Ontological Issues for Modelling Aspect Semantics

Hélène Włodarczyk[1]

CELTA - Centre de Linguistique Théorique et Appliquée

helene.wlodarczyk@paris-sorbonne.fr

Abstract

The CASK method (Computer-aided Acquisition of Semantic Knowledge) is being used at CELTA with the SEMANA (SEMantic ANAlyser) software, which was especially designed for this purpose. We use the notation of semantic feature structures as a meta-language for describing the category of aspect in Polish regardless of its pertaining to various levels of expression (morphological, syntactical or lexical). To model the category of aspect we treat types of situations as conditions for two relevant aspect parameters we call analysis of the situation (or internal aspect) and control of the situation (or external aspect). In order to cope both with the lexical diversity of aspectual morphemes (prefixes) and the grammatical (dichotomous and obligatory) character of aspect in Slavic languages, we have proposed to define the perfective aspect as a hypercategory. Our database of Polish aspect was analysed using KDD statistical tools (G. Sauvet 2008, http://celta.paris-sorbonne.fr/anasem/papers/). We obtained the first preliminary experimentally fixed semantic definitions of the perfective and imperfective values of the category of aspect in a Slavic language.

1. Ontology and Semantics

Multi-lingual contrastive studies need to refer to ontologies as *tertium comparationis*, including those which were especially designed in order to account for linguistic objects (expressions). We claim that this task cannot but be interactive (computer-aided), if we want to avoid the lack of precision and the variability of semantic parameters in traditional linguistic research, especially hazardous in the semantic domain.

We define the semantic content of a linguistic expression as a function mapping this expression on ontological concepts. Therefore, to describe the semantic content of aspectual expressions used in utterances we need to specify the ontology to which they refer. A formal cognitive description aims at giving an ontological account of a semantic category by treating its definition in different languages as a finite set of strictly defined feature structures. We aim at giving a two-fold account of semantic categories by:

- building a general set of ontological abstract structures necessary to interpret aspect in different languages

- choosing a specific set of semantic feature structures for the described category in each human language, in our case for Polish.

To describe aspectual values of verbs in context, we had first to define relevant aspectual semantic attributes and their values (AV). At this stage, linguists must use both

[1] Acknowledgment: the research on aspect with the SEMANA software has been conducted at CELTA (Paris-Sorbonne University) since 2005 by Hélène and André Włodarczyk (ontology and semantics of verbal aspect), Georges Sauvet and André Włodarczyk (conception of the SEMANA software), Georges Sauvet (analysis of the aspect database with SEMANA KDD tools). Doctoral and master students also take part in this research.

general knowledge of the category they are studying and specific knowledge of the language they are describing.

2. Computer-aided Acquisition of Semantic Knowledge

Our experience consists in using a software for the acquisition of semantic knowledge (SEMANA designed at CELTA) after years of traditional linguistic research on the topics of aspect. We have been conducting this research in the field of verbal aspect semantics in Slavic languages in contrast with Indo-European non-Slavic languages, mostly French and English. We turn to KDD methods in order to enhance linguistic research in two ways. 1) using a software as Semana can help grasping a large set of semantic features and make clear the relations between them, i.e. the structure or system they belong to ; 2) it provides the linguist with a universal (feature) language whose rules are logico-mathematical and with tools worked out by computer scientists to handle this language and perform calculation on it. And last but not least, a software offers the possibility to store large language data, to access them easily and share them with other researchers.

2.1. Towards Experimental Semantics

The method of Computer Aided Acquisition of Semantic Knowledge (CASK) is based on the idea that the meaning of linguistic units can be described only in context. In a given context (inside a text or discourse and in a particular speech situation) an expression is not ambiguous (it can be described by one feature structure). What is called ambiguity or polysemy is the possibility for an expression to be used with different senses in different contexts. For this reason, it is important to collect numerous samples of usages (build semantic databases), so that linguistic theories consist in describing and analysing these data with symbolic and statistical methods.

This approach differs radically from hypothetico-deductive linguistic theories (which use merely a few examples as illustrations) but it is confronted with the serious problem of meta-data input. In fact, we consider that bare facts do not exist as such, any description of linguistic “facts” relies on a chosen theoretical background; for this reason, we call the linguistic data we collect in our database “meta-data”.

At the stage of collecting and marking semantic data, the intuition of the linguist is irreplaceable although fallible. Interaction between man and machine (consisting of handling a list of fixed well defined monosemous features that demand conscious intervention to be modified or enlarged) can prevent from the subjectivity and variability of human appreciation of the meaning of expressions.

