Focus
in the Meta-informative Centering Theory

André WLODARCZYK[1] & Hélène WLODARCZYK[2]

Centre de Linguistique Théorique et Appliquée

Université de Paris - Sorbonne

1. Theoretical framework

The theoretical framework of this approach is the structural theory of language, stemming from the European tradition and based upon the study of various languages. Despite the distinction usually made between functional[3] and formal linguistics, the formalisation of structural (mostly functional) linguistic theories is feasible. In order to achieve this task the logical reconstruction of concepts, objects and methods underlying structural theories must be undertaken, i.e. formal models of these theories must be built. Our Meta-Informative Centering Theory is a formal model of linguistic information, that is to say the way information is linearised and expressed in human languages. In this model, as will be explained below, an important distinction is made between informative and meta-informative levels.

1.1. Grammar as interface between form and content

A significant achievement of structural linguistics consists in the comprehension of the multi-dimensional character of language structure which underlies the linear order of utterances. As a matter of fact, the distinction between Form and Content (as in Saussure's and Hjemslev's works) corresponds to different levels of language structure. In the framework of the present theory, **Content** comprises both semantics and pragmatics. Semantics is multi-dimensional because it reflects complex situations of the real world. Pragmatics concerns the way speakers represent situations, creating centres of attention and using what they suppose to be the common knowledge shared by the addressee. **Form** comprises phonology, morphology and "grammar". We call *grammar* (and not *syntax*) the device enabling the speaker to build "meaningful" messages out of "formal" elements in order to show that linearising morphemes into messages takes place not only in the *syntactic* dimension but also in the *paratactic* one. Needless to say that phonological and morphological levels of language are also structured in the two dimensions which were called syntagmatic and paradigmatic axes by Saussure.

---

[1] Professor at Université Charles de Gaulle (Lille 3).

[2] Professor at Université Paris-Sorbonne (Paris 4).

[3] The functional view of language as a communication means should not be confused with the functional approach to human behaviour in psychology.

CONTENT (signified)	**Pragmatics**	Centres of attention Honorific treatment of persons Relations between speaker and hearer, etc.
	Semantics	Participants with their semantic features, Situations and roles, etc.
GRAMMAR (interface)	**Syntax & Paratax**	Constituency relations Dependency relations, Paradigmatic relations, etc.
FORM (signifier)	**Morphology**	Lexical morphemes Grammatical morphemes
	Phonology & Prosody	Phonemes Intonemes, etc.

Table 1. Grammar as interface between form and content

Moreover, syntax uses at the same time both constituency and dependency relations, in spite of the fact that they were considered as two possibly distinct models of syntax[4]. The constituency model provides the concepts of *global* and *local* scope (the first noun phrase on the left of the constituency structure is the *global* constituent, the noun phrase which is part of the verb-phrase is the *local* one) while the dependency model uses the concepts of *main* and *dependent* member of the syntactic structure. Both kinds of relationships can be encountered in linguistic expressions.

1.2. Information and Meta-information

Most linguists refer to the problem of conveying "information" in utterances as the problem of *information structure*[5], but in fact it should be rather called "meta-information". Indeed, the semantic content of the utterance can be seen as the information itself, and the different forms that may be chosen to convey this information should be properly considered as *meta-information*.

Although meta-information is expressed by linguistic markers inside the boundaries of the utterance, it cannot be properly described unless the utterance is considered as part of a complete discourse unit, taking into account what precedes and what follows. In other words, the meta-informative structure of each utterance depends on its position within a discourse and is related to the coherence and cohesion of this discourse. Moreover, the extra-linguistic context in which a sentence is uttered also gives shape to the meta-informative structure of this utterance.

---

[4] Constituency was first used by Harris and Chomsky in generative grammar while the term Dependency has become known through the works of Tesnière and Mel'cuk..

[5] Also called communicative structure.

INFORMATIVE LAYER *(0-order)*	*typical semantic unit :* schemata
META-INFORMATIVE LAYER *(1st order)*	*linguistic unit uttered in a context :* (1) simple utterances (2) extended utterances
COGNITIVE LAYER *(2ND order)*	*organised set of utterances :* texts, dialogs

Table 2. Three layers of linguistic information

1.3. Hierarchical model of utterances

Let us consider the following utterance:

(1) Mary treats Peter.

The semantic content of this utterance, its *information,* can be represented by the following schema[6]:

$$
\begin{array}{l}
\text{SIT} : treat = \text{“treets”}\\
\qquad (treater \quad : \text{“Mary”})\\
\qquad (treatee \quad : \text{“Peter”})
\end{array}
$$

In such a schema, there is no hierarchy between the participants of the situation. But even such a simple utterance as #1 does not consist exclusively of information: it also comprises meta-information. Indeed, the informative content of utterances is always enriched by meta-informative pointers indicating the speaker's *centres of attention* (henceforth CA). Attention is that salient chunk of knowledge to which epistemic agents are attuned by periods of time. Whenever a speaker decides to communicate, usually he partitions his representation space into its Centre and Periphery. Such partitioning is necessary in order to linearise the salient information.

As regards utterance #1, the choice of "the treater" as SUBJECT of an active verb makes it the global CA of the message, i.e. this already belongs to the meta-informative level. Since no judgment may be built without selecting at least one CA, a linguistic utterance is always at least made up of both information and meta-information. Consequently, syntax must be considered as a form not only for *semantics* (relations between situation participants or information) but also for *pragmatics* (indirect reference to semantic situations or meta-information)[7].

The difficulty of analysing utterances comes from the heterogenous and compound nature of what is usually called "syntax" (assemblage of varied means). Indeed, in our theory, syntax (which works together with paratax) is the interface between *Form* and

---

[6] Such schemes are sometimes called "logical forms" of sentences but this term is extremely confusing because the logical theory of predicates, although originally derived by Aristoteles from the study of human language, should not be considered as a model of language but rather as a model of reasoning ; it belongs to logic, not to linguistics.

[7] Gutierrez Ordonez S. (1997) names these two concepts : "representar" (to represent) and "informar" (to inform) respectively.

