Izvorni znanstveni radovi

UDK 81’572:165
165.194:81
Izvorni znanstveni članak
Prihvaćeno za tisak: 4. studenoga 2015.

Vedran Galetić
The interdisciplinary postgraduate doctoral study programme “Language and Cognitive Neuroscience”, University of Zagreb, Croatia
Ericsson Nikola Tesla, d.d., Zagreb, Croatia
vedran.galetic@unizg.hr, vedran.galetic1@gmail.com

Towards the cognitive plausibility of conceptual space models[1]

This work is focused on formal approaches in cognitive semantics, namely, the formalisation of the conceptual level of representations as the intermediate level between the symbolic and the connectivist one. An account of a selection of existing models is given. It is argued that one of the most important shortcomings that keeps the existing models from being truly cognitively plausible is the fact that they do not properly address the correlations between objects’ perceptible features, which are argued to be causally linked to the underlying, essential properties. The argumentation is supported by empirical evidence, implying the existence and importance of the causal effects in categorisation and inductive learning. It is therefore claimed that any cognitively plausible model of semantic representations needs to be able to adequately describe these cognitive phenomena, which has not been achieved so far. The paper qualitatively sketches out a cognitively motivated semantic representation model based on Gärdenfors’ conceptual space theory, endowed with the capability of describing the correlation of surface properties, thus supporting the notion of psychological essentialism.

1. Introduction

Traditionally, the opposing views of describing the meaning can be summarised into two major groups: the realist and the cognitive one (Raubal 2004). The realist approach posits that meanings are “in the world”, i.e., they exist independently from the observer. On the other side, the cognitive approach views meanings as being “in the head” of the observer, i.e., inseparable from a reasoning agent’s interpretation. This paper adopts the cognitive view, thus focusing on the cognitive representations. The representations may be defined as natural or constructed substates of a cognitive system that support the system’s purposeful interaction with its environment (Aisbett, Gibbon 2001).

1 This research was supported by a NetWordS grant – 09–RNP–089.

A cognitively plausible model of semantic representations, describing the world's image and structure within the cognitive system, has been an elusive target from the aspects of the cognitive science, lexical semantics, and artificial intelligence, specifically its subdomains concerning natural language processing. Such a model, following the concepts of cognitive science, should have high explanatory power and provide more relevant results in tasks based on imitating cognitive abilities (e.g., language) than models that aim to circumvent the cognitive plausibility.

Gärdenfors' (2004) conceptual space model represents an important step towards that end and is taken as a base model in this paper. It is argued that a model of cognitive semantic representations needs to address the property correlations and underlying conceptual theories in order to cover empirically supported cognitive phenomena of implicitly observing property correlations as an indirect basis for inductive inference. Current formal semantic representation models take into account these phenomena superficially at most. In this work it is argued that incorporating these important, empirically supported, cognitive phenomena is an important part of conceiving a cognitively motivated model of semantic representations.

Section 2 describes Gärdenfors' conceptual space model and reports a selection of attempts of its formalisation and extension. Section 3 brings forward theoretical and empirical findings concerning property correlations and their role in humans' concept construing and inductive inference. Section 4 roughly sketches a model that would comply with the described cognitive phenomena, briefly discusses the challenges of semi–automatized extraction of its parameters, and lists some of its possible applications in natural language processing. Section 5 concludes the paper and suggests future work.

2. Conceptual space models

There are three levels of cognitive semantic representations with respect to the level of abstraction (Gärdenfors 2004): connectivist (associationist), conceptual and symbolic (figure 2.1).

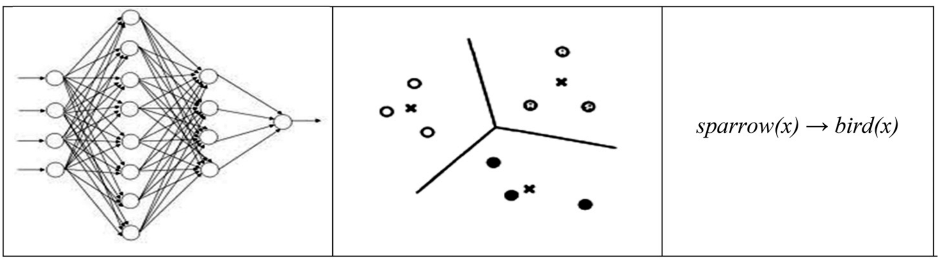


Figure 2.1. Illustration of abstraction levels of cognitive representations:

a) associationist b) conceptual c) symbolic

On the symbolic level, knowledge is abstract and cognitive processes are modelled as computing processes of the Turing machine. This level is most amenable for mathematical modelling of cognitive systems as it is appropriate for the application of formal, proposition or predicate, logic in symbolic manipulations representing cognitive processes.

