Network Theory and Terminology†

Ágota Fóris

Károli Gáspár University, Department of Hungarian Linguistics,
H-1088 Budapest, Reviczky u. 4 (Hungary), <aforis@t-online.hu>

Ágota Fóris, habil. (2006), Ph.D. (2002), linguist. She works at the Károli Gáspár University (Budapest, Hungary) Department of Hungarian Linguistics, Associate Professor, Head of Department and of the Terminology Research Group (TERMIK). Her field of research includes terminology, LSP lexicography, and researching technical and scientific vocabulary. She is a member of the Board of the EAFT; she is an elected member of the Dictionary Work-committee and the Applied Linguistics Work-committee of the Hungarian Academy of Sciences, member of the Board of the Hungarian Association of Applied Linguists and Language Teachers, and a founding member of the MaTT (Council of Hungarian Terminology).

Fóris, Ágota. **Network Theory and Terminology.** *Knowledge Organization. 40(6)*, 422-429. 38 references.

ABSTRACT: The paper aims to present the relations of network theory and terminology. The model of scale-free networks, which has been recently developed and widely applied since, can be effectively used in terminology research as well. Operation based on the principle of networks is a universal characteristic of complex systems. Networks are governed by general laws. The model of scale-free networks can be viewed as a statistical-probability model, and it can be described with mathematical tools. Its main feature is that "everything is connected to everything else," that is, every node is reachable (in a few steps) starting from any other node; this phenomena is called "the small world phenomenon." The existence of a linguistic network and the general laws of the operation of networks enable us to place issues of language use in the complex system of relations that reveal the deeper connections between phenomena with the help of networks embedded in each other. The realization of the metaphor that language also has a network structure is the basis of the classification methods of the terminological system, and likewise of the ways of creating terminology databases, which serve the purpose of providing easy and versatile accessibility to specialised knowledge.

Received and accepted 1 September 2013

† This study was supported by the Bolyai János Postdoctoral Fellowship (Hungarian Academy of Sciences).

1.0 Introduction

The new discoveries of network theory have proved that the world around us forms an enormous and elaborate network system. The most important characteristic feature of this network system is that "everything is connected to everything else" that is every node is reachable (in a few steps) starting from any other node. In microscopic or cosmic systems, the same general network laws operate as in the technical and biological systems or in the highly difficult system of human society. The scientific theory that developed following the recognition of this fact shapes our thinking, the new discovery transforms our worldview and provides us with new opportunities for learning and new targets appear in front of us. A part of the conceptual system that we developed earlier about biological systems, the difficult processes of healing, the laws of production and the social aspects of human society is being transformed on the basis of the knowledge we have concerning various networks.

The development of modern computerized ontology and its successful practical application in various fields is based on the recognition that obtaining knowledge is not enough; we can apply our knowledge in practice only if we know the connections among its various elements and are able to describe them. Network theory moves much farther than this—with the help of general laws, it systematizes the elements of and connections between various systems, determining the laws of its development and operation on the basis of which the consequences of the effects within and outside of the network can be calculated (anticipated). On the basis of network theory, we can understand the operation of enormous, coherent, and complex systems. This fact is especially important as the level of development at the beginning of the 21st century makes it necessary to organize the knowledge of smaller fields into a complex network, in a way that we can guarantee the possibility to understand and apply this many-sided and complicated system. The increasing integration of science, our technical equipment, economics, and society show the realization of this network-like operation. Therefore, we must get to know how these networks operate, and we must apply this knowledge.

The basic function of language is to encode, store, and transfer the knowledge accumulated by society. The structure of language and the complex processes of language use are approached in different ways by the various branches of linguistics.

Terminology research places the term into its focus and considers it as the basic unit (Sager 1990; Laurén and Picht 1993; Temmerman 2000; Budin 2001; Cabré Castellví 2003). Terms are part of the text. Term extractors, for example, rely on the role of terms in the text; they work with different methods, such as listing lexical units or their collocations based on their frequency of occurrence in the text, or take other features of the text into consideration while extracting terms.

Networks are present in texts at several levels. The structure of a text is the network that provides cohesion. The network between concepts establishes coherence, which appears as a configuration of knowledge formed by concepts. The networks present within the text are linked to external networks, such as the knowledge network of the author, the network of prior knowledge of the recipient, and their intertextual background.

Pierre Lévy's article on the responsibility of intellectuals drew the attention of researchers working in various fields to the importance of terminology issues. Lévy (2007) sees one obstacle to using the potential of collective intelligence in the great variety and fragmented nature of symbolic systems, one specific problem being the variety and incompatibility of classification systems in general and specifically of terminology.

Terminology plays a significant role in both theoretical and empirical research, and the precise development and description of terminology is a basis for scientific classification and scientific theories. Knowledge is conveyed through language, and technical texts not only convey knowledge but technical and terminological norms as well. The paper establishes the role of the terminological network in the representation of knowledge, and also brings examples to the possibilities for applying the network model successfully.

