
Ontology of Esperanto Documentation

Release

Epistemica

Jul 23, 2017

Contents

1	Project scope	3
1.1	A formal representation of Esperanto's grammar	3
1.2	The OWL-DL ontology of Esperanto	3
2	Ontology	5
2.1	Download the ontologies from GitHub	5
3	Paper «Esperanto Ontology»	7
3.1	Preface	7
3.2	Objectives	8
3.3	Ontologies and Description Logics (DL)	8
3.4	Esperanto	9
3.5	Approaching the project	9
3.6	Scheme of the ontologies	10
3.6.1	Ontology of grammatical analysis	10
3.6.2	Ontology of logical analysis	11
3.6.3	Ontology of period analysis	11
3.7	Considerations on the work done	11
3.8	Perspectives and hypotheses on the future of this type of ontology	12
3.9	Correlated products	12
3.10	Bibliography and webgraphy	12

The Tower of Babel, Pieter Bruegel the Elder, 1563

Ontology of Esperanto is a research project performed in 2006 by Epistemica.

Table of contents

A formal representation of Esperanto's grammar

This project comes from the question: *which could be the best natural language for vocal interaction between humans and computers?*

Esperanto was therefore considered, an artificial language officially born in 1887 for pleasure and thanks to the work of Polish Doctor Ludwik Lejzer Zamenhof, a passionate linguistics studier that wanted to develop a language purely international. From this, Esperanto was adopted by a significant community of speakers in all the world really able to maintain a serious conversation in this language. The choice of Esperanto was thanks simplicity and regularity of its grammar that could makes easier for computing systems to manage the structure of the language.

The project “Ontology of Esperanto” is proposed as one of the first attempts to produce a formal representation of a *non formal* language (natural language), based on the Semantic Web standards.

The goal was to develop an ontology usable into NLP software tools that can make easy a computer understand the logical form of varying complexity phrases expressed in Esperanto. From here we can imagine a scenario which on the Semantic Web becomes available innovative services that allow users perform queries in natural language for accessing data and documents, without having any expertise in formal query languages.

The OWL-DL ontology of Esperanto

Ontology is written in OWL-DL, a version of the language recommended by the W3C for the Semantic Web, using Protégé, the ontology editor developed by the University of Stanford (US), that uses the automated reasoning services of Racer, the semantic reasoner developed by the University of Luebeck (D).

The ontology verily are three: first for the *grammar analysis*, second for the *logical analysis* and thitd for the *analysis of the sentence*, where the highest level ontology refers to concepts at lower levels. In this way, once the various parts of the text is qualified at lower levels, so the higher levels can be automatically inferred.

Download the ontologies from GitHub

- Grammar Analysis
- Logical Analysis
- Period Analysis

Paper «Esperanto Ontology»

Authors Marco Romano¹ Luca Severini²

Version latest

Preface

Interest in Esperanto has gone through various phases in Italian culture during the last century. Initially it was of a humanistic nature and counted Bruno Migliorini, unforgettable president of the Crusca Academy and Alessandro Bausani, in the writer's opinion, the best Italian orientalist of all time, amongst its Honorary Fellows as well as many other lesser-known Honorary Fellows among linguists and men of letters. Over the last few decades, has emerged a new interest more orientated towards possible application of Esperanto in computing. In the 1980s there was intense activity in the formation of a dependence grammar of Italian for the Distributed Language Translation Project realised by a Dutch company, BSO, supported by the European Community and having Esperanto as its internal translation pivot language. Lecturers from the universities of and also participated.

Tullio de Mauro allures to this kind of possibility when he recently wrote "Esperanto today sees openings for real perspectives of use" in his introduction to a new edition of the Esperanto Manual by Bruno Migliorini (Milan: OCEDES, 1995). In the end, the same concept was represented by Francesco Sabatini, currently president of the Crusca Academy; he evokes a concept that the network magazine, summarises thus: "...the president of the Crusca Academy has a proposal in the interests of both people in computing and Esperantists. Opting for a common language means understanding each other better and accelerates development".