Content. Some parts of it are used by the speaker who selects linguistic items and organises them according to his centres of attention. Thus, syntax is first of all a device to put semantic information (which most probably is not always sequential) in an inevitably sequential order and at the same time add meta-informative pointers to it. Consequently, the utterance schema called valence covers a complex heterogeneous bag of varied relations which must be analysed on different levels of interpretation.

For example, let us consider the following utterance:

#2 Mary treats Peter with aspirin

As concerns its form, this utterance can be analysed on many successive levels : phonology, morphology and grammar. Because of this multi-level structure, models of utterances must be hierarchical. Leaving aside the phonological level, we recognise on the level of morphology a verb and three nouns with their specific morphemes. On the level of grammar we recognise syntactic and paratactic relations. The syntactic structure can be seen both (1) as the constituency structure

$S(NP_1(PN(Mary)), VP(V(treats), NP_2(PN(Peter)), PP(Prep(with), NP(Noun(aspirin)))))$

in which the first NP_1 is global, the second NP_2 is local (being a constituent of the VP), and (2) as the dependency relationship of the main noun phrase (SUBJECT : "Mary") with the two dependent noun phrases (direct OBJECT : "Peter", and indirect OBJECT : "with aspirin"). Though not immediately observable, the paratactic relations concern the identification that takes place in each noun constituent of the utterance : "Mary" as an element of the set of persons of feminine gender known to both the Hearer and Addressee, etc.

The content of the utterance #2 is made up of (semantic) information and (pragmatic) meta-information. Information concerns the situation type (an action: "to treat"), its participants (two persons "Mary" and "Peter" and a substance "aspirin") playing roles in the situation (Treater, Treatee and Means for treating), its spatial and temporal anchors, modality, truth-value etc. On the level of meta-information, we consider this utterance as predication with three CAs (SUBJECT and OBJECTS). Actually, we decided to reserve the use of the term "predication" for the pragmatic level[8]. This, indeed, matches the etymological meaning of "predication" : before being formalised by logicians, this term referred to what we defined as the *meta-informative structure* of utterances. As the matter of fact, the "predication" is what is being said about the "subject" of a message. We explain therefore We therefore explain that the Subject-Predicate relationship is a meta-informative bipartition of an utterance.

2. Centres of Attention

2.1. Simple and extended utterances

It must be stressed that no utterance is informatively neutral inasmuch as it occurs as a message and is part of a text or discourse (as short as it may be). Any utterance always contains at least one CA, although the CA may sometimes be only implicit[9]. We call CAs

---

[8] In logic, the term predicate is used as a synonym of a relation with n arguments and linguists often use this meaning of the term predicate in semantic theories (cf. Wlodarczyk A. 2004).

[9] in so called impersonal sentences the global CA is implicit as "anonymous subject" (Wlodarczyk H. 1996, Wlodarczyk A. & H. 2006)

those components of the semantic situation which are singled out by a pragmatic, meta-informative operation establishing a hierarchy between them. This hierarchy is expressed in linguistic utterances by different markers. One of the most important is the linear order of syntactic constituents: the global CA is usually at the beginning of the utterance. As a matter of fact, as soon as an utterance contains more than one CA one of them is *global* and the other(s) *local*. But not every utterance contains a TOPIC or a FOCUS. Thus, we distinguish two types of utterances which we call *simple* and *extended* (cf. table 2 above). On the first level, the level of *simple utterances,* SUBJECT and OBJECT serve as CAs. On the second level, the level of *extended utterances,* two other CAs — TOPIC and FOCUS — may be added to the first ones. Moreover, *simple* and *extended utterances* are part of texts. On the level of texts or discourses, we refer to CAs as GENERAL or PARTICULAR THEMES which play an important role in the coherence of the whole[10]. Thus, it must be stressed that, in our theory, TOPIC (of an utterance) and THEME (of a text) are not synonymous terms (cf. Table 3).

TYPE OF EXPRESSION	CENTRES OF ATTENTION	
	Global	*Local*
1.1. Simple UTTERANCE	SUBJECT	OBJECT
1.2. Extended UTTERANCE	TOPIC	FOCUS
2. TEXT / DIALOG	GENERAL THEME	PARTICULAR THEME

Table 3 Pivots of discourse

In the present theory, *utterances* are called *simple* if they are not divided into *new* and *given* information, if they are either entirely *new* or entirely *given*. In other words, in such utterances, SUBJECT and PREDICATE refer to the same kind of information which is either *new* or *old*. On the other hand, extended utterances consist in two contrasting parts: a CA and what is said *about* it, each pointing to information of different type. A TOPIC corresponds to *given* and is a CA of a *new* COMMENT. Conversely, FOCUS corresponds to *new* and is a CA of a *given* BACKGROUND. Thus, in our view, TOPIC is not a mere synonym of *given information*, nor is FOCUS synonymous of *new information.* As we have stated, a whole simple utterance may be either *given* or *new* (all the parts of its PREDICATION refer to the same kind of information) but an extended utterance always contains as well *given* as *new* reference to information (one part of its PREDICATION refers necessarily to *given* while the other part refers to *new* information). The extension of the utterance consists namely in that the Speaker intends to join, in the same utterance, a *given* chunk of information with a *new* chunk of it. The distinction between the TOPIC and the FOCUS is dependent on the kind of information as expressed in the other part of an utterance (COMMENT or BACKGROUND). We shall see below that the notions of *new* and *old* information, as they were first introduced into linguistics by the Theory of Functional Sentence Perspective (also known as the Prague School Theory of Information Structure), need to be redefined and further divided into two different types of *givenness* and *newness*.

---

[10] In this paper, we shall not dwell on the concepts of *general* and *particular themes* of a text, as an example, these concepts were used by Hélène Wlodarczyk 2003 in the analyses of a scene of a drama by Sławomir Mrożek.