2.0 Knowledge acquisition, terms and text

2.1. Terms in text

A text is created using the verbal and written signs of a language. The text not only encodes information, but also ensures it is distributed through space and time. Depending on the nature of information, texts can have different structures and length. The text always exists in some physical form, appears in a confined space and time, but its cognitive network of relations is unlimited in both space and time. The text encodes information on concepts through terms; therefore the role of terms in a text has to attract special attention when studying texts.

Texts can be studied from various viewpoints. A study focusing on a given aim can discover the general principles of the structure of texts, the links between a text and a natural language, the relation of text structure and the encoded information, or the relation of the text to other components of the communication process. The studies conducted in texts linguistics have discovered a large number of findings on these topics (e.g., Mel'cuk and Žolkovskij 1970; Dressler 1972; Halliday and Hasan 1976; Petőfi 1979; Beaugrande and Dressler 1981 and 2002).

Terminology and text linguistics both study the issues of encoding and distributing knowledge from different starting points and approaches. Terminology studies place the term into the focus, and it is considered to be the basic unit. These studies view the text created in the encoding process to be given, and determine the role of terms in handling information. Text linguistic studies focus on the text as the research subject. The text is studied within the complex network of relations of handling information, and the features of internal and external effects are described. In this approach, the term appears implicitly as the linguistic component that organises the cognitive content of information through the linguistic code. The two approaches have a common point: the study of knowledge storage and organization. Despite the different approaches and research methods, the findings can be incorporated into a common framework, they complement and strengthen each other.

During cognition, concepts are formed to map the elements of the world, and these concepts are structured into a system in the process of thinking to enable easy handling of the variety in the world. The linguistic sign for a concept is the term, and their system formed through classification is called terminology. Meaning is an inseparable part of the term, and it is described in terminology databases, dictionaries, and standards, etc. as definitions.

In the text linguistic approach of the features of texts, terms are not treated explicitly; however they are always present implicitly in the deciphering of conceptual relations (the relations between concepts). When examining features of the text, text linguistics unavoidably reaches issues that terminology needs to answer when laying its theoretical foundations; this is, based on Beaugrande and Dressler (1981, 85), the definition of the "concept" as a cognitive unit:

> A concept can be defined as a configuration of knowledge that can be recovered or activated with more or less consistency and unity... . the meaning of a concept is the sum of its possible uses (Schmidt 1978). Unfortunately, many concepts are so adaptable to differing environments that they remain quite fuzzy in regard to their components and boundaries. Therefore, defining concepts involves working with comparative probabilities.

The basic unit of terminology is the concept; its linguistic sign is the term. When these basic concepts were introduced, empirical findings (translation, contrastive linguistics, etc.) revealed that the characteristic features of the concept and the term form a fuzzy set. The free choice of classification features or the differences between the worldviews of various cultures, etc. lead to different categorizations of the same reality in the world. The features of a given concept are different not only in the various conceptual systems of different cultures, but also in different domains, and, at times even among different groups of experts working in the same field. The fuzzy nature of the concept can be described from a terminological approach: the actual meaning of the term can be different in different communication situations (or referring to the aforementioned, activates different configurations of knowledge).

2.2. Acquiring and organizing knowledge

When discussing the issues of text encoding and decoding, we encounter questions of data storage in the brain and searching the stored elements. In order to be able to encode knowledge in a text, the text producer needs to find the necessary linguistic elements and the rules of code formation. When decoding, the receiver has to peel off the knowledge elements from the linguistic signs. These processes that take place in the brain rely on the characteristic features that appear in the text through the standards of textuality (cohesion, coherence, intentionality, acceptability, informativity, situationality, and intertextuality). Different parts of the brain store linguistic signs and cognitive knowledge. The question here is how language code is formed to convey knowledge, and the other way around, in the process of decoding how cognitive knowledge is linked to the incoming linguistic signs.

Cognitive neuroscience calls this mental process, in which knowledge is stored in different parts of the brain and is retrieved in utilisation, active memory (Racsmány 2003; Gósy 2005, 27–71). The findings of cognitive neuroscience indicate that knowledge organised in the brain has a network structure (Gósy 2005, 193–200), meaning that the network structure enables a quick access to stored knowledge and linguistic signs. Based on my studies in the field of terminology, I presume that the mental processes undergoing in the brain occur in a special, so called scale-free network structure, the same way as the storage and retrieval of terms (Fóris 2007). These networks do not evolve randomly, but instead are created in a purposeful way, and their special characteristics enable very fast and reliable operation. Such a network structure is suitable to model several details of the workings of language and communication processes. The internal and external links of a text also form a network that can be traced within a short time. The reference systems of printed dictionaries or the search engines of online dictionaries lead through a network of dictionaries (Fóris 2008). Conceptual/terminological networks appear well beyond the physical boundaries of a text in intertextuality, in linkages that form cohesion and coherence, and among the relations between the term set of the text producer, the text itself, and the receiver. Knowledge and meanings are sensitive to context, they depend on it. Knowledge and meanings always occur through some kind of network. Knowledge can be organised into a network in different ways; if this network is well-organized, the text is considered coherent. Beaugrande and Dressler (1981) view coherence as a result of a network that contains concepts and relations.

