Modeling a software of semantic text analysis

Anca Christine Pascu[1[0000−1111−2222−3333]]

Université de Brest, Brest, France
Lab-STICC - Laboratoire des Sciences et Technologies de l'Information, de la Communication et de la Connaissance
anca.pascu@univ-brest.fr

1 Abstract

In this article, we present an analysis of the steps required to build a system to automatically locate expressions that represent a play on words in a text. We rely on the notions of *conceptual metaphor*, *semantic network* and *ontology* as the basic elements that can be considered as primitive in any machine model approach.

2 Résumé

Dans cet article, nous présentons une analyse des étapes nécessaires pour la réalisation d'un système de repérage automatique des expressions qui représentent un jeu de mots dans un texte. Nous nous appuyons sur les notions de *métaphore conceptuelle*, de *réseau sémantique* et *d'ontologie* comme étant les éléments de base qui peuvent être considérés comme primitives dans toute approche modèle-machine.

3 Introduction

There is as much beauty in truth as is truth in beauty.
Fred Sommers, The Logic of Natural Language

If we look up for the definition of "word" on Wikipedia [16], you can find four definitions following four fields of cognition: the first one labeled "All" gives a general definition in language; the second one is related to linguistics, the third in related to the theatre and the forth to computer science. These are the following[1]:

1. General definition: A single distinct meaningful element of speech or writing, used with others (or sometimes alone) to form a sentence and typically shown with a space on either side when written or printed. Example: *I don't like the word "unofficial".*

[1] These definitions are proposed by Oxford Languages.

2. Linguistics: A single distinct conceptual meaningful unit of language, comprising inflected and variant forms.
3. Theatre: The text or spoken part of a play, opera, or other performed piece; a script. *he had to learn his words.*
4. Computer Science: A basic unit of data in a computer, typically 16 or 32 bits long.

If we compare the definitional elements of definitions 2 and 4, we see that both refer to the notion of *meaning*, more in the linguistic definition than in the computer definition.
Why is that ?
A common feature of the "linguistic word" and the "computer word" is the fact that both are elementary units, (primitives), one in linguistics and the other in computer science. Another common feature of the "linguistic word" and the "computer word" is the fact that both carry "meaning". The line that separates these two types of "words" lies in the "meaning".
Therefore, in our opinion, the analysis must be conducted from the notion of "meaning".
In linguistics, a word can have a local meaning - lexical semantics takes care of this. In computing, the words that make up "data" have no meaning outside of the algorithm that manipulates them. It follows that there are different ways and means of arriving at meaning in "human understanding" versus "machine-understanding".
In [3], Jean-Pierre Desclés makes a comparison between compilation as a process defined by computer science and human understanding (arriving at meaning) thanks to cognitive structures:

"Thus, a natural language has a metalinguistic part and different metalinguistic systems of different interconnected levels inserted in a computational and cognitive architecture."

4 Words, Concepts, Objects

We claim that, in the man-machine relationship, all software is based on a conceptual model from which, through successive representations (models), we arrive at the software itself by a *generalized compilation* [3]. We do not have a formal proof for this assertion. It can be taken as a conjecture. That is for, in this section, we deal with the notions of *word, concept, object.* The three notions - *word, concept, object* - are primitive in any conceptual model. The logical relationship between these three notions can only be established by the defining features. For this purpose, we must "navigate" between linguistics, philosophy of language, logic and mathematics.
Linguistics states that *the word* is an elementary unit of language, a primitive of language. As for *the concept,* it is defined inside philosophy of language and inside formal logic.
The first philosopher giving a mathematical definition of concept is Gottlob Frege [4]. Frege begins with the opposition object - function. The object is a

saturated entity, while the function is an unsaturated entity ([5]. By analogy with a mathematical function, the notion of concept can be defined, according to Frege, as a function that applies to an object with the two truth values *true* and *false* ([5]). Continuing Frege's tradition, in the Logic of Object Determination (Logique de la détermination d'objets (LDO))([5]), the notions of *concept* and *object* are defined in such a way that one can express mathematically the properties that derive from a semantic network. In this way, we can give a mathematical status to a *semantic network*.
A formal description of LDO was given in papers as: [6],[7],[10].

