

Cognitive Linguistics and Information Technology

In search of common language

V. Dobrova

Samara State Technical University
Samara, Russia
victoria_dob@mail.ru

A. Zhuravlev

Samara State Technical University
Samara, Russia

V. Savitskiy

Samara State Technical University
Samara, Russia

V. Tkachev

Samara State Technical University
Samara, Russia

L. Nurtdinova

Samara State Technical University
Samara, Russia

Abstract—This paper searches for the points of contact between cognitive linguistics and information technology (IT) by drawing parallels between key concepts of cognitive linguistics and mathematical logics. Within this research, the notions of the metalanguage, the object language and the graph are considered not only as strictly logical concepts but as a tool that can be used to solve a range of interdisciplinary problems such as modeling the knowledge structure of a certain subject field (by developing principles that can be applied to every subject field). Since every natural language is both an object and a subject of cognition, such type of research seems to be essential in terms of the modern level of computer technology and artificial intelligence (AI). The paper also substantiates the reason to create interdisciplinary teams of software developers because such kind of intercommunication and cooperation between different areas of study seems to be vital to deal with the problems of modern educational and testing issues.

Keywords—*concept; category; meta-language; object language.*

I. INTRODUCTION

Research in the field of developing artificial intelligence has been conducted since the 1950's. Nowadays, thanks to the rapid rate of computer technology development, research activity in this area has become even more intensive. As a result, speech recognizing and synthesizing software has appeared, as well as translation programs. Quite often, however, this software produces output of disputable quality, especially when it comes to translation programs. This is caused by two factors. The first one is strictly technical aspects of programming. In order to figure out the second factor, there is no need to conduct a profound investigation – one can just check out the output of any popular translation program. It is obvious to any professional translator than no translation theory and practice experts were involved into the

process of software implementation of translation algorithm. Or (quite a deplorable fact) programmers and a translator, hired by the software developing company, just did not find common language. If it is the second option, the reason is likely to be as follows: translators tend to use rather abstract notions while programmers require quite specific terms and statements so that they can work out an algorithm to be implemented as a piece of software. This is not a problem of translator's or programmer's incompetence but rather an issue of interdisciplinary cooperation. Elena Kubryakova, a major Russian expert in the area of cognitive linguistics, writes that formation of cognitive science was connected with recognition of the fact that scientific community of the second half of the 20th century faced a big number of highly sophisticated problems dealing with human behavior planning, and that these problems can be solved not by any single science but in the course of interdisciplinary research [1].

All this is to show that, in the authors' opinion, one of the key factors of successful research in the artificial intelligence (AI) area is to provide close cooperation of information technology (IT) and cognitive linguistics. Teams of developers that include experts that work in various branches of science – programming, cognitive linguistics, etc. – appear to be the most effective. For example, one such interdisciplinary team of software developers was formed at the Department of Pedagogy, Psychology and Psycholinguistics of Samara State Medical University (Samara, Russia). This team consists of programmers, linguists and doctors-traumatologists. This team developed several programs for testing student's professional knowledge; these programs are based on direct interaction between a man and a computer. Currently the team is working on a system which will be able to recognize free-form answers of students and evaluate their knowledge by means of qualitative and quantitative criteria [2].

In the view of the above-mentioned information, the authors find it necessary to «translate» some basic notions of cognitive linguistics into the language that mathematicians and programmers can understand.

II. META-LANGUAGE AND OBJECT LANGUAGE

Notions that came from mathematics and logic can help us to “build a bridge” between linguistics and IT. These sciences overstepped their limits and have strong influence not only on related academic disciplines but also on those areas of knowledge that are traditionally considered to be far from mathematics. And there is no surprise as logical principals and statistical methods are applicable in all areas of science and in everyday life. Cognitive linguistics is not an exception.

Notions in question are a “meta-language” and an “object language”. They are quite widespread and used in many branches of science, from mathematics to philosophy. That is why they can be used to draw a parallel between such incompatible scientific disciplines as mathematics and cognitive linguistics. In fact, this incompatibility is illusory – it is necessary to remember just such scientific branch as mathematical linguistics.

For an objective mapping of a natural language, describing its facts and establishing significant links between them, it is required not only to create effective methods for its investigation, but also to conduct work related to the search for rational ways of isolating the elements to be studied, effectively presenting the results obtained, classifying and generalizing them. Since any knowledge must be expressed verbally or with the help of other signs, all these tasks will be possible only if the processes of natural language communication are provided with a sufficiently developed meta-language.