In the network of concepts, nodes have different functions. There are primary concepts and secondary concepts. The basis for classification is the extent to which the concept is suitable for the purposes of the governing hub that establishes mental continuity. A detailed discussion of this issue in the study of textual characteristics, especially of teaching materials, is highly important because terms that occur in a course book must be selected in a way that they activate the right concepts that ensure mental continuity (Fóris 2011). Organizing concepts and relations into a network does not only result in a conceptual network, but also a grammatical network that plays a significant role in cohesion.

Decoding knowledge that has been encoded by the linguistic code system of a text is a complex psychological process. Information retrieval is not simply an operation of a code key to decipher the meaning of codes. Many of the standards of textuality focus on the relation of the receiver and the text. The decoding of a text is a psychological process in just the same way as text production is, and the process is carried out along the same global patterns. In the process of information retrieval inference, supplementation, and being familiar with the conventions of text production have important roles.

The text can fulfill its purpose if decoding is completed. The success of decoding depends on the receivers being familiar with the signs and the code key, namely the accepted ways of expression. Knowledge transfer can only be effective if the receiver is able to understand the content of a text. This comprehension depends on whether the receiver has acquired the norms of the given text type, and within that terminological norms have key importance.

3.0 Networks

3.1. About networks in general

The concept of "network" has been known for a long time. In general, networks can be described as follows: elements of the set are represented by nodes, and the relations among the elements are represented by edges. The mathematical theory of networks was developed by Rényi and Erdős in the 1960s as part of their graph theory. In their model, they established random links among the nodes (based on throwing the dice in a specific way). Networks that are created in such a way are called random networks (Karoński and Ruciński 1997).

Research showed that networks evolved long ago in the development of nature and society, but their existence, features and key roles were only discovered in recent decades. It is now clear that there is a short chain of links between two seemingly distant nodes of a network, which shows that the relations between various things and their effects on one other differs from our earlier understanding. It is a fact that the complex systems of nature, society, technology, economics, etc., or better said, the elements in their networks, form simple "small worlds" (Watts 1999, cit. in Barabási 2002).

Barabási (2002) writes that the World Wide Web, for example, is a network with web pages as its nodes and hyperlinks as the edges, that is, possibilities to make contact between the web pages. This network works in a digital system; therefore it can be effectively used as a model-system. A map has been drawn of a fraction of the web, and the first studies have shown that this network differs from the previously mentioned random networks in many ways (see Albert et al. 1999). For example, if we examine the distribution of the number of edges going in and out of a node, we would expect that the number of nodes making 1, 2, 3, 4 … contacts increases up to a maximum, and thereafter decreases, meaning that a bell curve would illustrate the distribution of relations. We get such a distribution of the number of edges if we do our analysis on the network of motorways of a public road map. The actual number of links between web pages follows a different distribution: it does not have a maximum, instead it decays rapidly, meaning the distribution can be illustrated not by a bell curve but by the curve of a power law (Barabási 2002).

Thus, the difference between the two types of networks is that the World Wide Web has a large number of nodes that have a few links, and a few so-called hubs that have a very large number of links. These highly connected hubs play a special role in the evolution and function of networks. Complex networks that can be described by a power law distribution, and consequently their nodes cannot be grouped according to an internal scale, and are therefore called scale-free networks. A good example of a scale-free network is the airline route map of a continent, where airports are the nodes and the routes of flights are the edges. It is clearly visible on such a map that the various nodes have different roles in air traffic and that big airports play an evidently central role in the network (Amaral et al. 2000, cit. in Barabási 2002).

3.2 Main characteristics of scale-free networks

In the following section, I define some basic concepts of scale-free networks and describe their most important characteristics (for details, see Barabási 2002 and 2012; Csermely 2006; Fóris 2010 and 2012). In scale-free networks small worlds are always present. This means that seemingly distant points of such networks can be joined quickly, and therefore, although there seems to be a large distance between them, they have significant effects on one another.

3.2.1 Laws of the growth of networks

Scale-free networks self-assemble, evolving node by node. Over the course of evolution, every network, be it biological systems, social relations, or the networks of the economy, starts from a small centre and further nodes are added to this centre. In the beginning, when there is a small number of links between the elements, the system does not display the characteristics of the scale-free network; the evolving system starts to work as a network when every node has at least one link. This stage of network growth is called threshold (or percolation).

Self-assembled networks grow in a way that gives new nodes a higher probability of connecting to those nodes that have a large number of links. This is called the principle of preferential attachment, which is apparent in the centralised expansion of the networks of economics today (Barabási 2002; Csermely 2006).