The problem of data input comes partly from that linguistic expressions exhibit in context not only explicit meaning, but also entail as well presupposed as inferred knowledge. It is sometimes difficult to establish which part of the presupposed or inferred knowledge is pertinent in a given context and should be taken into account in the description.

2.2. SEMANA: a Software designed for semantic research

The problem of communication between linguists and computer scientists comes from the times when the latter used to analyze the formers’ needs in their specific field at a given stage of their research in order to offer them adapted tools. Most often such tools reflected the image (knowledge) of the field at the time of the programmer’s analysis but were not flexible enough to be easily modified when the specialist’s view of the field changes. The SEMANA

software[2], especially designed at CELTA for the CASK project (Computer-aided Acquisition of Semantic Knowledge) offers two sorts of tools: (1) tools for interactive intelligent and dynamic database designing; (2) tools for automatic KDD research.

SEMANA's Dynamic Database Builder (db builder) is highly interactive, relying on the linguist's expertise in a given domain; it makes it possible to change features and their values (and the structure they belong to) as soon as progress in research proves it necessary. Each card in the db builder contains a field for the specimen described (an utterance chosen from a corpus) and a field with a list of attributes and values from which the linguist chooses the appropriate values for the sample he is describing. The characteristic morpheme of the analysed expression is used as index. The db builder is completed by a Tree Builder Assistant which allows the linguist to organise in a tree structure the attributes and values chosen for the description of a semantic field. Any change in the feature tree of the Tree Assistant is transferred to each record of the database after the linguist is asked whether he accepts changes in the tree builder to be echoed in the database. Since the beginning of the research with Semana we could change several times the tree of attributes used to describe aspect in Polish. All specimens are automatically collected in a contingency table. The synthetic table has the form of a chart of attributes and values for each sample described in the database. In the synthetic table, linguists can observe which samples present the same value of the same attribute. For linguistic description it is an important assistance : it makes it possible to verify whether the same attribute and value were chosen rightly in different contexts and at different times of data description by the linguist. It may seem trivial but, in semantic annotation, the choice of attributes and values is very sensitive to narrow and broad context. This table is completed by tools which provide statistical information about the use of attributes and their values and suggest interactive restructuration of the attributes and their values: it checks objects with the same AV (duplicates), proposes to merge 2 or more attributes, shows types of objects by attributes or by values, checks the AV field (appearing in each card) and the feature tree (in the Tree Builder Assistant). Many automated checking procedures are available and help the linguist to check the consistency of his reasoning.

The second part of SEMANA consists in tools for KDD research integrating the following methods: Rough Set Theory (RST, Z. Pawlak), Formal Concept Analysis (FCA by R. Wille and B. Ganter), Statistical Data Analyses (by J.-P. Benzécri). These tools are described in André Wlodarczyk's article in the same volume: "Interactive Discovery of Ontological Knowledge for Modelling Language Resources".

3. The Categorial and Contextual Meaning of Aspect

As regards aspect, it is often a very delicate task to distinguish the explicit categorial aspect meaning of an expression used in an utterance from its inferences and presuppositions[3]. The latter are referred to globally as the "pragmatics" of aspect or its conversational implicature. In my view, what is called the pragmatics of aspect refers to two different kinds of Aspectual uses in context.

In the first place, properly aspecto-temporal senses of verbs in context must be taken into account. Such senses do not depend only on the categorial meaning of the aspect category in the language system but also on the utterance in which the verb is used. For instance, in some languages, uses of verb forms expressing that a process has reached its finish point in

[2] http://www.celta.paris-sorbonne.fr/anasem/papers/

[3] Much has been written about this problem by Jakobson 1932, 1936 (Gesamtbedeutung) , Bondarko 1971a, b, and others Primary and Secondary meanings of aspectual forms (Kuryłowicz 1977)

the past (before the speech time point) often allow in context to infer that the situation is in its *after* stage (in a new state resulting from the process expressed by the verb) thus producing what is called a *resultative* meaning. e.g. In Polish *Zmarzłem* (*I have got frozen)* means *I am now cold* (at the speech time) or *was cold* (at the time period serving as reference point). Such kind of inferred meaning properly relies on ontological knowledge and reveals important in translation. Since every utterance content is partial as regards the situation it refers to, the explicit/implicit part of an utterance content is not always the same in two different languages. When translating, it is sometimes necessary to replace the implicit inferred meaning of the original expression by an explicit expression in the target language. As a matter of fact, the aim of the translation is to produce an expression which refers to some ontological knowledge that is similar to that of the original expression.