Centering of Attention can be defined as a function. The first of its operands (the predecessor) is present in the expression, while the other (the successor) is absent. For this reason, it is said that the same segment may be a member of two relations (syntagmatic and paradigmatic) since the predecessor operand term of a paradigmatic relation is always a phrase (a syntagma). From a set-theoretical point of view, CAs are *distinguished* (selected, marked) elements of sets (or members of set families). From a topological point of view, CAs can be modelised as *distinguished places* in the Information Space, i.e. as **Meta-Information:** (a) SUBJECT and OBJECT are single topological spaces which represent terms referred to and (b) TOPIC and FOCUS are homotopies[11] from points to spaces (broadening) and from topological spaces to points (narrowing) respectively. From a logical point of view, CAs express both **reflexive identity** (as defined in predicate calculus) and **relative (analogical) identity** based on set theoretical relations[12]. CAs should therefore be defined with respect to both paradigmatic and syntagmatic dimensions at the same time, i.e. as (1) entities the identity of which is established on a given paradigmatic level and as (2) constituents of the syntagmatic level (syntactically determined) which correspond either to the SUBJECT or the OBJECT of the analysed utterance. The presence of an OBJECT (as local and dependent CA) in an utterance presupposes that of a SUBJECT (as global and main CA)[13]. In the same way the presence of a FOCUS presupposes that of a TOPIC. In other words, SUBJECT implies OBJECT, and TOPIC implies FOCUS. As a consequence, utterances containing a FOCUS often contain also a TOPIC. In such a case, TOPIC, as global, always precedes FOCUS because it is local (see examples in table 5 below). It must be emphasized once more that utterance extension has nothing to do with compound syntactic sentences: an extended utterance contains at least a TOPIC or a FOCUS or both. As will be shown below, (focalising) utterance extension may be marked by different prosodic, morphological and syntactic devices.

2.2. Centres of Attention of Simple Utterances and Semantic Roles

Most linguists consider that SUBJECT and OBJECT belong only to the syntactic structure and that the information structure exists only when the utterance includes a TOPIC and or a FOCUS. This approach differs radically from current linguistic theories because we claim that the so-called "syntactic" or "grammatical" notions of SUBJECT and OBJECT belong to the pragmatic level as CAs of non-extended utterances. Consequently, there is no hierarchy between arguments as parts of the semantic level but rather between CAs on the pragmatic level. SUBJECT is "higher" in the hierarchy than OBJECT because it is global. Nonetheless, in Indo-European (non-ergative) languages, there is a tendency to choose as SUBJECT (global CA) the active *animate*, most frequently *human*, participant of a situation. However, this is only a tendency, because the semantic level remains independent of the meta-informative structure. It is always possible to express the human agentive participant as an (indirect) object in a passive structure, or to keep it implicit in the passive, e.g. *this book is beautifully written (no matter by whom)* or in an utterance

---

[11] Cf. Wlodarczyk A. 2003.

[12] For more details on this point see Wlodarczyk André 2005.

[13] That is one of the reason why we refuse to consider as a SUBJECT the dative OBJECT in a so called *impersonal* utterance in Slavonic languages, e.g. in Polish: *Starszej pani zrobiło się niedobrze.* .(lit. *to the old-lady it became bad, The old lady felt bad*.). Such utterances contain what we call an anonymous subject of zero form in Polish, of the form *it*, called 'dummy subject' in Generative syntax. Cf. Wlodarczyk H. 1996 and Wlodarczyk A. & H. to be published.

with what we call an "anonynmous subject"[14], e.g. *napisano książkę (lit. was written a book)*.

In each language, syntax provides the speaker with different formal devices (diathesis, linear order etc.) making it possible to map in diverse ways the *pragmatic* CA onto *semantic* participants. These mapping possibilities are chosen in order to keep the text coherence. A story teller who chooses *cats* as his general *theme* will probably prefer the active utterance : *cats chase mice*. But if his story is about *mice*, he will prefer the passive form: *mice are chased by cats*. Thus, semantic roles are of very different nature than centres of attention (such as SUBJECTS and OBJECTS) they belong to two different, respectively informative and meta-informative, levels. Although there is no clear-cut distinction between semantic and pragmatic parts of linguistic content, we adhere to the widespread opinion that it is good to distinguish between some parts of content as semantic and other parts as pragmatic. Consequently, it should be stressed that the organisation of semantic roles[15] does not depend on the hierarchical arrangement of pragmatic CAs (which are syntactically global and local).

3. Centres of Attention of Extended Utterances

Let us now consider the utterances #3 and #4 below where successively the SUBJECT and the OBJECT of utterance #2 have been replaced by a TOPIC.

#3 <u>As regards Mary</u>, she treats Peter with aspirin.

#4 <u>As regards Peter</u>, Mary treats him with aspirin.

Utterances #3 and #4 consist of two parts: a TOPIC and a COMMENT which has the form of a simple sentence. In English extended utterances, the participants of the semantic situation which are introduced as TOPICS must be further referred to by anaphoric pronouns in the simple sentence that follows (eg. *she* in #3 refers to the TOPIC Mary, and him in #4 refers to Peter).

On the other hand, in utterances #5 and #6 below, the SUBJECT and the OBJECT of utterance #2 have been replaced by a FOCUS.

#5 <u>It is Mary who</u> treats Peter with aspirin.

#6 <u>It is Peter whom</u> Mary treats with aspirin.

The focalisation (or topicalisation) of a constituent of the simple utterance produces an extended utterance consisting of two parts: a FOCUS and a BACKGROUND (or a TOPIC and a COMMENT). In utterances #5 and #6 FOCUS is marked by cleaving which — on the level of syntax — splits the utterance into two parts: the FOCUS (as main clause) and the

---

[14] So called "impersonal" Sentences in Indo-European languages as a matter of fact always contain an "anonymous" (zero or dummy) subject. In Japanese, subject may remain implicit, cf. Wlodarczyk A. & H., (to be published in 2006).