3.2.2 Error tolerance of networks

Networks created by nature continue functioning even if some of the nodes are removed, while in the case of man-made constructions an error in just one of the components may impede the whole system. It has been proven that the removal of some of the nodes of a natural network has little to no effect on its operation. However, the moment that the number of errors reaches a critical level the network fragments into small, non-communicating pieces. A study of the map of the web in terms of scale-free networks produced a surprising discovery. Even if 80% of the nodes are removed, the remaining part still works. Further model examinations showed the key role of hubs in error tolerance: an error in a small number of hubs will cause the network to break into pieces. Hubs of scale-free networks have a fundamental role in their vulnerability. Every entity of our surroundings may be a node of several networks at the same time, due to their diverse characteristics. As a result, everything is connected to everything else around us, because the different networks are connected, and the phenomenon of small worlds is present in every one of them in a complicated way.

The structure and operation of interconnected networks can be simulated with a complex spatial model. The nodes inside the networks are connected to one another by "links," which represent the cohesion of the elements (and the relation between them). Most of the nodes have more links creating extensive networks, and within these networks the nodes not only come into contact with neighbouring nodes but also with ones located far away. This also explains the development of the small world phenomenon. Various types of connections are established among the nodes within a given network. In any network, the strength of links may be of two types: strong links and weak links.

The network may be spatial, temporal, conceptual, etc., depending on the set of elements containing the nodes and the features of the links. For instance, road accidents form a network in time where the relation of repetition and gravity is described by a power law—slight bumps are repeated several times over a short time span, while grave accidents are usually repeated after longer intervals. The strength of earthquakes or size of the lottery jackpot and their occurrence form the same temporal network.

One important characteristic feature of scale-free networks is the function that shows the degree distribution. This function is always a power law. The function does not have a maximum, and if the degree is increasing and there are only a few hubs with a lot of degree, its value continues to decrease steeply. The distribution of different f features of a scale-free network according to a v parameter is described by a power law: $f = C\ v^{n}$, where C and n are constants of the given network (see 4.1.). In the last few years a lot of publications have been written on the application of the scale-free network model. A summary of these can be read in English (e.g., Barabási 2002; Barabási and Bonabeau 2003; Csermely 2006; Newman et al. 2006).

4.0 Linguistic research based on network theory: language networks

There have been a great number of publications in recent years introducing the results of the successful application of networks in various fields. Concerning the preliminaries of network research, the earlier linguistic research that introduced the existence of language networks and the possibility of applying the knowledge of such networks in the field of linguistic research play an important role: "the human language is a complex network, where stability can be defined as the stability of meaning" (Csermely 2006, 219).

4.1. Zipf's law and power law

Before networks were widely researched, there were a great number of linguistic findings on natural languages that now lend themselves to new interpretation through network theory. Now that network theory has been widely interpreted and mathematically described, previous linguistic findings can be inserted into the system of networks (for details, see Fóris 2012).

The great number of studies conducted concerning the various language corpora indicates the existence of the semantic network, but beyond understanding the fact that they are connected to the network, these results do not enable us to learn the peculiarities of the networks and the laws functioning inside them. The results of quantitative linguistics, glottometrics and applied mathematics can take us closer to learning about the types of semantic networks and describing them more precisely. Several findings have been published that indirectly prove the existence of language networks covering the whole language. First of all, we should mention the various studies on word frequency, of which Zipf's law has been examined and applied the most. The detailed analysis of this can be found in many works (e.g., Balasubrahmanyan and Naranan 2002; Fóris 2007). The validity of Zipf's law was proved in relation to the corpus of several languages, as in the case of Chinese for example (Rousseau and Zhang 1992).

In Fóris (2012), I presented calculations on the numeric data obtained from previous publications on quantitative linguistic research. Zipf examined the frequency of occurrence of words in standard English and found that the product of the v rank of a word in the order of frequency, and the f value of its frequency in the corpus, is constant (C) (Zipf 1945 and 1949). The formula ($f\ v = C$) called

Zipf's law was thought to be a universal linguistic law. Other languages and dialects were also examined and it was found that the law is quite infrequent. After analysing statistical data, it was concluded that there is a more complex function that describes the relation of quantitative characteristics of a language. Mandelbrot (1953), Billmeier (1969), and Papp (1969) came to the same conclusion. In my previous studies, I have shown that Zipf's law ($f\ v = C$) is a special case of the power law ($f = C\ v^n$), where the value of n exponent is 1. My conclusion was that word frequency distribution can be described by the power law that is characteristic of scale-free networks. I reached this conclusion after a critical analysis of findings of quantitative linguistic research (Fóris 2010 and 2012).

Published research data prove that there are several cases where word frequency is described by a power law whose exponent is not 1, but another value. That is, if we put the data into a logarithm we get a straight line and the exponent of the function is not 1. As the complex networks of individual natural languages—and within those, the various parameters of one language—have different distributions, we get different results if we examine the same distribution for different parameters of a language or for different corpora, or if we examine the same parameter for different natural languages. This fact is demonstrated in the universal power law through changes in the constants, and more importantly, the value of the n exponent.

This statement is underpinned by my calculations with published data (Nagy 1985, which can be seen in Fóris 2012). The findings above show that the power law—a special case of which is the original form of Zipf's law—is suitable to describe the features of Hungarian corpora. The fact that the parameters of the power law have different values indicates that the analysed corpora are of different natures, and therefore the ranks of distribution are not the same. Further analysis of the relations of the network theory and the features of the examined corpora could provide an in-depth understanding of language networks.

