]

(4.7b)   *To SAMOLOT ↓ leci!*
             'It is a/the plane that is flying (and you thought it was a bird/balloon/blimp/glider/spacecraft etc.)

Both in (4.7a) and in (4.7b) the prominent element carries contrastive stress (represented by uppercase letters) , but of different nature: raising in (4.7a) and falling in (4.7b). It should be noted that both (4.7a) and (4.7b), being simple and short sentences, contain only one theme-rheme division, i.e. the first order division, between the noun and the verb. Usually the prominent division occurs in longer sentences and do not represent the first order division, but a secondary one.

(4.8a)   *ZIELONE ↑jabłka leżą na stole*
             'lit. the green apples are lying on the table'
             Intended meaning: 'As for the GREEN/UNRIPENED apples, they are on the table (but I have no idea as for other kinds'

(4.8b)   *Nie cierpię ZIELONYCH ↓jabłek*
             'I hate *UNRIPENED* apples (but nothing is said about other kinds)'

Here for (4.8a) the first order theme would obtain between *zielone jabłka* 'the green apples' and *leżą na stole* 'are lying on the table'. Within the first order theme the meaning of 'green/unrippened' would be singled out as a second order prominent theme. Similarly, in

---

[7] *"The theme of a sentence is hereby defined as the sentence element […] indicated by the speaker as the keyword or label under which the audience is supposed to file the information provided by this sentence"* [translation mine].

[8] The original Polish terms is *uwydatniony* . I decided to avoid translating it as *marked*, because this term is ambiguous enough in its current usage.

[9] This structure is extremely difficult to gloss, as the first variant, with emphatic *is* would correspond not to a prominent theme, but to prominent rheme, while the second variant with the expression *as for*, picks the indicates the prominent theme in a special way. I will discuss the *as for* formula at later stage.

(4.8b) the first order division could occur between *nie cierpię* 'I hate' and *zielonych jabłek* 'unrippened apples' with 'green apples' as a rheme, within which the second order prominent rheme would be 'unrippened'.

Another interesting feature of this approach is that it allows for both thetic sentences and for sentences in which the theme-rheme division is neutralized, i.e. those in which there is no exponent of either theme or rheme, or the division itself. The first type can be illustrated by

> (4.9)   *O, samolot leci!*
> 'Oh, there is a plane flying'

taken from [Huszcza 2000], while the second case is illustrated by (4.5a), which represents either the division into *Columbus* as theme and *discovered America* as rheme or to a complex theme *Columbus discovered* and *America* as *rheme*. Within this approach it would be wrong to posit a multilevel division for (4.5a), because the second order division needs to be a prominent one and (4.5a) is presented here as encoded without any marker of prominence. Interestingly, as already mentioned, this approach differs from others yet in allowing verbal themes, even in simple, un-cleft sentences, as for example MC announcements. In accordance to the classification carried out here Huszcza's proposal is sentence-internal but since it involves indicating how the sentence meaning should be processed, paradoxically it falls at the same time within the extralinguistic approaches.

Most functional sentence-internal and even sentence-external text internal approaches tend to present the information structure as binary (topic–comment, theme–rheme), but some functional extralinguistic approaches follow the initial formal approaches (discussed below) in so far that they posit both topics (as opposed to comment) and foci (as opposed to background). Some authors (including Erteschik-Shir and presumably Reinhart) use the pair topic–focus as somehow complementary elements of information structure. Yet other authors propose the existence of both topic vs. comment articulation and focus vs. background articulation.

> "The Focus function signals the Speaker's strategic selection of new information, e.g. in order to fill a gap in the Addressee's information, or to correct the Addressee information. The Focus function is assigned only in those cases in which this is linguistically relevant, i.e. when languages use linguistic mans to indicate that some part of a Linguistic Expression constitutes the relevant new information. The information not assigned the Focus function constitutes the Background. In Smit (fc)[10] Focus is defined as an update instruction to the Addressee" [Hengeveld, Mackenzie 2008: 89].
> "Another dimension of the organization of information structure is the Topic-Comment dichotomy. The Topic function, where relevant in languages, is not complementary to Focus, but part of this second dimension. Indeed, as we will show below, in certain circumstances a constituent can be simultaneously Focus (along the Focus-Background dimension) and Topic (along the Topic-Comment dimension). […] The information not assigned the Topic function constitutes the Comment. The linguistic marking of the Comment rather than Topic seems to be very rare. In Smit (fc) Topic is defined as the linguistic reflection of a 'retrieve' instruction to the Addressee".[Hengeveld, Mackenzie 2008: 92, emphasis mine]

Despite all the differences mentioned here and many fine-grained ones passed over in this brief survey, all the functional approaches tend to separate the information structure from semantics on the one hand and from grammar on the other. It is particularly evident from the fact how relevant terms are defined. This is not to say that functional approaches are not

---

[10] Actually [Smit 2007]

concerned with formal means of marking relevant sentence elements as topic, themes, foci, and so on. Just the description of formal means constitutes the second stage of the analysis.

By contrast, formal, i.e. mainly generative approaches, tend to invoke information structure to account for syntactic phenomena. The notion of topic and topicalization was first introduced to account for movement of some sentence elements, mainly noun phrases, to the beginning of the sentence. The complementary notion of comment is thus seen as referring to the remaining part of the sentence. By contrast, the notion of focus (complementary to presupposition) was associated with intonation [Ertheschik-Shir 2007: 27]. Interestingly, it is the early generative approaches that seem to have introduced both the two pronged distinction and the asymmetry between members of each pair. This is not surprising, since the notions were supposed to explain less than straightforward syntactic and phonological behavior of sentences. As the result, topics and focus have been invoked for such phenomena, among others, as fronting [Rizzi 1997], scrambling [Putnam 2007], interface between syntax and prosody [Fanselow, Lenertová 2011], split constituents [Fanselow, Ćavar 2003], to mention just the few. Within the Minimalist approach some linguists claim that the topic and focus features not only have to be present in the syntax, but need to be accounted for at the Numeration stage [Aboh 2010], or at least already in the syntax [Eretschik-Shir 2007: 55; Zubizaretta 1988: 30]. However, even those minimalist accounts that allow for functional Topic Phrases, Focus Phrases, like Rizzi [1997] and for topic and focus particles being a part of the Numeration, like Aboh [2010], seem to overlook one point that has to be made within the model they use: Within this framework not only functional phrases and particles need to bear the appropriate feature (topic or focus) but also lexical items that merge with them need to bear the same feature at the numeration as well; otherwise there is nothing to allow them to merge.

Broadly speaking, the main feature of formal approaches is that they rely on some formal characteristics of sentence elements (roughly and in minimalist terms: linear position at the Spell-Out stage and something that would trigger the appropriate intonation at PF) to distinguish topics from foci. An example of such distinction can be seen in Rizzi [1997: 286], where, as already mentioned, he rightly points out that in Italian and Spanish fronted foci do not need replicated clitics, while fronted topics do. However, it should be noted that within the generative approach syntactic phenomena associated with information structure, i.e. topicalization, clefting, pseudoclefting and answer fragments are standard tests applied to check if a given sentence segment is a constituent [Osborne, Putnam, Gross 2011: 233-234]. Nevertheless, it is possible to find fronted foci constructions used as evidence for constituency, as in [Cinque 2008a: 245].

