

A socio-cognitive approach to semantic analysis and categorization processes

Aleksander Z. Wassilew, Warsaw School of Economics, alex@sgh.waw.pl

Abstract

Information systems design, when viewed from a human-computer interaction perspective, needs to account for the cultural language semantics. Socio-cognitive approaches are helpful in understanding how the mind processes and models acquired knowledge in order to solve problems. Understanding needs knowledge of content and context for constructing a meaning, which like the communication has interactive structure. Interpreting as well as understanding is a subjective process. The relation perceived between compared objects, included in the same class, is a function of knowledge – the "theory" about a part of reality. The representations of perceived objects, called structure of qualis can be helpful in a minimal explanation of the meaning of the words. The classical goal of cognitive psychology is to solve, if the generalisation and assessment of similarity of the objects is important for identification and categorization. A modified method of pairwise comparison may be used for the assessment of qualis, the measurement of similarity between the objects for categorization, and in the research on "non-sharpness" of semantic fields, semantic spaces, psychological (and semantic) distance, for describing the membership functions, its measure, for quantification of information, in the research on the information's asymmetry etc..

Keywords: socio-cognitive, semantics, categorization, knowledge, similarity.

Introduction

Information systems (IS) are made for and by the people. IS are helpful tools for improving actions and processes within organizations; especially for routine, formal, and algorithmic tasks. For that reason, it becomes important to understand the essence of searching and processing information, as well as research the very essence of information, knowledge, and the art of using information (Wassilew, 2001). Information is defined as a change in the state of user's knowledge about something caused by the receipt of new, external raw data.

To better understand the role and function of information and knowledge, we should know: *Who knows and how understands it? How exactly it will be used? Who receives the message and what gives it a concrete meaning? Who is the information/knowledge for?* Knowledge is the result of processed information that has been put into context of a situation. (Wassilew, 2009).

From a human-computer interaction perspective, further research is needed to understand the semantics of cultural language differences as well as the interactive behavior between humans while designing Information systems. Language serves as a communication mean between people. It is also a perceptual filter of reality and a group memory of social experiences.

The basic model of knowledge acquisition relies on close group interactions where knowledge engineers mediate between experts and knowledge base systems or data mining. A knowledge engineer, interactively extracts data and information from experts and represents it in the

knowledge base. We should be very careful when we speak of data and information and the symbolic representations of the human intellectual processes.

The main goal of this paper is to present a socio-cognitive approach to how the mind acquires and processes knowledge. Socio-cognitive approaches are helpful for understanding the mind's processes, solving communication problems, and resolving questions connected with acquisition and modelling knowledge. Next, a modified method of consistency-driven pairwise comparison was proposed for the assessment of *qualis*, the measurement of similarity between concepts and for categorization. This can be helpful in research on "non-sharpness" of semantic fields, its measure, quantification of information, information asymmetry, psychological (and semantic) distance, semantic spaces, for describing the membership functions etc.

Socio-cognitive approach

According to Hjørland (2004), in traditional cognitive science the role of culture and society in cognition is marginalized; Understanding individual cognitive focus on symbolic representations is not enough. This increased an importance of semantic and pragmatic research to show how human and technological factors and organizational behaviors had been integrated together.

Cognitive and social properties are reciprocal. Thus, it becomes essential to understand how cognitive properties of the components of the socio-cognitive systems and invariant social interrelations are networked in order to determine their behavior. (Gadomski, 2002).

Gadomski (1997) explained systemic socio-cognitive metaphysics as focusing only on the observable (as well as by introspection) functions of the human brain in cognitive and social relationships. Van Dijk (2009) described a discourse as a socio-cognitive interface for the discursive interaction of relationships between the mind and society.

Again, Hjørland (2004) shows that socio-cognitive approaches emphasize the role of culturally produced signs and symbols and the way cultural, historical and socially constructed meanings mediate cognitive processes. Such approach is useful for describing integrated cognitive and social properties of systems, processes, functions, and models as well as socio-cognitive interactions (Hemingway & Gough, 1998).