4.2. Possibilities of application of the network theory in terminology research

The previously used terminological trees encoded a great deal of information about the relations of terms, such as hyponymy, hyperonymy, and co-hyponymy, etc. However, the language network, and within it, the terminology network of domains, is much more complex than the previously used form of terminology graphs. They could be modelled with a scale-free network.

Beside the aforementioned, Zipf also pointed out that there is a similar connection between word frequency and the number of phonemes in a word, and between word frequency and the number of meanings to that word. Studies in connection with the fulfilment of Zip's law have undoubtedly proven that the smaller the frequency of the word, the smaller the number of its meanings, that is, the more explicit its use. Studies of word frequency (such as Guiraud 1954; Papp 1969) showed that word frequency is connected to the semantic properties of the word: the rarer the word, the smaller its frequency and probability, the more defined its meaning, and the higher its informational value. These findings could be applied in various areas, for example the field of terms.

The calculations demonstrated in Fóris (2012) support the hypothesis that language networks can be described with the help of scale-free networks. It is also evident that the power law of distribution over the network not only describes the frequency of the use of words but the distribution of other properties as well, such as the distribution of the number of meanings to a word, the parts of speech, and the proper nouns in a corpus.

According to Cabré's 'theory of doors' model, terminological questions should be addressed from three viewpoints: from the cognitive (the conceptual), the linguistic, and the communicative sides (Cabré Castellví 2003). She introduced a third approach (pragmatic-communicative) alongside the two previously used approaches (the semasiological—that is, sign-based, and onomasiological—that is, concept-based). Consequently, the scale-free networks are appropriate to model language. In what follows, the model of terminological networks is elaborated starting from Cabré's model and using the findings of network theory. According to this model, the three components of the terminological approach—the cognitive, the linguistic and the communicative—form scale-free networks on their own, and the process of communication can be modelled by joining them. This is the model of terminological network (Fóris 2012).

Terminological networks are complex and multidimensional. The joined networks of the given levels of a language make up a cross-section of the whole network and these cross-sections allow for the optimal functioning of the language. The model is represented in a figure that has three layers: the top layer shows the cognitive, the middle layer shows the linguistic, and the bottom layer shows the pragmatic-communicative network. The model is simplified in the following ways: 1) the terminological network contains a great number of nodes and edges, and the drawing is just a small section of that; 2) the size of the network is infinite and it is made up of complexly overlapping spatial figures, while this model is a small, uniplanar section; 3) language may be segmented into more than the afore-mentioned three cross-sections, and consequently the model of its network structure may be more complex. For instance, phonetic aspects or the grammar words beside the terms cannot be ignored. Here these are overlooked in the interest of simplification. Further edges link the terminological units to other networks.

In this model, the role of terms in communication is determined by several factors. The cognitive component and its relations play a key role. The concept that is designated by the term determines the meaning of the term. It is widely known that the details of the meaning of a term are specified by its place in the terminological network. The relations of the communicational network allow the terms to create the links necessary for the articulation and transfer of information appropriate to the given communicational situation. The unity of these three components determines the communicational value of the term. Cabré calls this complex unit, formed around the term in this three-sided environment, the terminological unit. In the communication process the successful use of the terminological unit is only possible if users know their way around these three networks both during production (encoding) and during comprehension (decoding).

5.0 Summary

The discovery of the scale-free nature of language networks demands a new approach in terminology research. It is always observable in nature, the economy, microscopic, and cosmic systems that everything is connected with everything else. That is why, in conjunction with localised research, a special emphasis is currently laid on the approach to focus on the interrelations of complex systems. This leads to the compilation of modern computerised ontologies, as beyond providing factual knowledge, they are also practical to provide the environmental links of that knowledge. Within the complex system of language networks, sub-networks are also interrelated, and therefore it is necessary to study the relations of the various units.

In this paper, I first provided a brief summary of terms in text. The terminological unit is formed jointly by the three networks in language use. This means that text analysis is inseparable from the cognitive aspects. In sum, in this case the joint network of all three components—the cognitive, the linguistic and the communicative—needs to be analysed. Second, after the basics of networks, I summarized some of my calculations about language networks. Based on these, I suggested that language could be modelled with scale-free networks. Finally, I elaborated on some aspects of the application of network theory in terminology and drew attention to the possibilities of the application of terminological network model in research at the field of terminology.

References

Albert, Réka, Jeong, Hawoong and Barabási, Albert-László. 1999. Diameter of the World Wide Web. *Nature* 401: 130-1.

Amaral, Luis A. Nunes, Scala, Antonio, Barthélémy, Marc and Stanley, H. Eugene. 2000. Classes of small-world networks. *Proceedings of the National Academy of Sciences* 97: 11149-52.

Balasubrahmanyan, Vriddhachalam K. and Naranan, Sundaresan. 2002. Algorithmic information, complexity and Zipf´s law. *Glottometrics* 4: 1-26.