The next approach to be discussed here is represented almost uniquely by the Functional Generative Description [Sgall 2009]. It states explicitly that the topic-focus articulation is fundamental to language description (at par with dependency relations). In Sgall's [2009: 9] opinion:

> "Most theoretical approaches to natural language, especially those of the "main stream" appear to be mistaken in this crucial point, namely in ignoring or underestimating the fundamental position of TFA in the syntax and semantics of human languages"

Topic-Focus articulation is relevant not only for syntax, but for semantics (i.e. truth value of sentences) [Sgall 2009: 13]. It is connected to syntax inasmuch that it appears in the underlying representation of a sentence (called *tectagrammatical representation*), which takes the form of a dependency tree . The topic-focus articulation is a binary opposition between contextually bound nodes (CB) and non-bound nodes (NB); each node is appropriately marked in the tree itself.

The Functional Generative Description approach to information structure is directly related to its overall perspective on language and language analysis. The model itself strongly resembles the Meaning⇔Text Model [Žabokrtský 2005], in its basic tenets: dependency representation of syntax and the multilevel analysis. In general terms it differs from MTM by adopting mainly the decoding perspective [Lopatková, Plátek, Kuboň 2005; Lopatková, Plátek, Sgall 2007]. With respect to the information structure it includes the T-R articulation in the underlying representation together with syntax, and marks syntactic nodes as T or F. By contrast within MTM the corresponding level it is a separate structure belonging not to the semantics but to the communicative level [Mel'čuk 1988: 28; Mel'čuk 2001]. >From other generative approaches it differs by the decoding perspective, the adoption of dependency syntax and importantly by the symmetric relation between topic and focus. From all approaches discussed it differs by the way the Topic and Focus are defined: the notion of contextually bound item and unbound item are primes.

All the approaches mentioned so far tend to share one important feature: The elements identified as topics or themes generally have to be overtly present in the sentence, the only exception being so called *stage topics* in some accounts. Moreover, these elements have to be at least free forms if not constituents.

A radical departure from the types of analyses presented so far can be found in Bogusławski's [1977] model. The model itself is not widely known within the linguistic community as a whole, but has given rise to a linguistic methodology successfully applied by Andrzej Bogusławski himself and by numerous scholars of his school [Linde-Usiekniewicz 2008c] and at the same time has given rise to a heated debate on the nature of thematic-rhematic structure within Polish linguistics [Linde-Usiekniewicz 2008a]. Within this model the thematic-rhematic structure is a basic property of meaning, not only of sentences, but also of lexical units. This is made possible by the fact that the model proposes not only identifying themes and rhemes, but also introduces the notion of the thematic dictum [Bogusławski 1977: 38; 59-71], which corresponds roughly to the notion of what is presupposed. What is novel here is not noting the relation between what is presupposed and the information structure, since this can be find in several accounts (cf. Erteschik-Shir [2007]), but extending the thematic-rhematic structure to cover also the distinction between what is presupposed and what is posed in the meanings of words, in agreement with Fillmore [1971].

The theme-rheme distinction is connected to the aboutness criterion, but in a special way. The topics are not picked out by the asking what the sentence is about, but by the "contradistinction *thing spoken about* and *what is said about it*" [Bogusławski 1977: 142, original emphasis] that corresponds to the schema *say about ... that...*"[Bogusławski 1977: 143, original emphasis]. Thus

> "the mutual relation of the constructions *about* and *that*" (ibidem) picks a particular rheme associated with a particular theme. At the same time rhemes are defined as those elements that can be subject to eliminary contrasting" [Bogusławski 1977: 183-188]

Within the model the notion of theme is roughly equated with that of an object, however, these do not have to be real objects, and can perfectly be just concepts, as in categorical sentences [Bogusławki 1977: 23] and in cleft-sentences [Bogusławski 1977: 225]. Another interesting feature of this model is that it postulates a hierarchy of themes and rhemes (or hierarchy of theme-rheme divisions) within the sentence (Bogusławski 1977: 199-203).

What distinguishes this approach most emphatically from all other is the idea that the thematic-rhematic structure is intrinsic to sentences, even if there is no special markings. Linguistic devices such as intonation, particles, special structures, including cleft sentences and expression of the kind *as for*, are but exponents of the structure.

4.3. A Brief Selective Typology of Information Structure Elements

It is not surprising, given the diversity of approaches, that various information structure elements are identified in different works. Interestingly, a correspondence can be found between elements identified in non-binary approaches and in approaches that posit multilevel division and neutralization. For example, within Huszcza's model, Daneš's transition can be seen as corresponding to elements that are subjects to the neutralized theme-rheme division.

Another difference involves the number of dichotomies within the information structure, already mentioned in the previous section. Multiple dichotomies tend to be asymmetrical, each singling one element of the structure in contrast to unspecified remainder, as in topic-comment and focus-background division in early generative approaches. Single dichotomies tend to be symmetric, with some exceptions; one of them is the card-file model by Huszcza, which defines the theme but not the rheme[11]. However, one striking feature of some single and symmetric dichotomies is the fact that they are not exhaustive: the sentence does not divide into the two parts specified by the theory. A telling example can be found in the initial presentation of the notion of information structure by Erteschik-Shir [2007:1-2]:

> "This consensus [i.e. the general consensus concerning the information structure] extends to the status of question-answer pairs with respect to information structure. In (1), the constitutent that "answers' a wh-question is the focus (in upper case) and a constituent already introduced in the question is the topic of the answer (in italics)"

> (1)   a   Q: What did John do?
>   A: *He* WASHED THE DISHES

>   b   Q: What did John wash?
>   A: *He* washed THE DISHES

> Not all question-answer pairs are as easily analyzed as shown in (2):

> (2)   a   Q: Who washed the dishes?
>   A: JOHN washed *them*.

>   b   Q: What happened to the dishes?
>   A: JOHN WASHED *them.*

>   c   Q: What happened?
>   A: JOHN WASHED THE DISHES.

> In (2a) the predicate is presupposed. In some approaches the object (*the dishes*) is analyzed as topic, in other the whole predicate is viewed as a unit." [original emphasis]

Attention should be drawn to the fact that not only in (2a) the division into topic and focus is problematic, but that also in an apparently straightforward case of (1b) the verb appears as neither topic nor focus.

In some approaches, not only generative ones, a distinction is made between so called contrastive (or narrow, exhaustive, or exclusive foci [Erteschik Shir 2007: 29]. These would

---

[11] This has been the one of the points strongly criticized, both in writing and in spoken debate within Polish linguistics. While I find the criticism of his account for thetic, i.e. (in his terminology) theme-less sentences, unfounded [Linde-Usiekniewicz 2008a], I agree that even within the card-file model the rheme should be defined, though not in the way it is done by Reinhart [1981].

correspond roughly to prominent rhemes in Huszcza's account. Nevertheless, the distinction between wide and narrow focus does not imply, within the generative approach, the multilevel distinction: narrow focus is not necessarily a part of the wide focus. Thus for the apparent focus ambiguity [Erteschik-Shir 2007: 30-31] the distinction is not invoked:

> "Simplifying a great deal, this means that the sentence-final stress leads to focus ambiguity (capitals indicate stress):
>
> (38)   a   Maxwell killed the judge with [a HAMMER]$_{FOC}$
>   b   Maxwell [killed the judge with a HAMMER]$_{FOC}$
>   c   [Maxwell killed the judge with a HAMMER]$_{FOC}$

Although the author argues that while (38a) is the appropriate answer to the question *What dis Maxwell kill the judge with?*, (38b) is not the most appropriate answer to the question *What did Maxwell do?* while (38c) is not the most appropriate answer to the question *What happened*? In her opinion the appropriate answers would bear different stress assignments, all focus elements bearing additional stresses.

