

## **Agents, roles and other things we talk about : Associative Semantics and Meta-Informative Centering Theory**

André WLODARCZYK & Hélène WLODARCZYK

### **1. Introduction**

The term ‘information’ in the age of information societies should cover similar semantic fields in linguistics and in information science. The nature of what is usually called information in linguistics and how it relates to meaning and cognition should therefore be elaborated in agreement with other cognitive sciences. For this reason, we define information as knowledge about semantic situations, their participants (agents and/or figures) and spatio-temporal anchors<sup>1</sup>. Both semantics and pragmatics participate in the elucidation of the content (meaning) of utterance level signs (expressions). Relations and predication are driven by sequentially ordered functions, on the one hand, and Entities (internally, through comparison of their properties and, externally, through global and/or local centering) undergo parallel processes, on the other hand.

The importance of pragmatics for the study of human languages is still often underestimated notwithstanding the rise of research on pragmatics both in logics and linguistics over the last fifty years. Twentieth century linguistic research led to the conclusion that many linguistic categories (such as modality, aspect, tense, determination, etc.) concern not only the informative contents (semantics) but are also discourse-dependent. Taking these findings into account we developed the Meta-Informative Centering (MIC) theory in which pragmatics occupies an even more basic place in natural language.

Meta-information concerns the way information is ordered: to achieve the ordering of non-linear representations as texts (sequences of linguistic utterances), the speaker must select a center of attention (CA) and “predicate” about it. Hence, in the MIC theory, the term “predicate” is reserved for the meta-informative level and is used in its initial Aristotelian sense. Whereas in most contemporary approaches subject is regarded as a purely syntactic unit, in the MIC theory, it is defined as the global center of attention about which something is predicated in an utterance.

On the other hand, although it was generally admitted in linguistics that it

is impossible to generalize semantic relations due to their diversity the Associative Semantics (AS) Theory aims at modeling semantic situations, i.e. the information content of linguistic utterances. In this theory, roles and anchors are seen as isomorphic abstract tuples of concepts. Roles are defined ontologically as active and passive unary relations of associated semantic situations. As will be shown below, the associative combination of their realizations gives rise to median (instrumental, means etc.) roles through derivation. Importantly, one and the same participant (role filler) may play more than one role in the given semantic situation defined as a bundle of associated situations.

The Associative Semantics (AS) theory and the Meta-Informative Centering (MIC) theory modify and unify in a consistent way the "Predicate Argument Structure" theory with the "Information Structure" theory of classical structural linguistics.

## 2. Theoretical assumptions

The important assumption we adopt here is that languages are highly modular and massively multi-layered communication systems. In what follows, we will consider some of these modules. Roughly speaking, inside one of the basic modules, it is possible to distinguish at least the two following levels: a (morpho-phonemic) *word level* and a (message construction) *utterance level*. The units of both levels are defined as signs each having two components Form and Content. However, there are no externally motivated links between these components, consequently it is possible to model signs using only one of their components at a time. Namely, it is for this reason that phonetics and phonology could develop without thorough consideration of semantics. We elaborated therefore a *model* of the Content Core of the message level units separately (though not entirely) from their Form. Indeed, research done mostly on the signs from the level of words allowed De Saussure, F. [DES 13, 95] to make an essential distinction between the *syntagmatic (in praesentia)* and *paradigmatic (in absentia)* "dimensions" along which linguistic signs can be ordered in the language system and during speech acts. More generally, there are suggestive examples<sup>1</sup> of the fact that although expressions of human languages are "linearly" arranged, signs are not only combinable but also comparable.

Clearly, language can be defined as (1) a *system* of signs and (2) a set of *processes*<sup>2</sup> (functions). We claim that, in a similar way as in computational complexity theory, the two basic kinds of functions, *sequential* and *parallel*, underlie the saussurian *syntagmatic* and *paradigmatic* relations respectively.

---

<sup>1</sup> Note that, in psychological research, two distinct processes (**scenario creation** and **comparison of properties**) are presented as lying at the basis of human understanding of novel noun phrases [WIS 97].

<sup>2</sup> Cf. Hjelmslev Louis [HJE 53].

It is highly probable that it was this consideration that motivated the research project carried on by Jakobson, R. [JAK 59] (one of the founders of the “functionalist” structural tradition in language studies) who reinterpreted this basic distinction when defining **combination** (which deals with contrasts between phonemes or between morphemes) and **selection** (which deals with oppositions between phonemes or oppositions between morphemes) respectively.

Although, in the domain of Syntax (which can be seen as Form), it is generally assumed that - among other things - paratactic relations concern the juxtaposition of clauses within discourse structures, actually, generalizing the paratactic relation and applying the saussurian distinction of *syntagmatic* and *paradigmatic* relations to the units of the message level will enable us to distinguish between *syntactic* and *paratactic* relations (table #1) which again are due to the above mentioned *sequential* and *parallel* functions respectively. As a matter of fact, syntactic rules cannot be organized in a derivation tree but only *partially* (in a modular way) because of the linear insertions (in utterances) resulting from concurrent processes. Therefore, paratactic relations are supposed to provide the glue for establishing internal consistency of speech.

FUNCTION TYPE	SIGN FORM	
	Morpheme Level Sign	Utterance Level Sign
Sequential	Syntagmatic Contrast	Syntactic Combination
Parallel	Paradigmatic Opposition	Paratactic Selection

Table # 1. Linguistic signs (system) and function types (process)

In the domain of the Content of Utterance Sign Level (table #2), another important distinction enables us, first of all, to separate *semantics* (the level of the communicated information) from *pragmatics* (the level of those contents which refer to the speech act). The importance of pragmatics for the study of human languages is still often underestimated notwithstanding the rise of research on pragmatics both in logics and linguistics over the last fifty years. Twentieth century linguistic research led to the conclusion that many linguistic categories (such as modality, aspect, tense, determination, etc.) concern not only the informative contents (semantics) but are also discourse-dependent. In the Meta-Informative Centering (MIC) theory, pragmatics occupies an even more basic place in language because, due to the essentially linear nature of morpho-phonological Form of natural language, “non-linear” semantic representations, when uttered, must obligatorily be linearized. Meta-informative indexes are thus needed in order to indicate which segments of the linguistic (linearized) messages represent the centres of attention of the speaker.