Barabási, Albert-László. 2002. *The new science of networks*. Cambridge, MA: Perseus.

Barabási, Albert-László. 2012. *Network science book*. Available http://barabasilab.neu.edu/networksciencebook/.

Barabási, Albert-László and Bonabeau, Eric. 2003. Scale-free networks. *Scientific American* 288: 60-9.

Beaugrande, Robert-Alain de and Dressler, Wolfgang. 1981. *Introduction to text linguistics*. London and New York: Longman.

Beaugrande, Robert-Alain de and Dressler, Wolfgang. 2002. *Introduction to text linguistics*. Available http://www.beaugrande.com/introduction_to_text_linguistics.htm.

Billmeier, Günther. 1969. *Worthäufigkeitsverteilungen vom Zipfschen typ, überprüft an deutschem textmaterial*. IPK-Forschungbericht. Bonn: Magisterarbeit.

Budin, Gerhard . 2001. A critical evaluation of the state-of-the-art of terminology theory. *Journal of terminology science and research* 12: 7-23.

Cabré Castellví, Maria Teresa. 2003. Theories of terminology. Their description, prescription and explanation. *Terminology* 9: 163-200.

Csermely, Péter. 2006. *Weak links: the universal key to the stability of networks and complex systems*. Heidelberg: Springer. Available www.weaklink.sote.hu/weakbook.html.

Dressler, Wolfgang. 1972. *Einführung in die textlinguistik*. Tübingen: Max Niemeyer Verlag.

Fóris, Ágota. 2007. A skálafüggetlen hálók nyelvészeti vonatkozásai. *Alkalmazott nyelvtudomány* VII: 105-24.

Fóris, Ágota. 2008. Scale-free networks in dictionaries. In Bernal, Elisenda and DeCesaris, Janet, eds., *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. Sèrie activitats 20. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, pp. 1499–504.

Fóris, Ágota. 2010. Change of paradigm in terminology: new models in KO. In Gnoli, Claudio and Mazzocchi, Fulvio, eds., *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference 23–26 February 2010, Rome, Italy*. Advances in knowledge organization 12. Würzburg: Ergon Verlag, pp. 57-63.

Fóris, Ágota. 2011. Terminology and text linguistics. *Terminologija* 18: 51-65.

Fóris, Ágota. 2012. Terminology and the theory of scale-free networks. *Magyar terminológia* 5: 58-71.

Gósy, Mária. 2005. *Pszicholingvisztika*. Budapest: Osiris Kiadó.

Guiraud, Pierre. 1954. *Les charactères statistiques du vocabulaire*. Paris: Presses Univeritaires de France.

Halliday, Michael A. K. and Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.

Karoński, Michal and Ruciński, Andrzej. 1997. The origins of the theory of random graphs. In Graham, Ronald. L., Nesetril, Jaroslav, and Butler, Steve, eds., *The mathematics of Paul Erdős: algorithms and combinatorics*. Berlin: Springer, pp 311-6.

Laurén, Christer and Picht, Heribert. 1993. Vergleich der terminologischen schulen. In Laurén, Christer and Picht, Heribert, eds., *Ausgewählte texte zur terminologie*. Wien: TermNet, pp. 493-539.

Lévy, Pierre 2007. Nouvelle responsabilité des intellectuels. *Le Monde diplomatique* August. Available http://www.monde-diplomatique.fr/2007/08/LEVY/15006.

Mandelbrot, Benoît. 1953. An information theory of the statistical structure of language. In Jackson, Willis, ed., *Communication theory*. New York: Academic Press, pp. 486−502.

Mel'cuk, Igor A. and Žolkovskij, Alexander K. 1970. Towards a functioning meaning-text model of language. *Linguistics* 57: 10-47.

Nagy, Ferenc. 1985. *Kvantitatív nyelvészet*. Budapest: Tankönyvkiadó.

Newman, Mark, Watts, Duncan J. and Barabási, Albert-László. 2006. *The structure and dynamics of networks*. Princeton, NJ: Princeton University Press.

Papp, Ferenc. 1969. O masinnoj obrabotke odnojazücsnüh szlovarej. *Naucsno-technicseszkaja informácija* 3: 20-9.

Petőfi, János, ed., 1979. *Text versus sentence: basic questions in text linguistics*. Hamburg: Buske.

Racsmány, Mihály. 2003. Az emlékezet kognitív neuropszichológiája. In Pléh, Csaba, Kovács, Gyula and Balázs, Gulyás, eds., *Kognitív idegtudomány*. Budapest: Osiris Kiadó, pp. 459-84.

Rousseau, Ronald and Zhang, Qiaoqiao. 1992. Zipf's data on the frequency of Chinese words revisited. *Scientometrics* 24: 201–20.

Sager, Juan C. 1990. *A practical course in terminology processing*. Amsterdam and Philadelphia: John Benjamins.

