

Towards A Model of an Interactive Dictionary for Children

Wiesława Ferlacka

UAM, IFA, Poznań

Matejki 39a/5

72-600 Świnoujście

Poland

Abstract

Electronic dictionaries embedded into e-readers and interactive storybooks available online and on CD-ROMs and directed to children aged 5 to 12 years of age, and to many teachers using novel technologies for education, are still inadequate as they are wanting both in quality and quantity of data. And these two aspects of e-dictionaries integrated in narrations for children have been the focus of attention in the project just started in 2005. This paper aims at sharing the main thrust of the project as well as the ways in which a model of an interactive dictionary is being developed. Exploring the word properties that texts written for children subliminally impose and their dictionary counterparts is one of the main components of the project. The focus is on establishing dynamic (interactive-compensatory) relationships between the reader, the text and the dictionary by drawing on both form and meaning.

1 Introduction

Electronic dictionaries embedded into e-readers and interactive storybooks available online and on CD-ROMs directed to children aged 5 to 12 years of age, and to many teacher using novel technologies for education, are still inadequate (Vetulani et al. 1998: 73), as they are wanting both in quality and quantity of data. These two aspects of e-dictionaries integrated in narrations for children have been the focus of attention in the project just started in 2005 in connection with my PhD research.

One of the main components of the project is exploring word properties subliminally imposed by texts written for children and their dictionary counterparts. The focus, though, is on establishing dynamic (interactive-compensatory) relationships between the reader, the text and the dictionary by drawing on both form and meaning. Realizing that it is 'easier to criticize dictionaries than to do the work of compiling them' (Neubauer 1984: 121), the main objective of the project is establishing a Model of an Interactive Children's/Reader's Dictionary.

2 Conceptual Development of the Project

Behind the concept of such a dictionary stands a dissatisfied teacher, the author of this paper. Without her steadfast presence and lexicographic assistance, many of her learners would have given up reading interactive storybooks for lack of vocabulary explanations and comprehension problems. She had to stand for an electronic dictionary as it was on average

inadequate for the learners' needs. This is why her 'dictionary experience' may now, hopefully, be utilized in conceptualizing the above-mentioned Model.

2.1 Assumptions

One of the greatest difficulties of this project is to identify the areas in which the proposed Dictionary Model is feasible. The basic assumption is that the dictionary needs no longer to be a silent partner in reading and that it may take over the role of a teacher and companion in the process of language learning and reading, as well as having fun (see 'Tandem Learning' in Ferlacka 2005a). Another assumption is that mining the knowledge buried in the texts (e.g. semantic features of a word) and incorporating it into the dictionary formula will benefit readers. It will help readers' interact with texts more consciously, build their autonomy, as well as facilitate language acquisition (cf Ferlacka 2005b). Also word difficulty (Nation 2001: 7, Kuczyński 1997), its transparency and visualization potential, seen as critical components of the dictionary are, inter alia, well established assumptions of the project. This is why, Sobkowiak's (2005: 200) investigations into word phonetic difficulty (Phonetic Difficulty Index = PDI) are believed to be of assistance in establishing which words need specific phonetic assistance. Similar work on other difficulty indices, though, will have to wait for further research because this work is beyond the author's present powers. Therefore, other aspects of word difficulty within the framework of this Dictionary Model will be resolved more arbitrarily until further research is done.

2.2 Data Collection and Corpus Representativity

It has been essential for the project to acquire a corpus of children's fiction and non-fiction literature because if a reading dictionary is to be a good lexicographic assistant it needs to cover both types of texts for children. ProQuest LearningPage (USA), for whom the author has worked as a consultant, has proved remarkably supportive by providing her with full access to textual data ('Reading A-Z'), i.e. as many as 500 texts written for native children, which are doubly useful by being already divided into fiction and non-fiction, as well as graded into levels from A to Z. This makes them perfect texts also for ESL/EFL children by virtue of being gradable, authentic and varied.

Compiling a Textual Corpus is naturally only a beginning on the way to establishing a Dictionary Corpus. However, the main framework of the Dictionary Model being already established will make such work much easier. Unfortunately, much time needs to pass before the whole undertaking takes full shape.

2.3 Data Processing

The work of establishing lexicographese (the coded language) for the Dictionary, as well as the time-consuming coding and annotating of the files, is still in progress. Therefore, at the present moment, a pilot study involving a textual analysis of five so-called interactive storybooks for children has been carried out. This investigation into the content of these storybooks will help structure the dictionary and establish the ways in which concepts are built by these texts (Ferlacka 2006).

2.4 On The Quality of Data

The most striking conclusion is that definitions in children’s literature are scanty and inadequate, and not very infrequently lead to misinformation and miscomprehension. These critical remarks can be verified against the following definitions taken from electronic story-books by Disney (1995):

‘stoutness’ – ‘This is what makes Winnie The Pooh do exercises every morning.’ (‘Winnie The Pooh And The Honey Tree’)

‘gypsy’ – ‘Esmeralda’s friends. People who like to travel.’ (‘The Hunchback Of Notre Name’)

As one can easily see, the dictionary in these storybooks serves an entertaining rather than an educational role. The child can in no way get to the meaning of a word. This is why, the project involves investigating into the quality of such definitions as well.