[15] Wlodarczyk A. (to be published in 2006) defines semantic roles differently than Charles Fillmore in his deep cases theory. Instead of using such roles as *agent*, *patient*, *experiencer*, etc. Wlodarczyk 2006 uses terms representing elementary concepts that can be combined in different configurations: entities (divided into agents and figures) may particpate in situations in ACTIVE, MEDIAN or PASSIVE roles. Moreover, each participant (a named entity which fills a semantic role) of a situation may be described according to that situation by one or several semantic features (control features: *purpose* and *feedback*, emotive features: *desire* and *intention*, epistemic features: *belief* and *communication*).

BACKGROUND expressed as a dependent clause beginning with a relative pronoun. However, we must emphasize that focalisation or topicalisation always produces extended utterances even when thie latter remain syntactically simple. As a matter of fact, in English (as in many other languages), it is possible to mark FOCUS only by intonation, as in #7:

#7 __Mary__ *treats Peter with aspirin.*

In #7, *Mary* is a FOCUS as an answer to the question *Who treats Peter with aspirin ?* or when it serves to deny utterance #8:

#8 *The doctor treats Peter with aspirin.*

Just as utterance #5, utterance #7 is made up of two contrasting parts of meta-information: a FOCUS (new information) and a BACKGROUND (old information): what is supposed to be known to both speaker and hearer is that *somebody treats Peter with aspirin.* Thus, CAs of the second level (TOPIC and FOCUS) differ from those of the first level (SUBJECT and OBJECT) in that their meta-informative value contrasts with that of the COMMENT or BACKGROUND.

3.1. Two kinds of given/new distinction

Let us now consider with more detail the problem of information status. Meaning is information while predication is communicating __about__ selected chunks of it, i.e. what we have proposed to call 'centres of attention' (creating linguistic expressions in which the distinguished segments are highlighted). This 'aboutness' is of two kinds: either it is anaphoric (referring to what should be taken as known by hearsay during previous speech acts) or cataphoric (referring to what should be taken as unheard of in any previous speech act).

On the other hand, from a cognitive viewpoint, semantic situations (information) are typed as *generic, general, habitual* or *potential*, and the occurrences of these types can be defined as their dual counterparts[16] : *specific, particular, occasional* or *actual* respectively. The first two situation types can be easily organized hierarchically using the inclusion relation: *general < generic* and *particular < specific*[17].

Situation Types	Situation Occurrences
Generic > General *Habitual* *Potential*	*Specific > Particular* *Occasional* *Actual*

Table 4. Common knowledge-based informative motivation of aboutness

---

[16] The distinction *semantic situation type / semantic situation occurrence* can be compared to what is known as *continuant* and *fluent* entities in philosophical ontology, or *persistent* and *transient* information in information science. Needless to say that the *generic/specific* distinction recalls the Universal and Existential logical quantifiers.

[17] The duality described as *situation types* and *situation occurrences* is comparable to the distinction made by Von Wright in 1963 between "generic actions" and "individual actions", the former being performed repeatedly on different occasions, the latter concerning actions that are performed on a concrete occasion and at a certain instant.

In predication, the typed situations and their occurences are straightforward motivations for given and new meta-informative presentation of CAs. Thus, we consider today[18] that there are two kinds of given/new distinction[19]:

- (a) the first one is **speech-oriented** (**s**-*given* or **s**-*new*) because it corresponds directly to meta-informative contents such as the conceptual pair *anaphora* and *cataphora*

- (b) the second one is **information-oriented** (**i**-*given* or **i**-*new*) because it refers to what is motivated by *information* (semantic situation types and their occurrences) (table 4)

Consequently, two sorts of *given* or *new* predications should be distinguished (1) meta-information concerning *speech act orientation*, (2) *information about information*, hence also of a meta-informative character. We consider the above distinction as evoking the speech act oriented belief as opposed to the common belief (the latter including encyclopedia-like knowledge and that knowledge which characterises individual experience of the speech act participants).

Focalisation is always s-new and i-new: its markers normally combine with information on *semantic situation occurrence* adverbs (*only, even, once, twice, on that occasion…*) that is to say with expressions having a *specific, particular, occasional, actual* and/or *cataphoric* value. Indeed, it is worth emphasizing that our definition of FOCUS is more restrictive than the definition largely used in linguistics[20]. In our theory, not every piece of new information is necessarily a focus. We call FOCUS only that part of the utterance which is introduced as *new* information contrasting with *given* information in the rest of the utterance (the back-ground). In particular, we make a clear-cut distinction between extended and non-extended utterances. In simple (non-extended utterances), SUBJECT (as global CA) and PREDICATE have the same informative significance, and there is a sort of meta-informative 'agreement' between them.

3.2. Mapping TOPIC, FOCUS onto SUBJECT, OBJECT and Semantic Roles

As was the case with CAs of simple utterances, in extented utterances languages offer the possibility of mapping in many various ways semantic participants onto pragmatic centres of attention. Thus, if we consider utterance #2 once more, it is possible topicalise or focalise any of its constituents. In frame 1 hereafter, we enumerate only the most frequent possibilities of extending utterance #2 with a TOPIC, a FOCUS or both, using diathesis as a device making it possible to change the mapping from semantic participants onto CAs of the first meta-informative level. The change of linear order between (direct) OBJECT and indirect Object is also related to meta-informative operations[21]. As for

---

[18] Indeed, in the previous form of this theory (WLODARCZYK, A. & WLODARCZYK, H. – 1998 and later), we used a dual scale of validity with generic and anaphoric values as extremes on the one hand and with specific and cataphoric values as extremes on the other hand.

[19] Cf. In this volume the paper by Shatunovskiy I. "Two types of new – given and ways of introducing new objects into the world and discourse".

[20] Cf Lambrecht 1994

[21] For the sake of concision, we do not devote more attention to this problem in this paper. The linear ordering of direct and indirect object was discussed in Wlodarczyk H. 2004 in the contrastive perspective between French and Polish.

semantic participants, we use the terms defined in our theory of semantic roles[22]. We use the sign || to indicate that a constituent concerns at the same time a semantic role and one or two CAs (of both levels). In the frame below, it is obvious that first or second level CAs are independent of semantic roles. It must be added that they are also independent of each other, i.e. topicalisation and focalisation[23] may concern any NP[24] of the utterance, not only subjects or objects.