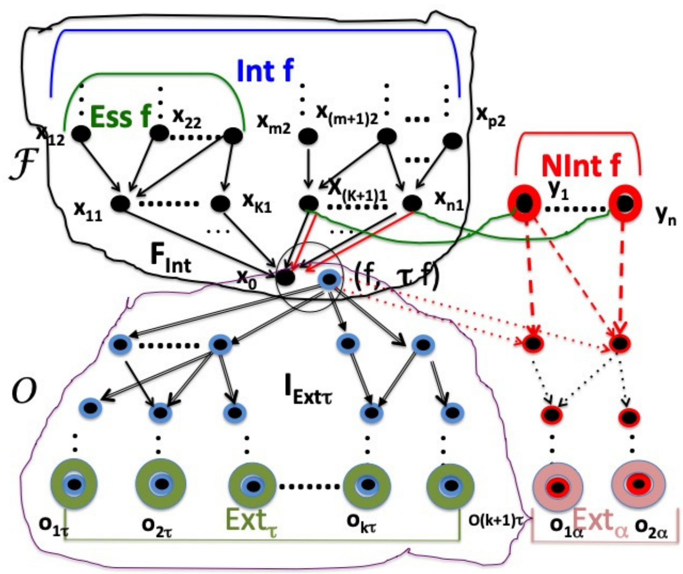

Fig. 1. The mathematical model by network of the LDO

In figure 1, we can see that a concept *f* belongs to a space $\mathcal{F}$. It is defined by a network of properties x_{ij}, by levels. Between two properties x_{ij} and x_{kl} it can be a relation of *inheritence*. All properties converges to the concept *f*. The space $\mathcal{F}$ contains two networks associated to *f*: Int *f* and Ess *f* such that Ess$f \subseteq$ Int *f*. The semantic of subnetwork Int *f* is: its properties x_{ij} are all the features of the concept *f*. The semantic of subnetwork Ess *f* is: Ess *f* contains all the properties x_{ij} neccessary for the concept *f* to be the concept *f*. In the figure 1, in the right side, there is a subnetwork NInt *f* of negation of some properties in Int *f*.
The objects associated to the concept *f* are displayed in the network $\mathcal{O}$ by levels. Objects are of two types from the point of view of their degree of determination: *more or less determinate objects* and *totally determinate objects.* The reasoning of the determination is the following: one associate to the concept *f* the object *to-*

tally indeterminate object τf. All more or less determinate objects are obtained from τf by succesive determinations. The totally determinate objects are saturated to determination. They form the set Ext *f.* The whole network in figure 1 highlights the quasi duality between concepts and objects. It can be seen that an primitive unit of a *semantic network.* This network was discussed deeply in [8], [10], [9].

5 Conceptual Metaphor

The notion of conceptual metaphor was introduced for the first time by George Lakoff and Mark Johnson in their works Conceptual Metaphor in Everyday Language ([12]) and Metaphor We Live By ([11]). They affirm that specific feature of the metaphor known from the literature are related to the language especially to the grammar. They also outline the idea that we can analyze metaphor as a notion related to philosophical notions such as the nature of meaning, truth, rationality, logic and knowledge. By its later developments, this notion becomes a notion belonging to the cognitive sciences. It is analyzed by postulating the concept of conceptual blending by Gilles Fauconnier and Mark Turner in ([15]). They state:
"Conceptual blending is a basic mental operation that leads to new meaning, global insight, and conceptual compressions useful for memory and manipulation of otherwise diffuse ranges of meaning." ([15] - page 57).
and
" Conceptual blending is described and studied scientifically in terms of integration networks. In its most basic form, a conceptual integration network consists of four connected mental spaces: two partially matched input spaces, a generic space constituted by structure common to the inputs, and the blended space. The blended space is constructed through selective projection from the inputs, pattern completion, and dynamic elaboration." ([15]-page 60).
A classical example for conceptual blending ([17]) is a blend of the conceptual space of house and the conceptual space of boat, yielding the concept of houseboats and the concept of boathouses as new emergent structures. Conceptual blending is the process of analysis of two conceptual spaces, a source space and a target space and the transfer operations leading from the concept in source space to new concept in the target space. The target concept is a new concept obtained from the source concept by transfer. Conceptual metaphor and conceptual blending have been studied as a systematic whole by ontologists in computer science for developing various ontology designs ([14]). Starting from the classical optimality principles of blending ([15]), ([17]), Goguen and Harrell explore more conditions for applying these principles. In the literature, Joseph Goguen et. al. extensively developed algebraic semiotics methods to describe the structure of complex signs and the blend of such structure. It is possible to capture the essence of the transformations between two different concept domains at the logical level, by the Institution Theory ([13]). The conceptual metaphor was modeled in the framework of the Logic of Determination of Objects (LDO)

in ([6]) in order to build a computational system for its analysis. This logic seems to be a good tool in the analysis of conceptual metaphor. Roughly speaking, the conceptual metaphor and conceptual blending represent a complex process to build from a semiotic space structured in some way (source space) a new concept in a new space (target space) with respect to its structure. This process is based by the two spaces, source space and target space, equipped each one of its own structure and a transfer (translation) component.

