

An ontology-based approach in the literary research: two case-studies

Daniela Alderuccio, Luciana Bordoni

ENEA – UDA/ADVISOR -
Via Anguillarese, 301, 00060 S. Maria di Galeria (Rome), Italy

{alderuccio, bordoni}@casaccia.enea.it

Abstract

The present paper suggests an application of ontologies in the analysis of a literary phenomenon. Starting from the needs of humanistic research and from the availability on the Web of Language Resources, the adoption of such methodology, whose reference tool is an ontology, has led to a deeper analysis of the “Dualism Truth vs. Propaganda” in Karl Kraus’s writings and of the novel “Cloud, castle and lake” of Vladimir Nabokov.

The aim of this paper is to put in evidence the role of ontologies in the literary research and text analysis as helping tools, able to make domain assumptions explicit. For this purpose, the two case-studies will show how the application of the WordNet ontology for word sense disambiguation can help the understanding of a literary phenomenon (in the first example) and the interpretation of the core of a novel (in the second example), thus facilitating the translation.

1. Introduction

The growing availability on the Web of humanistic and language resources is going to modify research in the Humanities both in terms of access and retrieval of literary documents and of interpretation of relevant texts.

Language resources play an essential role in Language Processing, “as the necessary common platform on which new technology and applications must be based” (Hovy & Ide, 1999).

Among the language resources presently available (i.e., lexical, terminological databases, corpora, grammars, lexicons) there are ontologies. Generally defined as a shared formal conceptualization of particular domains, ontologies provide a common understanding of topics that can be communicated between people and application systems (Guarino & Poli, 1995; Uschold & Gruininger, 1996). Many of them are already available in electronic form, as it is the case of WordNet (Miller, 1995), an on-line reference system designed at the Cognitive Science Laboratory of the University of Princeton to model lexical memory.

By defining a common vocabulary; by including definitions of basic concepts in the domain (also in a machine-intepretable form) and by identifying relations among them, a linguistic ontology allows experts to share information in a domain and to provide critical notes and comments on texts.

In order to show how to apply an ontology in the literary research, the “Dualism Truth vs. Propaganda” in Karl Kraus (Alderuccio, 2001) and the novel “Cloud, Castle, Lake” of Vladimir Nabokov (1941) have been analyzed.

2. Ontologies and Literary Research

Ontology is a representation vocabulary often specialized to some domain or subject matter. More precisely, it is not the vocabulary as such that qualifies as an ontology, but the conceptualization that the terms

in the vocabulary are intended to capture. Thus, translating the terms in an ontology from one language to another, for example from English to German, does not change the ontology conceptually.

The aim of this paper is to highlight the role of ontologies in the literary research as helping tools, able to make domain assumptions explicit.

The reason why we use such a reference tool is related to the complex nature of humanities texts, that need a fine discrimination. Natural language ambiguity (polisemy, ellipsis, etc.) poses a challenge to experts. In order to avoid misunderstanding in reading texts and to limit subjectivity in analyzing documents, the use of language resources in content analysis has revealed to be a viable idea.

Ontologies play different roles in information systems, natural language understanding and knowledge-based systems. They provide potential terms for describing our knowledge about the domain. What we propose here is an ontology-based approach in the literary domain. It will show how the application of the WordNet ontology for word sense disambiguation can help the understanding of a literary phenomenon (in the first example) and the interpretation of the core of a novel (in the second example), thus facilitating the translation.

3. Applications and Results

In the following paragraphs the explanation of the methodology underlying this research will follow.

In order to comprehend how apply existing language resources (LR) tools in the literary domain, the proposed methodology suggests:

- the selection of keywords, able to express the essence of the literary phenomenon under analysis;
- the consultation of a lexical data base, that allows to reach a semantic agreement on the meaning of key-words and to limit their semantic field.

The core of the methodology is the use of WordNet, a lexical data base (free available on line), working as a reference ontology.

WordNet was inspired by psycholinguistic theories of human lexical memory and organizes nouns, verbs, adjectives and adverbs into sets of synonym, the so-called synsets. Each synset represents one underlying lexical concept. Synsets are organized hierarchically by super and sub-classes (hyperonyms and hyponyms).

WordNet is one of the major steps in the direction of creating large and usable ontologies¹. It represents a valuable resource in the human language technology and artificial intelligence. The large coverage of English words (more than 138.000 English words²), and the creation of wordnets for several European languages³ within the EuroWordNet project⁴ extends its application from word sense disambiguation, information extraction and retrieval to trans-lingual text mining.

