THE PROBLEMS OF REPRESENTING NEAR-SYNONYMS IN A COMPUTATIONAL LEXICON

Hasanova Gulhida
PhD student, Uzbekistan State World Languages University.

Abstract: The given thesis makes an attempt to show the difficulties of representing near synonymy in computational lexicon base highlighting the dimensions of variation that near synonyms can present. In particular, denotational, stylistic, structural and expressive variations of near synonyms that are vital determining the differences of near synonyms are briefly overviewed to draw the conclusion on the topic. Based on the differentiation of near synonyms and conclusions drawn from them six main problems of representing near synonyms in a computational lexicon are stated. Knowledge base model that is supposed to be sufficient enough to solve the problems mentioned in the article is going to be a main topic of our further studies.

Keywords: near synonyms, dimensions of variation, denotational variation, stylistic variation, structural variation and variations in expressiveness.

Аннотация: В данной работе делается попытка показать трудности представления близкой синонимии в вычислительной лексиконе, подчеркивая размеры вариаций, которые могут представлять близкие синонимы. В частности, кратко рассмотрены денотационные, стилистические, структурные и экспрессивные варианты близких синонимов, имеющие существенное значение для определения различий близких синонимов, чтобы сделать вывод по теме. На основе дифференциации близких синонимов и сделанных на их основе выводов сформулированы шесть основных проблем репрезентации близких синонимов в вычислительном лексиконе. Модель базы знаний, которая должна быть достаточной для решения упомянутых в статье задач, станет основной темой наших дальнейших исследований.

Ключевые слова: околосинонимы, виды различий, денотационные различия, стилистические различия, структурные различия и различия в выразительности.

A word's core "dictionary" definition is not the only meaning, it has also its connotations, and attitudes. Also, a word frequently contains near-synonyms that are different from it only in above mentioned subtle semantic differences. Therefore, one must carefully weigh the distinctions between all of the possibilities in order to determine the appropriate word to employ in each given situation—that is, the phrase that properly expresses the desired meaning while avoiding undesired implications. Even for humans, selecting the correct term can be challenging, let alone for modern computer systems. For instance, with so many possible translations that are similar but somewhat different, how can a machine translation (MT) system choose the optimal English word for the French *bevue*? *Error, mistake, blunder, slip, lapse, boner, faux pas, boo-boo*, and so forth are among the options that the system may select; however, the best option depends on the context in which the word is used as well as the semantic differences between it and the other English terms. The algorithm needs to ascertain the subtleties that *bevue* conveys in the specific situation that it has been employed in. Alongside selecting other words simultaneously, it must also determine which English word (or words) most nearly express the same subtleties. Since *bevue* is most likely

as distinct from each of its potential translations as it is from each other, an exact translation is probably not achievable. That is to say, every translation option will, on average, either communicate or withhold some potentially unpleasant detail. As a result, accurate translation necessitates a complex lexical-choice procedure that can identify the closest or most suitable near-synonym for a word in another language among those offered by one language in a given circumstance.

More generally, a sophisticated lexical-choice mechanism is also necessary for a natural language generation (NLG) system to be fully articulate. The system needs to be able to consider the possible outcomes of each choice. Think about the possibilities of a brand-new thesaurus for a word processor that works with the writer by rating the terms based on how well they fit the context and the writer's overall preferences, rather than just providing a list of synonyms. Many writers would profit immensely from such an intelligent thesaurus, which would be a clear upgrade over the basic thesauri found in word processors today.

A thorough computational model of fine-grained lexical knowledge is required. However, while being a basic linguistic phenomenon that shapes lexicon structure, synonymy has received significantly less attention in linguistics, psychology, lexicography, semantics, and computational linguistics than polysemy, which is also a fundamental and well-studied phenomenon.

Whether for philosophical, pragmatic, or expedient reasons, synonymy is frequently regarded as a "non-problem": either there are synonyms, but they have exact same meanings and are therefore simple to handle, or there are none, in which case each word can be treated like any other. However, our research on near-synonymy reveals that it is an as intricate phenomenon as polysemy, and that it has an inherent impact on the organization of lexical knowledge.