A meta-language is a language, by means of which one describes and studies the properties of some other language [3], a language which denotative domain is the structure and functioning of another language. In the terminology of L. Hjelmlev, the meta-language is a sign system which content plan is another sign system – object semiotics [4].

The natural language studied by linguists is traditionally called a linguistic object or object language, and the language on which linguistic theory itself is formulated is a meta-language. The term “meta-language” is one of the key in modern logic and linguistics. It is used in the study of verbal languages, logical-mathematical calculi, for mapping the entire spectrum of connections of languages of different levels and for characterization of the relations of analyzed languages and the subject domains described with their help [5].

Since the notion of meta-language is used in different branches of science, its definition has become rather diverse over the years. For example, in accordance with the first understanding, any natural language is the union of the object language and the meta-language, in which one speaks about the object language [6]. This kind of approach to the natural language stipulates the existence in the meta-language of such expressions as meaning, denotation, signification, assertion, synonymy, presupposition, etc.

In accordance with the second interpretation, the semantic meta-language opposes the object language as another sign system, which allows more directly reflecting the structure of the expression of the object language, revealing and objectifying it.

So, the above-mentioned expressions under the second understanding of the meta-language will be considered as expressions of the object language, requiring explication in terms of the meta-language [7]. The meaning of the sentences of an object language is written using a semantic representation (representation, record, interpretation, etc.) composed of semantic meta-language units, which, according to the plan, not only conveys the same meaning as the object language, but its form also directly reflects the structure of this value [8].

“The task of representing knowledge is, first and foremost, the task of creating semantic meta-languages” [9]. Given that in our time there are meta-languages of phonetic and morphological levels, the semantics of the language still possesses a poor arsenal of means that allow explicitly fixing the content side of speech works. At the same time, the semantic meta-language is an integral tool for studying the object language in the communicative and cognitive aspects. It is a necessary part of the modeling applied to the language. The purpose of the meta-language is to model and compactly describe the content plan of the object language.

The meta-language of the semantic description of natural language is also called a semantic language that is used to represent the meanings of different units of language [10]. Traditional linguistics (phonetics, grammar, etc.) describes mainly a language expression plan – directly observable patterns. The meta-language of semantics explores what is not directly observable, describes the plan for the content of a natural language – an unobservable structure of the meanings of units of the language and the meaning of texts. This may be so-called “the grammar of semantics”, the identification of schemes that exist in terms of the content of the language, for example, through their modeling. Thus, the specific feature of this meta-language is that it is not the only possible way of describing an object language, so it admits alternative methods of description.

The various classifications of meta-languages used in the description of semantics have as their basis some significant oppositions. First of all, one should distinguish between nonverbal and verbal meta-languages. The first include languages in which not only words and other language units are used, but also drawings, various kinds of schemes, mathematical constructs (dependency trees, graphs, etc.).

Phrases in this language cannot be read, although in principle, like any images, they can be described in natural language. An example of nonverbal language can be considered in the semantic representation of I.A. Melchuk, which refers to the semantic language in the model “Meaning ↔ Text”. The complexity of the classification of non-verbal languages is explained by their large differences. They must be accompanied by some verbal comments or interpretations, that is, the introduction of a third-order language.

Verbal languages, in their turn, are divided into natural and artificial. Speaking about natural meta-languages (that is, when a natural language is used as a meta-language), it is necessary to take into account coincidences and discrepancies between the meta-language and the object language. As an example of coincidence, let us cite the language of interpretation, of which the definitions of words in explanatory dictionaries are composed. Examples of mismatch are presented in bilingual explanatory dictionaries, in which a word of one language is compared to its translation into another language – a word or a detailed formula. The possibility of a coincidence of the meta-language and the object language is a rather nontrivial fact.

After all, it is the natural language that can serve as a meta-language for describing oneself. In this case, as a rule, its two functions are under consideration, namely, the function of the object language and the meta-function. A special status was given to this phenomenon by Jacobson [11], who considered the meta-language function of the language as one of the main functions.

It should be mentioned that the coincidence of the object language and the meta-language can entail difficulties and problems related to the interpretation of words. The simplest of them is a tautology. Scientists who represent the meta-language in the spirit of analytical philosophy as ideal, insist on the elimination of tautologies [12; 13], which imposes the imposition of a number of restrictions on the natural language in the meta-function.