Basing on WordNet definitions (glosses) to the humanistic researcher is allowed the exploration of the semantic fields of keywords, claryfing in this way the conceptual domain of the literary phenomenon under analysis.

3.1. 1st Case Study: “Dualism Truth vs. Propaganda in Karl Kraus”

This methodology has been applied to the study of the “Dualism Truth vs. Propaganda” in the writings of the austrian intellectual Karl Kraus (1874-1936).

Kraus (1874-1936) was an Austrian intellectual and one of the bitterest satirists of fin-de-siècle Vienna. He was a journalist and the editor of the magazine “The Torch” - “Die Fackel” (Kraus, 1968) - for about 36 years.

He thought that a journalist had to inform the public rather than overwhelming it with propaganda.

Strongly believing in a language expressing the Truth, one of his major concerns was the German language and how it was abused by the press.

The aim of the present experiment is to extract from krausian works information, that allows to understand his behavior on journalism: he thought that the journalist’s duty was that of saying the truth and indicated in the german language the ideal linguistic system to express the reality of facts.

To synthesize the phenomenon under analysis four keywords (‘language’, ‘truth’, ‘journalism’, ‘propaganda’) have been selected. The consideration that each misunderstanding of the keywords would

have led to misinterpretations is the main reason why WordNet has been selected as a reference ontology.

The main goal of this experiment is to understand Kraus’ critical behavior against propagandistic journalism, that manipulate information.

Thanks to the lexical⁵ and semantic relations⁶ defined in WordNet, the above selected keywords have been disambiguated.

In WordNet *Journalism* has two senses:

- Sense 1: newspapers news media, fourth estate (newspapers and magazines collectively);
- Sense 2: the profession of reporting or photographing or editing new stories for one of the media.

Truth’s meanings can be resumed as follows. It is:

- either a *fact that has been verified* (sense 1) and a *true statement* (sense 3);
- or *conformity to reality or actuality* (sense 2) and *the quality of being accurate without error* (sense 4).

Propaganda has one sense in WordNet: *information that is spread for the purpose of promoting some cause.*

From WordNet definitions it emerges:

- the definition of *Information* as a *message received and understood that reduces the recipient’s uncertainty*;
- the purpose of *Propaganda*; i.e. that of promoting some cause by spreading information;
- and, something unexpected, such as the existence of two coordinate terms: fact and factoid.
Factoid : something resembling a fact; unverified (often invented) information that is given credibility because it appeared in print;
- Fact* : a statement or assertion of verified information about something that is the case of has happened.

As a result, thanks to WordNet concept disambiguation, it has been possible to explicit what it is linked to the concepts of Information and Propaganda:

- between Journalism and Propaganda it exists an antagonism, based on the faculty that news have to inform or to promote some causes;
- Journalism spreads verified information about something happened;
- Propaganda provides “factoids”, that is something resembling a fact; an unverified (often invented) information that is given credibility because it appeared in print.

Furthermore, consulting the definition of Journalism, it has been obtained an objective basis to the concept of journalism, as intended by Kraus.

¹ Friedman Noy, N. – Hafner, C.D. “The state of the art in ontology design”. In “AI Magazine”, American Association for Artificial Intelligence, Fall 1997

² Heyboer , K. “His idea to link words rewrote the dictionary”. In The Star-ledger, 22.1.2002. URL: www.wj.com/news/ledger/index.ssf?news/stories/0122wordnet.html.

³ Dutch, Italian, Spanish, German, French, Czech and Estonian. Wordnets are current developed for Swedish, Norway, Danish, Greek, Portuguese, Basque, Catalan, Romanian, Lithuan, Russian, Bulgarian, Slovenic.

⁴ EuroWordNet URL: <http://www.hum.uva.nl/~ewn/>

⁵ Lexical relations: synonymy, antonymy, polisemy.

⁶ Semantic relations: hyponymy, hyperonymy, meronymy.

