Acta Cybernetica **16** (2004) 601–609.

A Semantic Network Dictionary for Dysphasia Therapy*

János Bácsi†

Abstract

With the use of computer-assisted teaching and learning programs the speech therapy of children with dysphasia can be planned more effectively, than with the traditional, so-called prompted-image method employed by logopaedists. Our aim is the creation of a semantic network dictionary – on the basis of a huge quantity empirical database –, which predicts what other concepts will link to a concept in the STM of 4—7 year old children, i.e. what other concepts can the most easily be associated with an already existing one by the therapeut. Furthermore, the semantic network dictionary will enable the already literate child – on the basis of the idea of András Kocsor – to take a virtual tour through the connected words of the network dictionary by merely saying words into a microphone. Hence the software will help children develop and maintain semantic structures.

1 Introduction

The cognitive sciences, memory research, research of artificial intelligence, modern linguistics, computer linguistics, evolution psychology and modern learning theories have arrived at such results, which instigated therapist dealing with children with dysphasia, and researchers dealing with acquisition of written language to re-evaluate and reconsider their methods employed so far.

The development of children suffering from expressive language disorder cannot be overlooked, as language development with dysphasia is not language-specific, in all the languages of the world – where such survey has been performed – 4—7 % of the given population (4—8 year old children) is affected. The affected children must be taught separately, because without special development, acquisition of visual language (writing, reading) will cause problems. A child with dysphasia in the process of institutionalised education will show symptoms of dyslexia and/or disgraphy, which is not taken up at school like an infection, but as a result of expressive language deficit, the inappropriate level of necessary capabilities to acquire the visual language will surface it.

*Presented at the 1st Conference on Hungarian Computational Linguistics, December 10–11, 2003, Szeged.

†Juhász Gyula Teacher Training College Division, Teacher Training School, University of Szeged, H-6725 Szeged, Boldogasszony sgt. 8, Hungary, e-mail: bacsi@jgytf.u-szeged.hu

The most important characteristic feature of expressive disorder is limited vocabulary (see DSM-IV), hence the prime task of the therapist planning and maintaining development is the extension of vocabulary. This task is usually undertaken by logopaedists, who use those prompting images to extend vocabulary, which are employed during teaching how to read. The problem with these prompting images is, that a 4 year old child cannot put into words what he sees in the picture, and if he can, he will not find any semantic relations between the 40 images, so as a result of regular repetition for a log time he will learn them as 'discreet entities'.

Our aim in creating a semantic network dictionary is that on the basis of a large body of empirical data we could show what other concepts can the most easily be associated in the STM of a 4—8 year old child to an input concept, i.e. how a semantic structure evolves in the STM. The weighted relations of semantic structures will make it predictable what words can the most easily and the most swiftly be associated with the limited vocabulary of children with dysphasia, i.e. what new theraphic opportunities are offered by the semantic network dictionary.

In my paper I will attempt to show how the dictionary is developed, what are the already achieved results and what further tasks are there to be solved.

2 The Semantic Structure

Memory research has proved [3] that those concepts can be acquired more easily to which we can associate some preliminary knowledge, than those ones to which we cannot associate anything. One stage of learning a new concept can be, that the new idea entering the SMT retrieves 1—6 such ideas (preliminary knowledge) from the LTM, which can some way be associated with the new one. (If the new concept does not retrieve anything from the LTM, it is either deleted, or a 'go-between' concept should be sought for, which can associate the new concept with a concept stored in the LTM, so this triple relation can trigger off the learning of a word.)

I call semantic structures the relations evolving in the LTM (either retrieved by exterior or interior stimulus), and with it I mean such a network of concepts, which is created in the LTM by retrieving 1—6 such concepts from the STM, which can be further associated with the concept of the prompting stimulus [2]. (The semantic structure consists of the relation of 2—7 concepts, because at least the relation of 2 concepts may prompt a third, and the upper limit [7 concepts] is necessary, because the capacity of the STM is unable to deal with more [10]. Our research has up to now proved that the STM capacity of 6-year-old children is approximately 4—5 units.

If we assess with empirical examination what semantic structure evolves in the STM of a healthy 5—8 year old child in response to a prompting stimulus, then on the basis of the processed empirical data it will turn out what units a semantic structure consists of, and taking it into consideration, on the basis of it we can plan more effective (faster and shorter) therapies for children with dysphasia, than before.

3 The Process of Making A Semantic Network Dictionary

3.1 The frequency dictionary of word usage of 6 year old children

In order to ascertain the already evolved semantic structures in case of the given population we had to compile a frequency dictionary, which relevantly shows what words the given population uses and/or what words it comes across most frequently [1].

