Subjecthood and Topicality are both Pragmatic Issues

André Wlodarczyk & Hélène Wlodarczyk

CELTA, Université Paris-Sorbonne
wlodarczyk.andre@gmail.com, helene.wlodarczyk@paris-sorbonne.fr

Abstract. The concept of centre of attention (CA) is used in our Distributed Grammar framework focusing on its relevance for the syntax of human languages. Based on psychological evidence, this notion captures what is common between subject, topic and theme in an integrated system of concepts rather than as a disparate collection of them. We define respectively subject as the main CA of a base utterance, topic as the main CA of an extended utterance (containing both Old and New meta-informative status) and theme as the main (composite) CA of a text/discourse. When choosing an entity or an element of the semantic situation as the CA of the utterance, a speaker creates a common ground on which it becomes possible to communicate with the addressee.

Keywords: subject, topic, theme, pragmatics, common ground

1. The problem of Subject in Logic and Grammar: Subject or Argument?

In ancient logic, a proposition was considered to be composed of two terms: the subject and the predicate (SP). In Aristotle's metaphysics, this two-fold definition of a proposition led to the ontological interpretation: the subject is an entity (a substance) and the predicate is a property or quality (an accidence). In formal logic (since G. Frege), this two-fold schema SP has been replaced by the concept of an n-ary predicate (a relation) represented by the logical predicate formula $P(x, y, …, n)$ which is considered more universal and is used to represent all sorts of relations, not only the binary ones. Many logicians and philosophers however (e.g. Geach 1950) consider that besides the concept of n-ary predicate, the traditional view of proposition as composed of subject and predicate remains relevant in order to give account of the structure of natural language utterances.

  As a matter of fact, the linguistic notion of predication cannot be formalised by the formula $P(x,y,z)$ (i.e. 'predicate' of the first order predicate logic) because there is no hierarchy between arguments, whereas in linguistic utterances the subject is a privileged argument and the other arguments are secondary and therefore are called "complements". Moreover, the logical predicate (predicate with arguments) does not make it possible to distinguish between different linguistic utterances expressing the same semantic situation, and therefore the predicate logic fails to explain such variations as diathesis, word order, left dislocation, cleaving, among others. For instance, the unique formula $P(x,y)$, instantiated as *invite(Peter, Jim),* cannot account for as many different natural language utterances as below:

 # Peter invited Jim.
 # Jim was invited by Peter.
 # It is Peter who invited Jim.
 # As regards Jim, Peter invited him.

Associative Semantics (AS - Wlodarczyk A. 2008) and Meta-Informative Centering (MIC - (Wlodarczyk A. & Wlodarczyk H. 2006a,b, 2008a,b, 2013) theory now belong to the Distributed Grammar project (Wlodarczyk A. & Wlodarczyk H. 2016) in which we intend to explain the relevance of merely binary predicates for the semantic (information) interpretation of human language utterances and to introduce multiple level analysis in order to treat the meta-informative elements of discourse.

2. The Iceberg Metaphor

Signification consists in converting signs into infons (inner representations) and vice versa during the processes of analysis and synthesis of utterances respectively: extracting information from utterances or building utterances with signs. Signs in utterances have to be mapped onto sets of infons. The signification of linguistic expressions allows us only to encode/decode overt ("explicit") information. For instance, it is quite difficult to align texts written in two languages. Using an *iceberg* metaphor makes it possible to point to the two following important characteristics of human languages:

 (1) The content of linguistic expressions convey both *overt* and *covert* information. The implicit part of content is grounded in both contextual and cognitive representations. But interpretation consists of recovering the *covert* information from the linguistic expressions with the sole help of the *explicit* information they contain. There are various ways of recovering information. Let us mention just a few of them: all kind of presuppositions, paradigmatic functions and semantic hyponymy. For example, when the subject of a transitive utterance is not overt, its object can be recovered if it is defined as a dependent centre of attention: Object $\rightarrow$ Subject (the object entails the subject).