FUNCTION TYPE	UTTERANCE LEVEL SIGN CONTENT	
	<i>Semantics</i> (Information)	<i>Pragmatics</i> (Meta-Information)
<b>Sequential</b>	<b>Relations</b> – Frames, Roles and Anchors (Situation construction)	<b>Predication</b> (Basic & Extended Utterances)
<b>Parallel</b>	<b>Entities</b> (Property comparison)	<b>Centering</b> (Global & Local Centres of Attention)

Table # 2. Function types in Semantics and Pragmatics

In this approach, information is treated as the *core of semantics* because it concerns relations between entities. Indeed, we consider that the term “information” in our contemporary *information society* should cover similar semantic fields both in information science and in linguistics. The nature of what is usually called information in linguistics and how it relates to meaning and cognition should therefore be elaborated in agreement with other cognitive sciences. For this reason, we define information as knowledge about semantic situations, their participants (agents and/or figures) and spatio-temporal anchors<sup>3</sup>. Both semantics and pragmatics participate in the elucidation of the Content (meaning) of utterance level signs (expressions). Relations and Predication are driven by sequentially ordered functions, on the one hand, and Entities (internally, through comparison of their properties and, externally, through global and/or local centering) undergo parallel processes, on the other hand.

### 3. Associative Semantic

We proposed an associative model of semantic representations as *static* or *dynamic* situation frames [WLA 03] with their participants (animate agents or inanimate figures) playing different roles in situations and the spatio-temporal anchors of these situations [WLA 08].

The semantic situation is defined as a schema and its individuation, which contain three parts: relation(s), role(s) and anchor(s). The individuation of a relation is named frame (the scope of the frame is more restrictive than in Minsky, M. or Fillmore, Ch.), that of role - participant and that of anchor - support (anchor is more general here than in Situation Semantics by Barwise, J. and Perry, J. 1983). NB: the frames of semantic situations are closed with respect to inclusion and can be ordered: state < event < ordinary process < granular process (Włodarczyk, André 2003).

SEMANTIC SITUATION	ONTOLOGY
--------------------	----------

<sup>3</sup> See the paragraph “Anchor Component” below.

RELATION	FRAME	fragment of the world
ROLE	PARTICIPANT	entity
ANCHOR	SUPPPORT	location

We borrowed the generalized notion of *agent* from Artificial Intelligence and will call Agent any animate participant of a situation, whereas we will call Figure any inanimate participant. Both agents and figures may be explicit or implicit. Thus, all animate protagonists are agents but linguists generally call them “agents”, “patients” (or “contra-agents” or “themes”), “experiencers”, “beneficiants”, etc. Agents interact in particular situations and it is precisely this interaction (interdependency) that determines their abstract quality of agents in the generic sense.

Animate entities (Agents) are either *human* (+Hum) or *non human* (-Hum) and Inanimate entities (Figures) are material (+Mat) or immaterial (-Mat). It happens, however, that situation participants are transformed *from Figures to Agents* (agentivation, traditionally known as “personification”) or *from Agents to Figures* (figuration).

Characteristic Features	Control (autonomy)		Emotive (character)		Epistemic (reason)		Communication (language)	
	Feedback	Purpose	Desire	Intention	Belief	Cognition	Verbal	Visual
AGENTS								
<i>Human</i>	+	+	+	+	+	+	+	+
<i>Non human</i>	+	+	-/+	-	-	-	-	-

Table #3: The most characteristic features of semantic agents

Table #3 enumerates some characteristic features of agents on the highest level of abstraction. Agents of a situation are defined by several semantic features: (1) control features (autonomy): *goal* and *feedback*; (2) emotive features (character): *desire* and *intention*, (3) epistemic features (reason): *belief* and *cognition*, (4) communication features (language faculty): *verbal* and *visual*.

### 3.1. Frames, Roles and Anchors

In order to explain the linear ordering of basic utterances we propose to model the semantic situation as having three constitutive components<sup>4</sup>: frames, roles and (spatio-temporal) anchors. This model describes information contained in linguistic messages usually referred to as

<sup>4</sup> Note that in the proposal recently put forward by Pusteyovsky J. [PUS 95] the term "structure" is used instead of "component".

utterances. It suffices that static semantic situations be represented as spatio-temporally located frames (“spaces”, fragments of universe) but the representation of dynamic situations must include also the way situations internally develop in space and time. Moreover, roles determine the “places” (in the frames) that can be occupied by the entities (called participants when fulfilling some situation roles). We formalize this as a projection from mental roles of semantic situations into a set of entities which thereby become their (also mentally represented) participants. However, it is necessary to distinguish between *explicit participants* — that have to be recognized by the hearer at the stage of *signification*, i.e. *signified* participants — and *implicit participants* — that, in order to be understood, have to be added by the hearer at the stage of *interiorisation* or even deeper at the stage of *categorisation*<sup>5</sup>. This hypothesis is essential for the present theory of semantic situations which, as it takes into account the partiality of meaning of linguistic expressions, entails the need to use, for their formal representation, stratified structures (namely hypergraph structures) instead of simple tree structures. Thus, the *semantic partiality hypothesis* may be seen as an alternate theory to that of surface/deep structures because it makes it possible to elucidate incomplete contents, on the one hand, and to take into account their basically heterogeneous character (conveyed by linear language expressions), on the other hand.