The associationist (or connectivist) approach describes the system of knowledge as a densely interconnected network of primitive units. These units do not exhibit an explicitly expressible function; instead, they are relevant as a collection with emerging intelligence, the emphasis being on the types and strengths of their connections. The associationist approach is used in the artificial neural network modelling and computing paradigm.

The conceptual level lies between these two ends of the abstraction continuum (elaborated in Aisbett, Gibbon 2001). Cognitive semantics posits that meanings are mental entities, characterised by mappings from symbols to conceptual structures (Lakoff 1990). Thus, the concepts do have their explicit meaning yet they are abstract enough for formal description and computational treatment. Neither symbolic nor associationist approach succeed in explaining some important cognitive phenomena, such as concept acquisition, similarity description, and concept combinations; the conceptual approach provides a more appropriate basis for describing these phenomena. In addition, the conceptual level allows for geometric modelling of semantic representations, which arguably warrants its aptness for semantic representation from the point of neuroscience (e.g., Gärdenfors 2004: 67).

A geometrically organised representation space such as conceptual spaces are perceptually grounded and therefore a highly cognitively motivated model. In contrast, formal ontologies do not entertain cognitive plausibility, but instead they are oriented towards necessary and sufficient conditions that an instance needs to satisfy to be a part of a class, which is especially apt for application of formal logic (Schwering 2005), thus pertaining to the symbolic level of semantic representations. However, some recent exceptions exist, where attempts are made to extend the formal ontologies with the capability to represent the concepts with fuzzy boundaries and prototypical structure (e.g., Frixione, Lieto 2014).

2.1. Gärdenfors' conceptual space model

Gärdenfors' (2004) conceptual space theory yields a semantically interpretable and computationally amenable model of conceptual representations. A conceptual space can be represented as a geometrically organised vector space spanned by different quality dimensions, such as *colour*, *shape*, *size*, *mass*, *taste*, etc. A natural property[2] is defined as a convex region of a dimension (ibid.: 71). Convexity implies that if any two values are members of a set, then all values between them are also members of the set. (The relation of "betweenness" must also be defined for the observed system.)

2 The terms "quality dimension" and "property" will thus be used interchangeably throughout the paper.

A concept is a region within the conceptual space, representing a multi–property matrix. For example, a concept of *apple* is composed of certain regions of the quality dimensions of *colour*, *shape*, *taste*, etc. An object[3] is an exemplar of the pertaining concept and can be represented as a vector. E.g., a vector corresponding to "*this green, sour, medium–sized (…) apple*" has certain values for *colour*, *shape*, *taste*, etc., which are specific property values on the pertaining quality dimensions.

Similarity judgements are a prerequisite of differentiation and identification, both the fundamental components of reasoning (Aisbett, Gibbon 2001). In a geometrically organised conceptual space it is possible to define the relation of similarity between objects in terms of a suitably defined metric of the space (e.g. Euclidean, as in formula (R2.1), Manhattan, or other).

$$d_E(x,y) = \sqrt{\sum_i w_i \cdot (x_i - y_i)^2} \tag{R2.1}$$

Concepts as regions of conceptual spaces do not conform to Aristotelian view of categorisation, which presupposes that the membership to a category is of binary character, i.e., either something is or is not a member of a category. Instead, concepts are prototypically organised, which means that there are more and less typical objects of a concept.

The conceptual space theory is aligned with the tenets of cognitive semantics (discussed thoroughly in Gärdenfors 2004), hence the claim that conceptual spaces "provide an appropriate ontology for such a semantics" (ibid.: 159). Also, the conceptual spaces provide a formal elaboration of concept learning, a crucial cognitive functionality development, as well as concept combinations. These properties of Gärdenfors' theory arguably provide the legitimacy for the assessment of it being cognitively relevant, which was mentioned above as one of the primary goals of the conceptual approach of cognitive representation modelling generally.

2.2. Formalisations and extensions of Gärdenfors' conceptual space formulation

There are several noteworthy proposals of formalisation and extensions of Gärdenfors' conceptual space model. For example, Aisbett and Gibbon (2001) support and elaborate the claim that the conceptual level of representation is a link between the symbolic and connectivist levels. Furthermore, they provide a formal foundation for conceptual spaces modelled as vector spaces, geometrically organised based on different domains and the convexity of regions pertaining to concepts. They highlight the dynamics of conceptual spaces, manifested through context–dependent changes of attention areas and the ability of representing complex structures and relations.

3 The terms "object" will be used interchangeably with "instance" and "exemplar", following the notation in (Gärdenfors 2004).

Rickard (2006) also uses Gärdenfors' theory in devising a geometrically structured conceptual space model as an N^2–dimensional unit hypercube, where N is the number of quality domains spanning the space. Concepts are represented as points within the hypercube, while their coordinates are values of a graph connection matrix derived from co–occurrences of the concept's properties. A similarity calculus is proposed, both for between different observations as well as between an observation and a concept.