 (2) The content of linguistic expressions can be either concise (compressed) or precise (extended). A concise linguistic message contains fewer units (hence less information) and a precise one contains more linguistic units (hence more information)**.** Concision and precision are results of the fact that modal[1] equivalences can be ordered: $(x \equiv_{D_1} y) \leq (x \equiv_{D_2} y) \leq (x \equiv_{D_3} y)$ iff $D_1 \subseteq D_2 \subseteq D_3$.

3. Representation of Linguistic Information using Binary Predicates

Information is produced when *relations* are established between entities. In the framework of associative semantics (Wlodarczyk A. 2008), the universal ontological components of linguistically expressed situations are:

 (1) static or dynamic frames (states, events and processes),

 (2) their roles (enacted by animate agents and/or inanimate figures)

 (3) and anchors (indicators of spatial and temporal relations).

Each language provides speakers with linguistic means for expressing situation frames: verbal lexemes with aspectual properties and different types of valence opening sequentially ordered places in utterances. Entities playing roles in situations are classified with regard to such criteria as ±abstract, ±animate, ±human, etc. Human languages also use various sorts of adverbs and autonomous noun phrases to express spatial and temporal anchors.

 We define *information* as the content of utterances including the semantic roles as expressed

---

[1] Because they operate on a given domain D only.

by the noun phrases and the semantic situation frames with their participant roles as expressed by the verb phrase. Information is never exhaustively expressed in discourse (the iceberg metaphor gives a suggestive account of this). Information needs be reconstructed and completed by the hearer. In linguistic messages, information is always partial since speakers express only what is said to be **obligatory** (grammaticalised in their language) and what is pragmatically (due to the speakers' point of view, i.e. due to what they pay attention to) **relevant**.

Sets of binary predicates can be used to internally represent information (relations and their participants); they can be used as formal representations of semantic situations in which participants take part. On the level of information, the validation of utterances as true or false is a function mapping semantic frames of situations, their roles and anchors onto mental representations of states of affairs (or eventualities) and entities of the world. We will be keeping in mind that the binary formulas $P(x,y)$, in most general cases, do not establish hierarchical order between the arguments of P and do not constitute any foundation for the sequential (linear) order in which it is the verb valency that represents the (associated) situation(s) and the noun phrases show that their arguments are aligned in an utterance.

4. Meta-Information and Attention-Centred Phrases

Meta-information is information about other information. What is not taken into account in most syntactic theories of linguistic expression is the fact that the elementary syntactic structure of utterances corresponds to meta-information, not to information. What, in Generative Grammar (GB), is defined as a sentence by the rewriting rule: S --> NP + VP and what is represented by tree structures of immediate constituents, is in fact meta-information (not information).

The MIC approach differs from purely syntactic theories in that it defines the subject of an utterance not simply as a syntactic constituent but as an attention-centred phrase. We define the subject as that noun phrase which corresponds to the *global* centre of attention of an utterance. The syntactic properties of such a noun phrase may differ from language to language by different criteria: case-marker, word-order, agreement between subject and verb, etc. Thus, in the MIC Theory, centring is a structuring operation concerning not only texts but also as it were utterances. On the other hand, in the computational "centering theory" by Grosz et al. [1986, 1995], centres of attention are defined at the text level: one constituent of an utterance is treated as a *forward or backward looking centre* in order to maintain the cohesive flow of information from one utterance to its successor. Forward and backward looking centres make it possible to give an account of the relations which bind utterances together into a coherent text.

In the MIC theory, we consider that no judgment may be uttered without selecting at least one centre of attention (CA) among the participants of the situation spoken about; thus we consider centring as a structuring operation not only at the text level but also at the level of the utterance because of the necessarily linear (sequential) structuring of speech sounds in human languages. An utterance will be defined as a linguistic message having at least one Centre of Attention (CA). The general concept of centre of attention makes it possible to capture what is common between Subject, Object, Topic and Focus. In our theory, CAs are seen not as psychological phenomena but rather as those segments of linguistic utterances on which attention has been centred. In order to communicate semantic information in a non-linear manner, the speaker has to select one of the participants of the semantic situation and treat it as the global (primary) centre of attention, (i.e. the subject of the utterance) about which they predicate. Information corresponding to the *local* (secondary) centre of attention may be expressed as the object.

