From the UNL hypergraph to GETA's multilevel tree

Etienne BLANC
GETA, CLIPS-IMAG
BP 53, F-38041 Grenoble cedex 09
Etienne.Blanc@imag.fr

Abstract

In the UNL project of multilingual personal networking communication, the representation of a text is a hypergraph were nodes bearing "Universal words" are interconnected through arcs bearing semantic relations. In order to "deconvert" such a graph into a French text, we use the transfer and generation techniques of the Ariane system, where the inner representation of a text is a decorated tree. We compare both representations, and discuss the transfer from UNL graphs to GETA multilevel trees.

1 Introduction

The UNL (Universal Networking Language) project is a project of personal multilingual high-quality communication over the Internet initiated and coordinated by the Institute of Advanced Studies of the United Nations University in Tokyō. It is based on a pivot language, the Universal Networking Language.

Fourteen languages are presently tackled: Arabic, Chinese, French, German, Hindi, Indonesian, Italian, Japanese, Latvian, Mongolian, Portuguese, Russian, Spanish, Thaï. Each partner of the project develops a "Deconverter" (producing natural language texts from UNL graphs) and an "Enconverter" (producing UNL graphs from natural language texts) either using tools provided by the UNL center (UNU/IAS) or using his own tools.

The aim of this paper is to give a brief insight into the French part of the UNL project, with a particular stress on the general problem we were faced to, namely the transfer between two deep representations of a given text: the UNL hypergraph and Geta's multilevel tree.

2 The Universal Networking Language

In the Universal Networking Language, a text is represented as a hypergraph where nodes, bearing "Universal Words" (UW), are linked by directed arcs bearing "Relations Labels". The graph of Fig 1 is a representation of the English sentence "The cat catches a grey mouse in the attic".

The UW of a node, like *cat(icl>animal)* in Fig.1 is an English word *(cat)* followed by a list of constraints *(icl>animal)* specifying the word sense, and may be considered as an interlingual acception. A UW is generally followed in the graph by a list of attributes, which provide information about how the concept is used in the given sentence. The attributes *@entry.@pres* following the UW *catch* in Fig.1 express that the corresponding node is the entry point of the graph, and that the process *catch* takes place at the present time with respect to the speaker.

```
agt(catch.@entry.@pres,cat(icl>animal).@def)
obj(catch.@entry.@pres,mouse(icl>animal).@indef)
plc(catch.@entry.@pres,attic.@def)
mod(mouse(icl>animal).@indef,grey(icl>color)
```

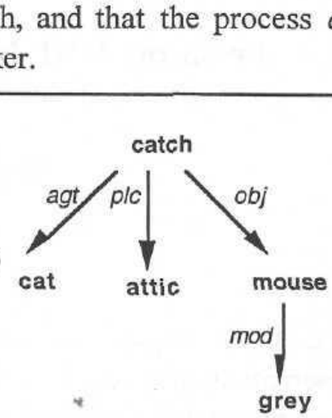


Fig.1 An elementary UNL graph.

A binary relation (conceptual relation) connecting two nodes of the graph denotes a deep case relation between these two nodes: in the graph of Fig.1, the relation *agt* expresses that the process *catch.@entry.@pres* is initiated by *cat(icl>animal).@def.* The set of the conceptual relations used in UNL is given in appendix at the end of the paper.

A detailed description of the Universal Networking Language and of theUNL project may be found on the Web site http://www.unl.ias.edu.

3 GETA's Ariane MT transfer system

The French deconverter is based on Geta's Ariane-G5 .