Even native language speakers may find it challenging to distinguish between near-synonyms well enough to utilize them consistently or to communicate those distinctions when they are understood. Furthermore, using the incorrect word could imply something you don't intend to. As a result, lexicographers have produced a large number of reference volumes that clearly distinguish between members of near-synonym groups (sometimes referred to as "dictionaries of synonyms").

Any study of near-synonyms must start with the idea of difference since, if two apparent absolute synonyms aren't actually the same, then they must be distinct. Difference, according to Saussure[31] (1916, page 114), is essential to the formation and delineation of meaning:

In a given language, all the words which express neighboring ideas help define one another's meaning. Each of a set of synonyms like redouter ('to dread'), craindre ('to fear'), avoir peur ('to be afraid') has its particular value only because they stand in contrast with one another.... No word has a value that can be identified independently of what else is in its vicinity.

Changes between near-synonyms are frequently remarkably complex. The near-synonyms vary not only in how they express different ideas and concepts, like blameworthiness and misconception, but also in how the ideas are implied, suggested, expressed, connoted, and stressed, how frequently they are expressed (common, occasionally, not always), and how strongly they are expressed (in strength).

[31] Saussure F. Cours de linguistique g ´en ´erale. Translated by Roy Harris as Course in General Linguistics. − London: G. Duckworth, 1983.

Dimensional variations. Generally speaking, any feature of the meaning of near-synonyms can have following variations[32] (Cruse 1986):

- denotational variations, in a broad sense, including propositional, fuzzy, and other peripheral aspects
- stylistic variations, including dialect and register
- expressive variations, including emotive and attitudinal aspects
- structural variations, including collocational, selectional, and syntactic variations

Variations in Denotation. Denotation, in its broadest meaning, is involved in several types of variation. While certain differentiae are easily described in terms of discrete abstract (or symbolic) qualities, like continuous/binary, many are not, according to DiMarco, Hirst, and Stede[33] (1993). Actually, the majority of differences in denotational variation are related to full-fledged concepts or ideas rather than basic facts; these distinctions pertain to roles and components of a scenario. For example, the notions that make up an *enemy* and a *foe* are not the same; the former emphasize animosity, while the latter emphasizes active warfare instead of emotional reaction[34] (Gove 1984). These indirect connotations are typically incidental to the major meaning that a word conveys, and it is typically difficult to determine definitely whether or not they were ever intended to be given by the speaker. Denotational differences can also be ambiguous as opposed to distinct.

Stylistic differences. Differences in a finite set of dimensions, on which all words may be compared, are what stylistic diversity entails. Stede (1993)[35], Nirenburg & Defrise (1992)[36], Hovy (1988)[37], and others have all put out a number of stylistic dimensions. Two of the most popular measurements are formal and informal.

Variations that are expressive. Numerous near-synonyms have different markings depending on the speaker's perspective on whether their denotation is positive or negative. Thus, the same person may be labeled as thin or slender if the speaker intended to be more favorable, thin or deprecating, or thin if he wanted to be neutral.

Structural differences. The final category of differences between near-synonyms consists of limitations on how they can be used as well as limitations that other utterance elements place on how they can be used. In either scenario, the limitations are unrelated to the definitions of the terms. The limitations can be syntactic, collocational, or selectional, meaning they rely on other ideas indicated or on other words or components in the speech.