As stated above, a segment which expresses a chunk of a semantic situation is "centred" (treated by the speaker as representing a CA) if it has been distinguished among other elements of one situation or many situations by linguistic meta-informative markers (syntactic, morphological, prosodic or any pragmatic marker). This view is very close to the concept expressed by Givón: "the subject and DO (direct object) may be viewed as the grammaticalised

primary and secondary topic of the discourse at the time when the clause in which they take part is being processed." (Givón 1994, 198). From the above quotation it is obvious that the author calls "topic of the discourse" what we call more generally the "centre of attention". However, in our theory, we use the word "topic" as a reserved term to refer to a constituent of an utterance which is prominent and which bears the old meta-informative status

Thus, the core of syntax does not map directly on information (semantics) but on meta-information (pragmatics) of the utterance. Syntax is a means of expression or a resource (along with morphology, phonology and prosody) used to linearise the information content communicated in an utterance.

5. Subject and Semantic Role

For the hearer, to understand (be able to reach a semantic interpretation of) the content to which the subject of an utterance points, it is necessary to interpret the semantic role played by the participant chosen for the subject. Since we consider that the subject belongs to the pragmatic module, it is independent of the semantic role enacted by the participant it refers to. However there exists a default relationship between the subject and one of the two main semantic roles (active or passive) depending on whether the syntactic structure of the language is either nominative or ergative.

In nominative (active) languages, in utterances with a verb in the direct, unmarked active voice, a default relation is established between the subject and the active role… and the object and the passive role.

Consequently, if we admit that the semantic interpretation of an utterance is realized in different successive steps, in the first step, by default, the subject is assigned the active role. When the entity chosen as the subject by the speaker is not an animate agent, in the first step of interpretation, the speaker and the hearer treat the subject as referring to a pseudo-agent, as it is the case in utterance #1 hereafter.

 # 1 A car hit a pedestrian.

Morpho-phonetic level	*A car*	*hit*	*a pedestrian.*
Meta-information level	global CA: subject	verb	local CA: object
Information level	pseudo-active role	action: process	passive role

In utterance #1, in the first step of semantic interpretation, the inanimate subject " a car" is assigned a pseudo-active role. In the second step, a more accurate semantic interpretation may be reached and expressed in utterance #2, in which the proper animate entity is assigned the active role and the subject of utterance #1 transformed into the instrumental PP "with his car".

#2 The drunken driver hit a pedestrian with his car.

Morpho-phonetic level	*The drunken driver*	*hit*	*a pedestrian*	*with his car*
Meta-information level	global CA: subject	verb	local CA: object	instrument
Information level	active role	action: process	passive role	median role

Thus the definition of the subject as the global centre of the utterance makes it possible to treat in a universal manner active and ergative languages as well as subject of active or passive verbs in active languages. This pragmatic definition frees the subject from any obligatory link with the active role (traditionally called the *agent*). It makes it possible to explain that the choice of a subject by the speaker is not simply a grammatical obligation but rather a pragmatic choice used in the discourse strategy.

6. The Old/New Meta-Informative Status of Discourse

Many linguists (Chafe - 1976 and Prince - 1981 among others) point to the importance of the distinction between old and new information. "Connected speech unfolds as an unbroken sequence of 'messages', in which the speaker is alternating between elements of given and elements of new; these map into the structures of the other grammatical units, most powerfully into those of the clause." (Halliday M.A.K. and Greaves W.S., 2008, p. 42). In the MIC theory, the well-known distinction between old and new information, is considered as "meta-informative old or new status" alternation, and - obviously - we agree so far with the mapping of old and new onto the units of the clause that we consider to reflect directly the structure of the utterance.

The old or new status of information conveyed by an utterance (or by one of its segments) depends on the discourse strategy chosen by the speaker. The speaker is free to introduce some chunk of information either with a new or old meta-informative status and to use it as a possibly major argumentation device.