The world of Esperanto speakers cannot but agree on many such opinions. Indeed, they are aware that all affirmations concerning the easiness and regularity of Esperanto, and man's use of it, are inferior to what Esperanto really offers. No other language studied in a formal manner and by adults can be mastered like Esperanto; its easiness and regularity quickly allow it to become a second mother tongue or at the least a second language that can be used as automatically as a mother tongue. This is what, more than people in computing and linguists, the poets have been saying for a century, by means of their contribution to the creation of considerable original and translated literature in Esperanto.

¹ PhD in Logics and Informatics

² Founder of Epistemica, holder of the research project

Personally, as an Esperanto speaker interested in the psycho-linguistic phenomena that occur in the mind of a speaker of this language and others, I cannot but wish great success to this new project on Esperanto Ontology which united logical rigour with extremely advanced technical knowledge.

Prof. Renato Corsetti - *Psycho-linguist, Department of Psychology, University of Rome Sapienza*

Objectives

The Esperanto Ontology project presents itself as one of the first attempts to produce a formal description of a non-formal language by means of innovative tools such as ontologies for the semantic web.

Indeed, by virtue of these it is possible to exploit the expressive richness of logic to obtain a symbolic description of any known domain onto which inferential-deductive automatic reasoning can be sustained. This type of reasoning on its own is not evidently sufficient to comprehend the whole functioning of a natural language, spoken by “semantically omnipotent” human beings and therefore capable of meaning anything through language, given that the human mind uses other forms of reasoning (capacity of abstraction, singling out pertinence between concepts, reasoning by images and association of ideas etc.) in order to master such a tool; it is, in any case, an important opportunity to increase the possibilities of recognising the logical form³ of a proposition on the part of a computer. The fact that this process makes queries expressed in natural language comprehensible to a computer is of particular interest to us for producing a formal description.

Our objective was exactly that of building a tool by virtue of which logical forms of phrases of varying complexity in Esperanto could be recognised by a computer. From here we can start to imagine an applicative scenario in which innovative services are available on the web which can be interrogated by users in a natural language, without the need for them to be competent in database interrogation languages.

A wider horizon for this work can be integrated in the *Epistemica* project, proposed as a high-level interface for computer communications managed by software reasoners.

Ontologies and Description Logics (DL)

Classic ontology is a formal description in any known domain that is capable of showing a hierarchy of concepts inherent to the matter (by means of taxonomy) and the relationships between them (by means of a series of properties defined for each concept): an ontology for the semantic web also has the characteristic of being written in a particular language that can be “read” (not) by human beings but by external computing applications.

The appropriate language for an ontology is a logical language capable of guaranteeing the formalism of the description to be produced; the language chosen by the World Wide Web Consortium (W3C) as standard for semantic web ontologies is the Ontology Web Language (OWL) of which there are 3 varieties: Full, DL and Light. OWL-Full is a “good translation” of the classic Logic of the First Order (FOL) language which, however, carries the limit of semi-decidability, that is to say, the impossibility of guaranteeing complete automatic reasoning owing to the fact that it can express propositions capable of sending the computer into a “loop” and blocking its reasoning. OWL-Light is the “poor” variety which does not cause problems to automatic reasoning but only permits very reduced expressiveness. OWL-DL however, is greatly plentiful in expressiveness which, even though reduced respect to OWL-Full, is enough necessary to contain expressiveness in the domain of decidability and therefore to guarantee complete automatic reasoning. The DL suffix stands for Description logic. Various description logics were developed during the 1970s, each one with different expressive characteristics and potentialities; it is from these that even more languages were built, distinguished for use of logic in the environment of knowledge management computing, an environment in which tools must be powerful enough not only to profoundly describe knowledge but also to guarantee a complexity of finished calculation to render knowledge really useful.

³ Logical form of a proposition means its logical structure, recognised by assigning appropriate roles of subject, predicate, object etc. to the parts that compose it.

Thanks to the OWL-DL language, we can realise an integrated system which uses ontologies as instruments for representation of knowledge and a programme named “software reasoner” for developing deductive reasoning beginning with encoded knowledge with the guarantee that reasoning will be complete.