Collocational variety refers to the concepts or words that a term might be paired with, sometimes in an idiomatic way. For instance, the collocational patterns of *task* and *job* are

[32] Cruse D. Lexical Semantics. − Cambridge University Press. 1986.
[33] DiMarco, Chrysanne, Hirst G., and Stede M. The semantic and stylistic differentiation of synonyms and near-synonyms. //In AAAI Spring Symposium on Building Lexicons for Machine Translation. − Stanford: 1993pages 114–121..
[34] Gove, Philip B. Webster's New Dictionary of Synonyms. Merriam-Webster. − Springfield, 1984.
[35] Stede, Manfred. Lexical choice criteria in language generation. //In Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics. – 1993.pages 119- 124.
[36] Nirenburg, Sergei and Christine D. Application-oriented computational semantics. In Michael Rosner and Roderick Johnson, editors, Computational Linguistics and Formal Semantics. − Cambridge University Press, 1992. pages 223–256.
[37] Hovy, Eduard. Generating Natural Language Under Pragmatic Constraints. − Lawrence Erlbaum Associates. 1988.

different: a person can encounter a difficult task, but not a difficult job. This is a lexical constraint; in contrast, the class of eligible objects is determined semantically, not lexically, in selectional restrictions (or preferences). Unlike *die*, for instance, *pass away* can only refer to people (or anthropomorphized pets); it cannot refer to plants or animals.

Different syntactic subcategorizations lead to variations in syntactic limitations. It is assumed that a group of words belong to the same syntactic category if they are synonyms or nearly synonyms. On the other hand, some near-synonyms may fall into a different subcategory than others. For instance, the adjective *open* can be used in any position, whereas the adjective *ajar* can only be used predicatively (The door is ajar; the ajar door).

Multilingual near synonymy. The rule in lexical transfer in translation is near-synonymy rather than synonymy: the word in the target language that is most similar to that in the source text may be a near-synonym rather than an exact synonym. For instance, the German word *wald* and the English word *forest* have similar meanings, but *wald* can also refer to a smaller and more urbanized area of trees than *forest*; in other words, *wald* includes some of the meaning of the English word *woods*, and in some cases, *woods* would be a more accurate translation of *wald* than *forest*. Therefore, we can talk about the distinctions in a group of cross-linguistic near-synonyms just as we can with a group of near-synonyms from a single language. In addition, reference books exist that offer guidance on cross-linguistic near-synonymy to translators and advanced second-language learners, much as they do for single-language near-synonym groups.

In summary, to account for near-synonymy, a model of lexical knowledge will have to incorporate solutions to the following problems:

• The four main types of variation are qualitatively different, so each must be separately modeled.
• Near-synonyms differ in the manner in which they convey concepts, either with emphasis or indirectness (e.g., through mere suggestion rather than denotation).
• Meanings, and hence differences among them, can be fuzzy.
• Differences can be multidimensional. Only for clarity in our above explication of the dimensions of variation did we try to select examples that highlighted a single dimension.
• Differences are not just between simple features but involve concepts that relate roles and aspects of the situation.
• Differences often depend on the context.

References:

1. Cruse D. Lexical Semantics. − Cambridge University Press. 1986.
2. DiMarco, Chrysanne, Hirst G., and Stede M. The semantic and stylistic differentiation of synonyms and near-synonyms. //In AAAI Spring Symposium on Building Lexicons for Machine Translation. − Stanford: 1993 pages 114–121.
3. Gove, Philip B. Webster's New Dictionary of Synonyms. Merriam-Webster. − Springfield, 1984.
4. Hirst G. Near-synonymy and the structure of lexical knowledge. //In AAAI Symposium on Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity. – Stanford: 1995, pages 51–56.
5. Hovy, Eduard. Generating Natural Language Under Pragmatic Constraints. − Lawrence Erlbaum Associates. 1988.

6. Inkpen D., Hirst G. Building a lexical knowledge-base of near-synonym differences. // Workshop on WordNet and Other Lexical Resources, Second Meeting of the North American Chapter of the Association for Computational Linguistics. – Pittsburgh: 2001, 47–52 p
7. Nirenburg, Sergei and Christine D. Application-oriented computational semantics. In Michael Rosner and Roderick Johnson, editors, Computational Linguistics and Formal Semantics. − nCambridge University Press, 1992. pages 223–256.
8. Saussure F. Cours de linguistique g ´en ´erale. Translated by Roy Harris as Course in General Linguistics. − London: G. Duckworth, 1983.
9. Stede, Manfred. Lexical choice criteria in language generation. //In Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics. – 1993.pages 119- 124.