Secondly, we claim that the properly *pragmatic* meaning of aspect is related to the meta-informative (sometimes called cognitive) *old* or *new* status of the utterance. We have devoted and continue to devote much attention to this part of the aspectual problem, which reveals very important when contrasting languages (cf. Włodarczyk H. 1997, Włodarczyk A. & H. 2008). In this paper we do not develop this problem because, for the time being, in our database, we limit the description of aspect uses to the first sort of uses, i.e. properly aspecto-temporal uses.

4. Aspect as a Hypercategory

Aspect in Slavic languages is based on the opposition of two values only (perfective and imperfective) but the aspectual opposition is expressed not only in aspectual pairs but also in what we call aspectual families (or derivational nests). In Slavic languages, all prefixed verbs derived from a simple verb become perfective[4]. Among them, two classes can be distinguished although the border between these two classes is not sharp : verbs with a new lexical meaning (lexical derivatives) and verbs, which keep the same lexical meaning but change aspect (aspectual derivatives). Among the aspectual derivatives, slavicists distinguish traditionally (since the beginning of the 20th century, cf. Agrell 1908) between a series of derived verbs called Aktionsart (or *lexical* aspect) and *one* derived perfective verb considered as *the grammatical* perfective partner of the simple verb. Recently, this distinction has been reappraised by many researchers (including, Sémon 1986, Paduceva 1996, Karolak 1997, Xrakovskij 1997). Moreover, much work has been done to prove that not all so-called "aspectual pairs" are semantically similar, i.e. the opposition of perfective and imperfective may be based on different sets of semantic features depending on the lexical meaning of verbs (more precisely, depnding on the ontological type of situation as defined below). The first pioneer work in that direction was that of Cezar Piernikarski (1969) who showed that there exist several different types of "aspectual oppositions".

Our ontological approach makes it possible to bring closer the two sub-categories of aspect and Aktionsart and to replace the notion of "aspectual pair" by that of aspectual nest when defining the perfective aspect of Slavic languages (Włodarczyk A. & H. 2001). Most of the meanings traditionally assigned to the category of Aktionsart are part of what we call **control** parameters (see below). In prefixed verbs, these meanings combine with those considered as strictly grammatical perfective meanings and are characteristic of the aspectual nest that can often be derived from a simple imperfective verb. In former theories of aspectual pairs, only prefixed perfective verbs with a *resultative* meaning were considered *pure grammatical* perfective partners of a simple imperfective verb. In our view, we consider as

[4] Exceptions are extremely rare and due to diachronical reasons.

aspectual pairs only those pairs consisting of a perfective verb and the suffixed imperfective verb that is derived from it, e.g. *przepisać* (perf.) / *przepisywać* (imp.), *to copy*.

Perfective as a hypercategory (a two-level category due to the derivational origin of apectual morphemes in Slavic languages) subsumes all meanings of so called *pure* aspectual partners and *Aktionsart*. The concept of hypercategory allows us to describe each derived perfective verb as inheriting several aspectual features. Following this hypothesis, no verbal prefix can be viewed as semantically void because different configurations of features are always at hand. As a matter of fact, none of the perfective partners of a simple imperfective verb can be considered as entirely "synonymous" to the root verb.

Using the database of the Polish Frequency Dictionary (SFPW) we studied the relative frequency of simple imperfective verbs and the different perfective verbs derived from them (Włodarczyk A. & H. 2001). It appears that the frequency of the so-called "pure perfective partner" (very often the one with *resultative* meaning) is much higher than the frequency of verbs considered as Aktionsart and this is probably the reason why this perfective partner was generally considered as *the only pure grammatical perfective partner* of the simplex imperfective verb. Our treatment of aspectual prefixed verbs can be regarded as one more contribution to the long-lasting discussion about "aspectual pairs" in Slavic languages. We fully agree with the opinion of slavicists who consider (although mostly on the ground of other arguments) that pairs are constituted only by an imperfective suffixed verb that is derived from a perfective verb, e.g. *przepisywać* imp. from *przepisać* perf. (both can be translated as "to copy"), *zamawiać* imp. from *zamówić* perf. (both can be translated as "to order"), etc.

The notion of hypercategory is derived from the idea that categories can be classified. As shown above, the aspectual meaning of a verb used in a given context can be described as a bundle of attributes. By using inheritance, we do not have to build disjoint classes of aspectual meanings (as it was the case with classes of Aktionsart verbs). In fact, aspectual meanings are irregularly linked to various superior nodes. Moreover, one and the same verbal lexeme may have different links depending on the context in which it is used. This is often the case of *po-* prefixed verbs that can have several meanings (cf. Włodarczyk A. & H. 2001, II).