Raubal (2004) adopts the cognitive approach to describing the meaning, which states that meaning is inextricable from the reasoning agent, as opposed to the realist approach, positing the objectivity of truthfulness regardless of the reasoning agent. Highlighting the semantic interoperability as a precondition of successful communication, Raubal develops a formal apparatus for description of conceptual spaces as vector spaces spanned by quality dimensions, as well as the mapping between conceptual spaces of communication participants (e.g., a machine and a human). Also, he uses the Euclidean distance as instance similarity indication to quantify their distinctiveness with regards to the prototypical region of a concept. This formal system is used in a navigation case study.

With a similar emphasis on spatial cognition and spatial information systems as a use case, Adams and Raubal (2009a) use Gärdenfors' formulation and devise formalisations for metric conceptual space definition, semantic similarity measurement, and concept combination. The main aim of their work is to provide a mathematical foundation for facilitating the construction of conceptual space knowledge bases. They model concepts as convex polytopes, claiming such structures to be especially appropriate for computational operations. The system also incorporates the dimension of context, represented as quality domain weights. An interesting result of one of the model's possible applications is a table summarising country similarities with respect to different contexts (ibid.: 15). Building on the developed algebra, the authors propose the Conceptual Space Markup Language (CSML), an XML based interchange format facilitating the creation and sharing of conceptual structures using geometric information (Adams, Raubal 2009b).

The μw–model (Galetić 2011) is aimed at facilitating typicality quantification. The key parameters of the model are the property weight (w) for a concept (e.g. weight of *colour* for *bird*) and the typicality (μ) of the property value for the concept (e.g. the typicality of *red* for *bird*). An object, representing an instance of the pertaining concept, may then be represented as a vector, as shown in (R2.2).

$$\vec{r_C}(c) = \sqrt{\frac{w_1(C)}{\sum_{j=1}^{n} w_j(C)}} \cdot \mu_{A_1(C)}(x_1) \cdot \vec{e_1} + \cdots + \sqrt{\frac{w_n(C)}{\sum_{j=1}^{n} w_j(C)}} \cdot \mu_{A_n(C)}(x_n) \cdot \vec{e_n}$$

(R2.2)

where c is the observed instance of concept C, $\vec{e_1} ... \vec{e_n}$ are basis vectors representing the concept's quality domains (e.g. *colour*), $x_1 ... x_n$ are the object's values of the respective quality dimensions (e.g., *yellow*), $w_1 ... w_n$ are the weights of the respective quality domains for C, and $\mu_{A1} ... \mu_{An}$ are the typicality measures of the actual property values for C. The choice of the letter μ as part of the notation as well as the A's in the indices underline the reliance on the fuzzy set theory (Zadeh 1965) – namely, the μw–model observes the concepts as fuzzy sets, meaning that each object belongs to different concepts to various degrees with respect to each of its property values (e.g., a certain *colour* assigns to an object c_x the amount of 0,69 for "apple–ness", 0,84 for "fire truck–ness", 0,07 for "sparrow–ness", etc.).

The typicality or representativeness (R) of a real or imaginary instance of a concept is calculated as the Euclidean distance of the vector from the coordinate system origin, according to formula (R2.3). The prototypical instance has the maximal typicality, equal to 1.

$$R_C(c) = \sqrt{\sum_{i=1}^{n} \frac{w_i(C)}{\sum_{j=1}^{n} w_i(C)} \cdot \mu_{A_i(C)}^2(x_i)} \tag{R2.3}$$

Similarly to Rickard's model, the μw–model is also an abstraction of Gärdenfors' conceptual space model, thus able to address the problem of representing quality dimensions that take nominal values, such as, e.g., *movement*, *consistency*, *diet*. (For comparison, Raubal (2004) focuses only on the numeric variables, such as *colour hue, saturation and brightness, shape factors, visibility*, etc.) These abstract models are applicable for nominal quality dimensions along with the ordinal, as they abstract the properties by representing them indirectly, e.g., via the w and μ parameters. However, a shortcoming of the μw–model arising from this abstraction is a possible loss of uniqueness of representation. Namely, suppose that a *red pigeon* and a *yellow pigeon* are the same in all respects except for their *colour*, and that the typicality of these two objects with respect to the property of *colour* is the same; then these two objects would be represented by the same vector r according to (R2.2), resulting also in the same typicality value R, according to (R2.3).