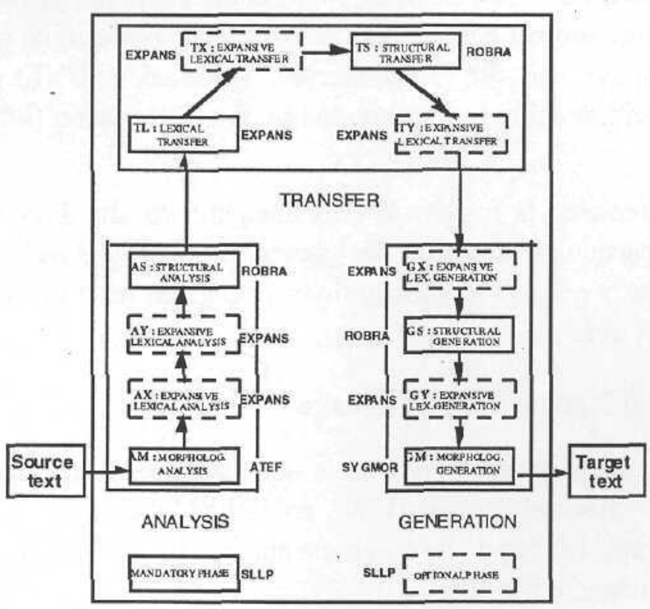


Fig.2 The Ariane-G5 system

Ariane-G5 is an environment for the development of MT systems. It is particularly well designed for the achievement of MT systems based on the transfer approach: the translation process is performed in three successive steps (analysis, transfer and generation), each step consisting in several lexical and structural phases. The lingware is written by the linguist using Specialized Languages for Linguistic Programming (Fig.2).

The unit of translation (which is not restricted to a sentence, but may include several paragraphs) is processed as a decorated "multilevel" tree. The name "multilevel tree" means that the complete representation of a sentence, as obtained after the analysis phase, is a tree bearing the three level of information: syntagmatic, syntactic and logico-semantic, the geometry of the tree being the syntagmatic one.

For the transfer from the source to the target language, only the deep level logico-semantic information of the multilevel tree supplied by the analyser is normally used to build the multilevel tree in the target language (the other levels being available as a "safety net"). Fig. 3 shows a set of such logico-semantic relations used in Ariane.

Fig.4 is for instance the result of the analysis of the English sentence "The cat catches a grey mouse in the attic". For the sake of clarity, we kept here only the deep information of the multilevel tree, which is the main information relevant for the subsequent processing of the text (beside of course lexical information, and attributes as tense, number, which we omitted too here). This tree is very similar to a dependency tree, the main difference being that each node is split into a couple, the "syntagmatic" node (VCL for verbal clause, NP for verbal phrase, AP for adjectival phrase), bearing the information relative to the group, and the "governor" node, bearing the information relative to the word itself.

Semantic relations:	
CAUSE, SOURCE, EMIT, COND ,AIM, FINAL, RECEPT, RESULT, CONSEQ, TOPIC, TOY, INST, ACCOMP, ANALOG, CONCES, QFIER, QOBJ, PROX, LOCAL, QUAL, QLOBJ, ID, GRP.	
The 4 kinds of LOCAL relations	The argumentary relations:
UBI, QUO, UNDE, QUA	ARG0 ,ARG1 ,ARG2

Fig. 3 Example of a set of semantic and argumentary relations used in Ariane.

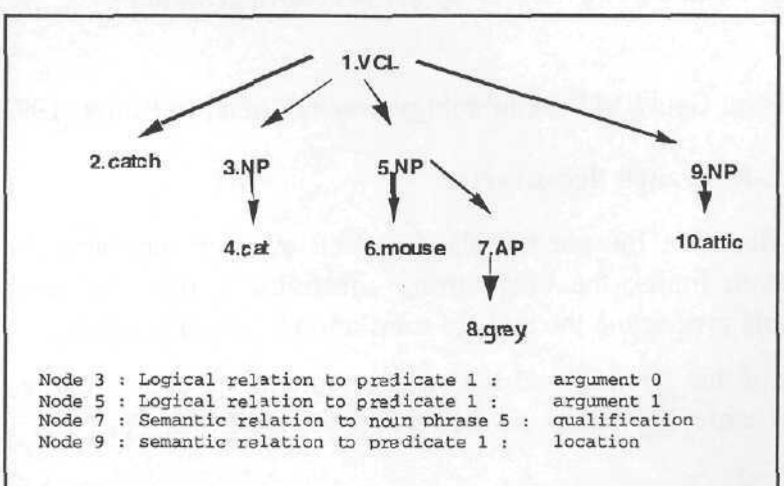
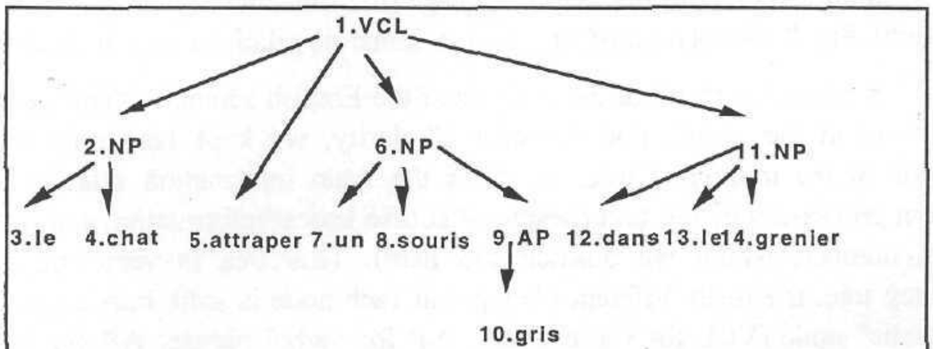