This approach allows us to take into account the lexical diversity in Slavic aspectual semantics and sheds light on the controversies about the grammatical or lexical nature of aspectual features. We conclude that in the aspectual derivation in Slavic languages there is no clear border between the lexical and the grammatical. Moreover, there is no need neither to consider this situation as "incomplete grammaticalisation". On the contrary, this endows Slavic languages with a very systematic way of expressing a broad range of aspectual nuances

5. Three Kinds of Aspectual Parameters and Formal Definition of Aspect

The aspectual attributes and values we use in the database are the result of previous research on aspect (Włodarczyk H. 1997) and are defined in the framework of the theory of aspect we outlined for interactive semantic research (Włodarczyk A. & H. 2003, 2006). We adopted the notation of *semantic feature structures* as a meta-language for describing aspectual meanings in various languages regardless of linguistic levels (morphological, syntactical, lexical etc). We propose to describe the meaning of the aspect category as a pair of feature bundles : *Analysis* and *Control*. The "analysis of a situation" (viewed as a whole or as one of its moments or stages) is considered as its *endocentric aspect,* concerning the internal development of a situation as time passes by. We call the "control of a situation" a set of operations such as iteration, modifications of flow or intensity, composition of two or more situations into one. This control parameters are imposed on a situation from outside of its

internal development in time and therefore we consider them as *exocentric aspect*. Moreover these internal and external aspectual features occur and combine diversely depending on the semantic type of the situation to which a verb is related in a given context. Each aspect use can therefore be described by a semantic feature bundle consisting of two parts: situation analysis and situation control. The situation type is considered as a condition for the use of aspect. This allows to propose a formal definition of aspect as follows.

Aspect = { Sit. Analysis , Sit. Control } condition : Sit. Type

5.1. Types of Semantic Situations

The classification we use (Włodarczyk A. 2003) differs from other classifications (Vendler Z. 1967, Laskowski R. 1998, among others) in that it does not incorporate features concerning situation participants and roles but only situation frames; it is based on four relevant features: space (three dimensions), time, progression and granularity.

SEMANTIC TYPES OF SITUATIONS				
Characteristic properties (dimensions)	**Static Situations**	**Dynamic Situations (ACTIONS)**		
	STATE	EVENT	Ordinary PROCESS	Refined PROCESS
Space (3D)	+	+	+	+
Time	-	+	+	+
Progression	-	-	+	+
Granularity	-	-	-	+

Table 1: Hierarchy of semantic situations (Włodarczyk A. 2003)

States differ from dynamic situations, in that they are not modified by time passing. Events have no progression: their beginning moment ("start") coincides with their ending moment ("finish") and it is impossible to observe an intermediary stage ("run") between them. In other words, events are dynamic situations without progression or development. Progression characterises processes, which develop in time from a *begin* stage to an *end* stage through an intermediary stage we call *run*. A granular process is made up of a repetition of many identical grains. Situations are hierarchically ordered: each type of situation inherits properties of the preceding type.

5.2. Internal Aspect: Situation Analysis

Any situation is preceded by a preceding situation (we call the "before" stage) and a following situation (we call the "after" stage). But only processes, i.e. dynamic situations with progression, may be further analysed into inner moments and stages (Fig. 1). Static situations may have only a start and a finish moment without any further analysis between them. Events are dynamic situations conceived as instantaneous (without progression) and therefore they are described as only one moment into which the start moment and the finish moment are merged.

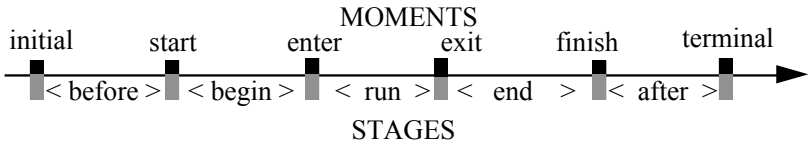

Fig. 1: Analysis of simple dynamic situations in moments and stages (constitutive parts)

The distinction that we make between moments and stages (represented on the figure as points and segments) may remind of the geometric metaphor of a "point" as opposed to a "line" often used in aspectology but, in our approach, we do not identify totally perfectivity with the point-view and imperfectivity with the segment-view. As a matter of fact, we consider the selection of a moment or a stage as only *one* of the semantic attributes of the aspect category. This attribute is combined in different configurations with other parameters in order to give account of different usages of the perfective and imperfective verbs.