2.3. The problem of correlations

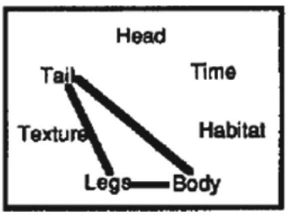
Although some of the described models arguably provide a certain level of cognitive plausibility – e.g., Gärdenfors emphasises many cognitive processes that his model covers (e.g., concept acquisition, induction); the μw–model is based on the prototype theory (Rosch 1973), a generally supported theory in cognitive science and cognitive linguistics – an important problem is present in each of them that prevents them from being fully relevant models of cognitive representations and processes. Namely, many models represent quality domains as orthogonal basis vectors, i.e., quality dimensions are independent from each other, whereas in reality the quality dimensions are often correlated within a concept structure. This problem is already mentioned by Raubal (2004: 10).

It is at this point fair to acknowledge Rickard's (2006) pioneering proposal of a conceptual space model providing description of correlations between different quality domains of a concept. The correlations are learnt from a prepared training set or set explicitly, and used for coordinates of concepts within highly abstracted conceptual spaces. Whilst this versatile model presents a very valuable resource of ideas for calculation and employment of property correlations, it may not be used as a state–of–the–art conceptual space model because the conceptual structures are reshaped as points instead of convex regions, thus sacrificing a certain amount of cognitive plausibility.

The current paper elaborates on the phenomena on the property correlations from the cognitive perspective, while also touching the notion of psychological essentialism. In the following section it brings forward a selection of empirical evidence emphasising these correlations as crucial for people's ability of concept construing and inductive inference. It also suggests a very rough draft of a model that would be in line with these cognitive phenomena, whilst sustaining all other cognitive plausibility criteria.

3. Property correlations

It has been demonstrated that people perform poorly at detecting isolated correlations, whereas they are excellent at detecting multiple correlations (Billman, Knutson 1996; Kornblith 1995; Jones, Smith 2002; Kloos, Sloutsky 2008; McClelland, Rogers 2003). The number of correlations is not the only parameter; systematicness of the correlations is also crucial for the ability of their detection. This is empirically confirmed by Billman and Knutson's (1996) experiment, where they presented to the subjects a sequence of imaginary animals and asked them to predict a missing property. The sequences were characterised by either bundled or independent correlated feature pairs, as in figure 3.1.



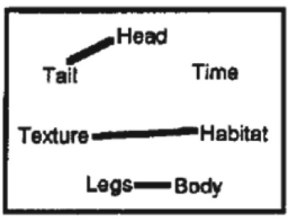


Figure 3.1. Examples of systematic and unsystematic property correlations (taken from Billman, Knutson 1996). Left: systematic correlations – the shape of tail predicts the shape of body and legs, and vice versa.
Right: unsystematic correlations – three independent correlation pairs: the shape of tail predicts the shape of head, the texture predicts the habitat and the shape of body predicts the shape of legs.

Furthermore, the natural categories (or natural kinds) are usually characterised by structured, bundled property correlations (e.g. Billman, Knutson 1996: 468; Rosch 1999; Gelman, Davidson 2013). E.g., considering the category "living beings", it is obvious that the value of the property "limb" is a good predictor of "locomotion" as well as "covering" (e.g., having wings is a good predictor of having feathers and the ability to fly, with a few exceptions, which need to be learnt explicitly).

Taking into account the observations that people are excellent at detecting structured correlations and that natural kinds are abundant with them, it can be concluded that people are adapted for successfully categorising natural objects. This is confirmed by Kornblith, claiming that "the way in which we detect covariation is precisely tailored to the structure of natural kinds" (1995: 102).

Natural kinds pertain to the categories of natural existence, as opposed to artefacts. Although artefacts also exhibit property correlations, they are not entertained in this theoretical development, since these correlations are of arbitrary nature, i.e., their properties are tailored to specific functionality that an artefact needs to fulfil (see, e.g., Ahn 1998). Thus for artefacts the analysis of property correlations and weights should rather be observed from the aspect of Gibson's (1977) "affordances" than reasoning agents' aptness for categorisation of objects based on cognitively effortless detection of property correlations.

3.1. Theory–theory

A crucial advantage of possessing the conceptual structure within the cognitive adaptive system is the ability to infer non–obvious properties based on induction originating from the information about the object's category (e.g., to infer that an object is dangerous provided it is a wolf; Ahn, Luhman 2005). In order to allow for such a capability, concepts need to be sound and well–structured, subsuming systematic correlations among features possessed by the pertaining objects.

One of the attempts to theoretically support the empirical findings that people are able to virtually effortlessly employ inductive inference in categorisation of natural kind exemplars via detected correlations is the "theory of theories", or "theory–theory". In the seminal work by Murphy and Medin (1985), theories are loosely defined as mental explanations relying on general world knowledge. Concepts are embedded in a complex network of existing knowledge (Ahn, Luhmann 2005), which can be represented as scientific theories. People use this preliminary knowledge for the conceptually influenced prediction of correct categorisation.

