

GRADED INFORMATIVE CONTENT OF LINGUISTIC MESSAGES

André WLODARCZYK
Université Grenoble 3, CAMS (CNRS, Paris-Sorbonne & EHESS)
Andre.Wlodarczyk@u-grenoble3.fr

Hélène WLODARCZYK
Université Paris-Sorbonne
Helene.Wlodarczyk@paris4.sorbonne.fr

Abstract: The analysis of many linguistic discourse-dependent phenomena (such as mediative, aspect, tense, determination, etc.) suggests that certain markers refer to what we consider to be various forces of informative validity of the utterance. Information may be taken as "given" or "new". Linguistic markers can be of various types (lexical and grammatical) and concern either the NP, the VP or the entire sentence. The global informative value of an utterance is the "result" of all those indices. There is practically never a bijective correspondence between linguistic indices and the scale of informative validity. We discuss from different Natural Languages.

Keywords: discourse, given and new, informative validity, plausible, implausible, potential, actual, habitual, occasional, general, particular, generic, specific, linguistic indices.

ISBN: 0 08 043 438X

0. INTRODUCTION: INFORMATION IN NATURAL LANGUAGES

Rather than attempt to imagine a new *theory* of language (this would restrict our position to an autonomous point of view with respect to the linguistic phenomena), we shall be concerned here with a *model* of understanding of linguistic messages. We must distinguish, however, between models (which abstract theoretically from Reality) and prototypes (which are experimental models).

We therefore adopt the cognitive approach. Indeed, we reject the behavioristic linguistic (language-oriented) philosophy that views language as a **black box**. Scientists claim today that "cognitive sciences start with the assumption that the analysis of observable 'intelligent' behaviours (...) should refer to mental representations." [Desclés J.-P., 1994] So, it is now emphasized that one should try to make the black box as *transparent* as possible, taking into account the capacity of a man to interact with his environment.

On the other hand, logicians emphasise on the necessity of introducing alternative possible worlds with regard to the Real World. But modal logics with «necessity» and «possibility» operators introduce so much abstraction that they are not really useful in linguistic studies. Nevertheless, the problem of the relation between linguistic messages and the reality they describe cannot be ignored.

Hypothesis # 1: Instead of atomizing the universe, we conjecture that every epistemic agent has his own representation of the world.

It is extremely difficult, however, to determine the degree of similarity between representations of two or more epistemic agents (given factors like social status, professional situation, religious beliefs, cultural background, etc.) in order to determine the background of what is known as "common knowledge".

Meaning is the content of linguistic units (words, phrases or sentences). Such units convey only partial information. We argue that partial encoding of information is totally unlike the Saussurian sign. Indeed, in our view, information has no *substitutive* character. The partiality of encoded information represents a vision (an aspect) of the World (be it "realistic" or "fictitious" for intelligent agents).

Perhaps the most difficult and, for this reason, very well-known problem that logicians encounter in their work is the problem of Truth. In our approach, truth is considered to be more closely linked to knowledge and beliefs than to Reality (or, semiotically speaking, to reference). Actually, there is no such thing as "absolute reference" (in physics there is no "absolute space" and in thermodynamics there is no *perpetuum mobile*). Information can be validated only by epistemic agents. However, even if the content of our knowledge may differ significantly (and, indeed, it does) from culture to culture or from individual to individual, control and memory management devices must have universal characteristics.

Hypothesis # 2: In linguistics, rather than on logical truth, we lay stress on the *informative validity* to which epistemic agents (the speaker and the hearer) can be attuned during the process of communication.

One of the aims of this research is to clarify the nature of information and to show how it relates to meaning and cognition. Usually, what we call information in everyday life is a recorded piece of knowledge in an encoded form. But let us start with an existing dictionary definition of information: "In a more restrictive sense, «information» is that which results when some human mental activity (observation, analysis) is successfully applied to data to reveal its meaning or significance. In this case, data is the vehicle and information is that which can result from its interpretation." (Galland, 1982).

ISBN: 0 08 043 438X

Hypothesis # 3: Information is composed of infons (fundamental structures of meaning). In order to be communicated to others (be it using linguistic messages or other ways of conveying knowledge, beliefs, experience etc.), information must be "formulated" beforehand as mental representations. ("... «situations» and «infons» [...] are not out there in the world, but rather they should be considered as mental representations" (Nakashima & Harada, 1995).

The analysis of many linguistic discourse-dependent phenomena (such as testimonial, aspect, tense, determination, etc.) suggests that certain markers refer to general informative content and to what we consider to be various forces (facets) of informative validity of judgments. In the framework of many-valued systems of logic, one can find many attempts to overcome the difficulties of truth values: anaphoric modality, degree of confidence, set-temporal truth, dimensions of truth, etc. Let us mention only two of them. (Belnap, 1977) proposed other parameters or combinations of parameters than the mere *per se* truth in order to enlarge the valuation capacities of logical formulae. From our point of view, his four-valued system of logic introduces *anaphoric* parameters into the "truth value" system. (Barth and Krabbe, 1982) distinguish between truth and validity. Validity may be of two kinds: s-validity (subjective validity) and i-validity (intersubjective validity) which are defined as follows: (1) a discourse is s-valid at a moment t if it does not entail contradiction; (2) a discourse is i-valid in relation to another discourse at a moment t if there is no contradiction between the two discourses.

We assume that the analysis of utterances should be conducted in two phases. Such treatment is in agreement with the unified theory of pragmatics by Robert Stalnaker (1970). This philosopher considers that the context determines what judgment is being uttered about and that the state of the world determines whether the given *judgment* is true or not (but see our hypothesis #2 concerning informative validity). Thus our theory takes for granted that judgment is intermediary between an utterance and its logical truth (in our approach: informative validity).

Hypothesis # 4: Information has degrees of validity; i.e., it can be evaluated in a graded manner. For this reason, the validity of information reveals its *pragmatic power*.

The power of informative validity is "oriented" (graded) and it may take one of the following values: in terms of **generalization** (plausible > potential > habitual > general > generic) and in terms of **specialization** (implausible > actual > occasional > particular > specific). See applications of this idea to the Slavic languages (Wlodarczyk Hélène 1997) and to the Japanese language (Wlodarczyk André 1982).