### 3.2. Role Component

The basic structure of a dictionary being roughly speaking *classification of usages* (i.e.: typified uses in their contexts), the usages being named *semions*, semantic situations are valid with respect to mental situations. The semantic part of semions can be represented symbolically using sets of infons<sup>6</sup>. Infons are triples  $\langle i, a, t \rangle$  where  $i$  = infon’s name (attribute or relation),  $a$  = list of arguments and  $t$  = truth value<sup>7</sup> (*true* and *false*). In this approach, we add a third truth value, namely the instantly *indeterminate* value (or *anonymous* value) in order to distinguish between the speaker’s and the hearer’s communication acts. When modelling the speaker’s activity, the truth value will be most frequently “true” unless he states it otherwise. In the understanding process of the hearer, that value will be “indeterminate” unless expressed overtly as “true” or “false” by the speaker. **Elementary infons** are unary infons. Sets of elementary infons with the same relation names are called **compound infons**. Thus, the names of compound infons are *generalisations* of the names of component infons, i.e.: each relation name of a compound infon being considered as the general name of a given semantic situation. In linguistic expressions, they

---

<sup>5</sup> The encoding and decoding stages are discussed in [WLA 08].

<sup>6</sup> Cf. Devlin, K. [DEV 91].

<sup>7</sup> Note that the presence of truth value makes the only difference between infons and predicates in classical logic.

correspond roughly to verbs.

Here is an example of a compound infon which represents the utterance: “*Brutus killed Caesar.*”

$$S_1 \models \langle \textit{kill}, (\textit{killer} : \textit{Brutus}, \textit{killee} : \textit{Caesar}), \textit{true} \rangle$$

The above compound infon can be developed as two elementary infons provided that both of them hold in the same mental situation  $S_1$ .

$$S_1 \models \langle \textit{kill}, (\textit{killer} : \textit{Brutus}), \textit{true} \rangle$$
$$S_1 \models \langle \textit{kill}, (\textit{killee} : \textit{Caesar}), \textit{true} \rangle$$

Now, if we want to represent more information which is associated with the main information contained in the above utterance, we may wish to add some other infons which hold in  $S_1$ .

$$S_1 \models \langle \textit{kill}, \textit{killer} : \textit{Brutus}, \textit{killee} : \textit{Caesar}, \textit{true} \rangle$$
$$S_1 \models \langle \textit{at}, \textit{loc.time} : \textit{Antiquity}, \textit{loc.place} : \textit{Rome}, \textit{true} \rangle$$

The mental entity ‘Caesar’ could be represented by infons such as :

$$\textit{Caesar} \models \langle \textit{name}, (\textit{family} : \textit{Caesar}, \textit{given} : \textit{Julius}), \textit{true} \rangle$$
$$\textit{Caesar} \models \langle \textit{position}, (\textit{rank} : \textit{consul}, \textit{city} : \textit{Rome}), \textit{true} \rangle$$
$$\textit{Caesar} \models \langle \textit{kill}, (\textit{killer} : \textit{Brutus}, \textit{killee} : \textit{Caesar}), \textit{true} \rangle$$
$$\textit{Caesar} \models \langle \textit{at}, (\textit{loc.time} : \textit{Antiquity}, \textit{loc.place} : \textit{Rome}), \textit{true} \rangle$$
$$\textit{Caesar} \models \dots$$

We call Roles the binary relations which correspond to elementary infons. Roles are component relations within the *frames* of more general situations. Roles either equal or are included in situations (Roles  $\sqsubseteq$  Situations), i.e.: roles are either relations or proper subsets of relations. Every semantic situation is therefore composed of *infons* which are partially determined by attributes from both (1) *upper ontologies* and (2) *domain ontologies*. Situation participants of upper ontologies are seen logically as *semantic arguments* and were represented in the theory of situation semantics [BAR 83] as functions from *roles* to *entities* named "anchors". Note that the term *anchor* will be used here in quite a different way.

In this semantic theory roles are defined as pairs of *participation* functions as follows:

- a **type participation** function from roles types to the high level ontological entities (TP: Role type --> Entity type)

- an **instance participation** function from the role instances to the individual entities (IP: Role instance --> Entity instance)

The instant roles inherit the properties from their types (abstractions: hypernyms, super-classes). But, due to the partiality principle mentioned above, when generating or compiling a particular (concrete) role, language users either may be unconscious of many inheritable abstractions or abstractions of entities may be taken into consideration whenever their instances are indeterminate.

The **Role Component** of semantic situations may contain from one to

three elements from the set of three kinds of roles whose generic names are: *active*, *passive* and *median*. This theoretical shift<sup>8</sup> has important consequences regarding the very nature of roles which are defined here qualitatively rather than substantially or relationally (as relations between terms). We must not forget that logical arguments are first of all terms. However, such terms are not order-free. If we want to make the sequential order of terms free, it is necessary to split each of them into two parts: the role and its filler (often referred to as participant). We will repeat this shift as regards our second component of semantic situations, which is Anchor Component.

There are two kinds of participant types: Agents (animate entities) and Figures (inanimate entities).

<i>Participants</i> \ <i>Roles</i>	ACTIVE	MEDIAN	PASSIVE
Agent (animate)	Initiator	Mediator	Terminator
Figure (inanimate)	Source	Means	Goal

Table #4 : Roles of Semantic Actions

Roles enumerated in the table #4 represent the most usual ones. Here are some agents or figures: (1) in *active roles* (Initiator, Causer, Enabler, Benefactor, Executor, Stimulant, Source, Instigator etc.), (2) in *passive roles* (Terminator, Causee, Enabled, Beneficient, Executed, Experiencer, Goal, etc.) and (3) in *median roles*, (Mediator, Instrument, Benefice, Motor, Means, Matter etc.). While active and passive roles are considered here as *primitive* roles, *median* roles are definable by introducing two embedded *associated situations* (a-situations) in which (a) the participants of median roles fulfil passive roles in the first of the associated situations and (b) they fulfil active roles in the second associated situation. We distinguish therefore *role associations* (see the above discussion about infons) as explicitly expressed by linguistic utterances from *situation associations* which are implicitly “responsible” for the creation of median roles (but note, however, that associations of semantic situations are also often marked in languages; i.e.: have explicit morphological markers, for example prefixing or suffixing verbs or using compound verbs). Let us also add that there may be more than one median roles in the same utterance, nevertheless this is rather rare. In general, linguistic information contains no more than maximum 3-4 roles and 3-4 anchors. This point is crucial for the present theory of *associative semantics*.