Additionally, the socio-cognitive approach, as an interdisciplinary perspective, presents the systematic relationship between fluctuations of the semantic category and the socio-demography of a community (Robinson, 2010).

Pragmatics, communication, meaning

The pragmatics of language deals with the relations between the language phrase and the user, taking part in a communication process, where the user can be a sender or a receiver of the generated message. The knowledge of the pragmatics of language is very important for understanding and explaining the functionality of information systems (Oleński, 2000). Pragmalinguistics are considered authentic events in a language, especially lexicon (vocabulary) and meaning, related to the context of a situation. The traditional logic appears not to be very useful in the reconstruction of the relationships between particular utterances in a colloquial

conversation. Thus, it is necessary for the consideration of specific units such as presuppositions, implications, conversational logic, etc. (Zgółka, 1996).

Description of the world's picture, closed in the semantics of the natural language, cannot be objective, because natural semantic units have no structure or function as scientific terms. The language is developed as an intersubjective tool for the mutual sharing (interchange) of subjective experiences of meaningfulness (i.e. a capability to transfer a meaning and widening the space of meaningfulness). "Interpreting" means "to understand", but understanding, as a subjective process is not one-sided. Understanding requires knowledge of the context (i.e. a knowledge of who the subject is and how he lives and behaves). Each interpretation is a hypothesis.

Communication is an interactive process. In the structuralist's opinion, the meaning is the goal. Therefore, it cannot be a means to achieving the goal. The meaning is discovered through the usage of meaning (Lester, 1971). Contents included in the information systems are translated through their usage. All we know about any culture is based on our perception, not on communication. The environment as an open system is difficult to describe and plays an important role in explaining the meaning of the words. The meaning has an interactive structure (Putnam, 1988). The mind creates some aspects of the reality through linguistic and cultural interactions.

The content is only one side in the process of constructing a meaning. Another, are the communicators and the context of the communication and the behaviour (Anderson, 1987). A signal may or may not be communicative, depending on the ability of choice (of the sender or the receiver). One of the basic principles in the semantics is the choice or the ability to choose a necessary condition, but it is not sufficient for the meaningfulness. Meaning assumes the ability to choose.

A misunderstanding between a sender and a receiver can have its source in the often applied simplifying assumption of the information's identity. This imputed in the signal by the sender as the choice of one of the existing possibilities, and this taken by the receiver through the choice of one of the (poorly assumed) possibilities. No word, phrase, or sentence, in any language is fully connected by a unique relation with the entity (object/event) or the expressed concept. After recognizing the concept, one needs to research the relationship between them for their classification to determine which category to include the word.

Neuro-semantic and categorization

Recognition should have the character of a relation, connecting one's perceptual category with another ("object", "up-down", "inside", etc.) and is not a result of a convention or an arbitrary setting. The development of the perceptual category does not demand the existence of a society using this language. The concepts include relationships with the real (outside) world, the trace in memory and of a previous behaviour. The meaning develops as a result of the memory's interaction "value-category" with active concept's areas and language fields. In interactions with the memory "I-others" (through categorizations of verbs) it makes possible to model the world, to create comparisons and assessments, anticipate results, and a modification of plans. The perceptual categorization has probabilistic nature (Edelman, 1998).

For the consistency of concepts, the models and the "theories", which people have about the world, have an important role. Perceiving the relationship between compared objects, which are included in the same class, is rather a function of knowledge, i.e. the "theory" about a given part of the reality. The conceptual organization of objects and events is dependent on accepted "theory" and for that reason a similarity between elements of a given class state a consequence of an accepted way of categorization. A given "theory" is accepted, because of given conditions, which allow the prediction of the future events, which are the best considerations for accepting these predictions in decision making.

It is necessary to assume that the way of categorization, as well as determining relationships in the setting of these compared events can be different, depending on possessed knowledge (i.e. the way of seeing the surrounding reality). Creating category and connected with these established relations between things, allows an optimal adaptation of human to the environment and this is closely related to the motivational aspects of the behaviour.