As judgments are *context dependent*, their meaning is intensional. The information contained in judgments can be viewed as either "given" or "new" regardless of its real support in memory (knowledge). For this reason, the epistemic agents, when communicating, may wish to lay stress on some portions of information, either they may suppose that the information is "new" (topicalization) or they consider it as "given" (focalization).

(1) **interpretation** which is defined as an *informative* operation aimed at grasping the *partial content* ("given" or "new" information) of a linguistic message as uttered in a *context* and

(2) **comprehension** which is defined as an *evaluating* operation aimed at filling empty slots or satisfying the background roles (cf. Nakashima & Harada, 1995) or deleting "unclear points" (cf. Sabah, 1993)

As we need to establish at least a partial similarity of informative contents of messages being communicated, not only shared cultural but also shared psychological conditions are necessary. In order to perform a successful act of communication the speaker must not only share some amount of "common knowledge" with the hearer, but also must be able to understand the hearer's psychological state, which depends on his individual experience (age, sex, social status etc.). Context (i.e.; speech situation) is distinguished from the representation of the world ("speech universe", i.e.: world being described). This does not suggest however that information concerning the context and information which describes the represented world may not correspond in some cases. In fact, they often do, e.g.: *I am here.*

ISBN: 0 08 043 438X

1. DEGREES OF INFORMATIVE VALIDY OF LINGUISTIC MESSAGES

Communicating agents in many cases need to specify whether the information which is being passed is to be taken as "given" (supposedly this information is shared) or "new" (supposedly the information should be shared since now on). The two kinds of information are roughly taken into account in most studies of communication pragmatics but we claim that each sort of information is further divided into five graded values. As we shall try to show, this graduation of "given" and "new" proves to be of great relevance in explaining several very different linguistic phenomena.

Table 1

"given" information	+5	Generic
	+4	General
	+3	Habitual
	+2	Potential
	+1	Plausible
(ambiguity zone)	0	no antinomy
"new" information	-1	Implausible
	-2	Actual
	-3	Occasional
	-4	Particular
	-5	Specific

The relations between discrete values of each domain are relations of gradation[1]. Values of the top of the scale (N°5) are the highest values : *genericity* is the highest degree of "givenness", *specificity* the highest degree of "newness". Going down each scale, we encouter decreasing values.

The *general* meaning can be viewed as a lower degree of the *generic* value. We understand *habituality* as a lower degree of *generality* : people first observe regularly recurring phenomena and then formulate the abstract idea that some phenomena can be considered as occuring generally. As regards the *potential* value, it can be viewed as a lower degree of informative power : the speaker does not say that something occurs generally or regularly but only that it may possibly occur. The *plausible* value means that the speaker whishes to treat the information as being supposedly already known.

The "new" domain is the domain of facts viewed as concrete events with the decreasing informative values: *specific* (concerning one unique fact), *particular*(as opposed to *general*), *occasional* (non occuring habitually), *actual* (as opposed to *potential*) and *implausible* or unheard of (i.e. never spoken about in the previous context and therefore not *plausible*).

The concept of graded relation makes it possible to capture the fact that the discrete information values of linguistic messages do not have any direct connotation in the real world. The choice of one or another neighbouring value on each scale depends on the speaker' strategy. Therefore when comparing translation from one language to another we find that, probably because of cultural background and linguistic conventions, the information value of the original message may be changed in the translation.

The lowest degrees of each scale are respectively the "plausible" value of the "given" domain and the "implausible" degree of the "new" domain. These terms are conventional names for the two opposed informative values we used to call in previous papers *anaphoric* and *cataphoric*.

ISBN: 0 08 043 438X

We decided however to change these terms because they are used in linguitics in a more technical sense (intrat-textual syntax) and therefore they could lead to misunderstanding. Nevertheless from *anaphoric* we keep the sense of referring to previously introduced information and from *cataphoric* on the other hand, the sense of introducing into the discourse information that appears for the first time and was not prepared by the context. *Implausible* thus must be understood not as surprising but only as not derivable from the beginning of the communicative exchange.

We can compare the opposition of "given" and "new" with that of Universal and Existential quantifiers in logics. The highest degree of our "given" domain is akin to the Universal quantifier, and the highest degree of the "new" domain to the Existential quantifier. But linguistic messages differ from logical propositions in that they can have different degrees of informative value. Another possible comparison is that to Gustave Guillaume's model of particularisation (our "new" domain values) and generalisation ("given") used to describe the different meanings of French articles in different contexts (Guillaume, 1964).

In our scheme, the highest values are those most distant from one another, the lowest ones are the closest. This reflects the fact that *generic* and *specific* messages differ from each other more radically than *plausible* and *implausible* messages. Therefore neutralisation can easily occur between the two lowest values : the information can be presented as *plausible* or *implausible* without changing the referent of the message in the represented world (see examples hereafter). Informative values can be seen as modulations of the linguistic message. What is traditionally called modality is only one means of expressing informative validity.

2. MESSAGES WITHOUT EXPLICIT LINGUISTIC INFORMATIVE MARKERS

Some linguistic messages do not entail any explicit linguistic marker of the informative value. The latter must be inferred from the context. The same message may be used with different informative values in different contexts.

The "plausible" value of the "given" domain can be marked only by the fact that an utterance is repeated a few times in a discourse.

#1) A. My son has caught a cold.

#2) B. Won't you go with us to the exhibition ?

#3) A. I won't go. My son has caught a cold.

Utterance #3 is made only to remind B of the situation he was first informed about in #1, but the grammatical and lexical form of #1 and #3 do not differ because the informative value of both utterances remains implicit. In a speech situation where the same utterance is repeated several times (eg Shakespeare "Brutus was an honourable man" in *Julius Ceasar*), this utterance has the lowest informative value of the "given" domain (Nr1 plausible) and it is uttered not for locutionary purpose but rather for some illocutionary or perlocutionary goal.

To give another example, the *generic* value can be marked only by the absence of any local or temporal specifier, the absence of determiners of the noun phrases and by the use of the present tense as denoting a permanent state:

#4 Whales are mammels.