Yet another way of distinguishing narrow, wide and sentence focus (i.e. not necessarily identifying the narrow focus with contrastive one) can be found in Nikolaeva [2001: 1-2, original emphasis]:

> „After Lambrecht (1994) [12,] I will further distinguish three types of focus structures, argument focus, predicate-focus, and sentence-focus. The ARGUMENT-FOCUS (or: narrow focus (Van Valin and LaPolla 1997)[13]) usually extends over one NP constituent. Operationally, it can be identified as a target of a wh-question, for example:
> (1)   What is John drinking?
>   He is drinking BEER.
> The function of the argument-focus is to provide the missing argument (beer) in a presupposed open proposition 'John is drinking X'. The pragmatic assertion evoked in (1) can be represented as 'X = beer'. The second type of focus articulation, the PREDICATE-FOCUS (or: completive focus in Dik[14] (1989: 277–85)) serves to increase information about the topic referent.
> (2)   What is John doing? or: What about John?
>   He is drinking BEER.
> The predicate-focus adds a predicate to a given argument (<u>John</u> in example (2)). The corresponding pragmatic presupposition can roughly be represented as something like 'John is doing X', while the pragmatic assertion is 'X = is drinking beer'. In this type of utterances focus can extend over several constituents which I will refer to as FOCUS DOMAIN, for example, over the VP <u>is drinking beer</u> in (2).[…]Finally, the SENTENCE-FOCUS type (also: 'presentational', 'all-new' or 'thetic'sentences) is pragmatically unstructured in the sense that it lacks a pragmatic presupposition. The focus domain extends over the whole proposition, cf.:
> (3)   What happened?
>   JOHN died"

Even more variety among approaches can be found in the elements singled out as themes or topics. Some authors recognize stage topics as topics of apparently thetic sentences. These do not appear in the sentence under analysis but are seen as "indicating the spatio-temporal parameters of the sentence (here and now of the discourse)" [Eretschik-Shir

---

[12] Lambrecht Knud, 1994.. *Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

[13] Van Valin Robert D, Randy J. LaPolla, Randy J., 1997, *Syntax. Structure, Meaning and function.* Cambridge: Cambridge University Press.

[14] Dik Simon C., 1989, *The Theory of Functional Grammar. Part 1. The Structure of the Clause*. Dordrecht, Providence: Foris

2007: 16]. Hovewer, at least in the author's opinion, they need to be given, arising either from the discourse situation or from the hypertheme[15].

While there is some strong evidence that themes do not have to be old (cf. supra), there is a tendency to distinguish between so- called "old topics" and "new topics" or between continued and shifted topics. The distinction can be carried out on functional grounds, i.e. the continued topic is the same as in the previous sentence, while the shifted topic is not the same, but to has to have been mentioned previously (or at least has to be available, in those accounts that allow for stage topics and permanently available topics). The two types of topics differ not only as to their functional properties, but also as to their syntactical ones. For example, in languages that allow elided topics (i.e. topic-drop languages) only the continued topics are dropped, while the shifted topics appear on the surface. In languages that would not allow for elision of some sentence part, be it subject or object, continued topics are represented by pronouns, with the risk of ambiguity involved. This can be illustrated by following sentences:

(4.10a)   Polish: *Anna ma nową sukienkę. Ø Kupiła ją wczoraj.*
(4.10b)   English: *Anna has a new dress. She bought it yesterday.*

where in Polish the continued theme ('Anna') is dropped in the second sentence, while in English it is represented by the pronoun. Interestingly, it can be argued that in (4.10a) *ją* represents a secondary theme, which is a shifted theme (corresponding to rhematic *sukienka* 'dress'.

A shifted theme cannot be dropped in Polish as seen in (4.11)

(4.11)   *Anna ma na sobie nową sukienkę, którą dostała od matki. Ta/*Ø*
          kupiła ją wczoraj.
          'Anna is wearing a new dress which she got from her mother.
          She bought it yesterday.

Interestingly, it seems that in some languages shifted topics tend to be identified not by virtue of not being droppable, but because they are fronted. Eretschik-Shir [2007: 13] illustrates it following Catalan sentences, taken from Villalba [1998]:

(4.12)   A: On va posar les coses?
                where PAST-3 put the things
                'Where did (s)he put the things?'
          B: Em sembla que *els llibres*, *els* va posar al despatx.
                to-me seems that the books them-MASC PAST-3 put in-the study
                'It seems to me that.(s)he put the books in the study

In this sentence 'the books' is a new topic, although it is derived from the hypertheme 'things' and thus cannot be dislocated to the left:

(4.13)   A: On va posar les coses?
                where PAST-3 put the things

---

[15] I have argued elsewehere [Linde-Usiekniewicz 2011] that actually the reverse is true: anaphora are identified by virtue of being themes. Thus I would claim that in Erteschik-Shir original example *John heard a beautiful concert. The composer directed it.* it is understood that *the composer* means 'the composer of the concert' because it is the theme of the second sentence, and not it analyzed as theme because the knowledge of the word would have it that concerts have composers.

'Where did (s)he put the things?'
B: #[16]Em sembla que   *els* va posar al despatx, *els llibres*.
to-me seems that them-MASC PAST-3 put in-the study, the books.
'It seems to me that(s)he put the books in the study

By contrast, if the expression *els llibres* 'the books' appears in the question, it cannot be fronted in the answer:

(4.14a)   A: On va posar els llibres?
where PAST-3 put the books
'Where did (s)he put the books?'
B: Em sembla que   *els* va posar al despatx, *els llibres*.
to-me seems that the books them-MASC PAST-3 put in-the study
'It seems to me that.(s)he put the books in the study

(4.14b)   A: On va posar els llibres?
where PAST-3 put the books
'Where did (s)he put the books?'
B: #Em sembla que *els llibres*, *els* va posar al despatx,
to-me seems that the books them-MASC PAST-3 put in-the study
'It seems to me that(s)he put the books in the study

The question arises if both left dislocated and right-dislocated elements should be recognized as topics. [Erteschik-Shir 2007:12] rightly points out that if the topicalization, i.e. fronting is taken as the criterion, the right-dislocated elements (occurring in Catalan, Spanish and other languages), will not be considered topics, in contrast to German, which can front the old information as well.

Finally, it should be noted that many frameworks recognize that a sentence can have more than one topic, each represented by a separate sentence element. For example in (4.10a, b) above in the second sentence not only 'Anna' represented by Ø in Polish and by *she* in English would qualify as topic (in Polish by virtue of being dropped), but also 'dress', represented by *ją* and *it* respectively. Similarly in (4.11) – (4.14a, b) not only 'the books' are topics, but also the person who put them in the study (The third person singular subject pronoun can be omitted in Catalan). Depending on the way topics or themes are defined, multiple topics can be associated with multilevel articulation into themes and rhemes or with different kinds of topichoods (continued vs. shifted, old vs. new, given vs. new, etc.). Normally only one of the topics is recognized as the principal one, but in some accounts no distinction between primary and secondary topic is made:

"Multiple Topics are possible in certain languages, Thus in Turkish, in which Topic can be expressed by initial placement, we find (194), from Kornfilt (1997: 205)[17] in which both *kitabı* and *Ali-ye* are Topic, as is evident from their placement in front of the subject constituent *Hasan*:
(194)   Kitab-ı   Ali-ye Hasan   dün   ver-di
Book-ACC   Ali-DAt Hasan   yesterday   give-PST
'Hasan gave Ali the book yesterday'
In this example *dün* has Focus function, expressed thorough placement in the preverbal position"
Hengeveld, Mackenzie 2008: 94]

---

[16] # indicates that the sentence is grammatically correct, but not acceptable in the context.
[17] Kornfilt Jaklin, 1997, *Turkish*, London: Rutledge

Additionally, in some accounts a distinction is made between so called (clause)internal topics and (clause)external topics. This distinction is presented with clarity and brevity in Nikolaeva [2001:5]:

> The possibility of multiple topics is suggested by a variety of grammatical evidence across languages. It is important to consider that syntactically we may be dealing with two different structures here. First, at least one of the topic expressions may have clause-external status. A clause-external topic expression is not integrated syntactically into the clause structure, and the relation of its referent to the clause is a matter of pragmatic construal only. Syntactic arguments for the extra-clausal status of the external topic are presented in Aissen[18] (1992) for Mayan languages. The external topic is base-generated and adjoined to the maximal clause. Although it may be coindexed with a coreferential pronoun within the clause, this is not necessary because the external topic position is not a landing site for movement. In this it differs from the clause-internal topic which is shown to be in a Spec of C' position and to bind a coindexed trace within a clause"

4.4. Presented Approaches and Their Results as Seen from the Encoding Perspective

It has been already mentioned in Chapter I that the very idea of trying to outline an explicitly encoding approach to language is the effect of noticing that at least some controversies surrounding the issue of the information structure arise from differences between the speaker's perspective and all other perspectives. I have argued [Linde-Usiekniewicz 2008a] that one way of resolving this controversy would be to propose an account of the relevant phenomena within the encoding perspective alone and sketched the initial way of defining themes and rhemes. I will elaborate my proposal in the following section. In this section I will limit myself to showing why none of the rich array of approaches existent in literature and briefly discussed in the two previous sections are none of them capable of being adopted as they are within the encoding grammar.

First of all, the encoding perspective limits itself to single sentences and what can be encoded into them. Thus it needs at first glance to reject all approaches that are sentence external and extralinguistic. On this grounds almost all functional approaches are rejected, with the sole exception of the Huszcza's version of the file-card approach. On the other hand formal approaches need to be rejected because, as already mentioned in Chapter I, they are not encoding approaches. Similarly, the Functional Generative Description and Bogusławski's proposal are thus rejected.

The question still remains if some of the approaches initially rejected on the grounds of not being explicitly encoding ones could not be modified to fit the encoding perspective. The answer is negative, but some evidence for this claim are necessary.

It seems that none of the extralinguistic approaches can be appropriately modified. For those that rely on previous discourse or knowledge common to both the speaker and the addressee the reason is two-fold. First of all the definitions or criteria employed in them would need to be completely reformulated in such way as not to depend on previous discourse or knowledge[19]. Moreover, it is easy enough to prove that even sentences without contents are divided into themes and rhemes. For example (4.1a, b) do not need context to show that in both of them the subject is the theme. A similar argument can be made for several of the examples presented so far. In addition, initial sentence of fictional narratives encode some of their parts as themes (and rhemes) by appropriate syntactic means although for them no

---

[18] Aissen Judith, 1992, "Topic and Focus in Mayan". *Language* 68, 43–79.

[19] As will be seen later on that is exactly what I propose to do.

shared knowledge of the world can be invoked. Here are just a few examples, from English narratives.

(4.15) *He was a very sick white man* (Jack London, *Adventure*)[20]

(4.16) *Harold Waring noticed them first walking up the path from the lake* (Agatha Christie, *The Stymphalean Birds*)

(4.17) *The wise traveller travels only in imagination* (W. Somerset Maugham, *Honolulu*)

(4.18) *One was called Mrs Richman and she was a widow.* (W. Somerset Maugham, *Three Fat Women of Antibes*)

(4.19) *The bar was crowded.* (W. Somerset Maugham, *Gigolo and Gigolette*)

Not only protagonists of the story can be thus introduced, but also the setting (4.19). Additionally, a categorial sentence at the beginning of a narrative can encoded as articulated into themes and rhemes (4.17).[21]

Examples can be found of sentences with even complex T-R structure appearing at the beginning of a story, as is the case of one of the tales of Pirx the Pilot by Stanisław Lem:

(4.20) *Aniel nie wrócił o czwartej, lecz nikt jakby tego nie zauważył* (Stanisław Lem, *Wypadek*)
'lit. Aniel did not come back at four, though nobody seemed to have noticed it. [22]'

Here, however, the picture is further complicated by the fact that it is complex sentence. In such sentences each of the clauses may have its own theme-rheme divisions encoded (cf. *4.5 The Theme-Rheme Division Defined Within the Encoding Grammar*).

It appears that besides a modified file-card approach, there are yet two good candidates among those discussed above: Bogusławski's [1977] and Functional Generative Description. Both of them are similar to what the Encoding Grammar needs in terms of being sentence internal and in terms of seeing the theme-rheme structure or topic-focus articulation as crucial to semantics. Nevertheless, both present several insurmountable difficulties for their reformulation in terms of encoding. While it would be possible to label syntactic nodes (i.e. roughly words) in the deep syntactic representation (cf. *2.2. The Architecture of the Encoding Grammar*) as either belonging to topic or focus, such labeling would be contrary to the tenet that the starting point for the Encoding Grammar is the Semantic Representation that is supposed to be made up not of words or their signifieds, but of simpler elements of meaning (cf. *2.2. The Semantic Representation*). Additionally, Functional Generative Description's solution of equating topic and focus distinction with that of context bound element and unbound element is untenable within the Encoding Grammar. Taking these terms literally and understanding them outside the FGD framework would make the account no longer sentence-

---

[20] I am indebted to Piotr Stalmaszczyk for this particular example

[21] It should be borne in mind that the analysis of the initial sentences of narratives presented here treats them as sentences and disregards the possible contexts created by story titles. Incidentally sentence elements co-referential with story titles can be encoded as rhemes in the initial sentences, cf. 4.16. It is only later that *them* is revealed as meaning the women referred to as the Stymphalean birds in the story title.

[22] In Louis Iribarne's translation ("The Accident", in *More Tales of Pirx the Pilot*, 1983, Harcourt Brace Jovanovich) the sentence is rendered as: "When Ariel wasn't back at four, no one thought much of it". I am indebted to Lea Sawicki for finding the translation for me.

internal. And accepting such primes would go against the very idea of the encoding perspective adopted here.

Boguslawski's account, by contrast, would be quite easy to adopt as a part of the semantic representation, in particular as it seems naturally applicable to the meanings of words. The obstacle here is of different nature: the basic terms are defined through analyzing exiting sentences. It would be absurd to propose, in order to account for the theme-rheme division into sentences, their semantic representation in the form of

> (4.21)   'It is said about … that …'

because there would be no way of distinguishing the procedure for encoding straightforward sentences of the kind quoted as for example (4.15) – (4.19), and their counterpart with the formula (4.21) explicitly encoded, as in the following putative paraphrases of (4.15) – (4.19):

> (4.15')   *It is said about him that he was a very sick white man.*
> (4.16')   *It is said about Harold Waring that he noticed them first walking up the path from the lake.*
> (4.17')   *It is said about the wise traveller that he travels only in imagination.*
> (4.18')   *It is said about one that she was called Mrs Richman and that she was a widow.*
> (4.19')   *It is said about the bar that it was crowded.*

Moreover, since no ambiguity is allowed in the semantic representation, an important feature of Bogusławski's approach cannot be made use of, because the results would be contrary to this tenet. Since the formula in (4.21) is supposed to work showing multiple theme-rheme division, such sentences as (4.16) would need to have several representations at the same time, or a multilayer one, with either one representation of one layer corresponding to:

> (4.16")   *It is said about them that Harold Waring noticed them first walking up the path from the lake.*

While it can be said that (4.16) can be roughly paraphrased as either (4.16') or (4.16"). (4.16') and (4.16") themselves are not mutual paraphrases.