Esperanto

Esperanto does not originate from a community of speakers, but was designed, it is an artificial language officially created in 1887 by the will and works of the Polish doctor Zamenhof, a passionate scholar who wanted to develop a purely international language. Since then Esperanto has been adopted by a consistent community of speakers (there are about 250 thousand Esperantists in the world) which have “brought it to life” from the artificial language that it was to huge enrichment and, inevitably, exposing it to the constant mutation of natural languages. In any case if on the one side, once built, Esperanto has been brought to life precisely because it was designed to behave like a natural language to all effects, on the other it conserves the elegance of total regularity of its grammar which is a rare case among natural languages

The reasons for choosing Esperanto for this project are therefore evident and pertain to at least two distinguished orders: on one side because Esperanto was born from an ambitious project to build a universal language which all men could use in equal competence, the language itself guaranteeing equal dignity to its speakers by putting them all on the same level as Esperanto is nobody’s mother tongue; on the other because it is particularly easy to describe in a complete manner thanks to its grammatical characteristics, both morphological and syntactic ones.

Approaching the project

The technology chosen to produce our basis of knowledge of Esperanto is therefore that of ontologies for the semantic web because we can flank to these a software reasoner, a programme born from research on Artificial Intelligence able to conduct logical inferences starting from expressions of Description logic. The ontology-reasoner combination generates the possibility of dynamically enriching the basis of knowledge, easily keeping it updates and, most of all, comparing the knowledge described in it to that coming from other bases. The tools that we used were precisely the freeware interface Protégé (version 3.1 with OWL-Plugin developed by Stanford University) for writing ontologies in OWL language and the software reasoner RacerPro (version 1.8 manufactured by RacerSystems) for verifying integrity and coherence of ontologies, which is also available as a reasoner for practical realisation of applications based on these ontologies.

There is no univocal, better or more correct way to describe a dominion of knowledge, therefore there may reasonably exist more ontologies dedicated to the same subject and each may show their own characteristics and advantages. In our case we tried to build an ontology that is functional to our purpose and therefore sufficiently slim and elastic in order for it to be easily used by a software reasoner, but it also need to be powerful enough to supply an interesting, thereby not exhaustive, description of Esperanto even to a human who observes taxonomic structure and accompanying properties.

In order to proceed with the decomposition and reconstruction of Esperanto in a systematic way, we used an approach “by level of analysis” from which three distinctive ontologies are derived: one for grammatical analysis which supplies a classification of the *words* according to the typical categories of the parts of speech; and another for logical analysis which exploits the sintagma concept to group single words into “logical groups”; and lastly the said level of analysis of the period destined to the interpretation of relationships between the larger components of the sentence, the *propositions*.

This approach has produced ontologies that are in a hierarchal relationship between them, so that the analysis for the period explains itself starting with the logical analysis which in turn is comprehended integrating it with the grammatical one; the use of this technique was permitted by the capability of OWL language to import ontologies that are distinct from each other.

Each of these three ontologies was created using the axiomatic method, this means that undefined terms are used in each one of these, terms left to intuition as one might say, or better, concepts that are defined somewhere else, in our case in the parser programme which, in a concrete application of this programme, must instance and accompany logical reasoning based on ontology. For example, the concept *word* is not defined by the ontology, but its definition is in the parser which will recognise and isolate each recognisable sequence of characters as a word according to the definition given by its programming; then, by means of ontology, it is possible to identify the various parts of the speech, all subsumed from the concept *word*.

The role of the parser to accompany the ontologies is not secondary because not only must it “read” the text at the beginning of the analysis but it must read it many times, and every time carrying the new knowledge about the text which the reasoner will have inferred on the basis of the combination of information it has received from previous readings and from the rules described in the ontologies. Therefore, three sequential reading and reasoning phases are to be considered, enriched from time to time by parser-reasoner interaction until the phrase is completely understood.

The approach followed in this project presents yet another aspect which merits attention: the construction of “layers” makes sure that only the lowest level, that of grammatical analysis, is completely configured on Esperanto, while the superior levels can easily be utilised for other natural languages by superimposing them onto a purpose-built grammatical level.

Scheme of the ontologies

Ontology of grammatical analysis

This is the ontology at the lowest level, it produces classification of words according to the traditional concepts of Substantive, Adjective, Pronoun, Verb, Adverb and Preposition which are all subclasses, as specifications, of *Word*. The concept of *word* is the axiom of this level and its operative definition is left to parser programming.