In general, active roles are filled by agents and passives roles by figures,

---

<sup>8</sup> Note however the rather exceptional idea of Tesnière, L. [TES 59] who theorized about roles using the abstract name “actant” (acting entity) and enumerated three of them: *first*, *second* and *third actants*. The most interesting issues in this domain may be however found in Sowa, J. [SOW 99] who proposed his theory of semantic roles following some original ideas of Somers, H. [SOM 87] and Dick, J. P. [DIC 91].

but this is not a rule. Initiator and Terminator are entities fulfilling Active and Passive Roles respectively. Agents typically fit the dynamic situations (Actions) whereas figures fit static situations (States). When it is not so, *shallow* level i.e.: partially specified semantic level is probably needed. When the Subject of a default (active or passive) diathesis sentence does not correspond to the default (active or passive in that order) semantic role of a given situation, the *shallow* level<sup>9</sup> of meaning must be introduced. In the cases of agentivation or figuration (see above), the shallow role names will be prefixed by the capital letter Q-... (as in quasi-). For example, "Q-initiator" "Q-source" will be said to designate the figure in an active role (which normally fit to agents) and "Q-source" will be said to designate the agent in an active role (which normally fit to figures). Research concerning the similarity of Proper Roles with Quasi-roles would probably require detailed exploration of a number of *analogies* in the ontological domain.

Hereafter is an example of transformation of the information contained in "The key opened the door" using participation role *types* only.

*Shallow level semantics:*

< open, (quasi-initiator-role: 'key', terminator : 'door'), true >

*Standard level semantics:*

< open, (initiator: **x**, means : 'key', terminator : 'door'), true >

Note that in standard level semantics, it is necessary to recognise that there is an indeterminate initiator **x** and that the former quasi-initiator has been transformed into a median role (because it is *passive* with respect to the initiator and again *quasi-active* with respect to the terminator).

### 3.3. Anchor Component

Indeed, it is possible to model the **Anchor Component** of semantic situations in a similar way as their **Role Component**. However, we must be aware of the fact that while the participation functions are projections into the entities, anchoring functions are projections into other situations.

<i>Localisation</i> \ <i>Anchors</i>	INITIAL	INTERMEDIARY	TERMINAL
Space	Start	Path (Itinerary)	Arrival
Time	Beginning	Course	End

Table #5 : Anchors of Semantic Situations

It is namely because both Components of semantic situations (Roles and Anchors) can be defined as pairs consisting of (a) the Role and (b) the Participant, on the one hand, and those of (a) the Anchor and (b) the Location, on the other hand, that in the past some linguists have had the

---

<sup>9</sup> Note, however, that due to the general principle of partiality of linguistic expressions both typified participants and their instances may happen to be *indeterminate*.

intuition of the localistic view<sup>10</sup> of the semantic structure of linguistic expressions.

#### 4. Centering: Global and Local Centres of Attention

The meta-informative presentation of information in an utterance requires to select centres of attention in a semantic situation. A segment of an expression is called « centered » (is a Centre of Attention or CA) if it contains meta-informative markers (syntactic, morphological, prosodic or any marker of pragmatic nature). 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 (at least a Subject), although the CA may sometimes be only implicit<sup>11</sup>. We call CAs 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 (the Subject) 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*: the object(s).

##### 4.1. Old and New Meta-informative Status of Utterances

Meaning is information while predication is communicating *about* selected chunks of it, i.e. what we call ‘centres of attention’ (creating linguistic expressions in which the distinguished segments are highlighted). This *aboutness* is of two kinds: either it is anaphoric or cataphoric (based on the order in which elements of information are introduced into discourse), or it is based on the way situations are viewed in cognition (as types or as occurrences). The above distinction concerns two different sorts of belief we consider to be at hand in the process of communication: the speech act oriented belief (we call it discourse based *old* or *new* meta-informative status:  $O_s/N_s$ ) as opposed to the common belief (we call *old* or *new* information-oriented distinction:  $O_i/N_i$ ).

The common belief relies on encyclopaedia-like knowledge and that knowledge which characterises individual experience of the speech act participants, it is based on the opposition between situation types and occurrences. In predication, as a matter of fact, the situation types and their occurrences are straightforward motivations for given and new meta-informative presentation of CAs. From a cognitive viewpoint, semantic situations are typed as *generic*, *general*, *habitual* or *potential*, and the occurrences of these types can be defined as their dual counterparts:

---

<sup>10</sup> Note that in this approach the analogy goes from Roles to Anchors rather than backwards.

<sup>11</sup> In so called impersonal sentences we treat the global CA as implicit and propose to call it “anonymous subject” (WLH 96, WAH 06a)

*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* (table #6).

Situation Types	Situation Occurrences
<i>Generic</i> > <i>General</i> <i>Habitual</i> <i>Potential</i>	<i>Specific</i> > <i>Particular</i> <i>Occasional</i> <i>Actual</i>

Table #6. Common knowledge-based informative motivation of aboutness

The distinction between semantic situation as *type* or as *occurrence* can be compared to what is known respectively 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. The duality described as *situation types* and *situation occurrences* is comparable to the distinction made by Von Wright, G. H. [WRI 63] 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 moment. The  $O_i/N_i$  opposition is grounded in information, in knowledge of the world; in other words it is related to ontological categorisation. This categorisation has been a matter of investigation since the dawn of philosophy and Aristotle himself already distinguished between substance and accident. In modern ontology as well, both entities and situations may be seen as types or tokens. Entities seen as types are permanent but they may be also transient as tokens and situations themselves may be either permanent or occurring.