A situation characterised as a process may be roughly analysed in three inner stages: *begin*, *run* and *end*. Moments serve as boundaries for stages, and we called them (arbitrarily) *initial, start, enter, exit, finish* and *terminal*. What is relevant in our theory is not the intuitive meaning of these words in English but the place they mark on the line representing the progression of a process in time

5.3. External Aspect: Control

Exocentric aspect ("control") consists in a set of parameters that are combined with the endocentric ones. The control may concern the repetition of the situation, the modification of the flow or intensity (*interrupt, resume, keep, off-and-on*[5]*, trans*[6]), the composition of two (or more) sequential or parallel situations into one complex situation. In Polish, the composition of situations is expressed in the case of verbs with prefixes indicating that the situation is composed of two or several situations. As an example we may quote the so called *distributive Aktionsart*: situations performed simultaneously or successively by different subjects or on different objects are composed into one complex situation, e.g.:

Pootwierałem wszystkie okna. (Lit. I opened all the windows one after the other.)

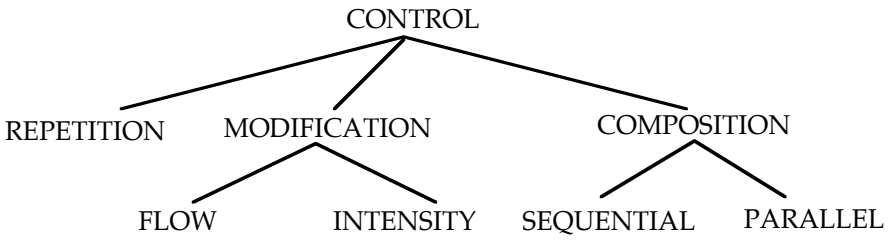

Fig.2 Aspectual Parameters of Situation Control

Many of what we call control parameters were previously described as *limitative*, *intensive*, *iterative*, *distributive*, *completive* etc. *"Aktionsarten"*. However, each type of aspectual meaning that was previously called an Aktionsart was generally described by only one single (often dichotomous) label; in our approach, such meaning pertains to more than only one parameter because we define aspect at least by the pair of two sorts of parameters: *analysis* of the situation and its *control*. Generally, in one verb usage, at least one control parameter (repeated or not repeated) or more (modification, composition) may be at hand, thus every aspect usage is defined by a structure of several semantic features.

6. Interactive Semantic Research on Aspect with the SEMANA Software

[5] The "off-and-on" flow modification concerns the unfolding of a situation intermittently.
[6] We call "trans" the unfolding of the whole situation from start to finish (regardless of its stages)

Hereafter we sketch out the interactive research on aspect both from the point of view of the linguist (aspect database building by Włodarczyk H.) and the computer-scientist (KDD analysis of the aspect database by Sauvet G).

6.1. From the work on the ontological tree of aspectual features

In the aspect database we use a tree of the attributes and values (AV) described above to annotate each aspectual specimen chosen in a corpus. This tree was first designed in the Tree Assistant as an ontology of aspect and modified several times as we collected more and more samples.

Let us quote just an example of the possibility of modifying the tree of ontological features during the collecting of data. In the first version of the aspect feature tree, the attribute ASPECT ANALYSIS was divided into *inner* and *outer* moments and stages:

```
ASPECT-*-ANALYSIS-*-MOMENT-------*-MINN------*-{AMI}=[enter|exit|finish|start]
       *           *               *-MOUT------*-{AMO}=[initial|terminal]
       *           *-STAGE--------*-SINN------*-{ASI}=[begin|end|run]
       *           *               *-SOUT------*-{ASO}=[before|after]
```

After we collected data, both *outer moments (initial, terminal)* that were never used in the db were deleted. This led us to simplify the attributes moments and stages as follows

```
ASP-*-ANLS-*-MOM--*-MOMI-*-{AMI}=[ent|exi|fin|str]

    *       *-STG--*-SI---*-{ASI}=[beg|end|run]

    *       *       *-SO---*-{ASO}=[bef|aft]
```

On the contrary both outer stages were frequently used in the description : we can understand that linguistic aspectual expressions of situations take into account the situation itself and its immediate bordering situations (before and after) but does not point at any dividing moment between the preceding situation (what we call the before stage) and another even more anterior situation because this would lead to an infinite regression (and the same for the moment we called terminal as end of the after stage. Thus, only the moments (schematized by points on a line) indicating the border between the outer stages and the situation itself can be expressed in verbal expressions of aspect : the moment we call *start* constitutes both the end of the *before* stage and the beginning of the *begin* stage of the situation, and the moment we call *finish* is both the last moment of the situation (of its *end* stage) and the first moment of the *after* stage.

6.2. Consistency checks

The field *index* of the db builder contains the aspectual morpheme of the described expression: a prefix or a suffix, or a periphrastic aspectual expression (e.g., an aspectual verb as *zaczynać*, "to begin", *kończyć* "to stop" or *nie przestawać* "do not stop", an adverb as *wciąż*, "continuously" etc).