In the opinion of A. Wierzbicka [14], it is necessary to impose a number of requirements on the meta-language.

First, it must be universal in order to be able to express a variety of meanings quite accurate.

Secondly, it should allow the inclusion of procedural elements that will allow to build on its basis models of cognitive processing of the natural language and implement them in computer programs.

Thirdly, it is desirable that such a language be sufficiently organic and intuitive.

But in general its basic meaning remains the same: meta-language is a natural or artificial language ("second-level" language) which describes other language ("first-level" language) which, in its turn, describes objects, properties and situations of the visual environment or some of its areas. It is generally recognized that the meta-language should be "richer" than the object language it describes, as it should not only contain designations for all names and expressions of the latter but also use its specific means to describe their properties and set various connections between them.

III. FORMALIZED LANGUAGE

Experience shows that the above-mentioned goals can be most effectively achieved by artificially created formalized languages, which also bear the names of formal calculi, formalized systems, abstract calculus, formalisms or logic systems, formal logics, etc. Having analyzed many definitions, the main characteristics of formalized systems must be singled

out. This is a noninterpreted calculus, which expression class is usually inductively specified, i.e., ("atomic" or "elementary") formulas and rules for the formation of formulas, and a subclass of provable formulas (theorems) – by specifying a system of axioms and the rules of transformation (derivation) of theorems from axioms and previously proven theorems that implies a complete abstraction from the meaning of the words of the language used and consideration of all the conditions governing the use of these words in theory.

Accordingly, the formalized language is used both for the purpose of representing the logical forms of the real contexts of the natural language, the reproduction of logical laws, and for expressing the ways of correct reasoning on the logical theories of the language [15]. The construction of such artificial language of logic begins when its alphabet is given – a collection of initial, elementary symbols, where it is customary to include logical symbols (signs of logical relations and operations, for example, quantifiers and propositions), non-logical symbols (i.e. parameters of descriptive components of the natural language), as well as technical symbols (for example, brackets). Further from the simple signs of the language, the rules for the formation of complex signs are formulated – different types of correctly arranged expressions are given, of which analogies of natural language pronouns – formulas – are the most important. The main advantage of a formalized language is the effectiveness of the definitions of any of its morphosyntactic categories: the problem of belonging to one arbitrary symbol or string of alphabet symbols to a class of language expressions can be solved with a finite number of steps.

Sometimes formalized languages, in addition to the alphabet and rules of formation, include a number of transformation rules – deduction procedures, rules for switching from one string of symbols to another. The observation that in such cases the formalized meta-language acquires the status of a logical calculus is essential.

In another interpretation, the formalized meta-language involves the adoption of rules for the interpretation of its expressions, which allows us to relate the semantic category to each syntactic category of signs. This is essential for distinguishing logical forms.

Thus, a formal language is the code on which a formal theory (formal system) is created, which is a text (or a collection of texts) on this code. However, this is an abstract theory (a scheme of interrelations between symbols), which is not applied to a specific area of reality (subject domain). To apply it to a specific subject area, it is necessary to subject it to a semantic interpretation, to give abstract symbols specific values from a given domain. Interpreted (i.e. content-filled) formal system is called calculus.

The calculus can be represented in the form of a formal apparatus for manipulating knowledge of a certain type, which is based on clear rules, and which allows an accurate description of a particular class of problems to be presented, or a solution algorithm for subclasses of a given class [16]. On the basis of some formalized language, a logical calculus is constructed. For this, the composition of the original symbols

is specified, and then, according to well-defined rules, these formulas form the formulas of the calculus in question. After adding an interpretation to the calculus that gives meaning to its original symbols and formulas, the calculus becomes a formalized language for describing a certain subject area.

The use of formalized languages in linguistic studies is justified, because the meta-languages make it possible to apply the reduction principle, which reduces the variegated variety of surface phenomena to a limited number of typical (invariant) schemes and in this way to reveal the deep regularities in the studied sphere.

The expediency of developing a formal description based on an artificial meta-language is determined by a number of reasons:

- 1) it reduces or eliminates the interference of the meta-language and the object language;
- 2) it significantly increases the uniqueness and accuracy of the object description;
- 3) it makes possible to determine the exact degree of frame homomorphism;
- 4) it makes such descriptions suitable for applied research.