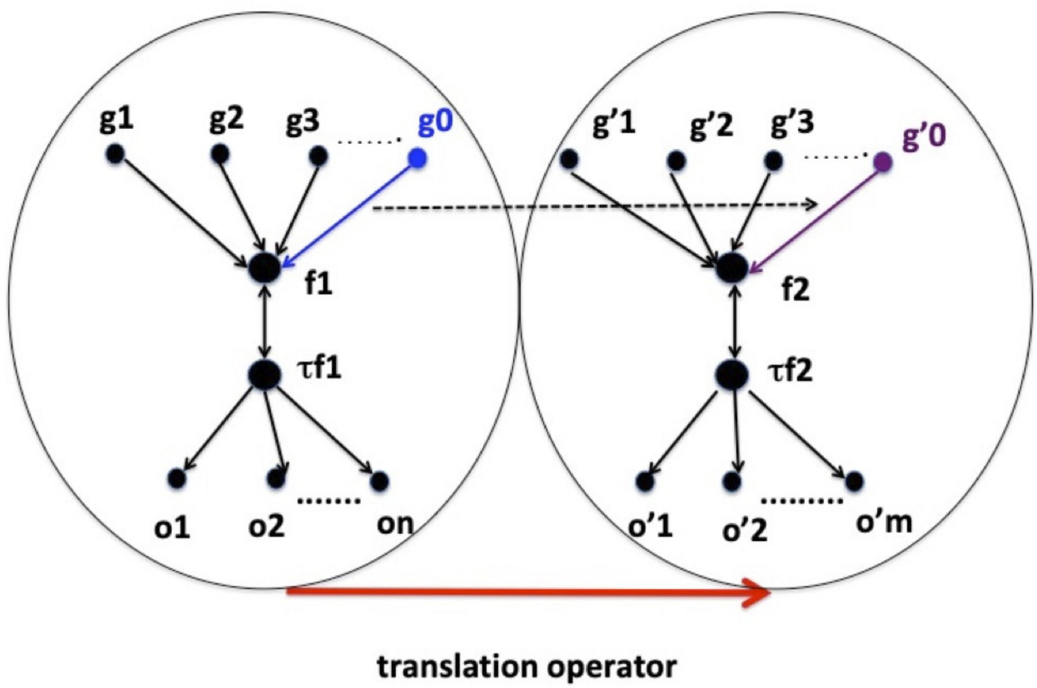

Fig. 2. Conceptual Metaphor by LDO

In figure 2, we show how to build a concept f_2 starting from a concept f_1 in the framework of LDO model. The idea is to translate a feature g_0 of the concept f_1 into a feature g'_0 of the concept f_2. The *translation operator* is a complex operator. Concepts f_1 and f_2 can belong to the same field or to the different fields.

6 Concept map, semantic network, ontology

In literature, there is no a formal definition of a *concept map,* a *semantic network* and an *ontology.* Some people speaks about concept map, some people about semantic network, the other about ontology. The only commun feature of them is that all are *formal representations,* and, mathematically speaking, they are *graphs.* There are many scientific and technologic fields where these notions are used.
The question is which are the features doing their difference, so what kind of graphs for each one. That is to know which of them is more appropriate in such and such application. To find the more appropriate is ,in other words, to find the most adequate that can be transformed in a mathematical model and then in a computational model.

Definition 1. *(Concept map)[18] A concept map or conceptual diagram is a diagram that depicts suggested relationships between concepts. Concept maps may be used by instructional designers, engineers, technical writers, and others to organize and structure knowledge.*

Definition 2. *(Semantic network)[19] A semantic network, or frame network is a knowledge base that represents semantic relations between concepts in a network. This is often used as a form of knowledge representation. It is a directed or undirected graph consisting of vertices, which represent concepts, and edges, which represent semantic relations between concepts mapping or connecting semantic fields. A semantic network may be instantiated as, for example, a graph database or a concept map. Typical standardized semantic networks are expressed as semantic triples.*