Each specimen is characterised by a set of AV and by its morpheme (used as index). It may be written as a rule: **if {given set of AV} then index**. This allows index consistency to be detected. As a matter of fact, in our database, the test of consistency detected several different prefixes that were described by the same set of AV. However the polysemous character of verbal prefixes in Slavic languages and the two-step categorial structure of aspect (the hypercategory) requires that the linguist check the detected inconsistencies and accept them under the following conditions: (1) in the case when different prefixes share a common bundle of semantic features (with the necessity for the linguist to add relevant distinctive features to give account of fine-grained semantic differences) or (2) when the same prefix is characterised by two or more different feature trees (polysemous prefixes).

6.3. Successive versions of the Aspect database

As we progressed by trials and errors in different versions of the aspect db we obtained gradually improved statistical reports. In Table X, column 2 displays the result of the automatic deletion of duplicates in the db. As we collect samples from text corpora it is obvious that this random access to data leads to the input of different samples having the same aspectual feature structure. In columns 3, 4 and 5, we can see that, as our work proceeded, we limited the number of attributes so that the number of theoretical combinations decreased and the possibility of merging attributes was reduced to zero.

DB version	Distinct objects	Number of attributes	Number of theor. combin.	Number of "merging attributes"
HW-Aspect-V1	61	12	2,064,384	9
HW-Aspect-V2	60	11	1,032,192	9
HW-Aspect-V3	77	11	829,000	6
HW-Aspect-V4	79	9	408,240	1
HW-Aspect-V5	79	8	136,080	1
HW-Aspect-V6	69	8	45,360	1
HW-Aspect-V7	74	8	61,440	0
HW-Aspect-V8	78	7	58,320	0

Table 2. Improvements reflected by statistical reports

6.4. Analysis of the first database of Aspect in Polish using SEMANA KDD tools

The feature tree used in the aspectual database (version 8) analysed by G. Sauvet was the following :

ASPECT-*-ANALYSIS-*-{ANA}=[enter|exit|finish|start|initial|terminal|begin|end|run|before|after|nonanalyzed]
***-CONTROL--*-FLOWMODIF-*-FLOW-------*-{MOD}=[interrupt|keep|resume|stop|trans|OffandOn|parallel|sequential]**
*** *-REPETITION-*-{CRE}=[defnb|indnb]**
***-INTENSITY-*-{ITS}=[increase|decrease|strong|weak]**

ASPVALUE-*-ASPVAL---*-{VAL}=[imperfective|perfective]
***-MRPHCOMP-*-{MCP}=[impPrf|prfImp|prfImpPrf]**

SITTYPE-*-{TYP}=[event|ordProcess|refProcess|state]

The multi-valued synthetic table corresponding to version 8 (fig. 3) was exported to the STA3 device of SEMANA and then automatically transformed into a one-valued table.

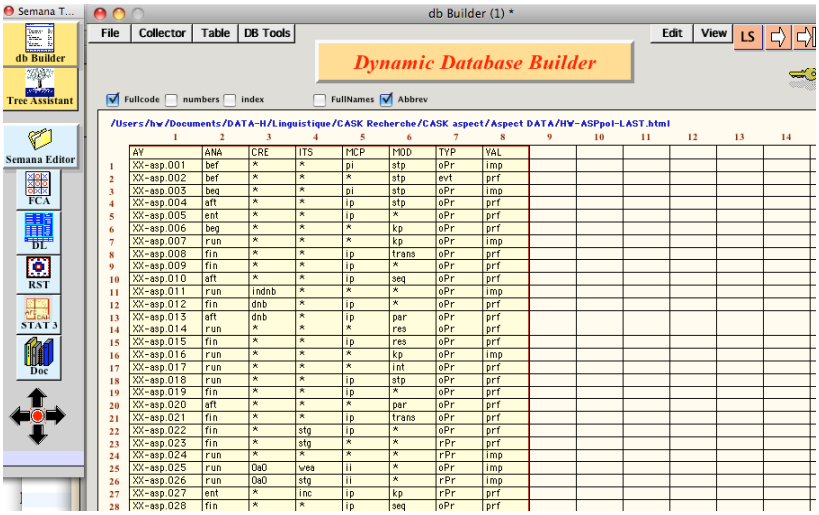

Fig. 3 Synthetic table of Aspect db Version 8

The Correspondence Factor Analysis (fig. 4) shows a clear partition of relevant features into two classes according to the attribute [VAL] = {perfective | imperfective}.

