

Text Translation by Means of Translation of Predicative Definitions.

Abstract.

It is here affirmed that it is possible to present a text as a summary of predicative definitions, arriving at this by means of a decomposition of the text into predicative definitions. Moreover, such a summary of predicative definitions constitutes the subtext and context of the text, which almost simultaneously allow one to identify the slang in which the text was composed.

Precise identification of the slangs in which a text was composed allows one, practically simultaneously, to identify a summary of predicative definitions in another language which will help to translate the text correctly with a computer.

A Slang.

The Merriam-Webster Online Dictionary defines a slang as follows:

Main Entry: slang

Pronunciation: 'sla[ng]

Function: noun

Etymology: origin unknown

1: language peculiar to a particular group; 2: an informal nonstandard vocabulary composed typically of coinages, arbitrarily changed words, and extravagant, forced, or facetious figures of speech.

I am completely in accord with the view expressed in Merriam-Webster: every area of human activity presupposes the existence of a specific means of communication -- a single one for all the subjects of a given activity -- a slang. Any geographically linked (now or in the past) group of people, united by a specialization in certain types of activity that are necessary for the physical and spiritual existence and development of those people, is the carrier of particular slangs.

I thus affirm that in order to translate a text into other languages, it is necessary to determine for communication in which area of human activity the author of the text created it.

Borrowings and Adaptations.

First of all, the author would like to clarify the meanings of the terms "predicative" and "non-predicative". As is well known, *praedicatum* in Late Latin means "what has been said (previously)". In Aristotelian and subsequent forms of traditional logic a predicate was understood to be one of the two terms for the judgment of a subject (the one in which something is said about the subject of speech). In his treatment the author counts as predicative any definition of a subject or object in which something is said about an observable subject or object as it changes. In addition, the author proposes as the sole measure of change the movement of a subject or object with acceleration: if a subject or object is immobile or moves evenly, it cannot be observed and, consequently, cannot be defined *predicatively*.

More: as is well known, Bertrand Russell introduced the notion of a "non-predicative" definition, in which what is to be defined is brought in through its relation to a class of which it is an element. For example: "the set of all sets that are not elements of themselves". It is said that the use of "non-predicative" definitions leads to paradoxes, so they should be dealt with carefully.

"Kernel Sentences" as Predicative Definitions.

It has been suggested earlier that we refer to minimal constructions carrying a basic load of meaning in texts as "kernel sentences" [20]. It has also been supposed that kernel sentences must provide an unambiguous, formal description arising from the subjects/objects of the given Names.

But I'm a philosopher. In this case, it seems more expedient to use a philosophical terminology – the terminology used by Aristotle, Moore, Russell and Wittgenstein, as the most significant philosophers to have dealt closely with the study of language. Those philosophers spoke of predicative and non-predicative definitions for the description of events and changes occurring in subjects or objects of the Universe. Moreover, predicative definitions and kernel sentences are one and the same thing.

The technology for extracting kernel sentences from a text -- as well as the method for formalizing kernel sentences -- did not exist previously and was created by myself. A more detailed description of this process can be found in US Patent 6.199.067.

The Text.

A text is made up of words. But what is a word? First of all, a word is made up of letters, which are, in practice, meaningless if separate. And yet a word, as the joining together of several letters, already, beyond any doubt, has a certain meaning. But the existence of synonymy makes the meaning of words, taken separately, vague and lacking in concreteness and separate words are declared to be non-predicative definitions. For example, the word "red", taken by itself, can mean anything: beginning with a colour and ending with a pejorative name for a Communist. In order to understand the "true" meaning of a word one must first identify in what minimal lexical construction of speech a given word is being used; where a minimal lexical construction of speech is a predicative definition: the articulation of three words, relating to three parts of speech - substantives, verbs and adjectives - in the context of a sentence, paragraph or text. All other parts of speech, with the exception of prepositions and interjections, can be, in some way or other, taken to be (or be reduced to) nouns, verbs, and adjectives, where:

1. A substantive has the meaning of the abstract Name of certain objects and subjects;
2. A verb defines the abstract Name of an action;
3. An adjective is the abstract Name describing objects and subjects in the process of change.

This triad of non-predicative definitions is indispensable to defining the subjective evaluation of a fact, when faced with the possibility and the need to include the objects and subjects of the fact within the certain context and subtext. In other words, a person has to evaluate a sandwich from all sides: he has to understand that it's a sandwich, and to decide whose it is, whether he should eat it or not, whether it's fresh and tasty, etc.

The presence of at least one predicative definition is absolutely necessary and sufficient for the creation of a sentence, even if it's missing one or more words from the substantive/verb/adjective triad. Such a word or words can be reconstructed on the basis of the context and subtext of the text; where:

1. The context consists of those predicative definitions where a substantive is used as the abstract Name of objects and subjects and abstractions;
2. The subtext consists of those predicative definitions where pronouns and interjections are used as the abstract Name of objects and subjects and abstractions.