The perceived and remembered objects receive an open mental representation. There is an assumption, which different operations (cognitive processes) are made of these representations, dependent on the task done by the subject. For understanding the task about the similarity of data, it is necessary to know not only the specification underlying representations of similarity, but also the cognitive processes acting on those representations (Nosofsky, 1992).

The cognitive models are conceptualized as a pair "representation-process" (Anderson, 1976). The representations called a structure of *qualis* can be considered as helpful in a minimal explanation of the meaning of the words. The linguistics' meaning is based on a non-linguistic experience. There exists a connection between a meaning appearing in the language and other behaviours, for establishing relationships between the content and the meaning, it is helpful to interpret the behaviour. The value of the word is possible to establish only by defining it in relationship to other words in its neighbourhood and the opposite. Only as a part of the whole word has meaning, only in the boundaries of the field, meaning exists (Trier, 1934).

In the traditional approach the meaning is seen as an imminent feature of the words, i.e. as "objects" (representations), on which it is possible to look at from different points of view, but stating a stable "semantic net". Its elements and distances between them leave invariable and independent of the observer's position. There does not exist a connection or an interaction between the subject and the object of observation. But after the theory of relativity and quantum mechanics, it is not possible to ignore the influence of the observer's participation during the investigation of different events even in the semantics.

In the reality the observer (subject) states for a difference of quantity as well as quality of the structure, in different situations, periods of life, physical states etc. A different way of constituting of the subject may be important for the choice, which objects will appear in his mind as meaning (semantic representations) (Stamenov, 1993).

The classical topic of the cognitive psychology is to solve, if the same principles of generalization and assessment of similar stimuli appear in the base of identification and categorization. A classification of the object is determined by the similarity to the individual elements in an alternative category.

Categorization can be seen as an exchange between accessibility of information and its predictive power. Knowledge has its base in expectations, it is a weak assumption for the environment, functioning as "working hypothesis" (preparing for uncertain future) (Winter, 1998).

The functional grammar (Halliday, 1985) is a unique way of describing how language is used in social situations. It is not a rule of grammar, but rather an analysis of language and its functions. The context is one of the keys to allow understanding how individual parts of the text are pacing to a wider perspective of the language as a whole. A net system is a theory of the language as a choice. It represents the language, or any of its parts, as a resource for creating meaning by choice. The meaning is determined by the context, and also by a lexical choice. It should be marked, that "social-semantic" (sociolinguistic) approach have a big heuristic value, but it is difficult to translate it to hypothesis for empirical verification.

Semantic analysis and categorization

In the analysis of a question, what a word means, it is possible to define component features. Another question may be differentiating the meaning of individual words, as well as in analysing the relations between similar words. They are a base for research the similarity between the sets of words. The two-sided dichotomy is one of the most important principle ruling the building of the language (Hatch, Brown, 1995).

Therefore, for describing the features, dichotomic scales are used. For example, <+male>|<+female>.t is observed, that features possess metafeatures: <+graduate>|<+continuous> (Sapir, 1944).

For defining a word we describe features, minimal set features, sufficient for the unique identification of the word's meaning. Usually there exists a tendency (relatively sufficient) to describe the word "generally" and classifying it "yes - no" or "0 - 1". The terms are subjective and not precise, not concerning to absolute place in dimension. More apparent oppositions are in reality extreme points on the continuous scale, especially for adjectives. Comparing two entities, considering the presence or the absence of one or more features, is made on the basis of similarity in other considerations. There is no evident way to share all those scalar points in the "plus - minus" category (Hatch, Brown, 1995).

By analyzing the semantic features there is no weight assigned to the primary meaning, and there needs to be another way to consider the fact, that some of the word's meaning appears more "central" than another word. Core (opposite to periphery) has been considered in the meaning of particular words, which are the most central, primary, or invariant. A prototype is the best representative instance of the concept. Empirical determinant for that event may be this, that it is categorized to a given class the most rapid, and often with minimal erroneous classifications.