During the compilation we processed the total vocabulary of all the 13 textbooks for 1^{st} grade students warranted by the Ministry of Education. Our preconception was, that the frequency dictionary compiled on the basis of the vocabulary of the textbooks will relevantly show what words occur the most often during the process of the acquisition of visual language in case of a 6-year-old child.

After processing the whole material contained in the textbooks, we deleted those grapheme sequences, which are not elements of our mother tongue, but they are unavoidable sound sequences during teaching how to read. Then we deleted the non-content words. Those words were considered to be content words, which carry all three sides of their meaning (lexical, grammatical and pragmatic).

Number of grapheme sequences	27.297
Number of content words	12.226
Number of words occurring 10 or more times	1.953

Table 1: Quantitative aspects of the dictionary.

Some of the quantitative aspects of the dictionary are shown in Table 1.

To create the semantic network dictionary we use the 200 most frequently occurring nouns. (We use nouns, because their acquisition is of primary importance as compared to the other words. We will process the verbs as well, but from another aspect.)

3.2 Compilation of data

Kindergarten nurses, teachers and high school students in 20 different settlements of the country currently do the compilation of data for the semantic network dictionary. The subjects of the compilation are third year kindergarten children, second and third year primary school children. We should like to collect data from 5,000 children by asking them what comes to their minds in connection with the 200 most frequently used nouns. We elaborated sheets on each of which 5 words occur, e.g. Father, mother, man, tree, word. The leader of the survey asks the question: 'What makes a tree a tree?' If the child has already acquired the visual language, he answers in writing, if not, he answers orally and the survey leaders takes down,

what he says. We do not use any quantitative limits, all tasks begin with the following command: 'Tell me . . . /Describe these concepts. . . '

The data compiled and the results processed so far show that children answer to a call word with 2.8 words or word sequences on the average, from which the following can be drawn: the semantic network dictionary will be competed on the basis of about 2,000,000 processed words. Naturally, these are not different words. On the basis of the processed material so far, we can say that about 350 concepts associate with one call word on the basis of material collected from 5,000 children. (The fluctuation is quite remarkable: the call word 'mother' retrieves 741 different associations, 'word' does only 178, so far these are the two extremities.) There will also be overlaps in the case of a call word associated with several concepts, that is why we cannot predict how many concepts will be included in the semantic network dictionary.

In the following I will describe five words, the processing of which on the given corpus has been finished: (the numbers in brackets show how many concepts are associated with the given word):

Mother:	she loves me	parent	dear	love	I love her
	(1588)	(1508)	(1263)	(1127)	(875)
Father:	parent	he works	love	he loves me	dear
	(1213)	(920)	(874)	(825)	(623)
Man:	being	life	he lives	head	clever
	(1813)	(1083)	(1046)	(934)	(930)
Tree:	plant	being	leaf	branch	air
	(2175)	(994)	(711)	(587	(514)
Word:	speech	letter	sentence	sound	he speaks
	(908)	(875)	(787)	(772)	(685)

3.3 Network Theories

In order to be able to compile our dictionary we studied the semantic network theories and their criticism [4,7,13], and the thesauri.

The problem with thesauri is that they do not provide any information on the relation of meanings of different lexemes. Furthermore, they combine the lexemes originating from different accents (regional, social, professional, etc.) without any remarks [5].

Still, the criticism of network theories and thesauri proved to me, why a network model resembling these theories may be the most appropriate conceptual framework to elaborate the most effectual therapies. Critics are right in saying that network models are to overpowering, that simply the evolution of the relation between the concepts is interesting, but they are incapable of showing the relationship between the concept and reality, which is the fundamental task of semantics.

However, from the viewpoint of the most effectual therapy that could be the most important point, that within the given population what relations between concepts already exist, and the large body of data is capable of showing this. It

can show on the basis of the frequency index what other concepts can the most easily and most effectively be linked with a given concept, and then what other concepts can be associated with these and so on.

Though no abstract semantic rule prevails in the evolution of semantic structures, which would be able to tell anything about the lexical meaning of a word, but a kind of subjective knowledge, still the total of empirically summed up subjective knowledges may provide some information on the words' organisation of meaning. This may not only present the free associations allocated to relevant words, but will unearth such relations between lexemes, which are created by each individual, i.e. it reflects the collocation.

Our semantic network dictionary will show these collocations, which is called by Landman intersubjective world-knowledge [9], so it will present not only the relations between concepts, but the new concepts generated by these relations, too.