The same argument can be made for the *as for* formula. First of all in many cases the phrase used in the semantic explanation cannot be seen as corresponding to its surfaced, i.e. explicitly encoded counterpart. In Linde-Usiekniewicz [2000: 23; 42] I argued against the explanation of the positive degree of adjectives in phrases of the kind *mądra kobieta* 'a wise woman' in terms of 'mądra jak na kobietę (wise for a woman)', since it is self-evident that *a wise woman* and *wise for a woman* are not equivalent. In the case of *as for* it can be argued that it cannot be the semantic formula for representing themes, because sentences with surfaced *as for* are not equivalent to the sentences without it. To use the same example: (4.16) and (4.16''') are not equivalent:

> (4.16)   *Harold Waring noticed them first walking up the path from the lake.*
> (4.16''') *As for Harold Waring, he noticed them first walking up the path from the lake.*

Moreover, if the 'as for' is reserved for specifying themes within the semantic representation, the question arises what represents the surfacing *as for* element.

Additionally, Bogusławski's account allows for themes not appearing in the actual sentences (not by virtue of ellipsis) [Bogusławski 1977: 16; 24-29], while the theme-rheme duality is a constant feature. Thus it does not allow for thetic sentences. Again, within the encoding perspective this claim is too strong or at least would give results contradictory to the idea of the encoding grammar: the semantic representation contains only the information to be encoded, thus it does not contain the 'here and now' or 'there and then' (although 'now' and 'then' are encoded by grammatical tenses in many languages.

As far as the formal approaches, discussed above, are concerned, although they tend to be sentence internal and semantic in nature, they cannot be adapted to the Encoding Grammar. Too often they rely on surface features of sentences, such as their syntax, prosody, or specific particles. While there is no doubt that these means of encoding are important in the study of languages, within the encoding grammar they correspond to the structure of theme-rheme division in different languages and not to the representation of the division, which is semantic in character. A telling example of not enough cross-linguistic character of formal features can be found in Eretschik-Shir [2007: 10-12]. She contrasts the Catalan facts, presented here in (4.12) – (4.14a, b) with Danish data, where both old and new topics can be fronted, although the latter tend to be fronted more often. She concludes [2007: 12]:

> The point of this comparison is to show that a linguist studying Danish rather than Catalan may reach quite different conclusions as to what topic is by generalizing over the elements in each language. Clearly, the optimal answer to the question posed above is that the definition of topics should be universal.

Paradoxically, while the Encoding Grammar cannot adopt any of the approaches discussed so far, it needs to include many of the elements recognized within them. First of all it needs the dichotomy between the theme and rheme to be binary and symmetrical, i.e. it cannot rely on defining any member of the dichotomy by default. Thus it cannot follow the classic topic—comment or background—focus approach. At the same time it needs to be able not only to pick some parts of the meaning of the sentence as either themes or rhemes, but also to be able to represent complex configurations of themes and rhemes, i.e. possibly multilevel and hierarchical divisions of meanings into thematic and rhematic elements that would make up some larger whole. It should take into account the speaker's intentions, and only them, as the basis of defining the appropriate terms, but at the same time should be able to explain the parallelism between themes and what is presupposed on the one hand, and the parallelism between rhemes and what is posed on the other. These similarities should also include the division between what is presupposed and what is posed in the meaning of lexical units, even though in their case such distinction is not imposed by the speaker, but resides in the semantic structure of a given language. On the other hand it should be able to present some sentences as thetic and others as instances of neutralized divisions, and explain the difference between the two cases. It should be also able to specify the difference between ordinary themes and rhemes and prominent or contrastive ones. And finally, it should take into account that the division into themes and rhemes, though appearing as a part of the semantic representation of the utterance, needs to be accessible to all modules of the Encoding Grammar, as the means of encoding it range from lexical, through syntactical to morphological and phonological, thus the division has to be present not only in the semantic representation, but in all subsequent representations as well.

Finally, a brief explanation is needed why the Encoding Grammar cannot make use of the question and answer pairs in its analysis of the relevant phenomena. The main reason is that the question and answer approach is sentence external. Secondly, it assumes that the

utterance produced as an answer to a question has exactly the same information structure as the utterance produced without such prompt. This seems too strong an assumption. Moreover, the potential match between what appears in the question and the two types of answer possible, i.e. the "full sentence" answer and the reduced answer is subject to cross-linguistic variance (including ellipsis and pronominalization, but also acceptable sentence fragments). Some sentence-types can never be matched with appropriate questions; such is the case of existentials. This can be illustrated by an example taken from Osborne, Putnam, Gross [2011: 323],

(4.22)   *Finite VP exists*

That cannot be produced even as an answer to the question:

(4.23)   *What exists?*

A more telling instances are provided by multiple questions. First of all, not all languages allow them [Hengeveld, Mackenzie 2008: 89]. Additionally, if they do, in some languages they are subject to superiority effects, and in others—not [Eretschik-Shir 2007: 174-184]. In some languages, however, for example in Polish, the two interrogative words can be joined in a coordinate construction, while this is impossible in the answer. Thus the multiple question quoted in [Hengeveld, Mackenzie 2008: 89]:

(4.24)   *Who recommended whom to whom*?

can be translated into Polish in several ways: maintaining the syntactic structure and the linear order of the English version (4.25a); maintaining the syntactic structure of the English version but fronting all the question words (4.25b); and conjoining some or all of the wh-words (4.25c)[23]:

(4.25a)   *Kto polecił kogo komu?*
(4.25b)   *Kto kogo komu polecił?*
(4.25c)   *Kto kogo i komu polecił*?

Nevertheless, what is important here is that even in the answer to a conjoined variant the answer words will never be conjoined.

Yet another example illustrates even better the shortcoming of the question-answer test, here in relation to themes of the answering sentence being mentioned in the question. The language is Wambon, originally described by Lourens de Vries [1985], and the example in question is quoted both by Erteschik-Shir [2007: 41] and by Hengeveld, Mackenzie [2008: 90]. All the authors agree that the *nde* particle is supposed to mark both focus (apparently simple and contrastive), and contrastive topics (Erteschik-Shir glosses it as focus, and Hengeveld, Mackenzie as contrastive.

(4.25)   A: Nombone   ndu-ngup   antle-ngup?
              This   sago-and   banana-and?
              'What about this sago and bananas?
         B: Wembane   ndu-nde   takhima-tbo

---

[23] There are several accounts of this phenomenon in the literature, none of them pertinent here. Interestingly, in different accounts the judgments about acceptability of conjoined variants tend to differ.

Wemba	sago-CONTR	buy-3SG.PST.FINAL
Wemba bought the sago		
Karolule	ande-nde	takhima-tbo,
Karolus	ande-CONTR	buy-3SG.PST.FINAL
Karolus bought the bananas'		

Neither account specifies if the sentence uttered by B would have the same form, with the *nde* element repeated, if it was not an answer to the question and even if it were, whether it actually should be interpreted as contrastive/focal particle or as a coordinator, since this another of its function recognized be de Vries [de Vries, de Vries-Wiersma 1992: 71-72]. Additionally, it seems that the English counterpart of the answer B from (4.25), i.e.