Moreover the interpretation of such a message as *generic* certainly depends on the common knowledge of both speaker and hearer. In a classroom with small children, the same utterance may have the value of "new" *implausible* information.

ISBN: 0 08 043 438X

3. DIVERSITY AND COMPLEXITY OF LINGUISTIC INFORMATIVE MARKERS

In the history of linguistic research, it was difficult to formulate the concept of informative validity, because in natural languages, informative validity is very seldom directly expressed by special markers. What we treat in some utterances as informative value markers have mostly a quite different litteral meaning, be they lexical items or grammatical categories. And the same lexical or grammatical items can appear in some utterances with a referential meaning and in some others only as information value markers. These markers (indices) are of various types (lexical and grammatical) and concern either the NP, the VP or the entire sentence. It is important to emphasise that there is practically never a one-to-one correspondence between linguistic indices and the scale of informative values.

3.1. From Temporal to Informative Meaning

As concerns lexical markers, we observed that in different languages time and place adjuncts may serve as informative value markers. We have made some tests on press articles[2] where we found that such adverbs as *always, often, sometimes* give the utterance not a temporal but an informative value. In journalistic prose, the adverb *often* may be used to express the *general* value : the phenomena written about are proposed to the reader not yet as scientific rules (it would be the *generic* value) but has having a satifactory degree of generality to allow some conclusions.

> *#5 Fr. Ces jeunes ont* ***souvent*** *le sentiment d'avoir affaire à une société fermée... (MD 1996)*
>
> *(Those young people often have the impression of being confronted with a closed society...)*
>
> *#6 Pol. ... powodem większości wybuchów (poza kilkoma spowodowanymi czyimś świadomym działaniem) jest ludzką nieuwagą i niedbałość oraz stan instalacji,* ***często*** *od lat nie konserwowanych ... (Don. 96)*
>
> *(Most explosions are caused by human carelessness and by the state of fittings which very often have not been maintained for years.)*

It is obvious that in the above utterances temporal adverbs "often" play an information evaluative role and can be glosed as "in most cases".

3.2. Evaluation and Modality

But in most utterances, both lexical and grammatical markers combine to express informative values. Linguists have already paid attention to the fact that verbal modality is only part of very diverse grammatical or lexical markers that are used to express the attitude of the speaker towards the situation he is speaking about. Jean Cervoni (1987, p. 78) proposed to call this very broad semantico-pragmatic field "evaluative modality". The conditional mode of verbs is often one of the markers of the *plausible* value.

In Polish utterance #7 we find three markers of the *plausible* value : the adverb *wiadomo*, the conjunction *jakoby*, and the conditional *wiedział.*

> *#7 Pol. Jak wiadomo, Józef Oleksy zaprzecza, jakoby wiedział, że jego długoletni przyjaciel Władimir Ałganow był agentem KGB. Don. 96)*
>
> *(As it is widely known, J. O. denies he would know that his old friend V. A. was a KGB agent.)*

ISBN: 0 08 043 438X

Some languages even possess a "mediative modality" which gives the utterance the *plausible* value (cf. for Bulgarian, Guentcheva 1990), but languages lacking such grammatical category make use of lexical (expressions like *probably, it seems,* etc.) or syntactical means (embedding and conjunctions).

In natural languages the *generic* value may be expressed not only as in logical propositions by a modal word ***it is neccessary that*** *x* but also by a temporal word ***x always***... Moreover nominal quantifiers contribute to the same evaluative modulation of linguistic utterances as modal words and it has been noticed that, in natural languages, it is possible to paraphrase (of course with lesser precision than in logical propositions) utterances containing quantifiers by utterances containing modal or temporal modifiers:

It is necessary that x do y ª all x do y ª x always do y ;

It is possible that x do y ª most x do y ª x often do y ;

There exist x that do y ª some x do y ª x sometimes do y.

3.3. Degree Words

As relations between informative values are relations of gradation degree words are very important to our purpose. In an article about degree words in English and Japanese (Kamei and Muraki, 1994), authors propose to match quantifiers *all, many, some, few, no* respectively with the following adverbs *always, often, sometimes, seldom, never.* They insist on the fact that degree words do not denote real world objects (although world is perceived as a continuum natural languages make use of discrete degree concept primitves) but are heavily dependent on pragmatic factors : the speaker's attitude, situation, etc. The difference between "literal" and "conversational" meaning is very important in the field of degree words (hence the difficulty of translation).

In the following utterance[3] the degree word *almost* modulates the *generic* value of adverb *invariably* and thus ascribes to the utterance the *general* value .

#8 I had saved him from starvation of that peculiar sort that is almost invariably associated with drink. This was all . (LJ)

In #9 *almost* weakens the *habitual* value of "any day".

#9 People said the white man could be seen with her almost any day; they walked side by side. (LJ)

In the following example (#10) the use of *almost* modulates the *implausible* value of an utterance which instead of being presented as *"new"* is introduced with the *potential* value :

#10 Pol. Włodzimierz Cimoszewicz prawie na pewno będzie premierem. (Don. 96)

(It's already almost sure that Wl. C. will be prime minister.)

Without modulation, the same piece of news would be introduced as "new":*W.Ci. will be prime minister.*

ISBN: 0 08 043 438X

4. MARKERS OF THE "GIVEN" INFORMATION

Generic messages may contain the following markers: *always, everywhere* or neither local nor temporal specification, in languages with the opposition of plural/singular, the subject noun phrase is in the singular denoting a whole class. Tense is present (in its so called gnomic or permanent use)

#11 Man is mortal.

#12 Pol. ... kobieta zawsze ludziom bliższa .

(A woman is always closest to people)

The *general* informative value is marked by such expressions as *always, generally, very often* and a singular (denoting a whole class) or plural subject nominal group :

#13 That's what you English always make (LJ)

#14 'It is always the unexpected that happens,' I said (LJ)

Habitual informative value is marked by lexical expressions meaning a regular repetition and occuring as adjuncts in the syntactic structure: *usually, regularly, on Sunday, every morning ...* In Slavic languages, such adjunct noun phrases occur together with imperfective verb forms. English has a semi-auxiliary verb of habituality: *used to*, the subjunctive form of the auxiliary *would* can also have an habitual meaning.

