The Role of Semantic Domain in the Field of Semantics

Karimova Iroda
Karshi State University, Teacher of Linguistics Faculty

ANNOTATION

This article is about Semantic domain which is the basic notion of linguistics. It is very important to know how to comprehend the words, sentences and texts and to investigate the linguistic games which express the concept and meaning of the text together. Its main task to clarify the difference between semantic domain in the word level and in the text level.

KEYWORDS: semantic domain, linguistic game, lexicon, corpus, subject field codes, word net domain, sentence net domain, word sense disambiguation, generic concept.

The notion of Semantic Domain improves that of Semantic Fields by connecting the structural approach in semantics to the meaning in use assumption introduced by Ludwig Wittgenstein in his celebrated "Philosophical Investigations [3:Wittgenstein, 1965:54-56].

A word meaning is its use into the concrete "form of life" where it is adopted, i.e. the linguistic game, in the Wittgenstein's terminology. Words are then meaningful only if they are expressed into concrete and situated linguistic games that provide the conditions for determining the meaning of natural language expressions. To illustrate this concept, Wittgenstein provided a clarifying example describing a very basic linguistic game: ". . . Let us imagine a language . . . The language is meant to serve for communication between a builder *A* and an assistant *B*. *A* is building with building-stones; there are blocks, pillars, slabs and beams. *B* has to pass the stones, and that in the order in which *A* needs them. For this purpose they use a language consisting of the words block, pillar, slab, beam. *A* calls them out; – *B* brings the stone which he has learnt to bring at such-and-such a call. – Conceive of this as a complete primitive language."[4:102]

We observe that the notions of linguistic game and Semantic Field show many interesting connections. They approach the same problem from two different points of view, getting to a similar conclusion. According to Trier's view, words are meaningful when they belong to a specific Semantic Field, and their meaning is determined by the structure of the lexicon in the field. According to Wittgenstein's view, words are meaningful when there exists a linguistic game in which they can be formulated, and their meaning is exactly their use. In both cases, meaning arises from the wider contexts in which words are located. Words appearing frequently into the same linguistic game are likely to be located into the same field.

In the previous example the words block, pillar, slab and beam have been used in a common linguistic game, while they clearly belong to the Semantic Field of building industry. This example suggests that the notion of linguistic game provides a criterion to identify and to delimitate Semantic Fields. In particular, the recognition

of the linguistic game in which words are typically formulated can be used as a criterion to identify classes of words composing lexical fields. The main problem of this assumption is that it is not clear how to distinguish linguistic games between each other. In fact, linguistic games are related by a complex network of similarities, but it is no possible to identify a set of discriminating features that allows us to univocally recognize them. "I can think of no better expression to characterize these similarities than family resemblances'; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. - And I shall say: games' form a family" [5:4]

We observe that linguistic games are naturally reflect in texts, allowing us to detect them from a word distribution analysis on a large scale corpus. In fact, according to Wittgenstein's view, the content of any text is located into a specific linguistic game, otherwise the text itself would be meaningless. Texts can be perceived as open window through which we can observe the connections among concepts in the real world.

Frequently co-occurring words in texts are then associated to the same linguistic game. It follows that the set of concepts belonging to a particular field can be identified from a corpus based analysis of the lexicon, exploiting the connections between linguistic games and Semantic Fields already depicted. For example, the two words fork and glass are evidently in the same field. A corpus based analysis shows that they frequently co-occur in texts, then they are also related to the same linguistic game. On the other hand, it is not clear what would be the relation among water and algorithm, if any. They are totally unrelated simply because the concrete situations (i.e. the linguistic games). They occur in general distinction.

It reflects on the fact that they are often expressed in different texts, then they belong to different fields. Our proposal is then to merge the notion of linguistic game and that of Semantic Field, in order to provide an objective criterion to distinguish and delimitate fields from a corpus based analysis of lexical co-occurrences in texts. We refer to this particular view on Semantic Fields by using the name Semantic Domains.

In our usage, Semantic Domains are common areas of human discussion, such as Economics, Politics, Law, Science, which demonstrate lexical coherence. The Semantic Domain associated to a particular field is the set of domain specific terms belonging to it, and it is characterized by a set of domain words whose main property is to co-occur in texts. An approximation to domains is Subject Field Codes, used in Lexicography to mark technical usages of words. Although this information is useful for sense discrimination, in dictionaries it is typically used only for a small portion of the lexicon.

Word net domains are an attempt to extend the coverage of domain labels within an already existing lexical database, word net. As a result word net domains can be considered an extension of word net in which sunsets have been manually annotated with one or more domain labels, selected from a hierarchically organized set of about two hundred labels. Word net domains represent the first attempt to provide an exhaustive systematization of the concept of Semantic Field and its connections to the textual.