The meta-informative status is *speech-oriented* ( $O_s/N_s$ , *s-old* or *s-new*) when it corresponds directly to the way information contents are introduced into discourse with reference to the conceptual pair *anaphora* and *cataphora*. We understand the anaphoric meta-informative value as referring to what should be taken as known by hearsay during previous speech acts and cataphoric value as referring to what should be taken as unheard of in any previous speech act. In such case information is validated as O or N not by the knowledge about it, but by the very way it is treated in discourse. The speaker is free to treat some information as O or N. Meta-informative validation consists in imposing the speaker’s point of view upon the addressee in spite of the opinion of the latter and notwithstanding generally admitted truth, as was the case with Copernicus’ utterance: “*The earth turns around the sun*”. Such an utterance may be *new* or *old* depending only on the circumstances in which it is uttered, in Copernicus’ times or in ours.

It is necessary to emphasize that the discourse based O/N opposition must not be confused with the syntactic sense of the concepts of anaphora and cataphora for which anaphora is defined as what has been already mentioned in the previous part of the text or discourse, whereas cataphora concerns items that are introduced for the first time. Our concept of

discourse-based O/N opposition is related to the speaker's attitude towards a fragment of information he wishes to propose to the hearer's attention as O or N – whatever may be the status of this information in common knowledge or in the previous parts of the text. Moreover, a noun phrase already quoted in a text may be treated as *new* in a subsequent part of the same text just because the entity it refers to enters a new situation (a new relation). For instance, in a text of which the general theme is Chopin's life, the name Chopin (or Frederic) may nevertheless occur with the *new* meta-informative status, although Chopin has been already quoted many times and is a well-known person: "*It was Chopin who invited Georges Sand to the party.*"

#### 4.2. Predication: Basic and Extended Utterances

The distinction between O and N meta-informative status lies at the basis of two different types of utterances which we call basic<sup>12</sup> and extended. In a basic utterance, both subject and predicate have the same (either O and O or N and N) meta-informative status. Moreover, in Indo-European languages, the concordance of meta-informative status is reflected on morpho-syntactic level by the agreement between subject and verb. However, even in languages without morphological concordance, it is possible to distinguish two different types of utterances depending whether subject and predicate have the same meta-informative status or not. As a matter of fact, in Japanese there is no grammatical concordance between subject and verb but nominal particles makes it possible to distinguish between subject and topic, i.e. between the global CAs of either basic or extended utterances.

Examples of basic utterances (#1 & #2):

#1 *Satellites turn around the Earth.*  
(Old Subject + Old Predicate)

#2 *A new satellite was launched yesterday.*  
(New Subject + New Predicate)

In what we call extended utterances there is a contrast between the meta-informative status of the CA and the rest of the utterance. Topic is characterised by its *old* meta-informative status contrasting with the *new* meta-informative status of the comment. The *old* status, as we mentioned above, must not be confused with the syntactic concept of anaphora. A topic is always O by nature, even if the NP it contains is introduced into the text or discourse for the first time. The contrast between a *new* CA and an *old* part of the utterance (background) creates a Focus [WAH 06b].

Examples of extended utterances (#3 & #4):

#3 *As concerns satellite X 05, it was destroyed yesterday by a meteorite.*  
(Old Topic + New Comment)

---

<sup>12</sup> A *basic* utterance (pragmatics) must be distinguished from a *simple* sentence (syntax).

#4 *It is satellite X 05 that was destroyed yesterday by a meteorite.*  
 (New Focus + Old Background)

Moreover, *basic* 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<sup>13</sup>. Thus, it must be stressed that, in our theory, *topic* (of an utterance) and *theme* (of a text) are not synonymous terms (cf. Table 6). To sum up, centering of utterances can be done on three syntactic levels of complexity: basic utterance, extended utterance and text or dialog. The hierarchical organization of different centres of attention of discourse is shown in Table #7.

TYPE OF EXPRESSION	CENTRES OF ATTENTION	
	<i>Global</i>	<i>Local</i>
1.1. Basic UTTERANCE	SUBJECT	OBJECT
1.2. Extended UTTERANCE	TOPIC	FOCUS
2. TEXT / DIALOG	GENERAL THEME	PARTICULAR THEME

Table #7. *Pivots of discourse*

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. The present 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 with what we call an “anonymous subject” [WLH 93, 94, 96].

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

---

<sup>13</sup> In this paper, we shall not dwell on the concepts of *general* and *particular themes* of a text, as an example, we make reference to these concepts in the analyses of a scene of a drama by Sławomir Mrożek [WLH 03].

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 in natural languages between semantic and pragmatic parts of linguistic content, we adhere to the opinion that it is necessary to distinguish between some parts of content as semantic and other parts as pragmatic. Consequently, it should be stressed that the organization of semantic roles (§2 above) does not depend on the hierarchical arrangement of pragmatic CAs (which are syntactically global and local). This is the reason why the classical semantic role theory as first designed by Fillmore, Ch. J. [FIL 68] had to be added such concepts as *subject* and *complements* in order to build efficient models of natural language utterances in HPSG<sup>14</sup>. However we claim that it is necessary to keep separated in our formal models what is inextricably combined in linguistic expressions, i.e. semantic roles (active, passive and median with their various particular realizations) and centres of attentions (*subject* and *object*). This separation makes it possible to give account for the possibility of matching any semantic role (or any anchor) with any CA (note that || means “processed in parallel”).