The theories are not of arbitrary nature, but rather the result of causal relations between the correlated properties. A similar theoretical account is brought forth in Rehder's (2009) work, stating that "ŠaĆlthough it may be vague or incomplete, one often has at least some general idea of the type of causal mechanisms that produce or generate a particular property". A question that naturally arises is the question of expertise, namely, whether it suffices to have an appropriate level of expertise in order to be able to create and employ the theories in inductive inference. For example, Quine (1969) reports that children rely exclusively on perceptually based similarity metrics in categorisation tasks (reported, e.g., in Murphy, Medin 1985: 290), whereas adults employ implicit knowledge that can be explained by the "theory–theory".

In addition, Murphy and Medin (1985) claim that children develop so–called "proto–theories" as early as aged two. Gelman and Markman (1986) concur by providing evidence that children as young as four tend to overrule the perceptual similarity in favour of the category membership information in inductive inference tasks. Gelman and Davidson (2013) empirically show that young children rely more on category membership than perceptual similarity, but only in case of natural kinds, which possess a well–developed conceptual structure. Furthermore, according to Murphy and Medin's (1985) "theory of individuality", children are able to correctly distinguish what kinds of objects receive proper names and which do not. A claim arises that people are biologically predisposed for developing theories, which is realised via perceptual and cognitive structures.

The "theory–theory" emphasises the role of underlying features that are essential for the concept coherence. Coherence of a concept indicates how easily one can visualise an object after removing one of its features (Sloman et al. 1998). For example, it is much simpler to visualise a bird that cannot fly than a bird that does not have a bird's DNA or internal organs. Feature immutability can be modelled as centrality of a feature in a network of features (id.).

3.2. Causal status hypothesis and psychological essentialism

In line with these accounts is the "causal status hypothesis" (Ahn et al. 2000a and 2000b), according to which features representing causal origins are more significant than those that are effects. The causal status effect "is expected to occur in categorization because people believe objects in the same category share the same essence which causes their surface features and because cause features are believed to have more inductive power than their effects" (Ahn et al. 2000b: 368). The causal status hypothesis has been proven by a behavioural task (ibid.) on adults, who evaluated identical features as having larger weight in categorisation when they served as causes than when they served as effects.

These theoretical accounts are shared by the philosophical notion of psychological essentialism, claiming that objects have essences or underlying natures that make them what they are. Kornblith (1995: 81) provides a summary of this idea, stating that "we are not at any time inclined to classify objects solely in virtue of their observable features, but instead take for granted that the observable features of an object are only an imperfect guide to their true natures".

According to this tradition, concepts are not just a list of characteristic features as represented by Gärdenfors' model and its described derivations, but instead naïve theories with causally related systems of features (Smith, Colunga 2012). Also, an argument in favour of conceptually modelled cognitive representations is them being a prerequisite for the ability of predicting non-obvious properties based on known category membership (Ahn et al. 2000b).

4. Towards a cognitively plausible conceptual space model

A system modelling cognitive representations and processes needs to be computationally tractable as well as take into account the property correlations as they are arguably s basis for concept learning and inductive inference. In other words, it needs to bridge the gap between the cognitive plausibility and computational amenability.

Whereas the surface quality dimensions span Gärdenfors' (2004) conceptual spaces and its reported derivations of it, a cognitively plausible conceptual space model needs to include the underlying quality dimensions that are causal origins of the surface dimensions. A general idea of such underlying quality dimensions can be provided with the following examples.

Take two objects of the concept *apple*, one *red* and *sweet*, another *green* and *sour*, as shown in figure 4.1a. If we observe more instances of the same concept, it is expected that they are approximately aligned along the line connecting the first two instances at the ends of the illustrated continuum. This would indicate that these properties are well correlated.

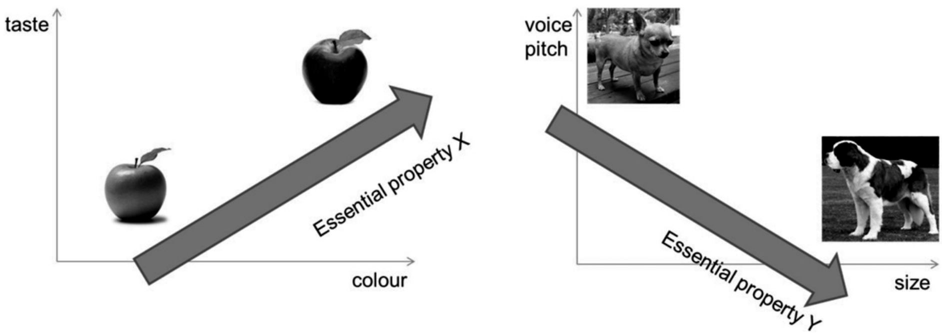


Figure 4.1. Surface and causally originating essential properties. Left example: a green and sour apple, and a red and sweet apple. Right example: a small dog with a high pitch voice, and a large dog with a low pitch voice

A well–known prerequisite for statistical inference is discerning between correlation and causation. The current qualitative analysis focuses on the natural kinds, whose variety is a result of the evolutionary development. From the ontological point of view, evolution brought about the correlations between natural kinds' features. Therefore, the correlations between properties of an object may be assumed to result from the evolution–based causation. In the current example this causality is modelled as an underlying, "essential property X", inherent to the natural concept, having the well–correlated surface properties as effect. This and other underlying properties are the underlying quality dimensions that should span a cognitively plausible conceptual space model.