4 Working of the software

When the whole word material is processed, computer-programmers will create a programme, which will present all the relations between the elements of the network with weighted bias. If we write, or say a word into a microphone, which is an element of the dictionary, then this lexeme will appear in the middle of the monitor, and the other five words with weighted bias, which is in the most frequent association with the word on the basis of the processed empirical material. If we click on, or say it in the microphone any of the five words forming a semantic structure with the original input word, then this will appear in the middle of the monitor as a call word together with the five most frequently associated concepts. As the software contains all the possible associations, you can go through the total vocabulary by evoking all the semantic structures in the middle of the screen. What we have said so far is shown schematically in Figure 2. taking the call word 'water' as an example.

The dictionary will be provided with a real-time speech-recognition system, which makes it possible for the children already capable of reading to take a virtual tour on the associated words of the network dictionary. This 'tour' will help enhance these semantic structures to develop and maintain in the memory of the child; i.e. it helps learning words.

We already possess a real-time speech-recognition system. Our research team recorded on computer 100 word inputs in eleven demonstration schools of the country, respectively from 500 subjects (6—7 year old children) on the basis of 1953 words (this is the number of words occurring 10 or more times) contained in the first grade textbooks. The recorded words were segmented and annotated. This material provides the basis for our software 'BeszédMester' [11], which can be employed for development in speech therapy and reading capability.

The basis for the speech-recognition module of our dictionary will be the nearly 250,000 segmented and annotated speech-sounds and the 50,000 recorded words (which includes the 200 most often used nouns serving as a base for the semantic

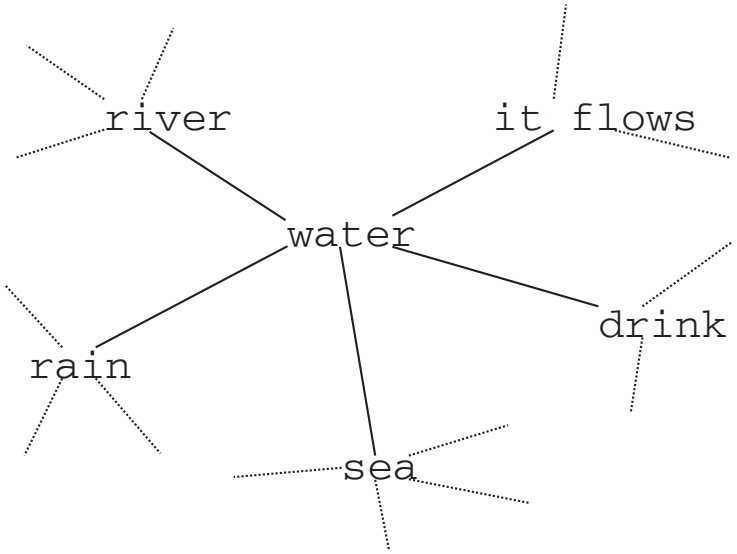

Figure 1: *A portion of the semantic network*

network dictionary).

We expect our semantic network dictionary to help therapist to plan development for children with expressive language disorder, as well as provide aid in expanding the vocabulary of healthy children.

5 Expected results of the research

The main purpose of the dictionary, as it was repeatedly mentioned, is to effectively aid the therapy of dysphasic children. In addition to the main purpose, several other aspects emerged during the project that may supply material to other researches. One such aspect is corpus linguistics since it would be impossible to process such an enormous database (nearly 3,000,000 words) in a short time without the advancement of informatics. It would be inconceivable to create all the empirically measured associations between the processed words. The complete survey and modeling of the associations may encourage further researches in child language, semantics, morphology and informatics as well as in pedagogy. The dictionary is innovative in both its form and functioning. There are already electronic dictionaries but this will be completely new in that it will allow unrestricted virtual walks through the whole network either manually or auditorily (by talking to the microphone) with immediate feedback about the relevant semantic structures. The Semantic Network Dictionary will seek its place in a cognitive theory not only because "association has its renaissance in cognitive psychology" [12] but also because a lot of cognitive processes are uninterruptible without postulating semantic