Thanks to the characteristics of Esperanto grammar, the definition of the speech parts is quite simple. Esperanto, in fact, is a productive language, that is to say it theoretically forms all types of words (i.e. name, adjective, verb etc.) starting from each semantic root and it characterises the grammatical type using termination so that for each name ends in *-o*, all qualified adjectives in *-a* and adverbs in *-e*. There are, however, particular cases like Pronouns and Prepositions which do not have such a general rule of production and therefore necessitate more specific treatment, thereby these words, which are a relatively small number, have been manually inserted as instances for respective class of pertinence without an automatic test.

Some of the above-mentioned concepts are furtherly specified: such is the case with the verb, equipped in Esperanto with a series of terminations which univocally characterise each combination of verbal mode and time, following which an efficient classification of all the verbal options is obtained; it is also the case of the Pronoun, of which all the categories existing in Esperanto are recorded, and of the Adjective which is defined in a mixed manner: in fact, two of its subclasses necessitate the same treatment used for the pronouns and prepositions (manual insertion of instances without the possibility of closing their classes with a general definition) while the other subclass – qualified Adjective – is properly defined exploiting the rule of production for which adjectives generally terminate in *-a*.

There is an unusual concept in the middle of this traditional classification of the speech parts, the “Accusative Word”. The utility of inserting this concept into the same level of the speech parts analysis will be clarified later, here it is sufficient to be aware that Esperanto does not have a complete flexible system but distinguishes an accusative case terminating in *-n*; the concept of the Accusative Word, therefore covers exactly the class of words that occur in accusative.

Behind the taxonomy of concepts, ontology of grammatical analysis also supplies tools for characterising words according to gender and number criteria: these are treated as two properties to be evaluated for each word that is recognised as Substantive, Adjective or Pronoun by the reasoner. Evaluation takes place with the parser programme operating, once again it will exploit Esperanto peculiarities to recognise words easily. In Esperanto, in fact, each plural word ends in *-j* and every female word is built with the *-in* suffix.

Ontology of logical analysis

This is a second level ontology as it exploits a few concepts of the ontology regarding grammatical analysis; it also introduces the “sintagma” concept as a new axiom for subtypes for which it supplies generally valid definitions (nominal, verbal and prepositional sintagma).

In this ontology there are also concepts of logical analysis on the supply of sintagma notions which are Verbal Predicate and Nominal Predicate, because we think these can produce operative advantages when an application is realised which will exploit this ontology to translate a query from Esperanto to a computing language.

By using ontology properties, the necessary links between Predicate and Subject, such as those of compliments dependent on a verb, are also introduced in this ontology.

Amongst the compliments the object compliment is outlined for two reasons: the particular role this compliment generally covers in the grammar of occidental and romance languages and the ease of recognition it has by virtue of the presence of the accusative case (examined at a level of grammatical analysis), the case in which the object compliment is always formed. Exactly for this reason it was deemed necessary to describe the accusative case on the level of grammatical analysis in order to have it as an available concept on this higher level.

Ontology of period analysis

The new axiom of this level is the *Phrase* concept of which the predicative function is taken into consideration, that is, to declare a state or relation regarding a subject, neglecting particular cases of phrases at the margins of linguistic analysis. The subconcept of Predicative Phrase is thereby immediately highlighted while the rest of the analysis concerns hierarchal relationships among which propositions can be maintained, i.e. semantically incomplete phrases which together form a period. In actual fact, a period is also a partially defined concept which requires a “raw” definition that can be utilised by the parser like the axioms of these ontologies.

In the absence of further analysis on the types of proposition (relative, causal, temporal, etc) which could also be distinguished in Esperanto, ontology at this level is simply implied as a tool for recognising hierarchal structure among larger components of the proposition so that the reasoning obtained from this level is mostly exploitable to improve the obtainable results on the lower level, that of logical analysis, the real heart of this project.