WordNet Results for "Coordinate Terms" search of noun "propaganda"

1 sense of **propaganda**

Sense 1

propaganda -- (information that is spread for the purpose of promoting some cause)

-> information, info -- (a message received and understood that reduces the recipient's uncertainty)

=> factoid -- (something resembling a fact; unverified (often invented) information that is given credibility because it appeared in print)

=> misinformation -- (information that is incorrect)

=> fact -- (a statement or assertion of verified information about something that is the case or has happened)

=> news, intelligence, tidings, word -- (new information about specific and timely events;

=> news -- (new information of any kind)

=> news -- (information reported in a newspaper or news magazine)

Figure 1: Extract of WordNet results for "coordinate terms" of "Propaganda"

Real Journalism has to be linked to facts, to the truth of an information depicting reality. So, the antagonism Journalism vs. Propaganda is based on truth.

Truth finds its way of expression in a pure linguistic expression.

Only using a language in its pure form, it is possible to spread the truth, informing the public, not influencing it by the use of interpretation of facts.

As far as the fourth keyword, namely Language, is concerned, WordNet gloss provides the link between writing and reading and bases in a correct writing the right perception of reality in the public of readers.

In WordNet *Language* (in the most frequent sense of the term) is a *systematic means of communicating by the use of sounds or conventional symbols*. In the linguistic process *the cognitive process is involved in producing and understanding linguistic communication*. The process of writing (producing a linguistic communication) and of reading (understanding a linguistic communication) are therefore strictly interrelated to the use of a common medium: the language.

Being such an important means, Language has to express truth, in order to permit to the receiver (the reader) to have a picture of the reality, not an interpretation. This is the only way to inform the readers without influencing them, as Kraus meant.

The validity of Kraus' position on information problems has been confirmed by the disambiguation of the keywords and the exploration of their semantic fields carried out using WordNet.

Disambiguation has allowed to identify the core of the antagonism between Truth and Propaganda in the creation of a propagandistic pseudo-reality.

3.2. 2nd Case Study: the Nabokov's novel "Cloud, Castle, Lake"

The second case study is based on the novel "Cloud, Castle, Lake"⁷ (1941) of the Russian writer Vladimir Nabokov (1899-1977).

7 In the review "The Atlantic Monthly" (1941) www.theatlantic.com/issues/41jun/clouds.htm; in "Nabokov's Dozen", translated by Peter Pertzov in collaboration with V. Nabokov, William Heinemann Ltd. (1959)

He was one of the most remarkable Russian émigré writers. (His best known novel "Lolita" brought him world-wide fame).

Born in Leningrad in 1899, he fled from the Russian Revolution in 1919 and completed his studies at Trinity College, Cambridge. Between 1923 and 1940 he published novels, short stories, plays, poems, translation in the Russian language, translations of Pushkin and Lermontov, and a study of Gogol. In 1940 the rise and establishment of Nazi regime in Europe forced him to move on to America. There he was a lecturer at the Wellesley College from 1941 to 1948 and then Professor of Russian Literature at Cornell University. In 1973 he was awarded the American National Medal for Literature. He died in 1977 in Montreux, Switzerland.

As a son of a rich and highly cultivated family, Nabokov had a trilingual household (Russian, French and English), and gained a magnificent command of English, developing an extraordinary narrative and descriptive skill.

"Cloud, Castle, Lake" (Marienbad, 1937) was originally written in Russian and was first published as "Oblako, ozero, bashnya" in the Russian émigré review "Sovremennie Zapiski" (Paris, 1937) under the pen name of V. Sirin and was included into two collections of short stories: "Soglyadatay" (Russkie Zapiski publisher, Paris, 1938) and "Vesna v Fial'te I drugie rasskazi" (The Chekhov Publishing House, New York, 1956).

The English version of this story was prepared by the author in collaboration with Peter Pertzov. "Cloud, Castle, Lake" came out in the "Atlantic Monthly" magazine in 1941 and appeared among the "Nine stories" brought out by New Directions in "Directions", 1947.

The subjects of the story are: tyranny and oppression, the opposition between private and individual experience against social coercion and the dehumanising effect of the collective.

The main character is Vasili Ivanovich, a modest young man, working as a representative, that told a story to his boss (the narrator). Vasili Ivanovich won a ticket for a pleasure trip at a charity ball given by Russian émigré in Berlin. This prize cannot be refused, so Vasili leaved with a group of tourists. On the train he is requested to stop reading his book, in order to join the group, for singing and playing cards.