We distinguish the three following kinds of motivation of old and new meta-informative status:

(a) The **communicative** motivation is explicit and speech bound. The situation spoken about is either connected to another one mentioned earlier (anaphoric) or to be mentioned (cataphoric) or it is a modal situation (either reported or to be reported).

(b) The **cognitive** motivation is related to the process of knowledge acquisition. Situations appear as already known (registered) or unknown (unregistered). This presupposes the existence of a kind of recent discourse memory (to be confirmed by neurological experiments).

(c) The **ontological** (referential) motivation depends on the knowledge stored in long term memory; the situation spoken about needs to be treated either as a type (generic, general, habitual or potential) or an instance (specific, particular, occasional or actual).

INFORMATION STORAGE	Type of MOTIVATION	Motivation of *Old* status		Motivation of *New* status	
Intermediate memory	**Immediate Communication**	anaphoric		cataphoric	
	Recently acquired information	known		unknown	
Permanent memory	**Ontological knowledge (mental reference)**	Σ	generic	Π	specific
			general		particular
			potential		actual
			habitual		occasional

Table 1. Motivations of the meta-informative old or new status

In the MIC theory approach, the old/new status alternation is relevant to the syntactic structure of clauses belonging to a coherent text, whereas the truth validity of an utterance has no direct influence on its syntactic structure and linear ordering. We propose therefore to pay more attention, in linguistic studies, to the old/new alternation within the syntactic structure of utterances. When the speaker changes the meta-informative status of the utterance or of one of its clauses, it has no effect on its truth-conditional validity. As noted by Kuroda S. Y. 1976, both utterances #3 and #4 which follow below have the same truth value, because the latter depends on the particular, widely known battle which the speaker can refer to in his/her discourse.

5

#3 The Greeks defeated the Persians.

as opposed to its passive form

#4 The Persians were defeated by the Greeks.

In the same way, the difference between subject and topic (or focus) is not relevant to truth-conditional validation of utterances. The following utterances #5 (with a topic) and #6 (with a focus) have the same truth validity as the first #3 and #4 utterances in the active and passive voices : the situation of the world which is the referent of this utterance remains the same and the truth validity depends only on the adequacy of the utterance to the state of affairs spoken about.

#5 As regards the Greeks, they defeated the Persians.

#6 It was the Greeks who defeated the Persians.

However, the choice between one of the four mentioned utterances (we call them "meta-informative paraphrases") has important consequences on discourse strategy and pragmatic felicity. In a discourse, in which the Greeks are the main theme, the speaker would rather choose utterance #3 than #4. Utterance #5 would be felicitous only in a discourse in which the speaker would not have been dealing with Greeks in the previous part of the text. Utterance #6 would be used either to contradict a previous utterance asserting that the Persians defeated the Greeks or in answer to the question *Who defeated the Persians?* Let us now explain the difference between utterances #3 and #4 on the one hand and #5 and #6 on the other. To achieve this, we need to distinguish between base and extended utterances.

7. Base and Extended utterances, the Definition of Topic

Base and extended utterances are defined as **pragmatic** units of discourse in contrast to simple and complex sentences understood as **syntactic** units. As a pragmatic unit, each utterance contains at least one centre of attention (CA). The CA phrases may have either ~~of~~ the same or a different meta-informative status (Old or New) than the rest of the utterance. In a base utterance there is no contrast between the status of the global CA and that of the rest of the utterance: it is either "all New" or "all Old". On the other hand, the CAs of extended utterances contrast with the rest of the utterance. The Topic bearing an Old meta-informative status is in contrast with the "New Comment", the Focus of New meta-informative status is in contrast with the "Old Background".