Schmidt, Siegfried J. 1978. Some problems of communicative text theories. In Dressler, Wolfang, ed., *Current trends in text linguistics*. Berlin: de Gruyter, pp. 47-60.

Temmerman, Rita. 2000. *Towards new ways of terminological description. The sociocognitive approach*. Amsterdam and Philadelphia: John Benjamins.

Watts, Duncan J. 1999. *Small worlds*. Princeton, NJ: Princeton University Press.

Zipf, George K. 1945. The meaning-frequency relationship of words. *The journal of general psychology* 33: 251-6.

Zipf, George K. 1949. *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.

KNOWLEDGE ORGANIZATION KO

Official Bi-Monthly Journal of the International Society for Knowledge Organization ISSN 0943 – 7444

International Journal devoted to Concept Theory, Classification, Indexing and Knowledge Representation

Publisher

ERGON-Verlag GmbH, Keesburgstr. 11, D-97074 Würzburg
Phone: +49 (0)931 280084; FAX +49 (0)931 282872
E-mail: service@ergon-verlag.de; http://www.ergon-verlag.de

Editor-in-chief (Editorial office)

Dr. Richard P. SMIRAGLIA (Editor-in-Chief), School of Information Studies, University of Wisconsin, Milwaukee, Northwest Quad Building B, 2025 E Newport St., Milwaukee, WI 53211 USA. E-mail: smiragli@uwm.edu

Instructions for Authors

Manuscripts should be submitted electronically (in Word format) in English only via email to the editor-in chief and should be accompanied by an indicative abstract of 150 to 200 words. Manuscripts of articles should fall within the range 6,000-10,000 words. Longer manuscripts will be considered on consultation with the editor-in-chief.

A separate title page should include the article title and the author's name, postal address, and E-mail address, if available. Only the title of the article should appear on the first page of the text.

To protect anonymity, the author's name *should not* appear on the manuscript, and all references in the body of the text and in footnotes that might identify the author to the reviewer should be removed and cited on a separate page.

Criteria for acceptance will be appropriateness to the field of the journal (see Scope and Aims), taking into account the merit of the contents and presentation. The manuscript should be concise and should conform as much as possible to professional standards of English usage and grammar. Manuscripts are received with the understanding that they have not been previously published, are not being submitted for publication elsewhere, and that if the work received official sponsorship, it has been duly released for publication. Submissions are refereed, and authors will usually be notified within 6 to 10 weeks.

The text should be structured by numbered subheadings. It should contain an introduction, giving an overview and stating the purpose, a main body, describing in sufficient detail the materials or methods used and the results or systems developed, and a conclusion or summary.

Footnotes are accepted only in rare cases and should be limited in number; all narration should be included in the text of the article. Paragraphs should include a topic sentence and some developed narrative; a typical paragraph has several sentences. Italics may not be used for emphasis. Em-dashes should not be used as substitutes for commas.

Italics should not be used for emphasis. Em-dashes should be used as substitutes for commas. Paragraphs should include a topic sentence and some developed narrative. A typical paragraph has several sentences.

Reference citations within the text should have the following form: (Author year). For example, (Jones 1990). Specific page numbers are required for quoted material, e.g. (Jones 1990, 100). A citation with two authors would read (Jones and Smith, 1990); three or more authors would be: (Jones et al., 1990). When the author is mentioned in the text, only the date and optional page number should appear in parenthesis – e.g. According to Jones (1990), …

References should be listed alphabetically by author at the end of the article. Author names should be given as found in the sources (not abbreviated). Journal titles should not be abbreviated. Multiple citations to works by the same author should be listed chronologically and should each include the author's name. Articles appearing in the same year should have the following format: "Jones 2005a, Jones 2005b, etc." Issue numbers are given only when a journal volume is not through-paginated.

Examples:

Dahlberg, Ingetraut. 1978. A referent-oriented, analytical concept theory for INTERCONCEPT. *International classification* 5: 142-51.

Howarth, Lynne C. 2003. Designing a common namespace for searching metadata-enabled knowledge repositories: an international perspective. *Cataloging & classification quarterly* 37n1/2: 173-85.

Pogorelec, Andrej and Šauperl, Alenka. 2006. The alternative model of classification of belles-lettres in libraries. *Knowledge organization* 33: 204-14.

Schallier, Wouter. 2004. On the razor's edge: between local and overall needs in knowledge organization. In McIlwaine, Ia C. ed., *Knowledge organization and the global information society: Proceedings of the Eighth International ISKO Conference 13-16 July 2004 London, UK.* Advances in knowledge organization 9. Würzburg: Ergon Verlag, pp. 269-74.

Smiraglia, Richard P. 2001. *The nature of 'a work': implications for the organization of knowledge.* Lanham, Md.: Scarecrow.

Smiraglia, Richard P. 2005. Instantiation: Toward a theory. In Vaughan, Liwen, ed. *Data, information, and knowledge in a networked world; Annual conference of the Canadian Association for Information Science … London, Ontario, June 2-4 2005.* Available http://www.cais-acsi.ca/2005proceedings.htm.