1a. Mary treats Peter with asprin.
 (Active verb + [Subject || Initiator] + [Object || Terminator] + [Indirect Object || Mediator])
1b. Peter is treated with aspirin by Mary.
 (Passive verb + [Subject || Terminator] + [Indirect Object || Mediator] + [Object || Initiator]).
2a. As for Mary, she treats Peter with aspirin.
(Active verb + [Topic || Subject || Initiator] + [Object || Terminator] + [Indirect Object || Mediator])
2b. As for Peter, he is treated with aspirin by Mary.
(Passive verb + [Topic || Subject || Terminator] + [Object || Initiator])
3a. As for Mary, it is Peter whom she treats.
 (Active verb + [Topic || Subject || Initiator] + [Focus || Object || Terminator])
3b. As for Mary, it is with aspirin that she treats Peter.
(Active verb + [Topic || Subject || Initiator] + [Focus || Indirect Object || Mediator] + [Object || Terminator])
3c. As for Peter, it is Mary who treats him with aspirin.
(Active verb + [Topic || Object || Terminator] + [Focus || Subject || Initiator])
4a. As for Peter, it is Mary by whom he is treated .
(Passive verb + [Topic || Subject || Terminator] + [Focus || Object || Initiator])
4b. As for Mary, it is Peter who is treated by her.
 (Passive verb + [Topic || Object || Initiator] + [Focus || Subject || Terminator])
etc.

Frame 1 Different mappings of semantic roles onto CA of simple and extended utterances

On the other hand, focalisation may concern any NP without or with a preposition, be it a CA (governed by the verb: SUBJECT or OBJECT) or any adverbial or adjunct NP. Here is an utterance (#10) where a prepositional NP indicating time is focalised.

---

[22] The specificity of this theory consists in defining semantic roles on the basis of a finite set of semantic features used to define elementary infons and making it possible to avoid arbitrarily creating a new role whenever the a priori list turns out to be insufficient (Wlodarczyk A. 2005).

[23] In the following part of this paper we shall concentrate exclusively on focalisation (to cope with the general theme of the present volume).

[24] Topicalisation or focalisation of the VP itself is possible under some conditions that vary from language to language. These operations seem to be possible when the verb is nominalised in the broader sense, it be gerund, infinitive, to + infinitive. Cf the oral communication of Pierre Cotte at Symposium on Focus at University Paris 4 in November 2003 : *Focus in some infinitive constructions in contemporary English*: "The constructions to be discussed are 'He saw flight to be impossible, He felt the plan to be all wrong, He is thought / believed / said to be over 100, He is sure / certain to win, He was unable to resign' and a few others." Pierre Cotte assumes that "the complex syntactic structure reflects meaning, i.e. underlying operations of cognition, and attempt to show that 'to' indicates focus on the predicate at the end."

#10 [...] **it is only at this point that** *two-dimensional picture-making becomes a powerful vehicle for the transmission of ideas and the prevailing cultural norms[...]* (Derek Hodgson)

In utterance #11, the focalised prepositional NP expresses spatial anchoring of the situation: "in our beautiful France".

#11 Alas! I have often since last night caught myself wishing for that fiend's death. But what you suggest is impossible! The laws of this country do not permit of murder! **It is only in our beautiful France that** *wholesale slaughter is done lawfully, in the name of Liberty and of brotherly love ;* (Baroness Orczy)

The possibility of focalising (or topicalising) an embedded subordinate clause is based on the fact that in complex sentences, embedded subordinate clauses occupy the same syntactic positions as NPs. For instance, it is possible to focalise an embedded time clause as in example # 12:

 # 12 She was in the mood when the sea has a saddening effect upon the nerves. **It is only when we are very happy, that** *we can bear to gaze merrily upon the vast and limitless expanse of water, rolling on and on with such persistent, irritating monotony, to the accompaniment of our thoughts, whether grave or gay.* (Baroness Orczy)

In the following utterance (#13) a causal subordinate clause has been focalised:

#13 **It is because** *there has been a lack of plausible attempts to interpret art within the larger framework of evolutionary history [...]* **that** *such proposals continue to be taken seriously [...]* (Derek Hodgson)

In utterances #12 and # 13, focalisation is expressed by cleaving, but in the case of embedded clauses focalisation can also be obtained simply by intonation and sometimes by changing the place of the embedded clause with regard to the main clause[25].

4. FOCUS in different languages

4.1. FOCUS marking

As concerns word order and informative structure, in most languages, given information is normally introduced at the beginning of the utterance, whereas new information is kept for the end. "Normally" means that this order prevails in utterances with the main intonative stress at the end. When this order is reversed, new information appears at the beginning of the utterance and this type of intonation is described as marked. In languages where word order indicates the relations between verb and first level CAs (SVO order in English or French) the non marked intonation with final stress can be used to focalise the OBJECT.

#14 — What did Mary buy ? — Mary bought **a hat.**

But in this type of language (with SVO order), the marked intonation with initial stress can be used only to focalise the subject.

#15 — Who did buy a hat ?— **Mary** *bought a hat.*

---

[25] This problem is discussed in this volume by Pasch and Dalmas for German and by Breuillard for Russian.

Otherwise, languages in which the relations of NPs to the main verb are marked by case endings (for example, Slavonic languages, except Bulgarian) make a more intensive use of word order to mark focus. In Slavonic languages, thus, it is possible to focalise the SUBJECT of a verb only by putting it at the end of the sentence.

#16 *Kapelusz kupila Maria.* (litt. hat bought Maria)

However, one must pay attention to the fact that in Slavonic languages not every SUBJECT appearing after the verb is the FOCUS of the utterance. In those languages, a subject appearing after a verb may also be part of an entirely new utterance[26].. To be a FOCUS the subject at the end of the utterance must provide new information and must contrast with the given informative value of the rest of the utterance.