#5. *Mary treats Peter with aspirin.*

(Active verb + [Subject || Initiator] + [Object || Terminator] + [Indirect Object || Mediator])

#6 *Peter is treated with aspirin by Mary.*

(Passive verb + [Subject || Terminator] + [Indirect Object || Mediator] + [Object || Initiator])

#7 *As for Mary, she treats Peter with aspirin.*

(Active verb + [Topic || Subject || Initiator] + [Object || Terminator] + [Indirect Object || Mediator])

#8 *As for Peter, he is treated with aspirin by Mary.*

(Passive verb + [Topic || Subject || Terminator] + [Object || Initiator])

#9 *As for Mary, it is Peter whom she treats.*

(Active verb + [Topic || Subject || Initiator] + [Focus || Object || Terminator])

#10 *As for Mary, it is with aspirin that she treats Peter.*

(Active verb + [Topic || Subject || Initiator] + [Focus || Indirect Object || Mediator] + [Object || Terminator])

---

<sup>14</sup> As an example of Argument Structure in HPSG, cf. [WEC 95].

#11 *As for Peter, it is Mary who treats him with aspirin.*  
(Active verb + [Topic || Object || Terminator] + [Focus || Subject || Initiator])

#12 *As for Peter, it is Mary by whom he is treated.*  
(Passive verb + [Topic || Subject || Terminator] + [Focus || Object || Initiator])

#13 *As for Mary, it is Peter who is treated by her.*  
(Passive verb + [Topic || Object || Initiator] + [Focus || Subject || Terminator])

....

Moreover, it may happen that the situation frame itself is selected as a CA, i.e.: the verb phrase also may be topicalized or focalized (example #20). In this case however, in general, nominalization must be applied.

#14 *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.** H.G. Wells, *The War of the Worlds*.*

The part of #14 in bold characters is focalized: it suggests that the focalized situation frame (verb phrase) is to be treated as a *new* information whereas the rest of the utterance (“*saved them*”) is the background with *old* meta-informative status due to the quasi-synonym *escape* previously used in the discourse.

#### 4.3. Universal Definition of Subject in Accusative and Ergative Languages

Only when defining Subject on the level of meta-information (pragmatics) as the global centre of attention of the utterance is it possible to give a universal definition valid for all languages. The identification of a participant that is treated as global CA is the first stage (*signification*) in the communication process but the *interpretation* of the Subject as a participant playing a semantic role is a more complex operation that takes place at a second stage. As a matter of fact, this interpretation becomes even more complex when we introduce metaphor and metonymy, as these tropes enable some figures to play agent-like roles (Q-roles).

Definitions of Subject which take into account only the level of syntax are inadequate for some types of languages: for instance, the definition of Subject as obligatory alongside the predicate for the grammaticality of the utterance [MAR 62] or the view of the Subject as non constitutive of the utterance but accommodating the predicate [KUR 64] are not suitable for such languages as Japanese where Subject is not obligatory because verbs have no personal conjugation and consequently have no morphological agreement with the Subject. On the other hand, the meta-informative theory

of Subject and Object as CAs of basic utterances makes it possible to explain ergative languages without binding Object or Subject to one semantic role only.

Some linguists observed that either (a) the semantic roles or (b) the centres of attention can be ordered using similar criteria. First of all, roles were ordered by their importance, e. g.: *Agent* < *Experiencer* < *Instrument* < *Patient* < *Goal/Source/Location* ... by Fillmore, C. J. [FIL 70] whereas Bresnan and Kanerva [BRE 89] proposed the following order : *Agent* > *Beneficiary* > *Goal* / *Experiencer* > *Instrument* > *Patient* / *Theme* > *Locative*. Grimshaw, J. [GRIM 90] using the criterion of prominence proposed this order : *Agent* > *Experiencer* > *Goal* / *Location* / *Source* > *Theme*. On the other hand, Walker, M., Iida, M. and Cote, S. [WAL 94] established the following order for what we call *centres of attention*: (*grammatical OR zero*) *topic* > *empathy* > *subject* > *object2* > *object* > *others*.

Shaumyan, S. [SHA 87], for whom the primary term of a passivized predicate is the secondary term of the active predicate, puts the terms of the Predicate Argument Structure in the following order: *Primary Term* > *Secondary Term* > *Tertiary Term* > *Oblique Term*. For instance in English, the mappings between the Surface Structure (S V O) and the Argument Structure ( $P T^2 T^1$ ) are often as follows: (S ||  $T^1$ , V || P, O ||  $T^2$ ).

In our Associative Semantics and Meta-Informative Centering joint theory, we consider the relation between *Situation Roles* and *Centres of Attention* as *paratactic* (due to a parallel function) and we use the logical procedure (that of *Default Logic*) in order to define the default mapping rules for active and passive roles in accusative and ergative languages as follows:

<p><b>Prerequisite:</b> Global Centre of Attention(X)  <b>Justification:</b> Active Role(X) in active voice utterance  <b>Consequent:</b> Active Role(X)</p>
--

*Definition of Default Mapping from Global CAs to Active Roles in Accusative Languages*

<p><b>Prerequisite:</b> Global Centre of Attention(X)  <b>Justification:</b> Passive Role(X) in passive voice utterance  <b>Consequent:</b> Passive Role(X)</p>
---

*Definition of Default Mapping from Global CAs to Passive Roles in Ergative Languages*

The above formulation of default mapping from global CAs to active roles in accusative languages should be read as “if *x* is selected as a **Global CA**, and it can be assumed that it fulfills an **Active Role** in a given semantic situation, then we can conclude that it really fulfills an **Active Role**”. Note that, in Default Logic, rules allow to derive consequents from prerequisites

but not vice versa. In the case of accusative languages, *Active* Roles are not always mapped into Global CAs and in that of ergative languages *Passive* Roles are not always mapped into Global CAs.