TYPE OF EXPRESSION	CENTRES OF ATTENTION	
	Global	Local
1.1. Base Utterance	Subject	Object
1.2. Extended Utterance	Topic	Focus
2. Text / Dialog	General Theme	Particular Theme

Table 2. Pivots of discourse (from Wlodarczyk A. & Wlodarczyk H. 2008a)

Table 2 shows that, in the MIC theoretical framework, the theme is not merely a synonym of topic. As a matter of fact, we define topic and theme by reference to the representation layer of linguistic information to which they respectively belong. We use the word "theme" as a term referring to texts or discourses as organised, linguistically coherent sets of utterances. Thus, making reference to the meta-informative status of base and extended utterances respectively, it is possible to capture and better explain the difference and, at the same time, the similarity which characterise the subject and the topic.

6

Informative Layer (0-order)	typical semantic unit : situations schemata
Meta-informative Layer (1st order)	linguistic unit uttered in a context : 1st m-inf. level: base utterances 2nd m-inf. level: extended utterances
Cognitive Layer (2nd order)	organised set of utterances : texts, dialogs

Table 3. Three layers of linguistic information (Wlodarczyk A. & Wlodarczyk H. 2006a)

The *second meta-informative level* is an extension of the predication: it consists of adding attention-centred phrases with contrasting status to a base utterance. Extended utterances consist of two contrasting parts each having an opposed meta-informative status; their centre of attention corresponds to an emphasized noun phrase contrasting with the rest of the utterance: an "Old" status phrase (topic) contrasting with a "New" status phrase (comment) or a "New" status phrase (focus) with an "Old" status phrase (background).

Thus, in order to define the topic, it is necessary to establish a contrast between the meta-informative status of information contained in discourse. In the topic position, the speakers place the constituent they wish the speaker (1) to pay attention to and (2) to consider as having the "Old" status of information.

Thus, in the MIC theory, topic is defined as a prominent or attention-centred phrase with an "Old" meta-informative status. It is only the comment part of the utterance which introduces information with a "New" meta-informative status. The comment itself may, in some cases, be further divided into two parts again: focus and background.

What is introduced by the speaker as a topic is supposed to be "taken for granted", presupposed to be known to everybody. Only the comment can introduce new information. Thus, the topic is (or is part of) the common ground making it possible for the speech participants to communicate new information.

8. Conclusion

On top of logical inference (reason), such psychological factors as attention, intention and emotion interplay as much in the processes of meaning creation as in that of communication. The Distributed Grammar is therefore a complex view of language which emerged as the result of a multi-level investigation into the sequential (linear) ordering of the constituents of linguistic utterances focusing on the fact that the sequential nature of language reflects the semantico-pragmatic overt (cf. explicature) and covert (cf. implicature) components of communicated information. It is an integrated framework for Associative Semantics (AS) and Meta-Informative Centering (MIC) theory.

In the MIC theory, the old/new status of a chunk of information depends on the strategy chosen by the speaker to enrich or even manipulate the hearer's knowledge. In an utterance, the subject corresponds to that noun phrase which represents the global centre of attention and the object — the local one. Depending on the attentional strategy adopted, the speakers need to choose among the utterance schemata (based on the verb valence) the one which allows them to assign the global and local centres of attention to the subject noun phrase and to the object noun phrase respectively assuming that there is a default mapping between the subject and the active participant (agent) of the semantic situation. Needless to say that in case the speakers wanted to talk about the passive participant (patient) paying more attention to it than to the active participant, they can use the selected utterance schema in passive voice. Thus, it is clear that any participant can be treated alternatively as the subject or the object of an utterance. Traditional grammarians were aware of this interchangeability of subject and object in the utterance. In our

framework, we treat the passive voice as one of the **meta-informative** devices which provide the speakers to express the distinction of salience (global/local) without changing the information content of the utterance.

The use of a topic (expressing the global centre of attention of the speaker) is very similar to that of a subject. It differs however in that the topic is used when the speaker wants to establish contrast between the meta-informative status of two chunks of information (contained in an utterance).

Thus, subject and topic are part of the common ground making it possible for the speaker and hearer to communicate: they are proposed to the hearer by the speaker as the global centre of attention about which something will be predicated in a base utterance or to which a comment will be added in an extended utterance.