structures evolving from associations. Without aiming at completeness, only a few cognitive processes will be listed here whose semantic structures are built up from associations. Barkochba, an elaborate form of the Twenty Questions quiz game, is very popular in Hungary. When one is attempting to find out the thought of entity continuously narrowing a huge conceptual network, they are looking for some relevant concepts (4-7 in our assessment) to build the semantic structure from the associations that will designate the given entity. A similar situation is when one has to understand and/or learn a concept of which one has no information in their mental lexicon. In that case, one turns to "external memory devices," which are written, printed or electronic information storage media accumulated by mankind since the acquisition of visual language. The cognitive and evolutional aspects of external memory devices are discussed in detail by Donald [6]. External memory devices are used in a process in which a person's own knowledge of the world is being adjusted to the knowledge stored in external memory devices until the elements of the two knowledge structures build a semantic structure that makes it possible to define, understand or learn the given concept and put it in the LTM. External memory devices are scanned for data until a relevant concept in the person's memory is connected with an element of the external memory to create a semantic structure. The unit or part of the structure that can be connected to a unit or part of the external memory will be referred to as mediating semantic structure. The empirical studies of the role of the mediating semantic structure in cognition have just begun but one might already assume that all three semantic structure types will be found in every person's semantic network dictionary, i.e. there will be semantic structures that will be equivalent with the whole semantic structure, there will be mediating semantic structures and there will be semantic structures that will only contain the semantic structures of the external memory devices. We assume that the virtual walks in the dictionary will help in the process of learning by reducing the number of the semantic structures in the dictionary that are unknown to the individual.

6 Summary

Computer assisted study – whatever the object of study is – is an enormous motivating factor. Nowadays children learn earlier to use the computer, than to read or write.

Computer assisted study requires such didactics, programmes, which have not been part of institutional education and speech therapy yet. The search for new didactics, creation of programmes may provide a new conceptual framework for a scientific research. Computer assisted study modifies the framework of science(s) in order to apply it to its own research purposes. This is what we achieved when we re-evaluated the semantic network theories, as our aim was not the creation of a new semantic theory, but the development of plans capable of being used in the field education. Whether there exists a semantic structure, or can the examination of the association network be regarded as a scholarly undertaking is not our task

to decide. Practice provides us with convincing evidence about the existence and build-up of semantic structure. Supposition of its conscious build-up is useful for the methodology of speech therapy and vocabulary development.

That beyond the practical use of our suppositions there might be some scientific foundation can be proved by the fact, that Pléh speaks about the renaissance of association in cognitive psychology [12], and suggests that '... association and structure or logic are equally to be used in modelling higher processes.'

References

[1] Bácsi J. and Kerekes J. (2003): Az elsőosztályos olvasókönyvek szóanyagából készült gyakorisági szótár. "Van szó". [Frequency Dictionary Based on the First Grade Textbooks. 'The word be'.] Módszertani Közlemények. [Proceedings in Methodology.] 2003. II. 52–57.

[2] Bácsi J. (2003): A hálószemantika szerepe a megkésett beszédfejlődés terápiájában. [The Role of Network Semantics in the Therapy of Speech Development.] Alkalmazott Nyelvészeti Konferencia Füzetei. [Proceedings of Conference on Applied Linguistics.] 274–283.

[3] Baddeley A. (2001): Az emberi emlékezet. [Human Memory.] Budapest, Osiris Publishing House.

[4] Collins A. M. and Loftus G. R. (1975): A Spreading Activation Theory of Semantic Processing. Psychological Review, 82. 240–247.

[5] Crystal D. (1998): A nyelv enciklopédiája. [An Encyclopaedia of Language.] Budapest, Osiris Publishing House.

[6] Donald M. (2001): Az emberi gondolkodás eredete. Budapest, Osiris.

[7] Jonson-Laird P. N., Herrmann D. J. and Chaffin (1984): Only Connections: A Critique of Semantic Networks. Psychological Bulletin, 96, 292–315.

[8] Kerekes J. (2003): BeszédMester. [SpeechMaster.] Alkalmazott Nyelvészeti Konferencia Füzetei. [Proceedings of Conference on Applied Linguistics.] 44–49.

[9] Landman F. and Veltman F. (1982): Varieties of Formal Semantics. Proceedings of the 4. Amsterdam colloquium, Dordrecht, Holland.

[10] Miller, G. A. (1956): The Magical Number Seven, Plus or Minus Two: Same Limits on our Capacity for Processing Information. Psychological Review, 63, 81–97.

[11] Paczolay D., Tóth L., Kocsor A., and Kerekes J. (2002): Gépi tanulás alkalmazása egy fonológiai tudatosság-fejlesztő rendszerben. [The Use of Machine-Assisted Study in a System for Development of Phonology-Consciousness.] Alkalmazott Nyelvtudomány [Applied Linguistics] II(2), 55–67.

[12] Pléh Cs. (2003): Az asszociáció renaszánsza a kognitív pszichológiában. [The Renaissance of Association in Cognitive Psychology]
`http://sol.cc.u-szeged.hu/~pleh/magyar/cikkek/regi/4asszren.html`

[13] Quillian, M. R. (1969): The Teachable Language Comprehender: A Simulation Program and Theory of Language. Communication of the ACM, 12. 459–476.

Received May, 2004