Thus, the proposed definition of Subject as global CA of an utterance and the default (“solidarity”) relation between CAs and semantic roles enables us to distinguish more appropriately between different types of languages:

(a) *accusative* languages in which (explicit or implicit) Subject by default corresponds to the Agent playing the Active Role in the semantic situation expressed by the unmarked (prototypically active) diathesis,

(b) *ergative* languages where Subject corresponds by default to the Agent playing the Passive Role expressed by the unmarked (prototypically passive) diathesis,

(c) *ergative* languages in which Subject matches by default the Median agent role and

(d) *ergative* languages with *split-intransitive Subject* [DIX 94] which discriminate between two sorts of Subjects: one whose morphology is like that of Subjects (prototypically matching the Passive Agent Roles) of transitive utterances and another one whose morphology is like that of Objects (prototypically matching the Active Agent Roles) of transitive utterances. The situation of languages of type (a) and (b) is shown schematically in table #8.

Mapping between cases and roles		<i>Centres of Attention</i>	
		<b>Global (SUBJECT)</b>	<b>Local (OBJECT)</b>
in accusative languages	Marker (case or syntactic position)	<i>Nominative</i>	<i>Accusative</i>
	Role (by default)	<i>ACTIVE</i>	<i>PASSIVE</i>
in ergative languages	Marker (case or syntactic position)	<i>Ergative</i>	<i>Absolutive</i>
	Role (by default)	<i>PASSIVE</i>	<i>ACTIVE</i>

Table #8. *Centres of Attention, Roles and Markers in Accusative and Ergative Languages*

Moreover, as concerns the internal description of one language, the meta-informative definition of Subject as CA explains different diatheses (including active, passive, “anti-passive” and “impersonal”) as offering the possibility to change the global CA of an utterance depending on the point of view on participants that is chosen by the speaker who seeks to preserve the coherence of the text or discourse [WA&H 06a, 06b].

## 5. Conclusion and Further Research

Both the Associative Semantics (AS) theory and the Meta-Informative Centering (MIC) theory were designed with the working idea that they could be used as a cross-linguistic framework for the description of essential functionalities of languages belonging to very different, historically unrelated, families. It is necessary however to correctly identify the language data the above theories can be applied to. This is rather a hard task especially because of the pre-existing traditional (local) solutions of the problems we deal with. We believe that what is being traditionally called “information<sup>ii</sup>” in linguistics will be recognized as corresponding to the linguistic phenomena treated in MIC as “meta-information” and we hope that henceforth they will be renamed “meta-information”. Indeed, presenting communicated information (or chunks of it) as *old* or *new* is presenting more than information.

We are well aware of the fact that *modeling* or *simulating* does not imply that reality is so or so. Modeling or simulating aims only at representing the given domain of interest using formal implications. For example, the distinction made in the field of research on the Form of linguistic signs between phonetics (the study of sounds) and phonology (the study of phonemes) was possible due to the introduction of mathematical notions of *types* and *instances* in linguistic studies by Trubetzkoy (1939).

Associative Semantics (AS) is now being formalized following certain solutions of the Information Flow Framework (Barwise J. and Seligman, 1997). As regards the MIC theory it has strong psychological support. According to Gurwitsch, A. (1957), the thematization which characterizes the field of consciousness is continuous (what is central becomes easily marginal). However, in the MIC theory, at present, only the communicative aspect of utterances (which are discrete language units) is developed. Indeed, in absence of well-defined universal concepts, any further exploration of communication phenomena (for example discourse sessions coherence and the like) seems inconceivable.

We hope that linguists, logicians, and computer scientists will investigate deeper (a) the relationship between the Old/New meta-informative statuses of information and the problems of truth validation, (b) the ways that determination (quantification, qualification, articles etc.) and modification (aspect, modality etc.) together contribute to the true/false and old/new validation of utterances (Włodarczyk A. & H. in press).

Last but not least: the MIC and AS theories are the result of research that the authors have been carrying on partly in cooperation, partly individually. At present, research continues with the purpose of defining formally semantic feature structures in order to create data for semantic databases as defined in the framework of the CASK<sup>iii</sup> Project. Indeed, such databases are crucial for conducting advanced typological and contrastive research<sup>iv</sup> on natural languages.

## 6. References

- Barwise, Jon K. and John Perry, J. 1983. *Situations and Attitudes*. Cambridge, Mass.: MIT Press.
- Barwise, Jon K and Jerry Seligman. 1997. *Information Flow, The Logic Of Distributed Systems*. Cambridge, New-York : Cambridge University Press.
- Bresnan, Joan and Jonni M. Kanerva. 1989. Locative inversion in Chichewa: A case study of factorization in grammar. *Linguistic Inquiry* 20:1–50.
- De Saussure, Ferdinand 1913, 1995. *Cours de linguistique générale*. Paris : Payot,.
- Devlin, Keith. 1991. *Logic and Information*. Cambridge, UK: Cambridge University Press.
- Dick, Judith P. 1991. *A Conceptual Case-Relation Representation of Text for Information Retrieval*, PhD dissertation, Report CSRI-265, Computer Systems Research Institute, University of Toronto.
- Dixon, Robert M .W. 1994. *Ergativity*. Cambridge, UK: Cambridge Studies in Linguistics, Cambridge University Press.
- Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67: 547-619.
- Fillmore, Charles J. 1968. The case for Case. In Bach Emmon & R. J. Harms (eds), *Universals in linguistic theory*: 1-90, New York: Holt, Rinehart & Winston.
- Fillmore, Charles J. 1970. The Grammar of Hitting and Breaking. In Jacobs R. and P. Rosenbaum (eds), *Readings in English Transformational Grammar*, Waltham, MA: Ginn.
- Fillmore, Charles J. 1977. The case for case reopened. In Cole, P. & J. M. Sadok (eds), *Syntax and Semantics 8*: Grammatical relations, 59-81. New York: Academic Press.
- Grimshaw, Jane, 1990. *Argument Structure*. MIT Press, Cambridge, MA.
- Gurwitsch, Aron, 1957. *Théorie du champ de conscience*. Paris: Desclée de Brouwer. Translated: *Field of Consciousness*, Pittsburgh, Pa.: Dusquesne University Press (1964).
- Hjelmslev, Louis. 1953. *Prolegomena to a theory of language*. Baltimore: Waverly Press.
- Jakobson, Roman. 1956. Two Aspects of Language and Two Types of Aphasic Disturbances. in Jakobson R. and Morris Halle *Fundamentals of Language - part II*. The Hague: Mouton. Rev. Ed. 1971.
- Kuryłowicz, Jerzy. 1964. *The Inflectional Categories of Indo-European*. Heidelberg: Winter,.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representation of Discourse Referents*, Cambridge: Cambridge University Press.