#15 He used to tell us his experiences in that line. (LJ)

#16 ... mad Matherson they generally called him ; the same who used to hang out in Haiphong, you know, before the occupation days... (LJ)

The *potential* informative value is marked by adverbs as *possible, maybe*, auxiliary verbs as *can, may, able to* :

#17 Such a consciousness may be wrong, or it may be right, or it may be condemned as artificial ... (LJ)

#18 There may be those who could have laughed at his pertinacity. I didn't. (LJ)

The *plausible* value is the less certain type of information: the speaker is not sure of what he is asserting because he has not enough evidence for it, or he has only heard about it without being able to verify it himself. In some cases this value corresponds to what is known by hearsay and this is the reason why we first called it *anaphoric*[4]. We rejected this term because this value of the *"given"* domain does not concern only an intratextual relation between one linguistic item and another one previously mentionned. *Plausible* is a variety of information quality that is based on the discourse strategy chosen by the speaker. Information is proposed to the hearer as *plausible* when it is presented as if it had been already mentioned, or as if it were considered as being already known by the hearer (because it is a piece of common knowledge or because the speaker considers that his particular hearer should know this information).

As we shall see below grammatical categories (aspect, article, Japanese nominal particles) are involved int the expression of the opposition "given"/"new". Let us mention first a few lexical markers. Polish has several adverbs that contribute to the plausible value *chyba* (it seems), *rzekomo* (one says), *podobno* (looks like).

#19 Pol. Podobno ludowcy poparli w komisji SLD w zamian za obietnic" powolania ministra obrony z PSL. (Don. 96)

ISBN: 0 08 043 438X

(It seems that members of the Populist Party (PSL) supported in the commission the Left Democratic Union (SLD) in exchange for promises that the minister of defense would be from their party.)

In French, we find *on dit, il paraît , on raconte, le bruit court, on a entendu dire, etc.*

The importance of the plausible value is attested by the fact that in the novel of J. Conrad *Lord Jim* we found 30 occurrences of 'it seems'. We can mention as evidence that the expression *it seems* concerns the pragmatic level the fact that it is in the present tense, even when the rest of the utterance is in the past.

#20 ... it seems he had not lost his appetite though, ... (LJ)

#21 The third man in, it seems, had been Tamb' Itam, Jim's own servant. (LJ)

#22 He was no stranger to the port, it seems, (LJ)

Another marker of the plausible informative value *probably* is used 19 times in *Lord Jim.*

#23 He had been stranded out East somewhere in Canton, in Shanghai, or perhaps in Yokohama; he probably did not care to remember himself the exact locality, nor yet the cause of his shipwreck (LJ)

Plausible information can be marked by the so called mediative forms in Bulgarian "mediative forms always encode that the speaker has a mediate perception of the spatio-temporal situations spoken about" (Guentcheva, 1990, pp. 179-196).

In languages without special mediative modality, the plausible value can be marked by the embedding of an utterance in another one. The embedding conjunction can be in English *since* (not with temporal but with argumentative meaning) associated with the progressive form in its presuppositonal function. Henri Adamczewski's theory of the *be+-ing* form in English (Adamczewski 1978) is that it has essentially a role of topicalisation. In an utterance with a *be+-ing* form, the focus is not on the verb which is itself topicalised. In the fairy tale *Goldylocks andThree Bears*, the question "Who has been eating my porridge" is put when it is obvious from the situation that the porridge has been eaten. It is interesting to point out that in Slavic languages in the same context we find the imperfective aspect also with a topicalising function (giving the utterance the plausible value of the "given" domain) : eg. Russian - *Kto pil iz mojej caski ? (Who has been drinking out of my cup ?*).

Using *Lord Jim* as a corpus we found two utterances where conjunction *since* is used with a *be+-ing* form and has a plausible informative value:

#24 .. he said that he had always done his duty by them ; up to that moment ; and even now he was not betraying their confidence, ***since he was leaving the ship*** *to as competent a seaman as could be found... (LJ)*

In the following utterance there are three markers of the plausible value (*since, could, possibly*)

#25 It was very much like a meeting in a wood, only more uncertain in its issue, since he could possibly want neither my money nor my life. (LJ)

ISBN: 0 08 043 438X

5. MARKERS OF THE "NEW" INFORMATION

The domain of "new" values in natural languages appears by contrast to the "given" domain. "New" messages seem to bear less explicit marks than the "given" ones. If an utterance does not entail any restriction of the type *probably, it seems,* etc. this can be interpreted as a message the speaker wants to propose as "new" information. The *implausible* value is the least marked because its value is determined first of all by the situation (what was said before, what is the common knowledge already available). Nevertheless, we can point to a few indices. In languages with articles the first occurrence of a noun phrase is normally accompanied by the indefinite article.

#26 Fr. Tout était calme. Soudain une voiture freina très fort.

#27 It was all quiet. around. Suddenly a car braked very hard.

In Slavic languages without articles and with declension marks, the subject noun phrase with a "new" value may follow the verb as in Polish:

#28 Pol. Było cicho. Nagle zahamował samochód.

Specific or particular informative value appears when the uniqueness of the "new" situation spoken about is marked.

#29 Only once in all that time he had again the glimpse of the earnestness in the anger of the sea (LJ)

#30 The captain kept moving here and there on the bridge; he seemed calm enough, only he stumbled several times; and once as I stood speaking to him he walked right into me as though he had been stone-blind. (LJ)

#30 Once I talked for three hours to a man in Auckland. (LJ)

#31 One day, coming ashore, I saw him standing on the quay ... (LJ)

The difference between specific and particular informative value is minimal

#32 Pol. Wczoraj we Wrocławiu zakończylo sie trwajce pięć dni Europejskie Spotkanie Młodych ... (Don. 96)

(Yesterday the European Youth Meeting which lasted five days ended in Wroclaw...)

The occasional informative value can be underlined by contrast with the habitual value:

#33 He was usually anything but free with his private store of liquor; but on that night he had departed from his principles... (LJ)

The actual informative value puts the stress on the fact that the situation arises in some concrete place and time. In English the insistance on the actuality can be obtained by using auxiliary *do* in non-negative sentences.