Definition 3. *(Semantic triple)[20] A semantic triple, or RDF triple or simply triple, is the atomic data entity in the Resource Description Framework (RDF) data model.[As its name indicates, a triple is a set of three entities that codifies a statement about semantic data in the form of subject–predicate–object expressions (e.g., "Bob is 35", or "Bob knows John").*

Definition 4. *(Ontology)[21] In computer science and information science, an ontology encompasses a representation, formal naming and definition of the categories, properties and relations between the concepts, data and entities that substantiate one, many, or all domains of discourse. More simply, an ontology is a way of showing the properties of a subject area and how they are related, by defining a set of concepts and categories that represent the subject.*

The définition 1 state just that a concept map is a *diagramme* showing relations between concepts. So, its translation inside graph theory is:a concept map a graph whose nodes are concepts and whose arrows are relations between concepts. There are no conditions about concepts and about relations.

For the semantic network, definition 2 specify the nature of graph elements taking into account eventually definition3 and saying that a concept map is an instantiation of a semantic network.

An ontology formalism is led further on, by developing more deeply its primitives by specifying the categories of the concepts and the categories of the objects and other general properties of them outside properties specific to their belonging field.

In [22], the difference between a simple network and a semantic network is given by:
"The main difference between simple networks and Semantic Networks is the kind of relations which combine the individual nodes. The relations bear a meaning by itself. This fact is a huge advantage of this data architecture, because software applications, e.g. search engines, are able to understand and process the relations. Especially abstraction relations (e.g. "is part of" or "is sub-assembly of") or inheritance mechanisms known from object oriented programming (e.g. "is subclass of") are very easy to represent."

In our opinion, in type linguistic applications as text annotation, automatic translation, text summarizing and other types of applications where we need a deep understanding of language, the ontologies model is not sufficient. We state that the basic representation of knowledge must be based on other primitives.

7 A conceptual modeling in a play on words

A play on words (word play, play-on-words) is defined by Wylipedia ([23]) by:

"Word play or wordplay (also: play-on-words) is a literary technique and a form of wit in which words used become the main subject of the work, primarily for the purpose of intended effect or amusement. Word play is closely related to word games; that is, games in which the point is manipulating words.

Word play is quite common in oral cultures as a method of reinforcing meaning."

This definition shows that both *word game* and *play on words* are words transformation in order to change the *meaning* of an expression. The difference between them lies in the fact that in word game is a transformation of words at linguistic level phonetical, morphological or syntactical as for the play on word to build the new sens as for play word the construction of the new meaning implies other linguistic reference systems (repositories) of knowledge such as, for example, sociological, political, cultural in general.

"An example of modern word play can be found on line 103 of Childish Gambino's III. Life: The Biggest Troll. ([24])
H2O plus my D (a die), a dice), that's my hood (hotte) , I'm living in it."

One can remark that the transformation lies at phonetical level.

Sometimes, play on words is given by wit, that is the play on words integrates wit, the humorous intelligence. A kind of definition of wit on Wikipedia ([25]) is:
"Wit is a form of intelligent humour, the ability to say or write things that are

clever and usually funny. Someone witty is a person who is skilled at making clever and funny remarks. Forms of wit include the quip, repartee, and wise-crack."

The great problem arising is: How to translates play on words from a language to an other. For analyse this problem we give a funny play on words from a humorous television show at French télévision. It was about answering questions.

Question:*Comment peut-on appeler la femme du maire de Bordeaux ?*

How can we call the wife of the mayor of Bordeaux ?

Answer: *La mère du bordel.*

The mother of bordel.

In English, this group of sentences taken literally has not even any sense, even more humor. In French, the play on words plays on three referentials: a social one, and a linguistic one with two components phonetical and morphological. The social referential is related to the French society and the linguistic referentials to French. This joke was produced when the problem of gender equality was making a lot of noise on social networks. Some people suggested that all adjectives should include the feminine and plural form in a written text where the discourse is addressed to a community of men and women. They argued that in this mode women are highlighted and this is a contribution to mark gender equality. Of course this is nonsense from all points of view. This change of spelling makes the writing heavier and does not solve anything of the problems of gender equality. The play on words makes fun of this aspect. This is the social context. The linguistic level has two referentials: a phonetic component and a morphologic one. The play on words plays on the prononciation of the words "maire" (mayor) and "mère" (mother) which is almost the same on a hand and for the morphological form of some adjectives at masculin and feminine i.e. beau-belle. In French the word "bordel" is associate also with "a great mess" from English. We can easily see the entanglement of conceptual metaphors.
The great question is :
How to build a system of interactive models that, implemented on the computer, will be able to recognize the "meaning" of such a text.