It is true that "such kinds of descriptions are programmable and can be specified in the modeling of intellectual systems by various means: in the language of frames, scripts, graphs, etc." [17].

The advantage of a formal description before the natural language is that it (a) is unified, (b) reduced, (c) generalized. It is the same for all situations of the class in question and reveals their invariant structure, which may not be visible when comparing various descriptions in natural language.

In cognitive linguistics, several sign systems have been developed that can be considered more or less formalized (i.e. reduced / unified / generalized) [18; 19], and which are based on different methodological grounds: situational semantics, graph theory, symbolic logic. This includes semantic networks, declarative and procedural semantic representations, as well as cognitive structures called configurations of memory organization, scripts, frames.

IV. A CATEGORY AS AN ELEMENT OF META-LANGUAGE

Now let us compare the above-mentioned definitions with what cognitive linguists write about cognitive processes and their linguistic representation.

Main notions of cognitive linguistics are «a concept» and «a category». N. Boldyrev says that "linguistic representation of knowledge is the result of conceptualization and categorization – two basic cognitive processes carried out with the help of a language" [20].

These closely interrelated processes build up a two-level cognitive system. The first level is accumulation of knowledge and its fixation in a human's mind in the form of concepts. This process is of statistical kind as accumulation of a certain amount of data causes shifting to a new qualitative level of cognition because a human needs to consider the received data

and put it in a certain order. This initiates the process of categorization which is "a mental activity to form the categories as ultimate concepts that generalize and classify the results of human's cognitive activity" [21].

Such words as «ultimate concepts that generalize and classify» explicate the idea that the authors employ in their attempt to «build bridges» between cognitive and meta-linguistic views of the knowledge structure. This is the parallel that the authors are looking for: an indication of the fact that categories serve as «nodal elements» of the hierarchical structure that combines and arranges the knowledge presented in form of concepts. Abstract nature of categories and their ability to build up hyponymic and hyperonymic connections between concepts are those very aspects that allow us binding them to meta-language and identifying categories as meta-linguistic elements and concepts as elements of the object language. Moreover, categories themselves are able to build up a hierarchically arranged structure; in this case, the point is that the next level of meta-language and meta-linguistic elements of the previous level (categories) would be considered as objects. Thus, the above described two-level process, which shifts from conceptualization to categorization and vice versa, flows in spiral manner and each turn of it brings cognition to a more abstract level.

V. GRAPHICAL PRESENTATION OF CONNECTION BETWEEN OBJECTS

As the authors have proved above, the meta-language is to be used as a basic tool (or one of the basic tools) for building a lexical-semantic model of a subject field. In the authors' opinion, its role is to arrange lexical units (terminological and non-terminological), that form a subject field, into a hierarchical structure. It seems that the most appropriate modeling method is an ideographic method developed by Yuri Karaulov on the basis of graph theory. Consequently, the authors will picture the model of the lexical-semantic structure of the subject field as a graph.

It is necessary to mention that within the current research, the authors will use a slightly modified version of the classical method of ideographic representation developed by Karaulov. This modification is caused by the fact that Karaulov's classical method considers only object language units (i.e. lexical units describing real objects) and a graph is built using these units only. Let us enhance the Karaulov's method by introducing the meta-language, which is a language of a higher level used to describe the object language. Within this research, let us consider lexical units describing the characteristics of bone injuries to be the elements of the object language, while the criteria, according to which these lexical units are grouped into the logical structures of a higher level (binary oppositions and scales), are considered as meta-linguistic elements.

So, a graph as a model of a subject field consists of two types of nodes. Nodes of the first type represent lexical units, or object language units; we will call them *leaf nodes*. According to the additivity concept, these nodes form oppositions consisting of two (binary opposition) or more elements. Criteria on the basis of which these oppositions are

formed are considered as meta-language units; they are represented by the second-type nodes. But it is necessary to mention that this scheme is ideal. In practice, not one but several levels of meta-linguistic nodes are required for appropriate representation of the lexical-semantic structure. Let us consider one of the grammatical categories of a noun – number. As is well known, this category is formed by two elements – singular and plural. The graph represents the structure of this grammatical category as follows: the node called “Number” is a meta-linguistic unit; it is considered to be the ‘parent node’ and forms the first level of the hierarchy. ‘Singular’ and ‘plural’ are the daughter nodes. They are also meant to be the meta-linguistic units and form the second level of the hierarchy. Each of these nodes, in its turn, describes lexical units of the next and final level of the hierarchy. These end nodes (leaf nodes) are object language elements: the ‘singular’ and the ‘plural’ nodes describe all nouns in singular and plural respectively. Thus, the lexical-semantic model of the number category, which consists of two meta-linguistic levels and one object language level, has been built.