#34 The Rajah wanted to know whether the white man could repair a watch? They did actually bring out to him a nickel clock... (LJ)

In Slavic languages, if we accept the concept of ***completed stopped process*** as the categorial definition of perfective aspect (Desclés 1989) we find many utterances with the *actual* informative value where the perfective verb points out at the result of the action. In English the only marker of the actuality of the result of action is the temporal adjunct.

ISBN: 0 08 043 438X

#35 Pol. Jan malował pokój przez godzinę.

#36 Jan painted his room during one hour.

#37 Pol. Jan pomalował pokój za godzinę.

#38 Jan painted his room in one hour.

6. GRAMMATICAL CATEGORIES AND INFORMATIVE VALUES (TWO STUDY CASES)

Most problems of interpretation of grammatical categories follow from the fact that they do not have solely categorial semantic referential meanings but also pragmatic conversational usages. **Some meanings that are generally ascribed to grammatical categories are in fact a matter of informative validity.** Terms of grammatical oppositions display complex behaviour as regards informative values of utterances (there is never bijective correspondance between categories and usages). The English progressive form, besides its categorial aspectual meaning, can be used as a marker of topicalisation.

6.1. Japanese Particles wa *and* ga [5]

For modern linguists, *wa* and *ga* particles constitute one of the most interesting and arduous problems of Japanese grammar. Particles *wa* and *ga* cannot be explained properly without taking into account the other particles that belong to the same classes of morphemes that *ga* and *wa* represent; i.e., case particles (kaku-joshi) and "concordance" particles (kakari-joshi), we shall rename the former as *argument* particles and the latter as *element* particles within the framework of our theory.

One should not forget that *ga* particle can refer not only to the subject function (which is not obligatory in Japanese) but also to many other syntactic functions (such as the object or the location) and *wa* particle, when attached to a subject constituent, is not always a topicalisation marker. As a matter of fact, we can observe the same opposition between *wa* and *ga* when they follow subject or object phrases and when they affect other phrases. Moreover, their meaning is closely related to their position in the sentence.

In our approach, we argue that *wa* and *ga* particles present historically motivated ambiguities, and that these ambiguities can be explained as the result of a *boomerang* relation of the mentioned classes of morphemes. This kind of relation is unknown in Structural Linguistics. Nevertheless, there is much evidence that in Japanese Grammar such a relation (which could possibly be defined as a *double privative relation)* should be added to the realm of "oppositions" that proved to be so useful in language studies.

Set-theoretical and Predicative Identity. We claim that in order to explain all these different uses of particles *wa* and *ga* we have to take into account that the latter classes of particles can mark two kinds of logical relations in different sentence positions (not only in post-nominal positions). These relations are of set-theoretical and predicative types. However, we do not take for granted that logical relations have their equivalents or one-to-one mappings in any Human Language.

Let us consider the two kinds of identity : (1) a is b. **p(a)**, (predicative identity of a with respect to b), where p = "is b" and (2) a belongs to A. **(a ∈ A)** ; i.e.: set-theoretical identity of a.

If we want to formulate both at the same time, we must consider that there are two different orders in sentences : actual (explicit) and virtual (implicit). These orders are called, in classical structural linguistics (F. de Saussure), syntagmatic and paradigmatic axis.

An attempt to formulate the two orders at once would look as follows :

ISBN: 0 08 043 438X

p(a ∈ A) where p may be the copula "to be" as a predicative relator (is_a) or any mono- or polyvalued predicate.

The meaning of such a formula would be something like this: *"a taken as an element of the set* A *is a...."*.

Example: *John is tall.* (John seen as an element of any class to which he may belong).

In classical logic (First Order Logic), this corresponds roughly to the so-called "Restricted Quantifier" whose meaning Q is indexed by x ∈ A where A is a domain that restricts the value of the quantified element x. The result of such a restriction is actually that of the abstraction operator, namely {x; p(x)} which describes all the elements x of A which satisfy the predicate p(x).

Such kind of "twofold" logical relations are characteristic of Japanese utterances. Let us consider the traditional concepts of Subject and Topic again : Subject and Topic are special cases of each of the above mentioned identities: "predicative" identity of an argument for Subject and set-theoretical identity of an element for Topic. Furthermore, Subject and Topic are often associated in speech processes as it is the case in Japanese. Let us mention that precisely for this reason it is extremely difficult to analyse the properties of *wa* and *ga* particles.

(1) First let us analyse the Japanese sentence *"A wa B da"*. Its English equivalent is *A is B*. But as a matter of fact, if we take into account all its nuances, it is possible to translate the above sentence in four different ways :

1. A exists as B.
2. A exists as being B.
3. As for A, it exists as B.
4. As for A, it exists as being B.

Versions 1 and 3 are felt as more usual ("natural") than versions 2 and 4. The unusual character of the latter versions can be explained by the fact that the copula *da* is still today replaced in written language by **de aru** (*de* "being" and *aru* "(to) exist"). Let us recall especially that **de** can be followed not only by the verb **aru** but also by a few different particles such as *wa, mo, koso, sae, dake* etc.

> *#39 O-tôsan ga go-byôki* ***de*** *wa, iro-iro to shimpai na koto deshô.*
>
> *(Your father being ill, you must be very much worried.)*

On the other hand, if we compare both versions 1 and 2 to versions 3 and 4, some Japanese native speakers might prefer the former couple. The reason is that one would rather use such sentences in speech.

(2) In order to understand the specificity of the Japanese predicative structure, let us consider the virtual equivalents of the English reading of the same logical formula *A is B*.

1. A wa B da.
2. A wa B de aru.
3. A ga B da.
4. A ga B de aru.

Two particles *wa* and *ga* occur alternatively in the four above sentences and it is necessary to distinguish their different values to explain the Japanese predicative structure "A wa B da" sentence. Before we do this, we must add that sentences 1 and 2 differ from 3 and 4 in that the former are considered "natural" whereas the latter may be used only in a specific context (emphasis or exhaustive enumeration).

ISBN: 0 08 043 438X

Let us first concentrate on the nominal phrase containing A. At first sight, nothing allows us to compare *wa* and *ga* with English articles *the* and *a* since we do not use "the A" or "an A" in the English reading of the formula. But we have to point out at a few similarities between some features of Japanese particles and those of articles in European languages. Both nominal and verbal phrases may be "given" (supposed/intended to be known to the addressee) or both may be "new" (supposed/intended to be unknown to the addressee). In the first case, Japanese speakers use version 1 and in the second case, version 2.