An important question at the core of the research is, if the polysemy of similar words has to be considered as different lexical items? Some of these meanings are more central or core-like in comparisons with others. The problem is to show to what degree the meaning is central or core-

like and how they have changed, when they go far from the core. This demands using dimensions.

The membership of elements into particular categories can be gradual. Categories should not have explicit boundaries. It is possible to observe gradual membership to categories, gradual power of connection within the central model, family similarity, not-hierarchical gradually in which central categories are dominating, and appear prototype effect.

Knowledge and categorization

The basic model in the acquisition of needed knowledge relies on group work, where a knowledge engineer mediates between experts, people, and a knowledge base. Knowledge engineers acquire knowledge interactively from experts and represent it in a knowledge base. An approximation of discrete relations through the rating of preferences reduces the order and probability of the expert's opinions. This is a useful tool for conclusion, applied to knowledge base systems. It is possible to determine the categorization between the concepts, as well as the roles, played, only by considering their given terminology definitions and structural comparison of adequate terms. This leads to the categorization of hierarchy of those terms, concepts and taxonomy of roles. Comparing to the databases, in knowledge bases terminology representation has a more open approach to the world.

In creating a knowledge base, it is possible to apply brainstorming sessions, the gathering of a mutual knowledge on a given topic. This can stimulate a more creative approach in searching solutions of the considered problem. In comparing a meaning of particular concepts we describe semantic space for seeking a solution. By a reduction of the inconsistency we can assess the metric distance between the concepts in that space, and in that way separate more general categories of used concepts. In theories of categorization there is assumed that the basis of classification of objects of a category is to conceive (depict) their relations of similarity.

Problems can be analysed on the base of cognitive representation P from research on categorization: $P = \langle A, c, d, m, p, r, s, w \rangle$ (Chlewiński, Falkowski, Francuz, 1995), where: A – area of cognitive process; c – a set of context features, knowledge, encoded in cognitive structure of man, created in the process of gaining an experience, ability to perceive similarity and categorization in the set of similar events, includes pragmatic aspect of cognitive situation of categorization; d – shows the similarity degree between two events (considered as a metric distance); m – membership function in $\langle 0,1 \rangle$; r – a set of syntactic and semantic rules; p – prototype event; s – a set of presuppositions; w – a set of possible worlds.

Similarity evaluation by pairwise comparison

Categorization is important to assign precise similarities to compared events. One of the most fundamental assumptions for exact evaluation is the use of precise scaling systems for all relations (often by expert's assessment). The goal is achieved by using pairwise comparison for all assessed relations. This method is also used to assess the relations pairwise, which is easier and more precise than considering all relations at the same time (Koczkodaj, 1996).

There is not a standard scale or metric system for semantic systems. The absence of metric systems may not be a problem when we can compare one object with another. The assessment of similar meanings, distances in semantic space of prototype to concrete elements, demand capability of the assessment of mental representations, qualities enabling perception of the outside world and the categorization and evaluation of observed events or objects – *Qualis*. For subjective comparing one can use prescaled table for assessment of similarity distance (Table 1). These improve the precision of assessment.

Code	Similarity relation or meaning	Explanation
0,1	There is no similarity	Two objects do not have similarity while compared
0,3	Weak similarity	There is very little similarity between objects (concepts)
0,5	Essential similarity	There is an essential similarity while comparing two objects
0,7	Big similarity	Very big similarity of meaning of two compared objects
0,9	Identity	Synonymy identity of meaning of two compared concepts
0,67 etc.	Compromise values	When compromise is needed

Table 1. Scale for assessment of similarity relation - comparing, identification.

Five words (concepts) were chosen for illustration of the idea: data, message, information, knowledge and wisdom. They are often used, but confused in the contemporary discourse.

The similarities are assessed accordingly to the context of the present articulation. The values of “distances” are evaluated by intuition (subjectively), still consistent with the considered material (Table 2):

	data	message	information	knowledge	wisdom
data	1	0,8	0,5	0,3	0,1
message	0,8	1	0,6	0,4	0,2
information	0,5	0,6	1	0,8	0,3
knowledge	0,3	0,4	0,8	1	0,5
wisdom	0,1	0,2	0,3	0,5	1

Table 2. Similarity evaluations by intuition.