For example, having said the word "unfresh", one can reconstruct the words "sandwich" and "exists" if we know in what context and subtext the word "unfresh" appears. And if we don't know the context and

subtext of a given predicative definition, then the word “unfresh” can be used with, for example, the words “fish” and “smells”. Only a text, being a collection of predicative definitions grouped together in meaningful sentences, can provide, more or less identically, the context and subtext of every one of these predicative definitions. That is, a text is considered to be completed in so far as its context and subject are, more or less identically, defined.

Processing the Text.

The task comes down to extracting all the predicative definitions from every sentence of the text. Such a collection is termed a summary of a sentence; the number of times each predicative definition occurs in the sentence, paragraph or text is referred to as its weight. (NLP uses the notion of predicative definitions not counting weights of them.) A summary, being an ordered list of predicative definitions, is susceptible to rapid processing by computer.

In principle, the extraction of predicative definitions could be called the "decomposition" of the text – "decomposition" for the purpose of creating a summary of the text's predicative definitions. I suppose that human reason is an instrument for a similar kind of text decomposition and composition in correlation with a certain type of activity.

Examples of Summary.

At the URL http://lexiclone.com/SummarySample_Fyodor_Dostoevsky.htm , the reader can see an extract from Fyodor Dostoevsky's summary (a part of which is reproduced below), created on the basis of his book *The Brothers Karamazov* in English (the numbers to the right represent the frequency of each predicative definition's occurrence in the text):

it - be - in : 1 466
i - be - in : 1 347
it - have - in : 996
you - be - in : 936
you - be - your : 798
i - have - in : 664
all - be - in : 657
it - will - in : 535
my - be - in : 496
all - have - in : 473

Clearly, it is subtext that predominates in Dostoevsky's text. In another summary - that of Patent #6.199.067 - it is context that predominates:

one - say - least : 1 447
segment - say - least : 1 124
datum - item - plural : 1 025
system - say - remote : 950
datum - say - plural : 888
computer - say - remote : 845
datum - item - linguistic : 845
system - say - least : 844

computer - say - least : 818

one - say - remote : 805

The Ethics and Aesthetics of a Text.

Thus it is supposed and demonstrated (cf. the fragments of summaries above) that a text is the integrated sum of a subtext and a context. Moreover, the context carries with it a cognitive load, displaying the knowledge that the subject who created the text has of the subjects-and-objects that are dealt with in the text. The subtext, meanwhile, carries with it an emotional load, demonstrating the interrelation of the subject who created the text with the surrounding world in general, and with the subjects-and-objects of the given text in particular. Indeed, the function of pronouns is to abstract from the Names that are signified by means of substantives and other parts of speech equated with them.

I go on to affirm that the subtext is the Ethical component of a text: Ethics relates to the evaluation of the consequences that would result from the relationship of the subject who created the text with the whole surrounding world, abstracted from concrete subjects-and-objects.

The Aesthetic component, meanwhile, is the context of a text: the knowledge of the concrete subjects-and-objects of a concrete text.

That is to say, in translating a text from one language into another it is necessary to translate the predicative definitions that correspond to both the Ethical and the Aesthetic components of the text.

Slang: More Details.

It is thus supposed that there is a unique slang for every type of human activity. Consequently, every slang uses a strictly specialized combination of predicative definitions, found only in that slang.

For example, the summary of a text dealing with Computational Linguistics looks as follows¹:

document-be-in: 649

question-be-in: 508

document-have-in: 505

document-turn-in: 492

document-result-in: 477

question-result-in: 416

document-retrieve-in: 408

answer-be-in: 406

document-recall-in: 398

document-be-retrieval: 398

question-be-retrieval: 368

question-have-in: 366

answer-be-retrieval: 366

1 Matthew W. Bilotti, Boris Katz and Jimmy Lin. What Works Better for Question Answering: Stemming or Morphological Query Expansion? *Proceedings of the Information Retrieval for Question Answering (IR4QA) Workshop at SIGIR 2004*, July 2004, Sheffield, England. In this case only an excerpt of the summary is shown.

Whereas business slang looks as follows²:

option-have-in: 83
option-say-in: 83
option-board-in: 83
option-will-in: 83
option-expense-in: 83
option-stock-in: 83
option-be-in: 83
option-price-in: 79
option-use-in: 79
option-require-in: 79
option-value-in: 79
option-account-in: 79
option-report-in: 79

Translation.

It is evident that computer translation of texts from one language into another is reduced to:

1. Identifying those slangs in the other language that are closest to the slang of the given text.
2. Identifying the summary of predicative definitions in the slang of the other language that correspond to the slang of the given text.
3. Identifying the group of words in the other language that relate to the slang of the given text.

Once one has established which slang in the other language the given text relates to, all three tasks can be considered to have been achieved. Indeed, knowing predicative definitions and words for a text it is easy to find predicative definitions and words in other languages that correlate with the meaning of the text.