The possibility of using word order to mark FOCUS is not completely ruled out in languages like English or French, particularly when other NPs than SUBJECT or OBJECT are concerned. Namely, it is possible to focalise embedded clauses (temporal, causal) simply by putting them before the main clause (without cleaving). This seems to be possible in utterances 12 and 13, although the resulting utterances would be less clear. When it is difficult to use only word order to mark FOCUS, cleaving is a device enabling the speaker to put at the beginning of the utterance any NP that needs to be focalised in the text coherence. In most of our previously quoted English examples (# 5, 6, 10, 11, 12, 13), FOCUS was marked by cleaving[27]. Furthermore, marking FOCUS by cleaving is easy to recognise in a written text (without data on intonation).

Nevertheless, cleaving is not the only means of expressing FOCUS even in English.

*#17 I think you will do as I bid, he said, turning to leave the cabin. Remember that I have your son — if you chance to hear the agonized wail of a tortured child it may console you to reflect that **it is because of your stubbornness that** the baby suffers — and that it is **your** baby… The Beasts of Tarzan by* Edgar Rice Burroughs

In utterance #17, there are two focusses, the first is marked by cleaving *"it is because of your stubbornness that…"*, the second is marked only by intonation: "it is **your** baby…"

In languages of very different types, FOCUS may be marked only by intonation. Most linguists describing English[28] claim that almost any word in an utterance may be "focalised"[29] only by intonation.

We must add to intonation, word-order and cleaving, the possibility to underline FOCUS by particles and "paradigmatising" adverbs[30]. As concerns particles[31], in Polish a

---

[26] As concerns the different informative types of utterances with the word order VS in Russian, cf Breuillard Jean, "À propos d'un type de phrases russes à séquence VSO - *Poshel starik v les"* in *Enoncer* 2004,.p.87-110..

[27] Cf. in this volume Linde J. "On the relative depth of cleaving in typologically different languages".

[28] Cf. Lilly & Viel in this volume.

[29] Many descriptions of English call *focalisation* in a broad sense the prosodic stress on a constituent and this stress corresponds sometimes to what we call topicalisation.

[30] Cf. in French "adverbes paradigmatisants" in Nølke 1983.

[31] In Japanese, particles *wa* and *ga* can be used to mark respectively topic and focus, but it is far from being their sole function in the utterance. *Wa* can also mark subject in a simple utterance of entirely *given*

meta-informative particle marking the frontier between the CA and the comment or background of an extended utterance has developped on the basis of the deictic pronoun *to* (it) which can serve for focalisation in cleft clauses with a zero form of verbform *jest* (is): *to (jest) Maria* (literally, *it (is) Maria))*. In spoken Polish, the *to* particle may be placed after the TOPIC as well as before the FOCUS, but both TOPIC and FOCUS can be expressed without this particle, as the particle's function is to emphasize the CA of extended utterances[32]. The FOCUS on *kot (cat)* in utterance #18 may be marked without the particle *to* only by a strong stress on this word.

> *#18 (To)* **kot** *zjadl ciasto. (It's the cat who ate the cake.)*

A paradigmatising adverb[33] (*only, solely, even, precisely*, etc.) may be used to strengthen a FOCUS and may be added to a cleft sentence.

> *#19 It was only when I got thus close to it that the strangeness of this object was at all evident to me. H.G. Wells The War of the Worlds*

> *#20 But that crowd of people had a far narrower escape than mine. Only the fact that a hummock of heathery sand intercepted the lower part of the Heat-Ray saved them. Ibid.*

In utterance #20 above, the FOCUS *(Only the fact ... Heat-Ray)* is of new information, and the comment part of the utterance *(saved them)* is given. This given informative value is due to the fact that *saved them* is a sort of synonym of *escape* in the previous utterance. "Paradigmatising" means that the NP appearing after *only* is a member of a set of other possibilities that are not explicitly quoted in the utterance. In ex. #20 *the fact* is considered as the member of the set of other facts.

FOCUS is often used to express contrast as in the following example #21 where *I* is opposed to *you*.

> *#21 — You would not do it!— cried the girl. — You would not — could not be so fiendishly cruel!*

> *—* **It is not I that** *am cruel,* **but you**, *he returned, for you permit a paltry sum of money to stand between your baby and immunity from suffering.* Edgar Rice Burroughs, Tarzan

This utterance clearly shows that the part of the utterance that is *given* may be simply repeated from the previous context, as is the case with the adjective *cruel*.

4.2. Weak or strong FOCUS

Different FOCUS marking devices enable the speaker to place greater or lesser emphasis on the FOCUS of an utterance[34]. A contrast between the new informative value of one constituent and the *given* informative value of the rest of the utterance (the

---

informative type and ga the subject of an utterance of entirely *new* informative type (For details, see Wlodarczyk André 2005).

[32] Keeping aside the different morphological origins, there are a few more differences between *to* in Polish and *wa / ga* in Japanese : *to* is a universal device for topic or focus, it is used only in extended utterances ; in Japanese, on the other hand, there are two different particles and they can be used as well in simple (for Subjects) as in extended utterances (for Topic and Focus), cf. Wlodarczyk A. 2005.

[33] Also called particles in Polish, cf. Grochowski.1986.

[34] We described also weak and strong topics, cf. Wlodarczyk H. 1999

background) suffices to turn this constituent into a Focus. When this constituent is not highlighted by cleaving, by a particle or an adverb, it can be called a weak Focus. Let us consider utterance #22.

> #22 *The next few days Tarzan devoted **to the weaving of a barkcloth sail** with which to equip the canoe, for he despaired of being able to teach the apes to wield the paddles, [...]*Edgar Rice Burroughs, Tarzan

In utterance #22, the phrase "to the weaving of a barkcloth sail" is a weak Focus. The direct Object appears at the beginning of this utterance, followed by the Subject and the verb, thus constituting a comment of *given* information, after which the indirect Object appears as a weak Focus of *new* information

Very often the mere contrast between a *new* Subject and a *given* predicate leads to the extension of the utterance, making a Focus out of what would only have been a Subject. In languages with definite and indefinite articles, the new value of the Subject may be marked by the use of an indefinite article.