Another example is shown in figure 4.1b. The two dogs represent two ends of the illustrative continuum of body size and voice pitch. If we observe additional dog instances and represent them in the same coordinate system, again, they would be roughly aligned along the line connecting the first two. These two surface properties are indicators of the underlying property, shown in figure 4.1b as "essential property Y".

As reported earlier, the selection of empirical findings supporting the existence and importance of underlying concept theories suggests that a sound concept possesses a coherent structure. Such a structure indirectly brings about the correlation among the concept's properties. The correlation and mutual prediction of the perceptually observable surface features can be used as indicators of the underlying essential properties that constitute the concepts' essence (e.g., "apple–ness" and "dog–ness") upon which the conceptual theories build. It is important to observe that within such a system concepts are still represented as convex regions, conforming with the postulates of conceptual space theory as well as cognitive semantics in general.

It is worth noting that in both examples the quality dimensions are named arbitrarily, by "X" and "Y". Namely, it can be argued that language is tailored to the surface properties and rarely do underlying properties have names. Counterexamples are mostly found in jargon of domains of expertise. For example, when classifying wines, experts are well–aware of naturally motivated causes that bring about the perceptual (visual and gustative) characteristics of a class of wine and they are able to competently predict a property based on another or a set thereof (e.g., if an opaque colour is observed, the wine is expected to be full–bodied; if told that a wine originates from a well–insulated terroir, a higher grape ripeness is expected).

4.1. Challenges of automated model quantification

Being aware of the plethora of currently available digital linguistic resources, a question arises whether it is possible to utilise them for automatic or semi–automatic qualitative and quantitative description of the hypothesised cognitively plausible model's parameters, e.g., definition of quality dimensions and quantification of their weights[4]. Undoubtedly, any such endeavour would be greatly challenging as, in the first place, the model would be based on quality dimensions that are essentially not explicitly referred to by language, as mentioned previously.

One may find detecting these underlying quality dimensions as evocative of the Latent Semantic Analysis (LSA), which is a well–established technique in lexical semantics, especially in document classification and similarity quan-

4 Represented, e.g., as conceptual centrality (Sloman et al. 1998).

tification. However, the dimensionality reduction as a crucial part of LSA is of questionable utilisation value for the current research as the interpretation of a low–dimensional result is often devoid of plausible interpretation.

However, despite these discouraging observations, there may be an avenue for research of automatic model quantification based on the available linguistic resources. Namely, since it is argued that the underlying quality dimensions bring about the correlations of causally resultant surface properties, it should be possible to use these correlations, found in corpora, as indications of the underlying quality dimensions. To that end, a combination of an array of complementary resources may be utilised, namely, corpora (e.g., Davies 2013) for finding correlations, and ontologies (e.g., OpenCyc[5]), Wordnet[6] (Miller 1995) and McRae's NORMS[7] for determining congruent properties and systematizing features as points or regions upon corresponding quality domains. Moreover, Rickard's (2006) work including property correlations is likely to serve as a useful guideline. An empirical research may also be necessary in order to validate the obtained model and provide a benchmark for model testing.

Such a research programme would bridge the gap between the two approaches of semantic modelling (Andrews et al. 2009): the experiential approach and the distributional approach. The experiential data refers to mental representations with an emphasis on objects' empirical descriptions, making these data perceptually grounded. Exemplars of the experiential approach of semantic modelling are McRae's (2005) NORMS and Gärdenfors' (2004) conceptual space model spanned by perception– and experience–based quality domains. On the other hand, distributional data is based on the linguistic use of lexemes connoting concepts or exemplars thereof, thus the distributional approach relies heavily on the corpus linguistics.