It allowed people to start an empirical investigation about the connections between the textual and the lexical counterparts of Semantic Domains. First it concentrated on collaborating a lexical-coherence assumption, claiming that a great percentage of the concepts expressed in the same text belong to the same domain. Lexical coherence is then a basic property of most of the texts expressed in any natural language and it allows us to disambiguate words in context by associating domain specific senses to them. Otherwise stated, words taken out of context show domain polysemy, but, when they occur into real texts, their polysemy is solved by the relations among their senses and the domain specific concepts occurring in their contests. [2:33]

Intuitively, texts may exhibit somewhat stronger or weaker orientation towards specific domains, but it seems less sensible to have a text that is not related to at least one domain. In other words, it is difficult to find a "generic" text. This intuition is largely supported by our data: all the texts exhibit concepts belonging to a small number of relevant domains, demonstrating the domain coherence of the lexical-concepts expressed in the same text. In particular, 34.5% of nouns in co-occurring in the same texts in Semantic Correlations are annotated with the same domain label, while about 40% refer to generic concepts.

The different senses of ambiguous words should be necessarily located into different domains, because they are characterized by different relations with different words. On the other hand, variability can be modeled by observing that synonymous terms refer to the same concepts, then they will necessarily belong to the same domain. Thus, the distribution of words among different domains is a relevant aspect to be taken into account to identify word senses. Understanding words in contexts is mainly the operation of locating them into the appropriate semantic fields. To corroborate these assumption we developed a Word Sense Disambiguation procedure relying on domain information only, named Domain Driven Disambiguation (DDD) [3:22] The underlying hypothesis of the DDD approach is that information provided by domain labels offers a natural way to establish associations among word senses in a certain text fragment, which can be profitably used during semantic correlations is a sub portion of the Brown corpus annotated by word net senses. DDD is performed by selecting the word sense, which Semantic Domain maximize the similarity with the domain of the context in which the word is located. For example, the word virus is ambiguous between its Biology and Computer Science senses, and can be disambiguated by assigning the correct domain to the contexts where it actually occurs. Results clearly shows that domain information is crucial for WSD(word sense disambiguation) allowing our system to improve the state-of-the-art for unsupervised WSD. The main conclusion of that work was that Semantic.

Domains play a dual role in linguistic description. One role is characterizing word senses (i.e. lexical-concepts), typically by assigning domain labels to word senses in a dictionary or lexicon. On the other hand, at a text level, Semantic Domains are clusters of texts regarding similar topics/ subjects. They can be perceived as collections of domain specific texts, in which a generic corpus is organized. Examples of Semantic Domains at the text level are the subject taxonomies adopted to organize books in libraries. The generality of these results encouraged us to extend the range of applicability of our assumptions, leading to the definition of a large number of techniques relying on the common theoretical framework provided by Semantic Domains in computational linguistics[1:27].

In particular we adopted an approach based on Latent Semantic Analysis to acquire domain models from corpora describing the application domain, and we assumed the principal components so acquired be mapped to a set of semantic domains. Latent Semantic Analysis has been performed on a term-by-document matrix capturing only co-occurrence information among terms in texts, with the aim of demonstrating our meaning-is-use assumptions. Then we exploited domain based representations to index both terms and texts, adopting a semi-supervised learning paradigm based on kernel methods. Empirical results showed that domain based representations performs better than standard bag-of-words commonly adopted for retrieval purposed, allowing a better generalization over the training data (i.e. improving the learning curve in all the supervised tasks in which they have been applied), and allowing the definition of hybrid similarity measures to compare terms and texts, as expected from the notion of Semantic Domain.

The conclusion of this experiment is that there exists a strong tendency for the lexicon in texts to be aggregate. Then we investigated the relations between Semantic Domains and lexical ambiguity and variability, the two

most basic and pervasive phenomena characterizing lexical semantics. Around a specific domain, such a tendency should be presupposed to allow lexical disambiguation.

References

1. Abdollahzadeh & Amiri (2009). The effect of semantic mapping as a vocabulary instruction technique on EFL learners with different perceptual learning styles. The Journal of Applied Linguistics, 2(1), 1-28.

2. Hayes. System der Sprachwisseenschaft. 1856, p.34

3. L.A. Novikov. The semantics of the Russian language, 1982, p. 33

4. L.Weisberg. The theory of field in Linguistics, 1939, p.102

5. Srinaowaratt, S. (2001). The Effects of Two Vocabulary Instruction Programs on Vocabulary Learning of Eleventh-Grade Thai EFL Students. Unpublished Ph.D. Dissertation, Illinois State University.

6. www.google.com