During a stay he experienced the vision of an idyllic landscape: he saw a blue lake, where a large cloud and a

castle were reflected in the water. The vision was so particular for its “inexpressible and unique harmoniousness” of the three elements, that Vasili Ivanovich had such an epiphany: he soon realized that in that place “life would at last be what he had always wished it to be (...) all around him were help, promise and consolation – so that there could not be any doubt that he must live”⁸ there for ever. But Vasili was not allowed to realize his dream. The tour leader was decided to bring him back alive or dead, and forced him to return to Berlin, as stated in the journey program. On the way back he is tortured by the members of the group. At the end of the story he said to the narrator, that he had not “the strength to belong to mankind any longer”.⁹

The starting point of the text analysis was a translation problem arisen by the ambiguous expression ‘*anonymity of all the parts of the landscape*’¹⁰: so the first word to be disambiguated was ‘anonymity’. More generally, in WordNet “anonymity” is indicated as “namelessness (the state of being anonymous). Starting from the consideration that BEING NAMELESSNESS leads to HAVING NO IDENTITY, therefore NO INDIVIDUALITY, two more words have been selected for the analysis using WordNet: ‘individuality ‘ and ‘identity’.

In WordNet “INDIVIDUALITY” has 2 senses:

1. Individualism (the quality of being individual);
2. Identity (the distinct personality of an individual regarded as a persisting entity).

<p>Individualism (the quality of being individual)</p> <p><u>Hyponyms</u>: singularity, uniqueness, peculiarity, specialness, distinctiveness</p> <p><u>Antonyms</u>: commonalty, commonness, (sharing of common attributes)</p> <p>Identity (the distinct personality of an individual regarded as a persisting entity)</p> <p><i>Personality</i>: the complex of all the attributes - behavioural, temperamental, emotional and mental – that characterizes a unique individual</p> <p><i>Personableness</i>: the complex of attributes that made a person socially attractive.</p>

Figure 2: Extract of WordNet results for “ Individualism” and “Individuality”

In other words, according to WordNet, IDENTITY springs out of individual characteristics by which a person is recognized or known, whereas INDIVIDUALITY opposes the above person to his social environment.

As a result, the translation of the phrase “anonymity of the landscape” depends very much on the translation of the term ‘anonymity’. ANONIMITY should be translated into other languages keeping in mind the main idea implied in it, namely the lack of peculiar traits.

The translator should thus look for a term in the target language that also implies ideas such as insignificance, triviality, indistinguishableness, etc.

The insignificance of the landscape is the reflection of a personal anonymity: the main character is a “refugee”, a stranger, a foreigner.

WordNet indicates for this term the following sense:

<p>Refugee (an exile who flees for safety) exil, expatriate, voluntary absent from house or country)</p> <p><u>Hyponyms</u>: displaced person, stateless person (a person forced to flee from home or country)</p> <p><u>Hyperonyms</u>: foreigner, alien, non-citizen, outlander (a person who comes from a foreign country).</p>

Figure 3: Extract of WordNet results for “ Refugee”

The main character is actually a non-citizen, an individual that aims to social integration, to assimilation, to minimize diversity to be accepted from the mass. This tendency to assimilation leads to anonymity (the narrator said that his name “may be is Vasilij Ivanovic”, but well remembers name or surname of the other members of the group).

This “anonymity of the landscape” reflects the private life of the main character and the relationship between him and the other journey mates.

The individual (the refugee) is in opposition with the mass (where persons lose their name and identity). This mass is dominated by a leader, who influences his followers, leading them to initiate an action. In this situation individuals act as an indistinct mass. Each disturbing element (such as personality) is perceived as an “alien presence” and attacked, as it is the case of Vasilij Ivanovic (the victim of the novel). At the end of the novel, on his way back home he is tortured and injured by some of his mates .

In this second case-study, we started from the translation problem arisen from the sentence “anonymity of all parts of the landscape” and obtained a better understanding of the underlying framework of the relations among characters.

Among various definitions in WordNet, the more relevant in the novel context seems to be the coordinated term “nowhere – an insignificant place”, that indicates a site without peculiar characteristics.

As a result, the disambiguation of the term “anonymity” has led not only to an enhancement of the translation but even to a deeper interpretation of the core of the novel, allowing to extract aspects hidden in text, and giving them much more visibility than earlier. Furthermore, by means of WordNet definitions, the structure of the relations among characters is made explicit through the exploration of meaning of co-related terms such as “identity”, “individuality”, “leader” and “mass”.