Considerations on the work done

This “ontological collection” intends to describe Esperanto in order to use this language in computing applications of various kinds which can range, as mentioned above, from a query translator which facilitates the man-machine relationship up to its integration in a more ample communications protocol between machine and machine. In any case, precisely because of the elasticity and available of use that permit a knowledge base like this, the strictly computing side of the construction of software applications which exploit these ontologies for specific purposes assumes considerable importance.

In fact, these ontologies for Esperanto, apart from the benefits of being built in an integrated manner, therefore exploitable on various levels of analysis and being the first to be produced, also have all the limits from an analytical point of view which derive from the decision to supply a description of Esperanto which is functional to combined use with a parser. Basically it is neither a grammatical proposal of using Esperanto for humans to learn, nor a model of complete and rigorous language but a description which outlines a good number of rules of Esperanto and, on higher levels, of the syntaxes in general that are deemed sufficient for an adequately configured programme to obtain the logical form of phrases written in Esperanto and therefore permit understanding of the phrases when complex systems dedicated to a particular context are produced; complex systems meaning the “translator” tool, complete with ontologies and parsers, applied over another knowledge base or even a simple database from which a specific ontology is supplied.

Perspectives and hypotheses on the future of this type of ontology

Ontologies written in OWL are tools belonging to the semantic web and their fate, according to the W3C, is that of granting a large step forward towards the web as foreseen by Tim Berners Lee where each website, each resource available on the Net will be described, even through ontologies, in a way that it will have a specific identification not only for its position which is the purpose of URIs, but also concerning its “being” as one might say in philosophical terms, or, to be clearer, its value within the context of the Net.

There will be a long wait for this project to come about, until all those giving out information of any kind on the web realise that by adding an ontology to the resources that they publish, they are contributing to the realisation of semantic web which will probably be more interesting and certainly more tidy and hospitable than it is today.

Yet, as this project has shown, ontologies seem to be quite useful for other purposes. As they are forms of descriptions of knowledge, we have used them to produce a knowledge base on Esperanto which does not accompany any resources on the web⁴, but is put forward as a nucleus and starting point for interesting and innovative applications which could considerably facilitate simply navigation and management, conservation and consultation of information on a web which is not yet semantic.

Correlated products

Result of this work are three ontologies in OWL-DL format which form a whole with the present paper. These are the ontologies:

Note:

- **EsperantoGrammarAnalysis.owl**
 - **EsperantoLogicalSyntagmaticAnalysis.owl**
 - **EsperantoPeriodAnalysis.owl**
-

Bibliography and webgraphy

Considering the huge amount of material published both in traditional and electronic format on each of the themes discussed in this document, the following bibliography is only for indicating some enlightening texts, possibly on an introductory level or alternatively originating from more authorised sources, on each argument identified by the paragraphs that examine them.

Esperanto:

- Amerio F., Bonvecchiato G., Fighiera G.C. (by), *Esperanto: Data and Facts*, 2nd edition, FEI- Milan 2002
- Migliorini B., *Manual on Esperanto*, CO.ED.ES – Milan 1995
- We recommend visiting the Italian Esperantist Federation website at <http://www.esperanto.it>
- For researching documents in Esperanto, research in this language is available on Google, <http://www.google.com/intl/eo>

Semantic Web:

- Gregoris A., Van Harmelen F., *A Semantic Web Primer*, The MIT Press – 2004

⁴ This could, however, be available on the web as a resource and perhaps accompanied by an ontology which describes it.

- And naturally World Wide Web Consortium documents available at <http://www.w3.org/standards/semanticweb/>

OWL-DL language:

- Costello R.L., Calvanese D., McGuinness D., Nardi D., Patel-Schneider P. (by) *The Description Logic Handbook: Theory, Implementation and Applications*, Cambridge University Press – 2003
- Grosz B.N., Horrocks Y., Volz R., Decker S., *Description Logic Programs: Combining Logic Programs with Description Logic*, in Proc. 12th Intl. Conf. On the World Wide Web (WWW-2003), May 20-23, 2003

In conclusion, the following are websites relative to the tools that we have used in this project:

- Racer, software reasoner - <https://www.ifis.uni-luebeck.de/~moeller/racer/>
- Protégé, ontology editor - <http://protege.stanford.edu>

-
- Project's repository on GitHub