On the other hand, the nominal phrase containing A may be considered as denoting some "given" information (A wa) or "new" information (A ga) while the verbal phrase is supposed to contrast with the above denoting the opposite kind of information (respectively "new" and "given"). Because of this contrast, the Japanese versions correspond in English to a topicalised utterance (As for A, it is B) or to an emphasised one (It is A that is B).

Element Particles (wa, mo, koso, sae). First of all we must describe the set of element particles to which at least one of the two particles (wa) belongs. Generally speaking, the element particles are markers of *absolute and relative identity* in the set-theoretical sense. The figure below shows how some of these particles can be classified according to the criterion of belonging of a chosen element to a virtual set.

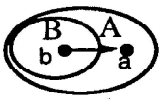
	absolute identity	**relative identity**
simple belonging	**a** wa  *indication*	**b** mo  *comparison*
complex belonging	**a** koso  *insistence*	**b** sae  *concession*

Fig.1 Particles marking identity of an element belonging to a set or to a subset

1) **wa** is the marker of belonging of the element a to the set A; i.e.: aRa, ($a \in A$)

2) **mo** is the marker of belonging of the element b to the set A, this belonging being established with respect to a which is another element belonging to the same set A; i.e.: bRa, ($b \in A$) & ($a \in A$)

3) **koso** is the marker of belonging of the element a to the set A, the latter set A being a subset of B; i.e.: aRa, ($a \in A$) & ($A \subseteq B$)

4) **sae** is the marker of belonging of the element b to the set B, the set B being a subset of A and the identity of b is established by contrast with a; i.e.: bRa, ($b \in B$) & ($a \in A$) & ($B \subseteq A$)

NB: In addition, **wa** and **koso** particles are markers of reflexive relations.

We have used here what has been called "attributive relations" (cf. Desclés 1987); i.e.: (a) as a relation of belonging and (b) as the subset relationship. As a matter of fact, the attributive relations can be seen as the result of a kind of *predicative projection* between elements of a virtual (paradigmatic) axis such as a Subject and those of the actual (syntagmatic) axis such as the attribute part of the Predicate. Consider the sentence where element particles may occur alternatively:

ISBN: 0 08 043 438X

wa as a marker of identity

Tori wa naku. (*wa* has no equivalent in English) /Birds sing./- "Birds belong to the set of beings that can execute the action of singing"

mo as a marker of comparison

Tori mo naku. (*mo* corresponds to "also", "even" etc.) /Birds also sing./ - "Birds belong to the set of beings that can execute the action of singing, this belonging is established with respect to another beings that have the same property"

koso as a marker of insistence

Tori koso naku. (*koso* corresponds to "exactly," "just" etc.) /It is birds that sing./ "Birds belong to the set of beings that can execute the action of singing, this set is seen as a subset of another set."

sae as a marker of concession

Tori sae naku. (*sae* corresponds to "even", "also" etc.) /Even birds sing./ "Birds belong to the set of beings that can execute the action of singing, this set is a superset of another set and the identity of birds is established by contrast with that of beings belonging to the subset.

Sentence as a Unit of Meaning. Reference to the World. Referring to the World (or more adequately to its mental representations) can be seen as the *descriptive* function (categorial meanings) of the language, which is perhaps its most important role (yet the most unclear), because human beings communicate their knowledge and beliefs. In this respect, the Japanese *wa* and *ga* particles play roles closely related to those of definite and indefinite articles because (a) *wa* and *ga* particles as well as (b) definite and indefinite articles co-occur in such basic operations like 'quantification' and 'determination' .

1) When both Subject and Predicate are given, the*wa* particle attached to the Subject displays the feature *+definite.*

#40 Sakura no hana wa taihen utsukushii desu.

(Cherry blossoms are very nice.)

2) When both Subject and Predicate are "new", the*ga* particle attached to the Subject displays the feature *-definite.*

#41 O-niwa no sakura no hana ga kirei desu ne.

(Cherry blossoms in your garden are beautiful, aren't they ?)

The difference between the definiteness of European languages and that of Japanese lies in the mark. While in European languages both definite (English: the, French: le/la/les, German: der/die/das) and indefinite (English: a, French: un/une/des, German: ein/eine) articles are marked (+), in Japanese only wa is marked (+definite), ga being unmarked (-definite). This difference is important, because the unmarkedness of *ga* particle sometimes causes ambiguity.

#42 Tarô ga kita. *(1) Taro came. (2) It's Taro who came.*

It is precisely because the particles *wa/ga* can be used in opposition to each other with the meaning of definiteness that Japanese logicians sometimes mention them when talking about the logical concepts of Universal and Existential quantifiers.

ISBN: 0 08 043 438X

Reference to the Discourse. Let us consider first that the meaning of the message may be either totally old ("given") or totally "new". In Japanese, *wa* as opposed to its counterpart which in this respect is ga for the noun phrase and the verbal auxiliary form -ru as opposed to -te iru for the verb phrase, can be used as markers of "given" and "new" respectively for each pair of *Noun & Verb* phrases. For instance:

Subject ("given" information) + Predicate ("given" information)

Ame *wa* fu*ru*. (lit. Rain falls = Rain is an atmospheric phenomenon)

Subject ("new" information) + Predicate ("new" information)

Ame *ga* fu*te iru*. (It is raining [now].)

However, we must emphasise that the main function of*wa* is not to denote "given information", nor is the role of*ga* to denote "new" information. As a matter of fact, the primary function of*wa* is related to the functions that are played by the above mentioned element particles such as: *mo, koso, sae, dake, bakari* etc. We argue that the proper categorial meaning of these particles today has nothing to do with their traditional description as element relations (kakari-musubi). These particles can follow argument particles such as (*wo, ni, de, kara, yori, made* etc.). The only exception[6] here is the particle *ga*. As we have already shown, the main function of element particles consists in marking set-theoretical relation on the paradigmatic (virtual) axis.