Conclusion.

The problem of computerized text translation into another language appears to be resolved.

References.

1. Andy Zuppann. A Connectionist Approach to Word Sense Disambiguation. <http://www.sccs.swarthmore.edu/users/03/azuppann/papers/wsd.pdf>
2. Celina Santamaría, Julio Gonzalo and Felisa Verdejo. Automatic Association of Web Directories to Word Senses. *Computational Linguistics*. Volume:29 Number:3 Page: 485-502
3. Eugene Charniak. *Taggers for parsers*, (with Glenn Carroll, John Adcock, Antony Cassandra, Yoshihiko Gotoh, Jeremy Katz, Michael Littman, and John McCann), Artificial Intelligence (forthcoming).

² Tech Firms Offer Plan for Options, Sept. 15, 2004 (TechNews). In this case only an excerpt of the summary is shown.

4. Choueka, Yaacov, Ehud S. Conley and Ido Dagan. A comprehensive bilingual word alignment system: application to disparate languages - Hebrew and English, in J. Veronis (Ed.), *Parallel Text Processing*, Kluwer Academic Publishers, 2000, pp. 69–96.
5. Cong Li, Ji-Rong Wen, and Hang Li, Text Classification Using Stochastic Keyword Generation, *Proc. of ICML'03*, 464-471.
6. Dagan, Ido. Contextual Word Similarity, in Rob Dale, Hermann Moisl and Harold Somers (Eds.), *Handbook of Natural Language Processing*, Marcel Dekker Inc, 2000, Chapter 19, pp. 459-476.
7. Dagan, Ido and Yuval Krymolowski. Compositional memory-based partial parsing, in R. Bod, R. Scha and K. Sima'an (Eds.), *Data-Oriented Parsing*, CSLI Publications, 2002, forthcoming (20 pages).
8. Dagan, Ido, Zvika Marx and Eli Shamir. Cross-dataset clustering: revealing corresponding Themes Across Multiple Corpora, in *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL)*, 2002, pp. 15-21.
9. Diana McCarthy and John Carroll. Disambiguating Nouns, Verbs, and Adjectives Using Automatically Acquired Selectional Preferences. *Computational Linguistics*, Volume: 29, Number: 4, Page: 639 – 654
10. Ehud Reiter and Somayajulu Sripada. 2003. Human Variation and Lexical Choice. *Computational Linguistics*, Vol. 28, Issue 4. p. 545 - 553
11. Escudero, Gerard, Lluís Marquez, and German Rigau. 2000. Boosting applied to word sense disambiguation. In *Proceedings of the 12th European Conference on Machine Learning*, pages 129–141, Barcelona.
12. E. Charniak, K. Knight, and K. Yamada. "Syntax-based Language Models for Machine Translation", *Proc. MT Summit IX*, 2003.
13. Geoffrey Towell, Ellen M. Voorhees and Claudia Leacock. Disambiguating Highly Ambiguous Words
14. G. Kondrak, D. Marcu and K. Knight. "Cognates Can Improve Statistical Translation Models". *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2003)*, companion volume, pp. 46-48, Edmonton, May 2003.
15. Hang Li and Cong Li, Word Translation Disambiguation Using Bilingual Bootstrapping, *Computational Linguistics* 30(1), 1-22, (2004).
16. Hang Li, Yunbo Cao, and Cong Li, Using Bilingual Web Data to Mine and Rank Translations, *IEEE Intelligent Systems*, Vol. 18(4), 54-59, (2003).
17. Ilya S Geller. The Role and Meaning of Predicative and Non-predicative Definitions in the Search for Information
18. Janyce Wiebe ; Theresa Wilson ; Rebecca Bruce ; Matthew Bell ; Melanie Martin. Learning Subjective Language. *Computational Linguistics*. Volume: 30 Number: 3 Page: 277 -- 308
19. M. Galley, M. Hopkins, K. Knight, D. Marcu, What's in a Translation Rule? *Proc. NAACL-HLT*, 2004

20. Noam Chomsky. 1957. *Syntactic Structures*.
21. Sabine Ploux and Hyungsuk Ji. A Model for Matching Semantic Maps between Languages (French/English, English/French). *Computational Linguistics*. Vol. 29, Issue 2, p. 155 - 178
22. Ted Pedersen, Discriminating Among Word Meanings by Identifying Similar Contexts (Purandare and Pedersen) - Appears in the Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04), July 25-29, 2004, San Jose, CA
23. Ted Pedersen, SenseClusters - Finding Clusters that Represent Word Senses (Purandare and Pedersen) - Appears in the Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04), July 25-29, 2004, San Jose, CA
24. Wessel Kraaij, Jian-Yun Nie, Michel Simard. Embedding Web-Based Statistical Translation Models in Cross-Language Information Retrieval. *Computational Linguistics*, Volume:29, Number:3, Page: 381-419