4.2. Possible applications

An operationalised cognitively motivated conceptual space model that is formally described and can be quantified in an automatic or semi–automatic manner would provide a valuable resource for computational linguistics, with an emphasis on lexical semantics. Its applicability and benefit may be assessed in problems such as:

- measuring the similarity between concepts, e.g., apple and orange. It should be noted that comparing two concepts will here be interpreted as comparing their prototypical objects. The traditional distributional semantics approach (Turney, Pantel 2010) compares the feature vectors representing each respective context obtained from the corpus. Incorporating the feature weights (and other parameters, e.g., typicality of a property value for an observed concept) in this comparison should yield a more precise and cognitively relevant result;

5 http://www.cyc.com/platform/opencyc (accessed on 8 June 2015)
6 http://wordnet.princeton.edu/ (accessed on 8 June 2015)
7 The human–generated feature database available at: https://sites.google.com/site/kenmcrae-lab/norms–data (accessed on 8 June 2015)

– distinguishing the relation of similarity (e.g. *airplane–rocket*) from the more general relation of relatedness (e.g. *airplane–pilot* are related, but not similar) – once a distributional method extracts a pair of related concepts, a conceptual space model may be used to refine the relation by elimination method as it covers only similarity, (thus far) not relatedness;
– detecting semantically opaque feature–concept combinations through low typicality assessment (e.g., using the μw–model);
– utilising the quantified parameter of quality dimension weight in an already existing formal semantic model, e.g., in dimensionality reduction process within the "distributional memory" model by Baroni and Lenci (2010) for the attribute–noun combinations.

5. Conclusion and future work

Gärdenfors' conceptual space model is a thoroughly developed theoretical semantic representation model focused on the description of conceptual structures and quality domains that constitute them. The paper brings forth an overview of this as well as other similar or derived models of conceptual spaces, highlighting their deficiency of cognitive plausibility due to their inability of adequate property correlation description.

It then elaborates the indispensability of property correlation awareness for humans' ability of effortless categorisation, supported by a selection of reported empirical evidence. An approach is proposed that would yield a conceptual space model taking a step further towards the cognitive plausibility by incorporating the correlation of properties realised via causally originating essential properties constituting conceptual coherence.

It is argued that the existing linguistic resources, such as ontologies, corpora, and human–generated feature lists, may prove beneficial in automatic or semi–automatic formalisation and parameter quantification of such a model. The future work will provide further evidence on the applicability of these resources with an emphasis on text corpora, thereby bridging the gap between the experiential and the distributional approach in semantic modelling.

Literature

Adams Benjamin, Martin Raubal (2009a). A metric conceptual space algebra. *Spatial Information Theory*. Springer Berlin Heidelberg, 51–68.

Adams Benjamin, Martin Raubal (2009b). Conceptual space markup language (CSML): Towards the cognitive semantic web. *2009 IEEE International Conference on Semantic Computing*. IEEE, 253–260

Ahn Woo–kyoung (1998). Why are different features central for natural kinds and artifacts?: the role of causal status in determining feature centrality. *Cognition* 69: 135–178.

Ahn Woo–kyoung, Christian C. Luhmann (2005). Demystifying theory–based categorization. Gershkoff–Stowe, Lisa, David H. Rakison, eds. *Building object categories in developmental time*. New Jersey, Lawrence Erlbaum Associates, Inc., 277–300.

Ahn Woo–kyoung, Susan A. Gelman, Jennifer A. Amsterlaw, Jill Hohenstein, Charles W. Kalish (2000a). Causal status effect in children's categorization. *Cognition* 76.2: B35–B43.

Ahn Woo–kyoung, Nancy S. Kim, Mary E. Lassaline, Martin J. Dennis (2000b). Causal status as a determinant of feature centrality. *Cognitive Psychology* 41.4: 361–416.

Aisbett Janet, Greg Gibbon (2001). A general formulation of conceptual spaces as a meso level representation. *Artificial Intelligence, 133.1: 189–232.*

Andrews Mark, Gabriella Vigliocco, David Vinson (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review* 116.3: 463.

Baroni Marco, Alessandro Lenci (2010). Distributional memory: A general framework for corpus–based semantics. *Computational Linguistics* 36.4: 673–721.

Billman Dorrit, James Knutson (1996). Unsupervised concept learning and value *systematicitiy*: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22.2: 458.

Davies Mark (2013). Corpus of Global Web–Based English: 1.9 billion words from speakers in 20 countries. Available online at http://corpus2.byu.edu/glowbe/.

Frixione Marcello, Antonio Lieto (2014). Towards an extended model of conceptual representations in formal ontologies: A typicality–based proposal. *Journal of Universal Computer Science,* 20.3, 257–276.

Galetić Vedran (2011). Nasrtljivi crvendać u dvorištu: Formalna kvantifikacija razine prototipnosti u okviru prototipne semantičke teorije kognitivne lingvistike. *Suvremena lingvistika* 71: 27–56.

Gelman Susan A., Ellen M. Markman (1986). Categories and induction in young children. *Cognition* 23.3: 183–209.

Gelman Susan A., Natalie S. Davidson (2013). Conceptual influences on category–based induction. *Cognitive Psychology* 66.3: 327–353.

Gibson James J. (1977). *The theory of affordances.* Hilldale, USA.

Gärdenfors Peter (2004). *Conceptual spaces: The Geometry of Thought.* MIT Press.