Fig.4 The deep level information in the decorated tree after the analysis and transfer steps in a classical English-French transfer system build on Ariane

Starting from this deep representation of the text, the structural generation step elaborates the "multilevel tree" of Fig.5, where all levels of representation are calculated. The decoration of each node is given in the lower part of the figure. The first column gives the French lemma, the second one the syntagmatic group, the third one the logico-semantic relation (of course the same as in Fig.5), and the last column gives the syntactic function (SUBJ: subject, DES: designator, OBJ1: first object, ATN: noun attribute, CIRC: adverbial phrase). The lexical and actualization attributes were again omitted here. The leaves of the tree represent the ordered sequence of words to be actualized by the morphological generation step (endings, elisions...).



1		K(PHVB)		
2		K(NP)	RL(ARG0),	FS(SUBJ)
3	UL(LE)	CAT(DET)		FS(DES),
4	UL('CHAT'),	CAT(CATN)		FS(GOV),
5	UL('ATTRAPER'),	CAT(CATV)		FS(GOV
6		K(NP),	RL(ARG1),	FS(OBJ1)
7	UL('UN')	CAT(DET)		FS(DES),
8	UL('SOURIS'),	CAT(N),		FS(GOV),
9		K(AP),	RS(QUAL),	FS(ATN).
10	UL(GRIS),	CAT(CATADJ),		FS(GOV),
11		K(NP),	RS(LOC),	FS(CIRC),
12	UL('DANS')	CAT(PREP)		FS(REG).
13	UL('LE')	CAT(DET)		FS(DES),
14	UL('GRENIER')	CAT(CATN)		FS(GOV),

Fig.5 The decorated tree produced by the structural generation

More details on Geta's MT methodology may be found in Boitet (1997).

4 The UNL-to-French deconverter

Ariane-G5 does not impose the classical MT transfer approach. Apart from some implementation limits, the only strong constraint is that the structures used for describing and processing the units of translation be decorated trees.

Making use of this flexibility of the Ariane system, we could build a UNL-to-French deconverter achieving the deconversion in the two following steps:

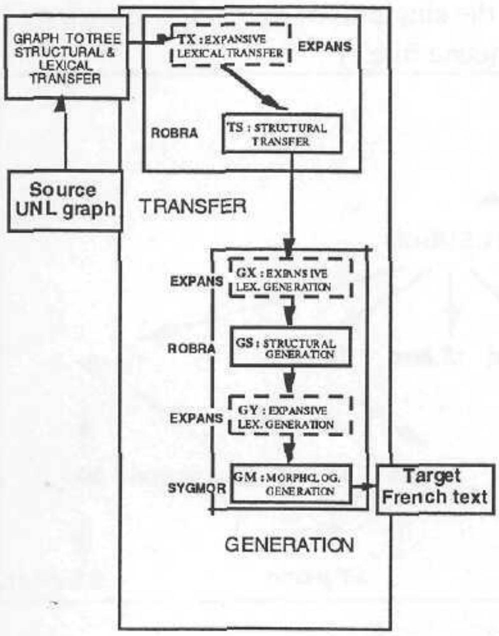


Fig.6 Principle of the Ariane based French deconverter.

1) - Structural and lexical transfer from the UNL graph to the internal representation used in Ariane for classical MT transfer systems (a tree like the one of Fig.4)

2) - Generation of the French text by means of a large French generator we had at hand.

As shown in Fig.6, the transfer step is performed part by a module external to Ariane (preliminary structural transfer transforming the graph structure into a tree suitable to the Ariane processing, followed by a lexical transfer using a UNL to French dictionary), part by a transfer module internal to Ariane (end of the structural transfer and transition from the UNL variables (semantic binary relations and nodes attributes) to the Ariane variables.

Such a transfer from a pivot language to a deep level representation raises non trivial questions. We will illustrate on a few examples the structural and semantic aspects of this transfer. A forthcoming paper will be devoted to its lexical aspect.