(4.26)   *Wemba bought the sago and Karolus bought the bananas*

would be an appropriate answer to both (4.27a) and (4.27b):

(4.27a)   *What about the sago and the bananas?*
(4.27b)   *What about Wemba and Karolus?*

while Polish counterpart of (4.27a):

(4.28)   *A sago i banany*?

can be answered both with an utterance with the names of the kinds of produce fronted and with the names of the buyers fronted, or even with the buyer fronted in the first clause an the produce in the second:

(4.29a)   *Sago kupił Wemba a banany – Karolus.*
(4.29b)   *Wemba kupił sago a Karolus – banany.*
(4.29c)   *Wemba kupił sago a banany – Karolus.*

4.5. The Theme-Rheme Division Defined Within the Encoding Grammar

As has been already mentioned, within the Encoding Grammar the respective pair of terms adopted is that of *theme* and *rheme*, together with the term *theme-rheme division.* All phenomena that have been singled out in other frameworks and are pertinent to the proposal presented here are treated as special cases of themes, rhemes or the division itself. The starting point for defining the terms and the phenomena in question is the file-card definition of the theme, as proposed by Huszcza [1983: 57]. Thus it is it is accepted here that the theme is, metaphorically, singled out as a kind of keyword or label on the appropriate card. As the response to the criticism that this approach leaves the rheme undefined, it is proposed here, still metaphorically, to treat it as the information that should be written on that card or, in other words, as the entry to be put on the card. Thus, the metaphor presented here diverges radically from [Reinhart 1981]) where entire cards were either topics of foci. Another feature that distinguishes the card-file metaphor applied here is that the division between labels and entries is not meant to reflect in any way the actual arrangement of the information provided by utterances by the audience. As I have written [Linde-Usiekniewicz 2008a: 134]:

> The audience arranges the information any way it wishes, even treating the contents of the message in a way that was not planned by the speaker. In fact, it may act in the way described in *Foucault's Pendulum,*[24] when one of the protagonists decides to become "a kind of private eye of learning" (p. 189):
>
> > "Still, I was accumulating experience and information, and I never threw anything away. I kept files on everything. I didn't think to use a computer […]. Instead, I had cross-referenced index cards. Nebulae, Laplace; Laplace, Kant; Kant, Königsberg, the seven bridges of Königsberg, theorems of topology… It was a little like that game where you have to go from sausage to Plato in five steps, by association of ideas. Let's see: sausage, pig's bristle, paintbrush, Mannerism, Idea, Plato. Easy. Even the sloppiest manuscript would bring twenty new cards for my hoard." (p. 199).
>
> > The quote is not accidental. There is nothing to prevent the audience from treating every piece of information present in the message as a source of file-cards for their mental catalogue and – should there be no drawer[25] for some esoteric piece of information – from creating a new one.

In addition, the division between the label and the entry is meant to be valid only within a particular sentence, or even clause (in complex sentences), and in case of multiple divisions, for just a sentence segment. This could be seen in (4.20), where the entire contents of the first clause (Aniel's not returning at four o'clock) was re-elaborated as a theme of the second clause. Thus it is impossible within this account to claim that some sentence element is a theme of that sentence just because it is the theme of the next. Consequently, within the Encoding Grammar it would be impossible to claim, as Fortescue [2004: 159], does, that in the Nootka sentence

(4.30)   *Quʔac-a'qλ*
        person-inside
        'There is a person inside'

"the topic is *person* (tracked as such in following discourse)". It is not that 'person' cannot possibly be the theme of (4.30), but within the approach proposed here some sentence internal evidence is necessary.

Nevertheless, the distinction between the label and the entry would not be the appropriate way of defining themes and rhemes within the Encoding Grammar, even though this distinction has originally been formulated in terms of the speaker's instructions to the hearer. The reason is that, as it has already been mentioned, the speaker has no guarantee that the hearer will heed the instruction, and no means of forcing the hearer to do it. Thus the distinction has to be re-formulated in terms of the speaker's encoded intentions concerning respective portions of the meaning of the sentence. In order to do so, the Encoding Grammar makes use of the notion of *cognitive environment* taken from Sperber, Wilson [1986] , and their thesis that "when you communicate, your intention is to alter the cognitive environment of your addressees" [Sperber, Wilson 1986: 46]. The idea of altering the cognitive environment is further developed here by distinguishing two kinds of changes that can be made to the environment: quantitative changes and qualitative changes. Following the reasoning present in the approaches that rely on knowledge and similar concepts, quantitative changes would correspond to introducing a new fact (or to follow Sperber and Wilson's formulations, making a new fact manifest) within the cognitive environment of the addressee. By contrast, the qualitative change involves re-arranging the cognitive environment of the

---

[24] The quotes and the page references come from the English version, translated from the Italian by William Weaver, and published by Ballantine Books, New York, 1990.

[25] In the initial version of the metaphor themes corresponded to headings on drawers in an old fashioned library systematic catalogue.

addressee by bringing a fact already present to the fore (which would correspond roughly again to making it more manifest within the cognitive environment, and thus more relevant.

Although this distinction seems to resemble the traditional contrast between *datum* and *novum* or old and new information, the similarity is false. As the authors rightly point out, the speaker can but make assumptions about the cognitive environment of their addressee. Thus our definitions of thematic and the rhematic part of the sentence can be stated as:

(4.31a)   A thematic part in a sentence is any part of its meaning intended by the speaker to introduce a temporally qualitative change to the cognitive environment of the addressee and encoded as such, either directly or indirectly, with respect to some other part of the sentence meaning

(4.31b)   A rhematic part in a sentence is any part of its meaning intended by the speaker to introduce a temporally quantitative change to the cognitive environment of the addressee and encoded as such, either directly or indirectly with respect to some other part of the sentence meaning

Several comments are necessary here. First of all (4.31a, b) deal with speaker's intentions and their encoding. No reference is made to speaker's success, or conversely, their failure, in changing the cognitive environment of the addressee. Additionally, if some part of the sentence meaning encoded as intended to introduce a temporally qualitative change fails to do it, because there is no corresponding item in the cognitive environment of the addressee, as happens whenever a completely new information is encoded a thematic part of the sentence, this part of the sentence meaning does not become rhematic, even if the change actually introduced, either temporally or permanently, is quantitative. Conversely, it is irrelevant that items already present in the cognitive environment of the addressee correspond to the part of the sentence meaning encoded as intended to introduce a temporally quantitative change. Thus there is nothing wrong if items already available to the addressee are encoded as rhemes or items not actually available being encoded as themes, since whatever set of beliefs, convictions and assumptions the addressee or the audience possess, they are immaterial to the theme-rheme articulation of the sentence.

Secondly, the distinction between quantitative and qualitative changes presented in (4.31a, b), concerns the sematic representation of the sentence only. That is why the reference is made there to parts of meaning, and not parts of the sentence. Thus there are no constraints on the nature of the part of the sentence meaning to be encoded as either rhemes or themes, including "verbal" themes, i.e. themes referring to events, at least at the level of the semantic representation. This division, however, affects the way the meaning of the sentence is encoded. For example if the structure of a given language would not allow to encode verbs as themes, either directly or indirectly, the verbal meaning is encoded in a different way, i.e. through nominalizations, as in the following examples:

(4.32a)   Polish: *Złoty medal zdobył – Carl Lewis.* [Huszcza 1983]
                    The gold medal (acc.) won Carl Lewis (nom.)
                    'The gold medal was won by—Carl Lewis.

(4.32b)   English: *The gold medal winner is Carl Lewis.*

This is not to say that the nominalization is the only way to encode verbal theme; another device is that of pseudo-cleft sentences, e.g.

(4.32c)   English: *Who won the gold medal was Carl Lewis.*

On the other hand, the nominalization strategy may be used in languages that allow verb phrases as themes, if other elements of the meaning to be encoded makes encoding the verb as the theme unwieldly, as can be seen in the following pair of Polish sentences.

(4.33a)   Polish: *Decyzja dyrektora, żeby…zmartwiła wszystkich.*
                       The manager's decision to… worried everybody
(4.33b)   Polish: *Wszystkich zmartwiło, że dyrektor zdecydował, że…*
                       'Everybody was worried that the manager had decided that…'

Nominalizations can also be used in analytical constructions, i.e. light-verb constructions [Linde-Usiekniewicz 2008b].