Jones Susan S., Linda B. Smith (2002). How children know the relevant properties for generalizing object names. *Developmental Science* 5.2: 219–232.

Kloos Heidi, Vladimir M. Sloutsky (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General* 137(1): 52–72.

Kornblith Hilary (1995). *Inductive inference and its natural ground: An essay in naturalistic epistemology.* MIT Press.

Lakoff George (1990). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

McClelland James L., Timothy T. Rogers (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience* 4.4: 310–322.

McRae Ken, George S. Cree, Mark S. Seidenberg, Chris McNorgan (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods* 37.4, 547–559.

Miller George A. (1995). WordNet: a lexical database for English. *Communications of the ACM* 38.11: 39–41.

Murphy Gregory L., Douglas L. Medin (1985). The role of theories in conceptual coherence. *Psychological Review* 92.3: 289–316.

Quine Willard V (1969). *Natural kinds*. The Netherlands: Springer.

Raubal Martin (2004). Formalizing conceptual spaces. Formal ontology in information systems, *Proceedings of the third international conference (FOIS 2004)* Vol. 114. 153–164.

Rehder Bob (2009). Causal–Based Property Generalization. *Cognitive Science* 33.3: 301–344.

Rickard John T. (2006). A concept geometry for conceptual spaces. *Fuzzy Optimization and Decision Making, 5.4: 311–329.*

Rosch Eleanor H. (1973). Natural categories. *Cognitive Psychology* 4.3: 328–350.

Rosch Eleanor (1999). Principles of categorization. Margolis Eric, Stephen Laurence, eds. *Concepts: core readings*. MIT Press, 189–206.

Schwering Angela (2005). Hybrid model for semantic similarity measurement. *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE*. Springer Berlin Heidelberg, 1449–1465.

Sloman Steven A., Bradley C. Love, Woo–Kyoung Ahn (1998). Feature centrality and conceptual coherence. *Cognitive Science* 22.2: 189–228.

Smith Linda B., Eliana Colunga (2012). Developing categories and concepts. Spivey Michael, Marc Joanisse, Ken McRae, eds. *The Cambridge Handbook of Psycholinguistics*. Cambridge University Press, 283–308.

Turney Peter D., Patrick Pantel (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37.1: 141–188.

Zadeh, Lotfi A. (1965). Fuzzy sets. *Information and Control* 8.3: 338–353.

Kognitivna plauzibilnost formalnih modela semantičkih reprezentacija

Gärdenforsov model konceptualnih prostora veoma je razrađen semantički model orijentiran opisu strukture koncepata i kvalitativnih domena koje ih sačinjavaju. U ovomu se radu daje pregled tog i izbor drugih sličnih modela konceptualnih prostora te se naglašava nedostatnost njihove kognitivne plauzibilnosti uslijed zanemarivanja ili kognitivno nedosljednoga opisa korelacija svojstava.

Postojanje i važnost korelacijskih učinaka u kategorizaciji i induktivnomu učenju potvrđeni su rezultatima mnogobrojnih empirijskih istraživanja temeljenih na kategorizacijskim i generalizacijskim procjenama ispitanika. Dakle, svaki kognitivno vjerodostojan model semantičkih reprezentacija mora uključivati fenomen kognitivnoga sustava koji se odnosi na automatiziranu detekciju strukturiranih korelacija svojstava. Ti se korelacijski učinci opisuju »teorijom teorijā« (engl. *theory–theory*), koja tvrdi da kognitivni sustav sadrži mikroteorije o konceptima, a te mu mikroteorije omogućavaju sposobnost automatiziranoga induktivnoga učenja kategorizacije prirodnih vrsta (engl. *natural kinds*) putem uočenih korelacija među njihovim svojstvima.

U ovomu se radu tvrdi da su svojstva dobro strukturiranih, koherentnih koncepata uzročno povezana, a ta se povezanost modelira dubinskim, uzročno ishodišnim kvalitativnim dimenzijama. Kao budući rad spominje se istraživanje mogućnosti automatiziranog ili poluautomatiziranog crpljenja parametara predloženoga modela uporabom digitalnih jezičnih resursa kao što su ontologije, Wordnet, liste svojstava koncepata koje su generirali ispitanici temeljem osjetilnih iskustava, odnosno korpusi. Time bi se ostvarilo kombiniranje iskustvenih podataka s podatcima temeljenim na jezičnoj uporabi, što bi tako dobiveni model činilo vrijednim resursom, i za kognitivnu znanost i lingvistiku, kao i za računalnu obradu prirodnoga jezika s naglaskom na semantiku.

Keywords: cognitive semantics, prototype theory, categorisation, psychological essentialism, property correlations

Ključne riječi: kognitivna semantika, teorija prototipova, kategorizacija, psihologijski esencijalizam, korelacija svojstava