5 Structural transfer

5.1 Duplication of nodes

The main structural difference between a UNL graph like the one of Fig 7 (representing the sentence "The old man cherished the girl as if she were his daughter") and a Ariane tree is that a graph node may have several father nodes (presence of cycles) whereas a tree node has one father node only.

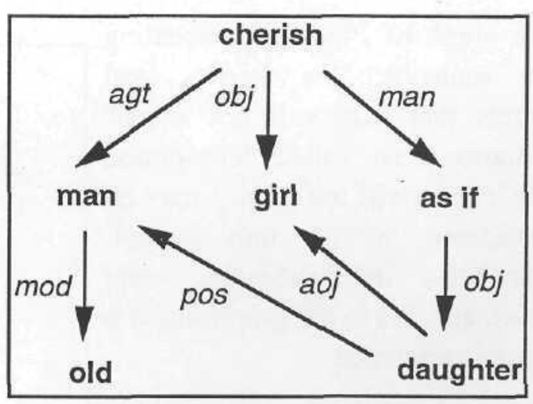


Fig.7 A UNL graph with cycles

In order to transfer such a graph into a tree (Fig.8), we duplicate the nodes having several father nodes, keeping the information that these duplicate nodes are in fact individual ones. As a result, the pronouns, implicit in the graph, may be restored in French. (The French output sentence reads "Le vieil homme chérissait la jeune fille comme si elle était sa fille" where elle is a pronoun representing the noun phrase "jeune fille", and sa is the possessive adjective representing the noun phrase "vieil homme". (Let's point out by the way that the single node "girl" was transferred into a tree describing the complex French word "jeune fille".)

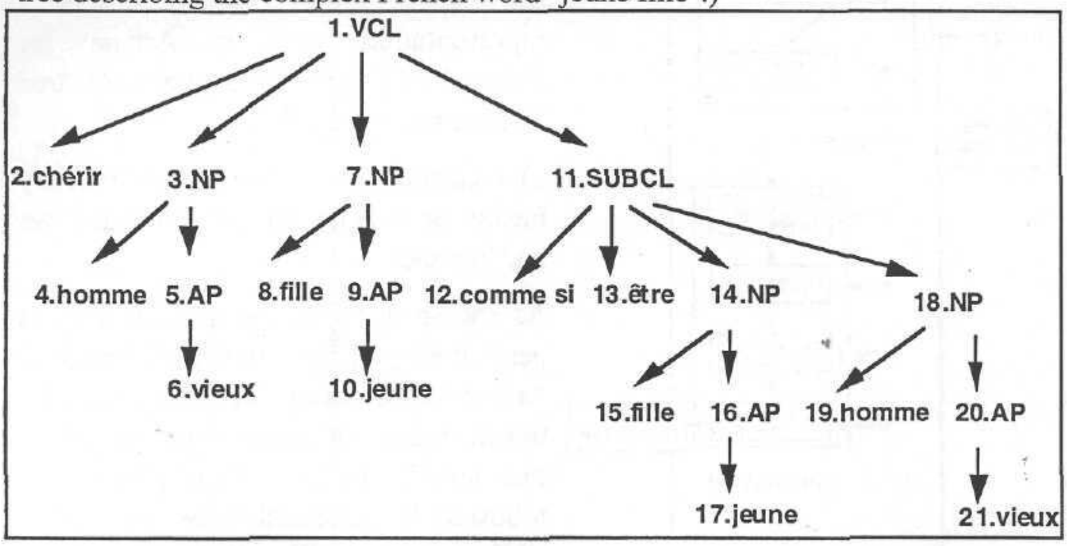


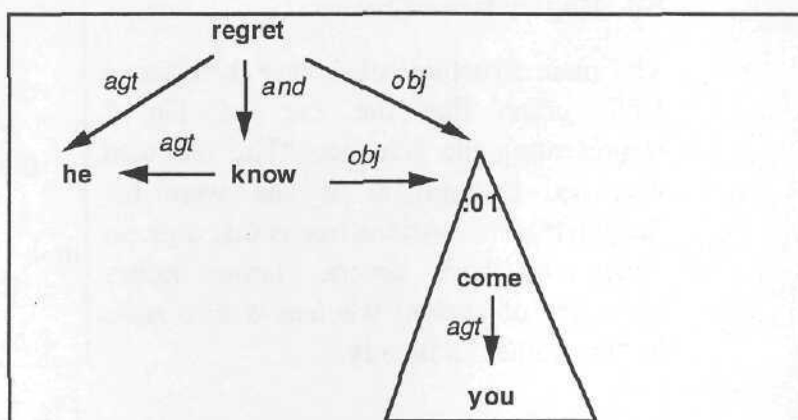
Fig.8 Transfer result for the graph of Fig.7

An interesting point in the graph of fig.8 is the representation of the semantic relation "as if" between the main and subordinate clauses, by keeping the conjunction as a node, having the subordinate clause as an object. This is a very convenient way of expressing a semantic relationship by the content of a node itself when this relationship does not appear as a conceptual relation in the UNL specification. We will come again to this point below when discussing the semantic aspect of the transfer.