Another example of alternative way of encoding the sentence to match the division into themes and rhemes is the use of either "compact", or more "extended" patterns. In compact patterns all the relevant elements are joined together in an appropriate syntactic pattern, in the extended pattern they are separated and not joined directly. Elements of meanings that correspond to a thematic part of a sentence might be thus encoded in a single or simple phrase, while elements of different character would be encoded in separate phrases. A straightforward example is provided by the way the same participant of the situation can be encoded either as an element of a noun phrase or within a separate noun phrase, i.e. with its rank downgraded or upgraded [Padučeva 1999: 223, original examples, emphasis removed]:

(4.34a)   *Статья Джона в "Times" рассердила мена.*
                       'John's article in "The Times" made me angry'
(4.34b)   *Джон рассердил мена своей статьей в "Times"*
                       'John make me angry with his article in "The Times".'

Within the framework presented here it can be said that the (4.34a) pattern makes it easy to encode 'John's article in "The Times"' as the theme, while the pattern in (4.34b) can be used to encode only 'John' as the theme and 'with his article' as a rheme.[26]

Thirdly, it should be noted that the thematic or rhematic character of an item can be encoded as such either directly or indirectly. Direct encoding may involve any device available within the language: thematic or rhematic particles, special syntactic patterns, linear order, intonation pattern or any combination of them. Indirect encoding concerns primarily those sentence elements that bear no overt mark of their themacity or rhematicity and are recognized as either thematic or rhematic by virtue of being complementary to an overtly marked element. A special case of indirect marking can be observed in languages with thematic particles, for example in Japanese. The particle *wa* normally follows the nominal head and directly marks it as the theme, but when accompanying complex nominals it indirectly marks the entire nominal phrase as the theme, or even, as in (4.35d) it may follow a phrase with a verbal head and mark the entire phrase as the theme:[27]

(4.35a)   *Ii   kuruma wa   takai.*
           good   car WA   expensive

---

[26] It should be noted that Padučeva 1999 does not relate the differences to the information structure but to different conceptualizations of the situation described.

[27] I am indebted to Bartłomiej Wojciechowski for the examples and his help with both structural and English glosses.

'Good cars are expensive'

(4.35b) *Shinsen na   yasai wa   oishii.*
fresh MOD   vegetable WA tasty

(4.35c) *Nogashita   sakana wa   ookii.*
Escape-PERF fish WA   big
'lit. the fish that we have let escape is big; fig. the grass is always greener on the other side (a proverb)'

(4.35d) *Hontoo ni nani   mo   wakaranai hito wa*
Truth OBL nothing   also   understand NEG WA
'[those who] know nothing
kono   tejun   doori ni   tsukuru no ga funan
this procedure   accordance OBL   safe OBJ NOM do
should follow the procedure'

This possibility of marking themes and rhemes either directly or indirectly is of particular importance when accounting for multilevel divisions into themes and rhemes, which postulates second order themes and rhemes in first order themes and rhemes (cf. 4.2. *A Brief Selective Typology of Approaches*, and particularly the analysis of (4.8a,b)).

The distinction between what is encoded as intended as a qualitative change and what is intended as a quantitative change can be extended to what is presupposed and what is posed. All presupposed elements of meaning correspond to items that should alter the cognitive environment of the speaker in a qualitative way, while the posed elements are meant to change it in a quantitative way. Thus themes are always presupposed and rhemes are always posed. The distinction holds both for elements of words meaning and for presuppositions introduced through appropriate syntactic device, as is the case of non-defining relative clauses [Keenan 1971: 14]. As it has already been mentioned, the difference between themes that are at the same time free forms within the sentence and thematic parts of meaning of words (lexical presuppositions or thematic dicta in Bogusławski's [1977] terms) is that while the division of sentence elements into thematic and rhematic depends mainly on the speakers, they cannot alter the distinction between thematic and rhematic components of meaning of a lexical item, thus the only recourse left to them is to choose a different word or an analytical expression.

Thus, since in Polish the evaluative distance meaning of dimension terms referring to height, as in *wysokie chmury* 'high clouds' contains the presupposition about the object x modified by the adjective [Linde-Usiekniewicz 2000: 197-198, translation mine]

(4.36)   'it is impossible not to think that x is located in a special way[28] in relation to the ground'

which prevents the adjective *wysoki* from being used in reference to high flying birds, airplanes, etc., as already mentioned in *2.2. The Architecture of the Encoding Grammar*, the adverb or the noun will have to be used:

(4.37)   *wysoko lecący ptak*
           'a bird which is flying high'

while *wysoki ptak* 'a high bird' can be used only in reference to the bird's stature, as in:

---

[28] This is a simplified version of the explication proposed; "special way" is further specified in the definition in terms of empty space between x and the earth, and this space being measured vertically (i.e. from top to bottom).

(4.38)   *Jak się nazywa taki wysoki ptak, który brodzi w wodzie?*
          'What is the name of the high bird that wades in the water?'

An apparent counterargument against identifying both presuppositions and themes with the intended quantitative change in the addressee's cognitive environment can be made out of non-defining relative clauses. By virtue of their being presupposed they should be encoded as intended to alter the cognitive environment in a qualitative way. Nevertheless, it is obvious that they are actually used to bring a quantitative change. If we take the famous Keenan's [Keenan 1971: 14] example:

(4.39)   *The Tiv, who respected Bohannon, are (are not) generous people.*

it is obvious that the fact that the Tiv respected Bohannon is mentioned there in order to change the cognitive environment of the addressee in a quantitative way. Again, the account proposed here seems to be facing the problem of the same item apparently being at the same time thematic and rhematic in character. The response to this conundrum lies in the fact that in sentences containing non-defining relative clauses what is presupposed in any non-defining clause is its truth-value, and not its actual contents. Thus the truth-value of the clause is encoded by making it a non-defining one, and not connecting it another way to the remaining part of the sentence, and the contents is encoded by choice of clause-internal lexical items and their syntactic arrangement. Therefore no contradiction is really involved.

The multi-level division of themes and rhemes is not the only way to encode the fact that several elements of meaning are thematic or rhematic. If no prominent themes or rhemes are involved and there is not specific narrative intent, similar to presented in (4.20), the simplest way to encode such division is to link all the thematic elements into one constituent, and all the rhematic element into the other, as could be seen in (4.34a) as opposed to (4.34b). Thus in (4.34a) 'John's article in *The Times*' is such thematic whole, and in (4.34b) the theme is 'John' and 'being made angry by his article in *The Times*' is a rhematic whole[29]. Similarly, in (4.32a) 'the winning of the gold medal' is the thematic whole. On the other hand, within the proposed framework, there is no way of postulating both 'John' and 'the article in *The Times*' in (4.34b) as a discontinued theme. Just the contrary, within a single division themes and rhemes have to be continuous and constitute a syntactic structure (a phrase) if not a constituent. In order to encode both 'John' and 'the article in the *The Times*' as themes, one would need to encode something like:

(4.40)   *Своей статьей в "Times"Джон рассердил мена*
          'With his article in "The Times" John make me angry'

where 'with his article' would be the first order theme, and John the second order theme, as presented below:

(4.41)   [*Своей статьей в "Times"*] $_{T^o}$ [[*Джон*] $_{T'}$ [*рассердил мена*] $_{R'}$] $_{R^o}$

Within the theoretical framework proposed for the division into themes and rhemes within the Encoding Grammar the distinction between continued themes and shifted themes becomes immaterial: Since the division affects a single sentence, it is irrelevant what was happening in the preceding discourse. In either case there is no encoding of the sentence

---

[29] The possessive here is not a means of encoding themacity, but serves to identify the article.

element as being intended to introduce a quantitative change of the cognitive environment of the addressee. On the other hand, if an element completely absent from the cognitive environment of the addressee is marked as theme, as often happens both in sentences constructed for the purpose of grammatical analyses, language teaching and in beginnings of narratives, a quantitative, instead of qualitative, change occurs. But this again is immaterial within the encoding perspective, as well as the fact that in larger narratives, themes of subsequent sentences can be used as elements of cross-reference to previous sentences to assure textual cohesion.