According to traditional ontology, the distinction Generic/Specific indicates the belonging not only of objects to classes but also to the properties viewed as classes. In our opinion, it is preferable to consider the above distinction as a discrete simplification of the continuum of values contained between two poles (Generic and Specific) with a common point of antinomies (cf. table N°1). This common point (Zone 0 in the table) makes it possible to explain the ambiguity of*wa* and *ga* . When referring to Discourse, the role of these particles is to mark the type of information (*"new"* or *given*).

Notice that the informative validity increases/decreases according to a well-defined order. This kind of reference is indirect in the case of the Japanese wa/ga particles. In general, such is also the case of some other morphemes (tenses, aspects etc.). In Japanese, however, there are many morphemes which are specialised in referring to the Discourse (ex. sentence final particles: *yo, ne, zo, kanaa* etc.)

According to the statistical information (cf. Hayashi et al., 1982) about the usage of wa and ga particles in Japanese press, *wa* often occurs in texts about politics, while *ga* often occurs in texts about social life. From our point of view, this observation suggests that political subjects are aimed at providing mostly *"given"* information (i.e.: assumed to be obvious to everybody), and that social texts indicate first of all *"new"* events (i.e.: instances that were not foreseen).

6.2. Verbal Aspect in Slavic languages[7]

It is important to emphasize that one should not confuse the categorial meaning of verbal aspect and its textual values, which depend on the informative value of the utterance as a whole. In Slavic languages, it is perhaps easier to point at the textual values of verbal aspect because in these languages this category can occur separately from tense.

From our observations of several Slavic languages, we could conclude that perfective verbs occur mostly in "new" utterances, whereas imperfective verbs rather suit "given" utterances. Nevertheless, in its neutralised use (as the non-marked term of the aspectual privative opposition) the imperfective aspect can appear in "new" utterances (with actual durative or conative values) or just to avoid the categorial meaning of the perfective (limit or result). On the

ISBN: 0 08 043 438X

other hand, in stylistically marked utterances, a metaphorical use of the perfective aspect (as the marked term) occurs in "given" utterances (general value in proverbs, expressive potential value, sporadic repetition) . The metaphorical use consists in using a category in a context, which is in contradiction with its typical meanings. Thus, the relations between the two members of the aspectual category and the informative values of utterances are anything but simple as it appears in table N°2.

Table 2: Categorial Meaning and Informative Values of Perfective and Imperfective Verbs (in Russian and Polish)

	PERFECTIVE	IMPERFECTIVE
Categorial meaning (Out of context)	*marked:* limit, result	*non marked*: neutral
1. Typical Contextual Meanings	**specific** **particular** **occasional** **actual** **implausible**	**generic** **general** **habitual** **potential** **plausible**
2. Non Typical Contextual Meanings	*metaphorical meanings* **general** (in proverbs) **habitual** (sporadic repetition) **potential** (expressive potential)	*neutralised meanings* **particular, occasional and implausible** (without result) **actual** (without limit : duration) (without result: conation)

Just to give a very brief explanation of this table, lets us recall that perfective verbs are used to introduce for the first time new situations, and therefore are said to confer a dynamic character to the narration. Imperfective verbs, on the other hand, can occur in generic utterances:

#43 Pol. Ryby oddychają skrzelami.

(Fish breathe with their gills.)

The plausible value of the imperfective occurs when an imperfective verb is used as a piece of argumentation and the information conveyed by the verb is considered as already known by the hearer. In a novel by Witold Gombrowicz *Murder with Premeditation* , the person who is leading the investigation about the death of the father of the family asks the son, whether he has locked the door of his father's room on the night of his death.

#44 Pol. Czy może będzie pan utrzymywał, ze pan nie zamykał ?

- Zamknąłem!

- A dlaczego, ***w jakim celu pan zamykał?*** [8]

(Will you perhaps pretend that you did not lock the door?

- I did close it.

- But why, for what purpose did you lock it?)

ISBN: 0 08 043 438X

As an example of the complex relation between aspect and the "given"/"new" opposition, we can point at ritual situations where speakers of Slavic langages can formulate commands either in the perfective or imperfective aspect depending on cultural factors. In Russian, to be polite, one says rather *sit down please* in the imperfective

#45 Rus. Sadites' pozalujsta!

with the *plausible* value (it is not a real order but just what the speaker is expected to pronounce), whereas in Polish to be polite and formal you avoid the plausible value of the imperfective (that would be interpreted as too familiar) and choose the *new implausible* value of the perfective:

#46 Pol. Proszę, niech pan siądzie.

Moreover, in North Slavic languages there exists a special grammatical class of motion verbs that possess two different imperfectives : one of them appears mostly in utterances with the "given" informative values (so called undetermined imperfective, eg. rus. *xodit'*, to walk or go) the other one (so called determined imperfective, eg. rus. *idti*, to walk or go) appears with the "new" informative values.

7. CONCLUSION

Our approach provides a basis for comparing utterances with similar global informative validity in different languages. For instance, what is expressed in the noun phrase in languages with articles has to be transferred to the verbal phrase in other languages. The "given"/"new" opposition is reflected in the use of articles in French. In most utterances, the indefinite article is an index of "new" values (very often: particular and implausible) but it can also occur in generic utterances. On the other hand, the definite article is fit for generic and plausible (traditionally "anaphoric") values but can also mark the particular value in appropriate contexts. As concerns the contrast between French and Slavic languages, the habitual or plausible values can be marked by the imperfective in Polish and by the definite article in French. For instance, in the following utterances, habituality is marked in French by the definite article and in Polish by the imperfective undetermined motion verb:

#47 Fr. Le samedi, nous allons au cinéma.

#48 Pol. W sobotę chodzimy do kina.

(On Saturday we go to the movies.)

In different languages, the choice of one or another neighbouring value on each scale of informative values depends on the speaker's strategy. Therefore, when comparing an original text with its translation (even between languages of the same family as Slavic languages) we find that probably because of cultural background and linguistic conventions, the information value of the original message may be slightly changed (cf. examples in Polish and Russian in [Wlodarczyk H., 1997]).

The problem we have to solve is how to evaluate the global informative value of an utterance. A complex procedure is required to calculate the relative weight of different markers in an utterance. Different indices may have contradictory values. The subject noun phrase may be particular, whereas the modality of the whole utterance is plausible as in #49.