- Martinet, André. 1962. *A functional View of Language*, Oxford: Clarendon.
- Mathesius, Vilém. 1975. *A Functional Analysis of Present Day English on General Linguistic Basis*, Josef Vachek, ed. Mouton: The Hague, Paris, Academia, Prague. Translation of *Obsahový rozbor současné angličtiny na základě obecně lingvistickém*, originally published in 1961.
- Minsky, Marvin, 1975. A framework for representing knowledge, in *The Psychology of Computer Vision*, Winston P. H. (ed.), McGraw-Hill.
- Pustejovsky, James. 1995. *Generative Lexicon*, Cambridge: Bradford Books, MIT Press.
- Shaumyan, Sebastian. 1987. *A Semiotic Theory of Language*. Indiana University Press.
- Somers, Harold. 1987. *Valency and Case in Computational Linguistics*. Edinburgh Information Technology Series 3, Edinburgh: Edinburgh University Press.
- Sowa, John F. 1999. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Pacific Grove, CA: Brooks Cole Publishing Co.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*, Paris : Klincksieck.
- Trubetzkoy, Nikolai. 1939. *Grundzüge der Phonologie. Travaux du cercle linguistique de Prague 7*, Prague.
- Walker, Marilyn A., Masayo Iida, & Sharon Cote. 1994. Japanese Discourse and the Process of Centering. *Computational Linguistics 20/2*:193-232, MIT Press.
- Wechsler, Stephen. 1995. *The Semantic Basis of Argument Structure*. Stanford, California: CSLI Publications.
- Wisniewski, Edward J. 1997. Understanding Novel Noun Phrases, *Proceedings of IWHIT 1997*: 97-101, The University of Aizu.
- Włodarczyk, André. 2003. Les Cadres des situations sémantiques. In *Études cognitives / Studia kognitywne 5* : 35-51. Warszawa: Slawistyczny Ośrodek Wydawniczy, Polska Akademia Nauk.
- Włodarczyk, André. In press. Roles and Anchors of Semantic Situations, *Études cognitives / Studia kognitywne 8*. Warszawa: Slawistyczny Ośrodek Wydawniczy, Polska Akademia Nauk.
- Włodarczyk, André & Hélène Włodarczyk. 2006a. Subject in the Meta-informative Centering Theory. in *Études cognitives / Studia kognitywne 7*: 39-64. Warszawa: Slawistyczny Ośrodek Wydawniczy, Polska Akademia Nauk.
- Włodarczyk, André & Hélène Włodarczyk. 2006b. Focus in the Meta-informative Centering Theory. In *La Focalisation dans les langues*, A. & H. Włodarczyk (eds), Paris: L'Harmattan, Collection Sémantiques.

- Włodarczyk, André & Hélène Włodarczyk. In press. The Pragmatic Validation of Utterances. In *Études cognitives / Studia kognitywne 8*, Warszawa: Sławistyczny Ośrodek Wydawniczy, Polska Akademia Nauk.
- Włodarczyk, Hélène. 1993. Sentences without a Nominative NP in Polish. In Laskowski, R. and Hentschel (eds), *Studies in Polish Inflectional Morphology and Syntax - Synchronic and Diachronic Problems*. München : Otto G. Sagner Verlag.
- Włodarczyk, Hélène. 1994. Les Phrases à sujet anonyme et verbe réfléchi en polonais et en russe. In Sémon J.-P. & Włodarczyk H. (eds) Organisation sémantique de la phrase dans les langues slaves, *Revue des Études Slaves* T. LXVI /3 : 669-680. Paris : Institut d'Études Slaves.
- Włodarczyk, Hélène. 1996. Les Traits sémantiques du sujet 'anonyme' en polonais, russe et français. In *Semantyka a konfrontacja językowa* : 179-198. Warszawa : Sławistyczny Ośrodek Wydawniczy, Polska Akademia Nauk, Instytut Sławistyki.
- Włodarczyk, Hélène. 2003. L'Interprétation dynamique des centres d'intérêt dans les dialogues français et polonais. in *Études linguistiques romano-slaves offertes à Stanisław Karolak*: 511-528, Kraków: Oficyna Wydawnicza Edukacja.
- Von Wright, Georg H. 1963. *Norm and Actions*, London: Routledge and Kegan.

---

<sup>i</sup> See the paragraph "Anchor Component" below.

<sup>ii</sup> See the concept of *information structure* in the Prague School tradition (Mathesius. 1975). Cf. also Lambrecht (1994).

<sup>iii</sup> CASK – Computer-aided Acquisition of Semantic Knowledge. The Project started at CELTA (Centre for Theoretical and Applied Linguistics, Paris-Sorbonne University) in 2006. The high-tech methods of KDD (Knowledge Discovery in Databases) are extensively used aiming at building databases for research on selected grammatical categories such as Aspect and Modality in different languages. The most advanced KDD algorithms are implemented in the SEMANA (Semantic Analyzer) platform. More about the CASK Project can be found at: <http://www.celta.paris-sorbonne.fr/anasem/>

<sup>iv</sup> Post-experimental examination of data from consistent semantic databases using the SEMANA platform.