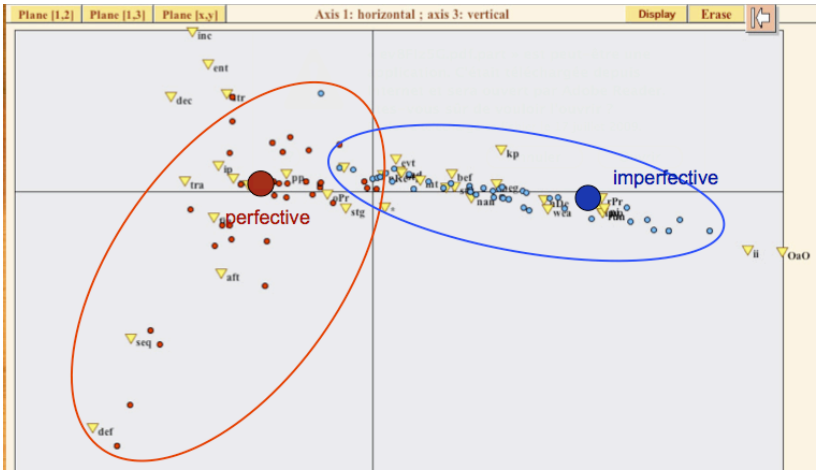

Fig.4 Two classes of values of attributes around the pefective and imperfective

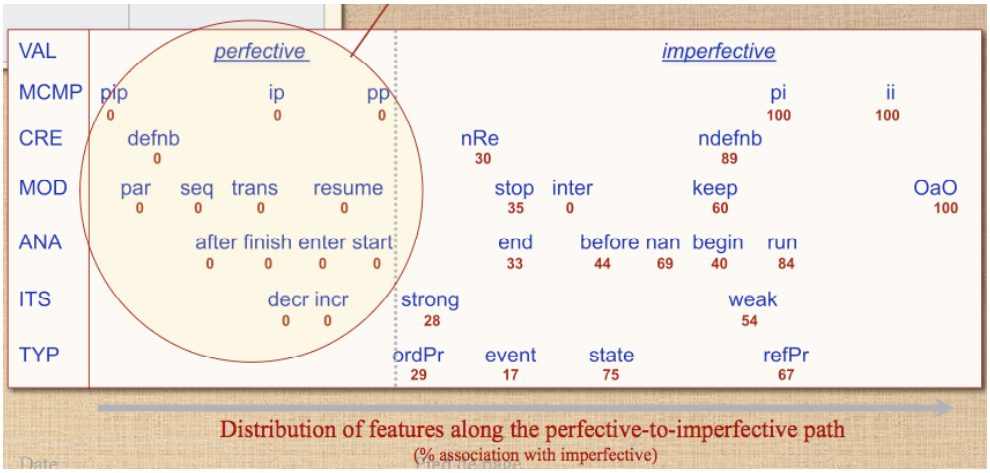

Fig.5 Polish Aspect: Correspondence factor Analysis

The distribution of features along the perfective to imperfective path in Correspondence Factor Analysis (Fig.5) shows that a group of features imperatively requires the perfective value of aspect. Among them we find the *definite number of repeated situations* opposed to

the *non definite number of repetitions* clearly situated in the imperfective zone. Three moment values of the attribute *situation analysis* are clearly associated mostly with perfective verbs: *start, enter* and *finish*; this captures the traditional view on perfective aspect as denoting either the end or the beginning of a process. On the other hand the values *nonanalysed* (nan) and the value of stage *run* of the same attribute are characteristic mostly of imperfective verbs which are know as able to feature situations as non analysed wholes or in progress in their run stage without taking into account any border moment, neither at the beginning nor at the end. As concerns situation types: ordinary processes are situated between the two zones (such type of situation may be expressed as wll by an imperfective as a perfective verb) whereas refined processes appear clearly in the imperfective zone. Among the different values of the *flow modification* attribute the *stop* and *interrupt* values are closer to the perfective whereas the *keep* and *off-and-on* values are closer to the imperfective.

These first promising results will have to be improved by collecting a larger amount of samples and defining some extra features in order to capture the nuances introduced into the perfective hypercategory by different prefixes. This task is carried on in the years 2008-2010 by a group of master students working on Polish prefixes at CELTA of Paris-Sorbonne.

7. Experimental Semantics and Language Contrasting by Alignment of Ontological Structures

With the possibility to use the techniques of knowledge discovery in databases (KDD), provided that the latter contain meta-linguistic information, our theory of the category of aspect in Polish can be seen as the first attempt of applying computational approximation-based methods in order to determine the relevance and relative importance of the semantic parameters used to model aspect. Only such detailed work with databases may be supposed to offer formal, experimentally tested and comparable cross-language definitions of semantic categories.