 # 49 A student asked for you, it seems.

ISBN: 0 08 043 438X

Modal (verbal or lexical) markers seem to have a larger scope than nominal ones. Embedding structures impose their informative value on embedded ones.

The ultimate meaning of utterances is evaluated information on «situations» (traditionally called "speech universe") which is **a binary partial function from Contexts and Default Knowledge and Beliefs into validity degrees.** Let us mention also that the appropriate formalism for research in the direction described above is founded on the formalism which constitutes the basis of Extended Situation Semantics [Nakashima H. & Harada Y., 1995].

8. REFERENCES

Adamczewski Henri (1978) *Be+ing dans la grammaire de l'anglais contemporain*, Thèse d'Etat Paris VII, Atelier repr. thèses Lille III, Diff. Libr. Honoré Champion, 7, Quai Malaquais, Paris.

Barth E., Krabbe E. W. (1982) From Action to Dialog, De Gruyter, Berlin

Belnap Nuel D. Jr. (1977) A useful four-valued logic, *Modern uses of multiple-valued logic*, edited by M. Dunn and E. Epstein, pp. 5-37; Dordrecht: Reidel

Cervoni Jean (1987) *L'Enonciation*, PUF, Paris.

Desclés Jean-Pierre (1987) Réseaux sémantiques, *Langage* N° 87, Éditions Larousse, Paris.

Desclés Jean-Pierre (1989) State, Event, Process and Topology, *General Linguistics* **vol. 29 Nr 3**, The Pennsylvania State University Press, University Park and London.

Desclés Jean-Pierre (1994) *Études Cognitives*, **vol. 1**, p. 11, Slawistyczny Osrodek Wydawniczy, Warsaw

Galland Frank J. (1982) *Dictionary of Computing*, John Wiley & Sons, Chichester - New York - Brisbane - Toronto - Singapore

Guentcheva Zlatka (1990) L'énonciation médiatisée en bulgare, *Revue des Etudes Slaves* T.62/1-2, 179-196, Paris.

Guillaume Gustave, *Langage et science du langage,* Paris-Québec 1964.

Hayashi Ooki et al. (1982) *Zusetsu Nihongo - Gurafu de miru kotoba no sugata* , Statistical data from Kokuritsu Kokugo Kenkyûjo (The Japanese Language Research Institute), Kadogawa-shoten, Tokyo.

Kamei Shin-ichiro and Muraki Kazunori (1994) A Discrete Model of Degree Concept in natural language, *Coling 1994 Proceedings* **vol. II**, 775-781, Kyoto.

Milward David (1996) Integrating Situations into a Theory of Discourse Anaphora

Nakashima Hideyuki and Harada Yasunari (1995) Situated Disambiguation with Properly Specified Representation, *Semantic Ambiguity and Underspecification*, Kees van Deemer and Stanley Peters editors, CSLI Publications

Nęcki Zbigniew (1966) *Komunikacja międzyludzka*, Wydawnicwto PSB, Kraków

Sabah Gérard (1993) Vers une Conscience artificielle ?, *Modèles et concepts pour la science cognitive*, hommage à J.-F. Ny, PUG

Stalnaker Robert C. (1970) Pragmatics, in *Synthese*, **22**, Reidel B.V. (Dordrecht), pp. 272-289

Tokarz Marek (1992) *Elementy pragmatyki logicznej*, Wydawnictwo Naukowe PWN, Warszawa

Wlodarczyk André (1980) Syudai kara syugo e, sosite syugo kara syudai e - *wa* to *ga* (From Subject to Topic and from Topic to Subject - *wa* and *ga*). In: *GENGO ("Language" monthly review)*, **vol. 9**, No. 8/80, Editions Taisyuukan, Tokyo.

Wlodarczyk André (1982) Entre le thème et le sujet - wa et ga, in *Travaux de linguistique japonaise*, **vol. 6**, Université de Paris VII, Paris.

Wlodarczyk André (1996) La Thématisation en japonais et l'article français, in *Semantyka a konfrontacja językowa*, ed. Koseska-Toszewa V. and Rytel-Kuc D., Editions Slawistyczny Ośrodek Wydawniczy, 131-140, Varsovie

Wlodarczyk Hélène (1994) "L'Aspect verbal slave et les domaines du donné et du nouveau", *Etudes cognitives* **vol.1** *Sémantique des catégories d'aspect et de temps*, ed. Desclés J.-P., Guentcheva Z., Karolak S., Koseska-toszewa V., PAN, Inst. Slawistyki, SOW, 113-130, Warszawa.

ISBN: 0 08 043 438X

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

9. CORPUS

(Don. 96) *Donosy, Dziennik Liberalny*, 1996, Red. L. Białkowska, Donosy-Redakcja@fuw.edu.pl
Gombrowicz Witold, Zbrodnia z premedytacją in *Bakakaj*, Wyd. Literackie, Kraków 1993.
(LJ) Conrad Joseph, *Lord Jim* , Oxford Text Archive, archive@vax.oxford.ac.uk
(MD)*Le Monde Diplomatique* 1989-1996, CEDROM-SNI, Québec, www.cedrom-sni.qc.ca

---

[1] as defined first by N. S. Troubetzkoy for phonology in 1939.

[2] *Le Monde Diplomatique* (MD), *Donosy, Dziennik Liberalny,* year 1996. (Don. 96)

[3] Conrad Joseph, Lord Jim (LJ).

[4] Indeed the necessity to accept a broader concept of anaphora is put forward in most recent work. David Milward (1996) stresses the necessity of taking the situation into account in order to resolve anaphora in a text. For instyance, to understand the message *I got home late last night since the bus was ten minutes late*, "we use world knowledge that a bus is likely to be involved in someone getting home" since the bus has no linguistic antecedent in the text.

[5] For a more exhaustive presentation, cf. (Wlodarczyk André, 1980, 1982, 1996)

[6] This problem is discussed in (Wlodarczyk A. 1980, 1982, 1996).

[7] This presentation will be very brief, for more details cf. (Wlodarczyk H., 1997)

[8] Gombrowicz 1993 pp. 47-54.

ISBN: 0 08 043 438X