Illustrations should be kept to a necessary minimum and should be embedded within the document. Photographs (including color and halftone) should be scanned with a minimum resolution of 600 dpi and saved as .jpg files. Tables and figures should be embedded within the document. Tables should contain a number and title at the bottom, and all columns and rows should have headings. All illustrations should be cited in the text as Figure 1, Figure 2, etc. or Table 1, Table 2, etc.

The entire manuscript should be double-spaced, including notes and references.

Upon acceptance of a manuscript for publication, authors must provide a wallet-size photo and a one-paragraph biographical sketch (fewer than 100 words). The photograph should be scanned with a minimum resolution of 600 dpi and saved as a .jpg file.

Advertising

KO is published bi-monthly by ERGON-Verlag GmbH.

- **The price for the print version is € 229,00/ann. including airmail delivery.**
- **The price for the print version plus access to the online version (PDF) is € 258,00/ann. including airmail delivery.**

KO KNOWLEDGE ORGANIZATION

Official Bi-Monthly Journal of the International Society for Knowledge Organization ISSN 0943 – 7444

International Journal devoted to Concept Theory, Classification, Indexing and Knowledge Representation

Scope

The more scientific data is generated in the impetuous present times, the more ordering energy needs to be expended to control these data in a retrievable fashion. With the abundance of knowledge now available the questions of new solutions to the ordering problem and thus of improved classification systems, methods and procedures have acquired unforeseen significance. For many years now they have been the focus of interest of information scientists the world over.

Until recently, the special literature relevant to classification was published in piecemeal fashion, scattered over the numerous technical journals serving the experts of the various fields such as:

philosophy and science of science
science policy and science organization
mathematics, statistics and computer science
library and information science
archivistics and museology
journalism and communication science
industrial products and commodity science
terminology, lexicography and linguistics

Beginning in 1974, KNOWLEDGE ORGANIZATION (formerly INTERNATIONAL CLASSIFICATION) has been serving as a common platform for the discussion of both theoretical background questions and practical application problems in many areas of concern. In each issue experts from many countries comment on questions of an adequate structuring and construction of ordering systems and on the problems of their use in opening the information contents of new literature, of data collections and survey, of tabular works and of other objects of scientific interest. Their contributions have been concerned with

(1) clarifying the theoretical foundations (general ordering theory/ science, theoretical bases of classification, data analysis and reduction)
(2) describing practical operations connected with indexing/classification, as well as applications of classification systems and thesauri, manual and machine indexing
(3) tracing the history of classification knowledge and methodology
(4) discussing questions of education and training in classification
(5) concerning themselves with the problems of terminology in general and with respect to special fields.

Aims

Thus, KNOWLEDGE ORGANIZATION is a forum for all those interested in the organization of knowledge on a universal or a domain-specific scale, using concept-analytical or concept-synthetical approaches, as well as quantitative and qualitative methodologies. KNOWLEDGE ORGANIZATION also addresses the intellectual and automatic compilation and use of classification systems and thesauri in all fields of knowledge, with special attention being given to the problems of terminology.

KNOWLEDGE ORGANIZATION publishes original articles, reports on conferences and similar communications, as well as book reviews, letters to the editor, and an extensive annotated bibliography of recent classification and indexing literature.

KNOWLEDGE ORGANIZATION should therefore be available at every university and research library of every country, at every information center, at colleges and schools of library and information science, in the hands of everybody interested in the fields mentioned above and thus also at every office for updating information on any topic related to the problems of order in our information-flooded times.

KNOWLEDGE ORGANIZATION was founded in 1973 by an international group of scholars with a consulting board of editors representing the world's regions, the special classification fields, and the subject areas involved. From 1974-1980 it was published by K.G. Saur Verlag, München. Back issues of 1978-1992 are available from ERGON-Verlag, too.

As of 1989, KNOWLEDGE ORGANIZATION has become the official organ of the INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (ISKO) and is included for every ISKO-member, personal or institutional in the membership fee (US $ 55/ US $ 110).

Rates: From 2013 on for 6 issues/ann. (including indexes) € 229,00 (forwarding costs included) for the print version resp. € 258,00 for the print version plus access to the online version (PDF). Membership rates see above.

ERGON-Verlag GmbH, Keesburgstr. 11, D-97074 Würzburg; Phone: +49 (0)931 280084; FAX +49 (0)931 282872; E-mail: service@ergon-verlag.de; http://www.ergon-verlag.de

Founded under the title International Classification in 1974 by Dr. Ingetraut Dahlberg, the founding president of ISKO. Dr. Dahlberg served as the journal's editor from 1974 to 1997, and as its publisher (Indeks Verlag of Frankfurt) from 1981 to 1997.

The contents of the journal are indexed and abstracted in *Social Sciences Citation Index, Web of Science, Information Science Abstracts, INSPEC, Library and Information Science Abstracts* (LISA), *Library, Information Science & Technology Abstracts* (EBSCO), *Library Literature and Information Science* (Wilson), *PASCAL, Referativnyi Zhurnal Informatika, and Sociological Abstracts.*