Our ontology-based semantic approach is appropriate for contrastive studies: the complete tree of ontological features used in different languages (language specific and universal features) can serve as intermediary comparison language (tertium comparationis). The interactive work with SEMANA consists in describing each language independently of the other(s) and exploring original text corpuses (not translations). For instance, we propose to describe aspect uses in a language L1 (collecting a database 1), obtain types (usages) defined by their feature structure (partial tree), compare these types with those obtained independently in language L2 (in database 2): language specific semantic feature structures are partial trees of the general complete ontological tree of aspect features. Computer-aided translation methods would thus consist in bringing together expressions from L1 and L2 with identical or similar feature structures. Approximation methods (RST and FCA) make it possible to compare not only totally *identical* but also *similar* feature structures.

8. References

Karolak St. (1997) “Arguments contre la distinction: aspect / modalité d’action (Aktionsart)” in *Etudes cognitives / Studia kognitywne,* Vol.2, SOW, Warszawa, 175-192.

Laskowski R. (1998) „Uwagi o znaczeniu czasowników“ in *Gramatyka współczesnego języka polskiego*, *Morfologia*, wyd. 2 zmienione, T. 1, P.W.N., Warszawa. p. 152-171.

Paduceva E. V. (1996) *Semanticeskije issledovanija*, Moskva.

Piernikarski C. (1969) *Typy opozycji aspektowych czasownika polskiego na tle słowiańskim*, Ossolineum, Wrocław.

Sauvet G. (2008) "Symbolic and statistical Analyses of meta-data using the “Semana” platform — a bundle of tools for the KDD research", CASK SORBONNE 2008 (Language Data Mining)

International conference, June, 13th-14th, 2008, Université Paris-Sorbonne – Paris 4 *http://celta.paris-sorbonne.fr/anasem/papers/*

Sémon Jean-Paul (1986) «Postojat' ou la perfectivité de congruence, définition et valeurs textuelles» *Revue des Etudes Slaves*, T. 58/4, Institut d'Etudes Slaves, Paris.

SFPW: *Słownik frekwencyjny polszczyzny współczesnej* (1990), red. Kurcz I., Lewicki A., Sambor J., Szafran K., Worończak J., PAN, Instytut Języka Polskiego, Kraków.

Vendler Z. (1967) « Verbs and Times », in Z. Vendler, *Linguistics and Philosophy,* Ithaca, New York: Cornell University Press., pp. 97-121 (revised version of Vendler Z. « Verbs and Times », *The Philosophical Review*, 66 (1957), 143-160)

Włodarczyk, A. (2003) « Les cadres des situations sémantiques », *Etudes cognitives / Studia kognitywne* V, SOW, Polish Academy of Sciences, Warszawa.

Włodarczyk, A. & Włodarczyk, H. (2001) « La Préfixation verbale en polonais I. Le statut grammatical des préfixes, II. L'Aspect perfectif comme hyper-catégorie », *Etudes cognitives / Studia kognitywne* IV, SOW, Warszawa p. 93-120.

— (2003) « Les paramètres aspectuels des situations sémantiques », *Etudes cognitives / Studia kognitywne* V, SOW, Warszawa, p. 11-34.

— (2006) « Semantic Structures of Aspect (A Cognitive Approach) », *Od fonemu do tekstu, prace dedykowane Profesorowi Romanowi Laskowskiemu*, Instytut Języka Polskiego Polskiej Akademii Nauk, Lexis, Kraków, p. 389-408.

— 2008, « The Pragmatic validation of Utterances », in *Etudes cognitives / Studia kognitywne VIII, SOW, Warszawa,* 117-128.

Włodarczyk, H. (1997) *L'Aspect verbal dans le contexte en polonais et en russe*, Institut d'Etudes Slaves, Paris, 240 p.

— (1998) « Wykładniki wartości informacyjnej wypowiedzenia w j. polskim i francuskim (aspekt, okreslonosc, modalnosc) », Congrès des Slavistes Cracovie 1998, *Revue des Études Slaves* T. LXX/1, p. 53-66, Paris.

— (2003) «L'Aspect perfectif comme hypercatégorie (approche cognitive) », communication au XIIIe congrès des slavistes à Ljubljana en août 2003, *Revue des Études Slaves* LXXIV/2-3, p.327-338 Paris.

Xrakovskij V. S. (1997) "Mul'tiplikativy i semel'faktivy (problema vidovoj pary)", *Semantika i struktura slavjanskogo vida*, red. S. Karolak, Wyd. Naukowe, Kraków.