“Information is the most central (core) conception in the semantic space. The high values of some similarities indicate the interchangeability of the established words. Very often “data” and “information” are treated like synonyms, and also “information” and “knowledge”. “Message”

and “wisdom” are not concepts used in the work, but they would have the presented values, if they were.

The effects of the above method were compared with one arbitrary chosen program for similarity evaluation – WNSim, developed by Cognitive Computation Group at University of Illinois – of the same words (concepts) (Table 3.):

	data	message	information	knowledge	wisdom
data	1	0	0,3	0,09	0
message	0	1	0,3	0	0
information	0,3	0,3	1	0,3	0,027
knowledge	0,09	0	0,3	1	0,09
wisdom	0	0	0,027	0,09	1

Table. 3. Similarity evaluations by WNSim (<http://cogcomp.cs.illinois.edu/page/software>).

It seems that “data”, “message” and “knowledge” are used like synonyms of “information”. Like a base of investigations, they are used as stationary knowledge sources - dictionaries, thesauri, etc., but these resources are inflexible and not taking into account the concrete context (Do & Roth, 2010).

Summary and conclusion

Information systems are made for people to help them to improve, especially routine, formal and algorithmic tasks. The design of human-computer interaction systems and interactions between human groups requires research into the semantics of languages. The socio-cognitive approach is helpful for understanding the mind's processes and for understanding and solving problems with acquisition and modelling of knowledge.

The meaning like the communication has interactive structure. It is discovered throughout the usage of the meaning. Interpreting and understanding are subjective processes. Understanding needs knowledge of content and context for constructing a meaning.

The relation perceived between compared objects, included in the same class, is a function of knowledge – the "theory" about a part of reality. The representations of perceived objects – like open mental representation – called structure of *qualis* can be helpful in a minimal explanation of the meaning of the words. The words have meaning only as a part of the whole.

The classical goal of cognitive psychology is to solve, if the generalisation and assessment of similarity of the objects is important for identification and categorization.

A modified method of pairwise comparison may be used for the assessment of *qualis*, the measurement of similarity between the objects for categorization, and in the research on "non-sharpness" of semantic fields, semantic spaces, psychological (and semantic) distance, for

describing the membership functions, its measure, for quantification of information, in the research on the information's asymmetry etc.

Internet search engines, like library systems, need more precise tools for semantic evaluations. It is desirable to develop "intelligent" software, which would be comparable to human assessments in different contexts. More investigation is necessary in the analysis of the basis of understanding, and the level-sense of the meaning of the concepts. It is worth to compare the "subjective" and "objective" methods to check for their potential convergence.

References

- Anderson, J. A. (1987). *Communication research: Issues and methods*, NY: McGraw-Hill
- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NY: L. Erlbaum Associates.
- Chlewiński Z., Falkowski A., & Francuz P. (1995). *Wnioskowanie przez analogię w procesach kategoryzacji. Psychologiczne badania biologów, fizyków i humanistów*. Lublin: Towarzystwo Naukowe KUL.
- Do, Q. X., & Roth, D. (2010, October). Constraints based taxonomic relation classification. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 1099-1109). Association for Computational Linguistics. Retrieved from <http://cogcomp.cs.illinois.edu/papers/DoRo10.pdf>
- Edelman, G. M. (1998). *Przenikliwe powietrze, jasny ogień. O materii umysłu*. Warszawa: PIW.
- Gadomski, A. M. (1997). Global TOGA Meta-Theory. The page since 1997, Last small updating May. 2007, 2008. Retrieved from <http://erg4146.casaccia.enea.it/wwwerg26701/Gad-toga.htm>
- Gadomski, A. M. (2002). High-Intelligence Paradigms. Retrieved from <http://erg4146.casaccia.enea.it/HID/Hi-def.htm>
- Halliday, M. A. K. (1985). *The introduction to functional grammar*. NY: Oxford University Press.
- Hatch, E., & Brown, C. (1995). *Vocabulary, semantics, and language education*. Cambridge University Press.
- Hemingway, C. J., & Gough, T. G. (1998). A socio-cognitive theory of information systems. *RESEARCH REPORT SERIES-UNIVERSITY OF LEEDS SCHOOL OF COMPUTER STUDIES LU SCS RR*, (25), All-All, Retrieved from https://www.engineering.leeds.ac.uk/computing/research/publications/reports/1998/1998_25.pdf
- Hjoerland, B. (2004). Domain Analysis: A Socio-Cognitive Orientation for Information Science Research. *Bulletin of the American Society for Information Science and Technology*, 30(3), 17-21. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/bult.312/full>
- Koczkodaj, W. W. (1996). Statistically Accurate Evidence Of Improved Error Rate By Pairwise Comparisons. *Perceptual And Motor Skills*, 82, 43-48.