> #23 *The telephone rang. Mary picked up the receiver. An unknown man was speaking.*

In languages like Polish (where there is no article but case marking) the newness of the Subject (and in the same way, its weak focalisation) may be marked by placing it after the verb.

> #24 *Zadzwonił telefon. Maria podniosła słuchawkę. Odezwał się głos nieznanego mężczyzny.*

Languages which cannot easily change the order of constituents around the verb use complicated syntactic devices such as cleaving and anaphoric pronouns. Another way to focalise a part of an utterance are *pseudo-cleft* constructions[35] consisting in placing the part to focalise after a topicalised part of the same utterance at the beginning in the form of an embedded interrogative clause.

> #25 ***What** he saw **was the giant figure** of an almost naked white man emerging from the bush. Tarzan*
>
> #26 ***Wha**t is important **is the conclusion.** (from a French original by Henri Poincaré, in his book* La Science et l'hypothèse *: Ce qui importe, c'est la conclusion).*

In Slavonic languages the possibility to put any constituent (with its case mark) at the end of the utterance avoids using such pseudo-cleft constructions.

> #27 *[...] na razie znana jest tylko wersja 'nieoficjalna' (Donosy 1996) (Literally "for the time being known is only unofficial version")*

But, in order to translate such an utterance in which it is the Subject that appears at the very end, English can use a psuedo-cleft sentence :

> #28 *For the time being, what is known is only an 'unofficial' version.*

4.3. Isomorphism of CA in Simple and Extended Utterances

---

[35] Pseudo-cleft constructions are also called WH-cleft, cf. Lambrecht 1994 p. 123, see also Linde-Usiekniewicz J. in this volume.

The isomorphism between the CAs of simple utterances and that of extended utterances deserves to be pointed out. The analogy of TOPIC and SUBJECT as global CAs[36] is comparable to that of FOCUS and OBJECT as local CAs. Moreover, for the same reason, weak TOPIC is very similar to SUBJECT and weak FOCUS to OBJECT. This isomorphism can be observed not only when comparing utterances within one langage, but also in the practice of translating from one language to another.

Within one langage, it is possible to compare utterances where the same participant occurs either in the position of an OBJECT in a simple utterance or in that of a FOCUS in an extended utterance[37]. As an example, let us consider a simple (non extended) utterance in which a participant (the agent playing the role of the animate initiator), *the student,* occurs in the position of the SUBJECT as the global CA in an active construction, whereas the inanimate participant (the figure playing the role of the terminator, i.e. *the book*) occurs in the position of the OBJECT as the local CA.

#29 A student took the book.[38]

The meta-informative structure of this simple utterance can be changed if the text cohesion requires representing the student as local CA. Two possibilities are then available:

(1) The utterance remains simple but the verb changes into the passive voice. In that case, *the student* occurs as the local CA in the position of the passive object and the book becomes the global CA in the position of the SUBJECT:

#30 The book was taken by a student.

(2) The utterance is extended when the student occurs as a FOCUS, but the verb remains in the active voice and *the book* remains in the position of OBJECT as local CA :

#31 It is a student who took the book.

The choice of one or the other type of utterance depends on the coherence of the text (discourse) in which the utterance occurs. In #31, there is a contrast between the FOCUS (introduced as *new* in the discourse) and the background which is introduced as *old*. In #29 and #30 (with no special stress) there is no contrast between new and old information within the utterance. The choice of #31 presupposes that there is a contrast *in absentia* (on the paradigmatic level) between *the student* and members of a set of people to which he belongs (the set of people who might take books : students, professors, readers in general).

The isomorphism of CAs of simple and extended utterances has an important application in translation. In translation practice, it can be observed that a weak FOCUS is often replaced simply by an OBJECT. For instance, in Polish utterances a NP can be identified as a weak FOCUS when it occurs in the final position and when there is an informative contrast between this NP and the rest of the utterance. The weak FOCUS may be either an OBJECT or a SUBJECT. At the end of the utterance, it is easier to identify a

---

[36] see Wlodarczyk H 2004 and (to appear)

[37] In a previous paper (HW 2004) we compared the subject of a passive verb and the topicalised object of the same active verbe.

[38] In order to be interpreted as a simple utterance #29 must be uttered without special stress on either of its constituents.

SUBJECT as a weak FOCUS because the neutral non focalised position of the SUBJECT is at the beginning. In #32, the proper noun *Marek* is both SUBJECT and weak FOCUS.

#32 Dziś lody funduje Marek.

The litteral English translation which keeps the same word-order is unacceptable:

*#33 *Today icecream offers Marek*

It is however possible to translate the weak FOCUS of Polish original into a strong FOCUS:

#34 Today it is Marek who offers icecream.

But, in Polish there exists also an utterance with a strong FOCUS, it would be :

#35 Dziś to Marek funduje lody.

In order to translate the two different Polish focuses into two different English utterances, it is possible to translate the weak FOCUS in Polish by the OBJECT of a passive construction in English:

#36 Today icecream is (will be) offered by Marek.

Another way of rendering a weak FOCUS expressed as a postverbal SUBJECT in Polish is to use a lexical conversion[39] enabling the speaker to introduce the participant as an OBJECT in English. In ex. #37 the verb *podejmowac decyzje (to take a decision)* is replaced in the English translation by the locution *the decision belongs to*.

*#37 Wchodzi w życie rozporządzenie Rady Ministrów określające procedury w razie konieczności zestrzelenia samolotu porwanego przez terrorystów. Decyzję, czy samolot zmusić do lądowania, czy zestrzelić, podejmuje **minister obrony lub dowódca Sił Powietrznych**.. ” Metro” 13 Januarz 2005 p. 5.*

#XX A decree of the Council of Ministers specifying the procedure to be used in the event of the necessity to shoot down a plane hijacked by terrorists goes into force. The decision to force the plane to land or to shoot it down is the responsibility of the Defense Minister or of the Chief of the Air Force.