We propose that the primitive of this system is the conceptual metaphor represented in the LDO. This system starts from a subsystem which is a *logical model* based on conceptual metaphors. Then, by generalized compilation, in the mind, it can be transformed within graph theory and, finally, into a computational system. Its functionality is supposed to "capture" the "meaning" of this type of texts.

8 Conclusions

This paper is intended to be an epistemological analysis of a way to build a computer system for "texts comprehension (understanding)". This system must

be based by a sequence of models: the conceptual model, the logical logical model, the mathematical model, the computational model. The conceptual model is supposed to identify concepts and objects related to the problem. The logical model must be a LDO representation of the whole problem. The mathematical model might be a graph theory model of all networks issued from the logical model. The computational model is the transformation of the mathematical model in order it can be processed by the computer.
This approach is a way to express modeling of a "text problem" from the step of "problem identification" to its computer processing.

References

1. Pascu, A: Jeu de mots et réseaux sémantiques. In: Le jeu de mots-De la construction esthétique à la déconstruction transgressive. pp. 67 - 83. Université de Brest, France (2012)
2. Fu, T.K., Pascu, A.: Conceptual Metaphor in Teaching Logic. In:,Shih, J. L.et al.(eds) 27th Conference on Computers in Education. Taiwan Asia-Pacific Society for Computers in Education (2019)
3. Desclés, J-P.: Intersémiotique et langues naturelles. In: SIGNATA Annales des sémiotiques/Annals of Semiotics, https://journals.openedition.org/signata/745
4. Frege, G.: Basic Laws of Arithmetic (exposition of the system, University of California Press (1967)
5. Pascu, A.: Logique de la détermination d'objets : concepts de base et mathématisation en vu d'une modélisation-objet, ANRT (2003)
6. Desclés, J.P., Pascu, A.: Logic of Determination of Objects (LDO) : How to Articulate Extension with Intension and Objects with Concepts. Logica Universalis, Springer, vol. **5** nr **1**, 75 – 89 (2011)
7. Desclés, J-P., Pascu, A.C.: Logique de la Détermination des Objets (LDO); structuration topologique et quasi-topologique des extensions. Conference La logique en question / Logic in Question, Paris-Sorbonne, (2016)
8. Pascu, A. Ch., Desclés, J-P., Biskri, I.: A topological approach for the notion of quasi topology structure. South American Journal of Logic, **Vol. X, n. X** 1-18 (2019)
9. Desclés, J-P. Pascu, A. Ch.: The Mathematical Model of the Logic of Determination of Objects (LDO) in the Soft Set Theory. unpublished
10. Desclés, J-P., Pascu, A. Ch.: Logic of Typical and Atypical Instances of a Concept—A Mathematical Model, Axioms, **8** 104 (2019)
11. Lakoff, G., Johnson, M.: Metaphor We Live By, Chicago, University of Chicago Press, (1980)
12. Lakoff, G., Johnson, M.: Conceptual Metaphor in Everyday Language. The Journal of Philosophy. **77** 453 – 486 (1980)
13. Goguen J. A., Burstall R. M.: Institutions: Abstract Model Theory for Specification and Programming. Journal of the Association for Computing Machinery **39** 95–146 (1992)
14. Kutz O., Mossakowski T., Dominik, L.: Carnap, Goguen, and the Hyperontologies., Logica Universalis, Special Issue on "Is Logic Universal?", **4(2)** 255-333 (2010)
15. Fauconnier, G., Turner, M.: Conceptual Blending, Form and Meaning. https://tecfa.unige.ch/tecfa/maltt/cofor-1/textes/Fauconnier-Turner03.pdf,(2003)

16. Wikipédia, https://www.google.fr/
17. Goguen, J., Harrell ,D. F.: Style as a Choice of Blending Principles. https://pdfs.semanticscholar.org/6c9d/4c14549451e3b892f3af6e73e29d1580da69.pdf, (2004.)
18. Concept Map.Wikipédia. https://www.google.fr
19. Semantic Network.Wikipédia. https://www.google.fr
20. Semantic Triple.Wikipédia. https://www.google.fr
21. Ontology. Wikipédia. https://www.google.fr
22. Conrad, J. and al.: Comparison of knowledge representation in PDM and by semantic networks. International Conference on Engineering Design, ICED 07, Paris, (2007)
23. Word play. Wikipedia.. https://www.google.fr
24. Word play example. Wikipedia.. https://www.google.fr
25. Wit. Wikipedia.. https://www.google.fr