5.2 Compound UWs

The graph of Fig. 9, representing the sentence "He knows and regrets that you will not come" contains a so called "compound UW": "you will not come" may be considered as a unit-concept, described independently, and which appears in the graph under a numeric reference.

The result of the transfer of this graph is shown on Fig. 10. The compound UW had to be duplicated, but the information that they represent in fact the same node has been kept as a particular variable, so that the morphological generation may give




```
agt(regret(equ>be sorry).@entry,he(icl>human))
obj(regret(equ>be sorry).@entry,:01)
agt:01(come(agt>human,goal>place).@entry.@future.@not,you)
and(regret(equ>be sorry).@entry,know(agt>human,icl>#event))
agt(know(agt>human,icl>#event),he)
obj(know(agt>human,icl>#event),:01)
```

Figure 9: a graph containing a compound UW.

the correct translation "Il sait que tu ne viendras pas et il le regrette", and not "Il sait que tu ne viendras pas et il regrette que tu ne viennes pas".

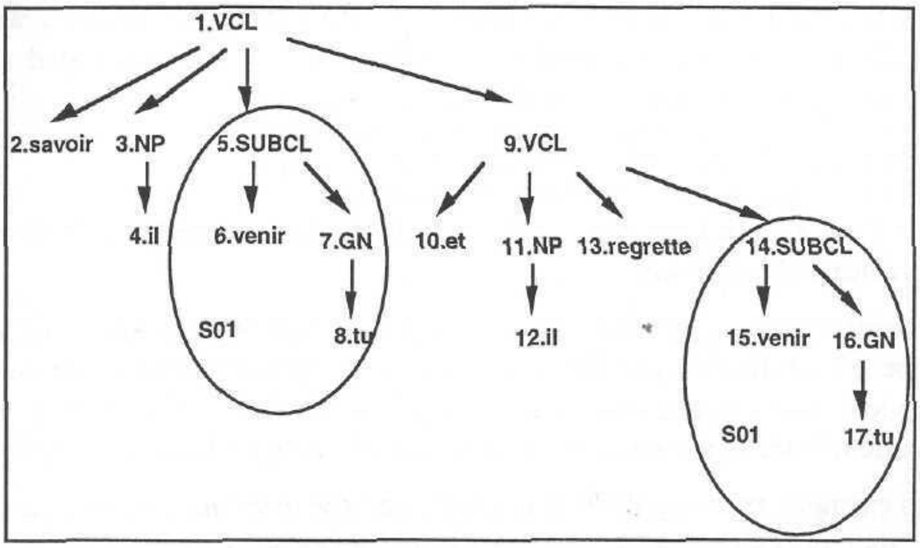


Fig.10 Transfer result for the graph of Fig.9

In conclusion, the structural transfer is not a critical aspect of the processing, and does not lead to any loss of information.

6 Semantic transfer

Roughly speaking, the semantic transfer consists in transferring the set ot UNL conceptual relations given in the appendix into the set of semantic relations of Fig.5. The first set seems at first sight more complete than the second one. In fact, the use of the argumentary relations of the predicates compensates the lower number of semantic features: knowing the argumentary relation between a predicate and its dependent, we don't need to know the fine semantic relation between both in order to get the correct syntactic relation for the generation.

In fact, the UNL interlingua makes use of the argumentary relation, but in an implicit way. Let's for instance consider an example illustrating the aoj relation ("thing with attribute") in the UNL specifications available on the UNL site mentioned in section 2. The sentence "John knows" is encoded as:

```
aoj(know(aoj>thing,obj>thing), John(icl>person))
```

The restrictions `aoj>thing` and `obj>thing` express clearly the two arguments of the predicat know, but the fact that John is effectively the first argument is not explicitely indicated in the binary relation. One of the drawbacks is that the distinction between an argument and an adverbial phrase is not clearly indicated. Let's again consider an example taken from the UNL specifications, "John gives to Mary":

```
agt(give(icl>do), John(icl>person))
ben(give(icl>do), Mary(icl>person))
```

Here the argumentary relation is not indicated, even in the in the restrictions of the UW. A difficulty will arise if we want to code the classical example "John gives Mary a book for Peter". How to distinguish what is usually considered as an argument (Mary) from what is an adverbial phrase (for Peter).