Prominent themes and rhemes are themes and rhemes respectively, by virtue of (4.32a, b), but their prominence serves as the means of encoding yet something else. In the case of prominent themes the speaker also encodes that the quantitative change affects what has been encoded as the theme and nothing else, or in other words, that they do not intend to alter in any way those elements of the cognitive environment of the addressee that may be somehow connected with the elements corresponding to themes. Thus in (4.8a) the speaker is explicitly encoding that they do not want to affect the cognitive environment of the addressee with respect to any other apples and in (4.8b) they are explicitly encoding that the change of the cognitive environment of the addressee in respect of the speaker's dislikes must concern only the green or unripened apples, and not other kind of apples.

Since the Encoding Grammar approach invalidates all reliance on context, previous discourse and even assumptions on the part of speaker about the actual cognitive environment of the addressee, it is clear that the analysis of utterances must be based on what is actually encoded in the utterance, either directly or at least indirectly, thus any thematic-rhematic division properly encoded must be observable in the surface form of the utterances. As it will be discussed in further chapters of this book, it may be not always possible to encode the theme-rheme division, because other encoding considerations may override the necessity of its encoding or the language structure may lack adequate means of encoding. Thus it is possible for the speaker to refrain from encoding the division and to count on appropriate changes being made (either temporarily or permanently) to the cognitive environment of the addressee at a later stage. However, although the speaker may count on further re-arrangements in terms of qualitative and quantitative changes, there is no way the division of the initial sentence into its thematic and rhematic part can be thus retrieved. This happens in the instances of so-called neutralized division. In such sentences some elements are still encoded as belonging either to the theme or the rheme, but other elements may lack such encoding. That would be the case of (4. 42), taken from [Erteschik-Shir 2007: 31-32):

(4.42)　　*Maxwell killed the judge with a HAMMER.*

where the final stress marks 'with a hammer' as undoubtfully rhematic, but it is not encoded whether the entire verb phrase should be considered rhematic or not. It should be noted that while in the sentences with neutralized division some part of it is still encoded as theme, in the thetic sentences the entire sentence is encoded as the rheme, as in (4.9)

If the division, i.e. the boundary between the thematic and the rhematic part of the utterance, is of paramount importance and never expendable, it is represented as a prominent division in the semantic representation. The division may be prominent because it affects a prominent theme or a prominent rheme, as has been proposed by Huszcza [1986; 1991b]. However, the Encoding Grammar allows for the division being prominent on its own, without involving marking either a theme or a rheme as prominent. This seems to be the case of the cleft and pseudo-cleft sentences, as illustrated by (4.32c), and discussed in [Linde-Usiekniewicz 2006: 90]. Cleft constructions are obviously not the only way of marking the division as prominent, since phonetically is always marked by pause.

As it have been repeatedly insisted on, the Encoding Grammar on the one hand limits itself to single sentences, i.e. it does not take into account any meaning encoded in the previous or the subsequent sentences of a longer utterance or passage. On the other hand it necessarily accounts for everything that happens to be encoded in a single sentence, however complex. The division into themes and rhemes in complex sentences presents a serious difficulty for the Encoding Grammar. Since themes and rhemes are identified by the way they are encoded, each of the clauses constituting a sentence may have its own theme and rheme, i.e. an element of meaning encoded as intended by the speaker to alter the cognitive environment of the addressee in a qualitative or a quantitative way. On the other hand, if a eventual complex sentence contains various instances of encoding the same chunk of meaning, such chunks may intended to be encoded as theme within one clause and as rheme within another[30].

Moreover, if the theme-rheme divisions have to respect the articulations into clausal elements, and particularly, if neither theme or rheme can correspond to a sequence that does not constitute at the same time a syntactic whole, multiple theme-rheme articulations may occur even when encoding a single clause. That was actually the case of (4.41), where the meaning 'with his article John' could not be encoded as a single theme.

The solution comes from the way the theme and the rheme have been defined (cf. 4.31a, b). Both the qualitative and the quantitative changes are meant to be introduced temporarily. This was invoked to argue that what is encoded as rheme in one sentence can be encoded as theme in the next. Thus it can be argued that when the *as for* formula is used for encoding, it encodes the coordinated phrase as a theme, but only for the extraclausal element. Within the clause, several different arrangements of themes and rhemes is possible as can be seen in (4.43-4.45)

(4.43) *Co do sago i bananów, sago kupił Wemba a banany – Karolus.*
„As for sago and bananas, (the) sago was bought by Wemba and (the) bananas by Karolus

(4.44) *Co do sago i bananów, Wemba kupił sago a Karolus – banany.*
1. 'As for sago and bananas, Wemba bought (the) sago and Karolus – (the) bananas'
2. 'As for sago and bananas, it was Wemba who bought (the) sago and it was Karolus who bought (the) bananas'

(4.45) *Co do sago i bananów, Wemba kupił sago a banany – Karolus.*
 'As for sago and bananas, it was Wemba who bought (the) sago and Karolus bought (the) bananas'

 In (4.43) in the clause following the introductory *co do* the two kinds of staples are encoded as separate themes; depending on the intonation pattern (4.44) can encode either the buyers as two different themes and the staples as rhemes or the buyers as two different rhemes with staples as themes; (4.45) encodes the buyers as two different rhemes (and does it through different procedures) and the two staples are encoded as themes.

So far this section has been concerned with the basic notions of the theme-rheme divisions and the way they interact with other elements of the semantic representation of a sentence. As I have repeatedly mentioned in previous chapters, there is a great cross-linguistic variation as to means languages have to encode these representations. However, it seems impossible to postulate the existence of a theme-rheme division structure in each

---

[30] This issue is different than that presented by non-defining relative clauses; there two different elements of meaning were identified i.e. a sort of „blanket" truth-value and the actual contents.

language, in parallel to the distinction between the semantic representation of sentences and the semantic structure of language. This is due to the fact that the theme-rheme division of the semantic representation has no corresponding encoding level and in many instances the divisions are encoded indirectly. Even languages that have special thematic particles, like Japanese and Korean [Huszcza 1974; 1979; 1983; 1987] encode rhemes by default, and conversely, languages with rhematic particles, such as Gungbe [Aboh 2010], encode only nominals rhemes, while phrasal rhemes, i.e. those corresponding to verb phrases, are encoded as rhemes only indirectly or by other means [Pawlak 1984: 196-213]. Other differences, already discussed concern the already mentioned possibility of presenting verbal phrases as themes, and special syntactic devices, such as availability of cleft and pseudo-cleft constructions, and different left dislocation patterns for fronted themes and fronted rhemes in some Romance languages. It should be noted, additionally, that most of these syntactic devices do not work on its own, but usually in concert with yet another means of encoding the theme-rheme division, i.e. the intonation. In these instances the syntactic pattern of the sentence cannot be said to encode the theme-rheme division on its own. On the other hand, other means, normally associated with syntactic functions, like differing agreement patterns, different case markings, etc., serve as indirect markers of either themacity or rhemacity of an element or of the presence or absence of a division. Some of these cases will be presented in Chapter VI.