By using WordNet definitions, the analysis of a few terms of the story has led to a better understanding of the role and relations among the characters involved in action, thus achieving a translation closer to the writer’s purposes.

⁸ Nabokov, V., “*Nabokov’s Dozen*”, p. 96.

⁹ Nabokov, V., “*Nabokov’s Dozen*”, p. 97.

¹⁰ Nabokov, V., “*Nabokov’s Dozen*”, p. 91.

WordNet Results for "Coordinate Terms" search of noun "identity"

4 senses of **identity**

Sense 1

identity, personal identity, individuality -- (the distinct personality of an individual regarded as a persisting entity)

-> personality -- (the complex of all the attributes--behavioral, temperamental, emotional and mental--that characterize a unique individual)

=> identity, personal identity, individuality -- (the distinct personality of an individual regarded as a persisting entity)

=> personableness -- (the complex of attributes that make a person socially attractive)

Sense 2

identity -- (the individual characteristics by which a thing or person is recognized or known)

-> recognition, identification -- (the process of recognizing something or someone by remembering)

=> identity -- (the individual characteristics by which a thing or person is recognized or known)

Sense 3

identity, identity element, identity operator -- (an operator that leaves unchanged the element on which it operates)

-> operator -- ((mathematics) a symbol that represents a function from functions to functions)

=> linear operator -- (an operator that obeys the distributive law)

=> identity, identity element, identity operator)

Sense 4

identity, identicalness, indistinguishability -- (exact sameness)

-> sameness -- (the quality of being alike)

=> identity, identicalness, indistinguishability -- (exact sameness)

=> similarity -- (the quality of being similar)

=> equality -- (the quality of being the same in quantity or measure or value or status)

Figure 4: Extract of WordNet results for "coordinate terms" of "Identity"

4. Conclusions

The applications of this ontology-based approach in the literary research has led:

- from one hand, to the confirmation of the validity of Kraus's position on information problems. The disambiguation of the selected keywords and the exploration of their semantic field (carried out using WordNet) has allowed to identify the core of the antagonism between Truth and Propaganda in the creation of a propagandistic pseudo-reality;
- on the other hand, a better understanding of the Nabokovian novel has been reached. The examination of WordNet definition for the term "anonymity" has given an important key to the interpretation of the role and relationships among characters, allowing a translation closer to writer's purposes.

Insofar as word sense disambiguation of literary documents is concerned, the achieved results show the validity and effectiveness of language resource applications, namely ontologies.

This methodology seems to allow a better analysis of literary phenomena or texts, so you reach a deeper understanding, which in its turn, leads to a better translation.

5. References

Alderuccio D., 2001. Proposta metodologica per un'applicazione integrata di risorse informatiche all'analisi di un fenomeno letterario. *Workshop on "Artificial Intelligence for Cultural Heritage and Digital*

- Libraries" - AI*IA 2001, 7th Congress of the Italian Association for Artificial Intelligence, September 25th-28th 2001, Bari, Italy,*
- Arntzen, H., 1975. *Karl Kraus und die Presse*. Muenchen: Wilhelm Fink Verlag,
- Guarino, N. & Poli, R., 1995. The role of Ontology in the Information Technology, *Int'l J. Human-Computer Studies*, Vol. 43, Nos. 5/6: pp. 623-965.
- Gruninger, M. & Lee, J., 2002. Ontology applications and design. *Communications of the ACM*, February 2002, Vol. 45, No. 2
- Hovy, E. & Ide, N., 1999. *Multilingual Information Management: current level and future availabilities*, Hovy & Ide Editors. (URL: <http://www-2.cs.cmu.edu/~ref/mlim/index.shtml>)
- Kraus K., 1968. *Die Fackel*. Koesel Verlag
- Miller G.A. (1995) WordNet: a lexical data base for English. *Communications of the ACM* 38:11, 39-41. (URL : www.cogsci.princeton.edu/cgi-bin)
- Nabokov, V.,1941. Cloud, Castle, Lake. *The Atlantic Monthly* (URL: www.theatlantic.com/issues/41jun/clouds.htm)
- Nabokov, V., 1959. In *Nabokov's Dozen*, Cloud, Castle, Lake, translated by Peter Pertzov in collaboration with V. Nabokov, William Heinemann Ltd.
- Ushold, M. & Gruinger, 1996. Ontologies: principles, methods and applications. *Knowledge Engineering Review*, vol. 11, Number 2, The University of Edinburgh.