- Lester, M. (1971). *Introductory transformational grammar of english*. NY: Holt, Rinehart & Winston, Inc.
- Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual review of Psychology*, 43(1), 25-53. Retrieved from <http://users.ecs.soton.ac.uk/harnad/Papers/Py104/nosofsky.sim.html>
- Oleński, J. (2000). *Elementy ekonomiki informacji*. Warszawa: Uniwersytet Warszawski.
- Putnam, H. (1988). *Representation and reality*. Cambridge: Mit Press.
- Robinson, J. (2010). Awesome insights into semantic variation. *Advances in Cognitive Sociolinguistics*, 45, 85. Retrieved from <http://books.google.com>
- Sapir, E. (1944). Grading, a study in semantics. *Philosophy of science*, 11(2), 93.
- Stamenov, M. (1993) *Problemi na znaceniето v subektivnata semantika*. Sofia: Universitetsko Izdatelstvo "Sv. Kliment Ohridski".
- Trier, J. (1934). Das Sprächliche Feld. Ein Auseinandersetzung. *Neue Jahrbücher Für Wissenschaft Und Jugendbildung*, 10, 428-449. (Available from Neufeind, L. (2004) *Die Wortfeldtheorie*, Munchen: GRIN Verlag).
- Van Dijk, T. A. (2009). Critical Discourse Studies: A Sociocognitive Approach. Retrieved from <http://www.discourses.org/OldArticles/Critical%20discourse%20studies.pdf>
- Wassilew, A. Z. (2001). Aspekty (meta)językowe komunikacji - efektywność współpracy informatyków z użytkownikami systemów informatycznych. In Grabara, J. K., & Nowak, J. S. (Eds.). *Efektywność Zastosowań Systemów Informatycznych*, Warszawa: WNT.
- Wassilew, A. Z. (2009). Technologie Informacyjno-Komunikacyjne a Sieciowość we Współczesnym Społeczeństwie. *Gospodarka Narodowa*, 11-12, 2009, Warszawa: Szkoła Główna Handlowa.
- Winter, S. (1998). *Expectations and linguistic meaning* (Vol. 71). Lund University. Retrieved from <http://www.lu.se/People/Simon.Winter/Thesis/> (visited 2001).
- WNSim Retrieved from <http://cogcomp.cs.illinois.edu/page/software>
- Zgółka, T. (1996). Lingwistyka w paradygmatach. Kontynuacje i przełomy. In: Pogonowski, J., & Zgółka, T. (Eds.) *Przyczynki Do Metodologii Lingwistyki*, Poznań: Uniwersytet im. A. Mickiewicza.

Biography

Aleksander Z. Wassilew graduated from the Warsaw University of Technology in 1975. After that, he worked about 3 years in the Central Institut of Computing in Sofia. Since 1981 he has been working in Warsaw School of Economics in the Institut of Information Systems and Digital Economy. Author of works on socio-cultural aspects of ICT, networking, quality of life, gift economy, psychological assessment, evolutionary approach to language.