References

Chafe, W. L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: Li. C. (Ed.), *Subject and topic*. New York: Academic Press.

Geach, T. 1950 Subject and Predicate, *Mind* (1950) LIX (236): 461-482.

Givón, Talmy. 1994. The Pragmatics of Voice: Functional and Typological Aspects. in T. Givón, (ed.), *Voice and Inversion, Typological Studies in Language* 28. John Benjamins: Amsterdam.

Grosz B. J., Sidner C. L. 1986. Attention, Intentions and the Structure of Discourse, *Computational Linguistics*, Vol. 12, Nr 3, 1986, p. 203-226

Grosz B. J., Joshi A.K. and Weinstein S. 1995. Centering: A framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, Vol. 21(2), 175-204.

Halliday, M.A.K. and Greaves, W.S. 2008. *Intonation in the Grammar of English*, London: Equinox.

Hayashi A., Tomlin R. S. and Yokota T. Attention detection and Japanese (Japanese Fish Film Paper) http://logos.uoregon.edu/tomlin/in_progress.html

Kuroda, Shigeyuki. 1965. Subject. In M. Shibatani, ed. *Japanese Generative Grammar - Syntax and Semantics Vol. 5*, pp. 1-16, New York: Academic Press.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In *Radical Pragmatics*, pp. 223-255. New York: Academic Press.

Włodarczyk André. 1980. 「主題から主語へ、そして主語から主題へ－ハとガ」 *Shudai kara shugo e, soshite shugo kara shudai e –'wa' to 'ga'* [From Topic to Subject and from Subject to Topic: 'wa' and 'ga']. *Gengo* 9(8), 78-85. Tokyo : Taishuukan.

Włodarczyk, André. 1998. The Proper Treatment of the Japanese "wa" and "ga" Particles, *Proceedings of the International Workshop on Human Interface Technology.* Aizu-Wakamatsu.

Włodarczyk, André. 2005. From Japanese to General Linguistics - starting with the "wa" and "ga" particles. in *Paris Lectures on Japanese Linguistics*, 27-62. Tokyo: Kurosio Shuppan.

Włodarczyk, André. 2008, Roles and Anchors of Semantic Situations. *Études cognitives/Studia kognitywne VIII,* 53-70, Warszawa: SOW.

Włodarczyk André and H. Włodarczyk. 1998. Graded Informative Content of Linguistic Messages. *CD-ROM: Proceedings of the 16th International Congress of Linguists*. Paper N° 0431. Oxford: Pergamon.

Włodarczyk André and H. Włodarczyk. 2006a. Focus in the Meta-informative Centering Theory. In A. Wlodarczyk et H. Wlodarczyk, *La Focalisation dans les langues,* Coll. Sémantiques, 27-45. Paris: L'Harmattan.

Włodarczyk André and H. Włodarczyk. 2006b. Subject in the Meta-informative Centering

Theory. In *Études cognitives / Studia kognitywne 7*, 39-64. Warszawa: SOW.

Włodarczyk André and H. Włodarczyk. 2008a. Roles, Anchors and Other Things we Talk About: Associative Semantics and Meta-Informative Centering Theory. In I. Kecskes, ed., *Intercultural Pragmatics*, 5 (3). Berlin/New York: Mouton de Gruyter: 345-366.

Włodarczyk André and H. Włodarczyk. 2008b. The Pragmatic Validation of Utterances. In *Études Cognitives / Studia Kognitywne* 8, SOW, Warszawa: 117-128.

Włodarczyk André and H. Włodarczyk (eds.), 2013, *Meta-Informative Centering in Utterances – between semantics and pragmatics*, SLCS 143, John Benjamins: Amsterdam: 325 p.

Włodarczyk André and H. Włodarczyk. 2016. Trójwarstwowa struktura informacji w treści wypowiedzi (szkic o programie Gramatyki Rozproszonej) [Three-tier Structure of Information in the Content of Utterance (essay on the Distributed Grammar program)], paper in Polish with English abstract. *Investigationes Linguisticae*,Vol. XXXIV, Poznań: 73-112.