Other grammatical categories of a noun – person, gender etc. – can be represented in the same manner because they all comply with the additivity concept. Eventually one can build a model representing all grammatical categories of a noun as a hierarchically arranged graph. This method of modeling is universal; it can be equally efficiently applied to building lexical-semantic models of all kinds of subject fields. Furthermore, this method of representation of a subject field structure serves as a comfortable basis for developing the qualimetric software, where graph-like models are ideal for qualitative and quantitative estimation of professional knowledge.

Now it is necessary to show how both the structure of knowledge and the principle of the cognition process can be represented graphically. A graph – instrument used in mathematical logic – is an ideal option for this. In the semantic research, one can find a number of graphs describing various phenomena [22; 23].

The reasons for this choice are obvious – being a comprehensible and extremely powerful tool, the graph is a very useful instrument for describing connections between objects.

It is evident that this graph starts with the concepts “a cat”, “a dog” and “a cow”. Let us consider these concepts to be the elements of object language; quite obviously, they form a category “Domestic animals”. This is the first level of the meta-language.

At the second level of the hierarchy, the above mentioned category together with another category of the same level called “Wild animals” forms a second-level category named “Animals”. Thus, at this stage, the first-level category becomes an object-language element for the second-level meta-language.

Category “Animals”, in its turn, is opposed to the “Plants” category; at this point they are considered as the elements of the object language so they build up a third-level meta-language category which is “Animate nature”.

Finally, two third-level categories form a fourth-level category “Nature”.

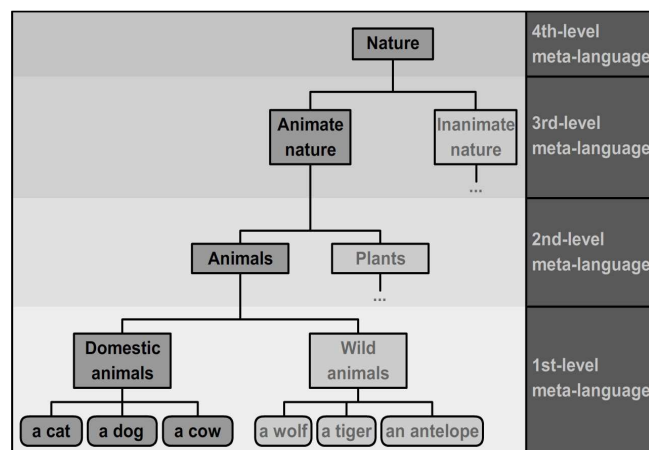


Fig. 1. Lexical-semantic group “Nature”

It should be mentioned that one can go along this graph both upwards (which is natural for the natural cognitive process) and downwards.

The above-shown example may look too simple and obvious, but it does its job well as it illustrates how logical concepts and structures can serve as points of contact between cognitive linguistics and programming.

VI. CONCLUSION

Such lexical-semantic structure as a graph demonstrates the two-level principle of human cognitive activity. After reaching, certain “critical mass” concepts form categories, and these categories, in turn, build up higher-level categories. Methods for program realization of graphs and graph related operations have been developed and successfully implemented a long time ago. That is why lexical-semantic structures (like the one shown above) can be easily actualized by means of programming and used as a basis for educational software and programs for testing.

The research conducted in this paper shows that any lexical-semantic group, as a rule, is a strictly arranged and hierarchically organized corpus of lexical units (terminological and non-terminological). Description of such lexical-semantic group should include not only the description of terminological and non-terminological elements that form the object language of the certain subject area, but also the description of the meta-language – and first of all, criteria used to form any kinds of semantic subgroups. A transition from semantics description to its modeling and visualization leads to building a graph consisting of two types of nodes; these nodes are represented in speech by the elements of the object language and the meta-linguistic elements respectively. This is what makes this kind of graph different from the classical graph structure proposed by Yuri Karaulov.

In conclusion, it is necessary to emphasize that building lexical-semantic models of subject fields is a massive amount

of research work. This research requires cooperative efforts of experts in various branches of science.