Using argument relations as a complement to the conceptual relations would release the difficulty. "John gives Mary a book for Peter" could thus be encoded

```
agt.@a(give(icl>do), John(icl>person))
obj.@b(give(icl>do), book(icl>volume)
ben.@c(give(icl>do), Mary(icl>person))
ben(give(icl>do), Peter(icl>person))
```

where `@a, @b and @c` denote respectively the first, second and third arguments of the predicate give(icl>do).

Moreover the use of arguments would make the enconversion easier: one would avoid a sometimes difficult evaluation of the conceptual relation between a predicate and its dependent. This is quite analogous to the use of explicit conjunctions when the exact semantic relation is not represented in the set of conceptual relations (cf Fig.7).

As an example, encoding the English sentence "The tower reaches the heaven" as

```
obj.@a(reach(icl>do(plt>thing)).@entry,tower.@def)
plt.@b(reach(icl>do(plt>thing)).@entry,heaven(icl>region))
```

would even prevent the consequences of an incorrect choice of the conceptual relations, like

```
agt.@a(reach(icl>do(plt>thing)).@entry,tower.@def)
gol.@b(reach(icl>do(plt>thing)).@entry,heaven(icl>region))
```

if priority is given to the argumentary relations in the deconversion process.

We introduced the processing of such a "decoration" of the binary relations into our deconverter, but we are aware of the fact that using arguments of the predicates would assume an agreement from each user of the interlingua about the number and order of the arguments for each predicate.

7 Conclusion

The main problem we were faced with in developing the French deconverter lies in the transfer of the UNL semantic relations. As pertinent as a set of semantic relations may be, it seems difficult that it could satisfactorily describe the semantic characteristics of all predicative concepts. It is why we are convinced that introducing argumentary relations beside (and of course not instead of) a set of semantic relations would be very helpful into the UNL project, as in any such interlingua based system.

8 References

Boitet C., Sérasset G. (1999) UNL-French deconversion as transfer & generation from an interlingua with possible quality enhancement through offline human interaction. MT Summit 1999

Boitet C. (1997) GETA's methodology and its current developments PACLING'97, Meisei University, Ohme, Japan, sept 97, Proceedings 23-57.

9 Appendix: the UNL conceptual relations

agt	a thing which initiates an action
and	a conjunctive relation between concepts
aoj	a thing which is in a state or has an attribute
bas	a thing used as the basis for expressing degree
ben	a not directly related beneficiary or victim of an event or state
cag	a thing not in focus which initiates an implicit event which is done in parallel
cao	a thing not in focus which is in a state in parallel
cnt	an equivalent concept
cob	a thing not in focus which is directly affected by an implicit event which is done in parallel or an implicit state in parallel
con	a non-focused event or state which conditiones a focused event or state
coo	a co-occurred event or state for a focused event or state
dur	a period of time during which an event occurs or a state exists
fmt	a range between two things
frm	an origin of a thing
gol	a final state of object or the thing finally associated with an object or an event
ins	the instrument to carry out an event
man	a way to carry out event or the characteristics of a state
met	the means to carry out an event
mod	a thing which restricts a focused thing
nam	a name of a thing.
obj	a thing in focus directly affected by an event or state
opl	a place in focus where an event affects
or	disjunctive relation between two concepts
per	a basis or unit of proportion, rate or distribution
plc	a place where event occurs or state is true or thing exists
plf	a place where an event begins or a state becomes true
plt	a place an event ends or a state becomes false
pof	a concept of which a focused thing is a part.
pos	a possessor of a thing
ptn	an indispensable non-focused initiator of an action
pur	a purpose or objectives of agent of an event or the purpose of a thing exists
qua	a quantity of a thing or unit
rsn	a reason that an event or a state happens
scn	an virtual world where an event occurs or a state is true or a thing exists
seq	a prior event or state of a focused event or state
src	an initial state of object or the thing initially associated with an object or an event
tim	a time when an event occurs or a state is true
tmf	a time when an event starts or a state becomes true
tmt	a time when an event ends or a state becomes false
to	a destination of a thing
via	an intermediate place or state of an event