This possibility of translating a weak FOCUS by an OBJECT is based on the isomorphism between FOCUS and OBJECT: Both are local and dependant CAs.

5. Conclusion

The Meta-informative centering theory that has been sketched here is purposely based on cognition because this is the only method whichmakes it possible to analyse and explain text coherence within one language and to compare different means of CA marking in various languages. We must stress the idea that the concept of information in our contemporary information society should cover similar semantic fields as well in computer science as in linguistics. For this reason, we have defined information, first of all, as a semantic relationship between agents and/or figures, and what has been called "information" in linguistics up until now, we have proposed to rename "meta-information".

---

[39] Cf. Apresjan 1980 p. 329-362

In our theory FOCUS is considered as a dual partner of TOPIC, and this assumption enables us to build a consistent model of both "predicating" and "uttering", in particular to explain the isomorphism between SUBJECT and OBJECT on the one hand, and TOPIC and FOCUS on the other. Needless to say, this theory requires further elaboration with respect to its applicability to different linguistic families, namely ergative languages. It may also be applied to language teaching and to an approach to translation which takes into account not only semantic content of the original text but also its pragmatic force and shape.

6. References

APRESJAN Ju. D. (1980) Semantyka leksykalna. Synonimiczne środki języka (Leksićeskaja semantyka. Sinonimiceskije sredstva jazyka, Moskva 1974), Ossolineum, Wrocław.

DUSZAK A., 1987, "A composite view of sentence topics", in *Od kodu do kodu, Prace ofiarowane Prof. Olgierdowi Adrianowi Woltasiewiczowi,* Wyd. Uniwersytetu warszawskiego, Warszawa.

COTTE P., DALMAS M., WLODARCZYK H. ed. 2004, *Énoncer – l'ordre informatif dans les langues*, Coll. Sémantiques, L'Harmattan, Paris

GROCHOWSKI, M., 1986, *Polskie partykuły. Składnia, semantyka, leksykografia,* Ossolineum, Wrocław.

GUTIÉRREZ ORDÓNEZ S., 1997, *Temas, remas, focos, tópicos y commentarios,* Arco / Libros, Madrid.

   2002, *Forma y sentido en sintaxis,* Arco / Libros, Madrid.

LAMBRECHT K.., 1994, *Information Structure and Sentence Form. topic, focus, and the mental representations of the discourse referents*, Cambridge University Press, Cambridge, p. 388.

NØLKE H., 1983, "Les adverbes paradigmatisants : fonction et analyse", *Revue Romane* Numéro Spécial 23. Copenhague : Akademisk Forlag

Von WRIGHT G. H., 1963, *Norms and Actions,* London :Rouledge and P.Kegan, New-York : The Humanities Press.

WLODARCZYK A., 2003a, "Les Homotopies du topique et du focus", *Ordre et distinction dans la langue et le discours*, Actes du colloque international de Metz (18,18 et 20 mars 1999, publiés par Bernard Combettes, Catherine Schnedecker, Anne Theissen, Paris, Honoré Champion éditeur, p.513-526.

   2003b, "Linguistique dynamique: évolution du discours dans le temps", *Etudes linguistiques romano-slaves offertes à Stanisław Karolak*, Oficyna Wydawnicza Edukacja, Cracovie 2003, p. 497-510.

   2004, "Centres d'intérêt et ordres communicatifs" in *Enoncer, l'ordre informatif dans les langues,* réd. par Cotte P., Dalmas M. et Wlodarczyk H., L'Harmattan, Paris.

   2005, "From Japanese to General Linguistics : Starting with the *wa* and *ga* Particles", in *Paris Lectures on Japanese Linguistics*, edited by Wlodarczyk A., Kurosio Pub. Co., Tokyo.

   To appear in 2006 "Les rôles des situations sémantiques", *International Conference "Cognitive Studies, Semantics and Contrastive Linguistics*, Polish Academy of Sciences, Institute off Slavistics, Warsaw, April 4-5th 2005.

WLODARCZYK A. & WLODARCZYK H., 1998, "Graded Informative Content of Linguistic Messages", CD-ROM: *Proceedings of the 16th International Congress of Linguists.* Pergamon, Oxford, Paper N° 0431

To appear in 2006, "Subjects (explicit, implicit and anonymous), a cognitive and typological approach"; *International conference "Cognitive Studies, Semantics and Contrastive Studies",* Polish Academy of Sciences; Institute of Slavistics, Warsaw april 2005.

WŁODARCZYK H., 1996, « Les Traits sémantiques du sujet « anonyme »,en polonais, russe et français », *Semantyka a konfrontacja językowa,* Akademia Nauk, Instytut Slawistyki, Slawistyczny Ośrodek Wydawniczy, pp. 179-198, Warszawa.

(1999) « Les marqueurs de la validation des énoncés en français et polonais », *Etudes cognitives / Studia kognitywne III*, SOW, PAN, Warszawa. p. 135-162 (Erratum : dans cet article, il faut lire Σ pour Π et Π pour Σ).

2003, "L'Interprétation dynamique des centres d'intérêt dans les dialogues français et polonais" *Études linguistiques romano-slaves offertes à Stanisław Karolak*, Oficyna Wydawnicza Edukacja, Cracovie p.511-528.

2004, "Les Centres d'intérêt de l'énoncé en polonais et en français", *Énoncer – l'ordre informatif dans les langues*, dir. Pierre Cotte, Martine Dalmas et Hélène Wlodarczyk, Coll. Sémantiques, L'Harmattan, Paris, p.33-48.

Sources

Donosy, Dziennik Liberalny (1996) Red. Lena Bialkowska

http://www.fuw.edu.pl/donosy

Hodgson Derek, "Art, Perception and Information Processing: An Evolutionary Perspective" http://mc2.vicnet.net.au/home/cognit/web/hodgson.html

OXFORD TEXT ARCHIVE

Tarzan of the Apes by Edgar Rice Burroughs,

The War of the Worlds by Herbert George Wells

The Scarlet Pimpernel by Baroness Orczy