References

- [1] E. Kubryakova, "On place of cognitive linguistics among other cognitive studies and on its role in research of conceptualization and categorization processes," *Cognitive study of language*, vol. 7, pp. 15-19, Tambov, 2006.
- [2] A. Krasnov, A. Zhuravlev, and E. Sloeva, "Problems of ideographic representation of taxonomic-type expressions semantics," *Bulletin of Samara scientific center of the Russian Academy of Sciences*, vol. 15, , pp. 453-458, Samara 2013.
- [3] E.P. Stabler, "Meta-meta-linguistics," *Theoretical Linguistics* 37.1/2, pp. 69-78, 2011.
- [4] L. Hjelmslev, "Structural analysis of language", *Studia Linguistica*, vol. 1, Is. 1-3, pp. 69-78, June 1947.
- [5] V. Dobrova, and O. Kistanova, "Meta-languages: Nature and characteristic features," *Voprosy Kognitivnoy Lingvistiki*, Iss. 1, pp. 114-117, 2017.
- [6] H.B. Curry, "Foundations of mathematical logic," McGraw-Hill Book Co., Inc., N.Y., 1963.
- [7] A. Wierzbicka, "Making sense of terms of address in European languages through the Natural Semantic Metalanguage (NSM)", *Intercultural Pragmatics*, vol. 13, No. 4, pp. 499-527, 2016.
- [8] P.V. Grashchenkov, and I.M. Kobozeva, "Semantic classes and government of adjectives," *Komp'yuternaya Lingvistika i Intellektual'nye Tehnologii*, vol. 2, Iss. 16, pp. 134-149, 2017.
- [9] B.Y. Gorodetskiy, "Computer linguistics: Modelling of language communication," *Novoe v zarubezhnoi lingvistike*, Iss. 24, pp. 3-26, Moscow, 1989.
- [10] M. Mccord, and A. Bernth "A Metalogical Theory of Natural Language Semantics," *Linguistics and Philosophy*, vol. 28, Iss. 1, pp. 73-116, February 2005.
- [11] R. Jakobson, "Linguistics in its relation to other sciences", *Eight decades of general linguistics: The history of CIPL and its role in the history of linguistics*, pp. 265-304 November 2012.
- [12] A. Bazzoni, "Pure quotation, metalanguage and metasemantics," *Linguistics and Philosophy*, vol. 39, Iss. 2, pp. 119-149, April 2016.
- [13] T.J. Taylor, "Languaging, Metalanguaging, Linguistics, and Love," *Language Sciences*, vol. 61, pp. 1-4, May 2017.
- [14] A. Wierzbicka, "Language and Metalanguage: Key Issues in Emotion Research," *Emotion Review*, vol. 1, Iss. 1, pp. 3-14, 2009.
- [15] V.A. Bocharov, E.K. Voishvillo, A.G. Dragalin and V.A. Smirnov, "On problems of the evolution of logic," *Soviet Studies in Philosophy*, vol. 18, Iss. 4, pp. 31-52, 1980.
- [16] S. Edelman, "Language and other complex behaviors: Unifying characteristics, computational models, neural mechanisms", *Language Sciences*, vol. 62, pp. 91-123, July 2017.
- [17] I.S. Ladenko, "Methodology of semiotic modeling", Novosibirsk, 1987.
- [18] J.C. van den Herik, "Linguistic know-how and the orders of language," *Language Sciences*, vol. 61, pp. 17-27N, May 2017.
- [19] N. Love, "On languaging and languages," *Language Sciences*, vol. 61, pp. 113-147, May 2017.
- [20] Boldyrev, "Cognitive schemas of linguistic interpretation," *Voprosy Kognitivnoy Lingvistiki*, Iss. 4, pp. 10-20, 2016.
- [21] K. Abisheva, "Categorization and its basic principles", *Voprosy Kognitivnoy Lingvistiki*, vol. 2, pp. 21-30, Tambov, 2013.
- [22] N. Asher, and E. McCready, "Were, Would, Might and a Compositional Account of Counterfactuals," *Journal of Semantics*, 24 (2), pp. 93-129, 2007.
- [23] G. Jäger, "Using Statistics for Cross-linguistic Semantics: A Quantitative Investigation of the Typology of Colour Naming Systems," *Journal of Semantics*, vol. 29 (4), pp. 521-544, 2012.