Naturally, children can be expected to infer some meaning from the context (Nunan 1991) but their sensitivity to it depends on a complex interplay of contextual factors and their potential of cognitive gains. The acquisition of the ability to derive such implicatures is of great interest here, the more that they are reported to contribute to semantic content.

2.4.1 Text-Mining: Pragmatic Implicatures and Semantic Content. Text-Derived Definitions

The already-mentioned investigation into texts for children has shown a path to challenging Wiegand’s (1984: 18-21) statement that a lexical paraphrase can always be interpreted as an explanation of the meaning or as a description of an object! Defining ‘barracks’ as ‘a set of buildings for the permanent housing of troops’, can neither be treated as a sufficient explanation of meaning, nor as a sufficient description of the object due to providing the user with insufficient characteristic properties of ‘barracks’. This lexical paraphrase does not build any particular picture in the mind of the user and the learner may still fail to know what ‘kind’ of buildings ‘barracks’ are.

Text-derived lexical paraphrases may serve as an example of an intervention into the child’s world of concepts and seeding it with information about properties that ‘have to’ be appropriated, as well as those that are ‘arbitrary’ and connotative rather than denotative, yet very useful for a much-telling explanation of meaning. Let us analyze a handful of examples of what one can derive from reading texts for children carefully:

Word or phrase taken from the Storybook	Textual Implications
(When you are) a bachelor	you need a mate
	you have an eye on a pretty lady
	you arrange for you and her to meet

Figure 1. ‘101 Dalmatians’, Disney, 1995.

The representation of *'bachelor'* is made up of "adult," "unmarried," and "male" (cf Quine 1951). When compared with dictionary definitions:

- (1) *'a man who is not married'* (Cambridge Learner's Dictionary Online)
- (2) *'a man who has never been married'* (Onelook Dictionary Online)
- (3) *'unmarried man – (a man who has never been married)'* (WordNet 2.1)

one can conclude that the 'definition' derived from the above texts conveys these aspects, though in a different way. This fact might be made use of to start an exchange of information between the computer and the learner about the meaning of the word (interactive reading). The above online dictionaries do not mention 'adult' as a necessary property of the word. 'Not married' and 'male' are the only properties the dictionaries provide, which implies that the reader should be capable of deriving the meaning of the word 'bachelor' from the context, if appropriately guided to deduce it from the text. This is where the Model of Interactive Dictionary is to take over in interaction, the dictionary being seen as the guide towards meaning, rather than its provider only. The question is if the dictionary ought to remain a reference tool or become a tool for meaning construction AND a teacher of autonomy.

2.4.2 *The Usability of a Text-Derived Definition*

The dictionary definition in order to be 'usable' ought to state 'the properties which predicates belonging in the lexical paraphrase must have' (Wiegand 1984: 25). Text-derived semantic units required for meaning construction define 'a bachelor' as a person looking for a pretty lady, needing a mate and wanting to arrange a meeting. One could now emphasize the inferential relationships between these concepts and help the reader realize that "a person who needs a mate, has an eye on a lady, and wants to arrange a meeting" implies a male in the unmarried/single state. This text-derived meaning might assist the reader just as well as a regular dictionary definition, or might at least support it. Naturally, not every word can be supported so. What can provide child-readers with a wedge into unknown lexicon and teach them autonomy is extracting such supportive 'definitions' and promoting looking for them in the text.

According to definitional theory of meaning, the meanings of words can be fractured into component parts in a set of levels. On an unsophisticated level the problem of meaning can be approached as the perception of certain physical (e.g. sounds) and non-physical, abstract (feelings about the object) entities which might lead the reader to associative thinking. 'Words are not isolated units of language, but fit into many interlocking systems and levels. Because of this, there are many things to know about any particular word and there are many degrees of knowing' (Nation 2001: 24). The operational content of text-derived definition leaves scope for prosodically marked complementary and supplementary word meaning description. This is why the project involves an empirical search for correlations between text-based word meaning and dictionary definition (cf Neubauer 1984: 121) as well as the ways into which such correlations might improve the interaction between the reader, the text and the dictionary.

One may only hope that such statements about semantic correlations could also apply to other texts in a larger corpus. But this requires further research into children's literature (Ferlacka 2006).

2.4.3 Quantity: Access Units – The Target Words vs All Words

The pilot analysis also shows that the number (Biber – Conrad – Reppen 1998) of highlighted glossed words in storybooks is very limited (mean: 17 glossed words in 100). According to research glosses should cover 2-5% unknown tokens (cf. Nation 2001: 147) because pleasurable reading is claimed to take place when 95-98% of running text is understandable. Yet, it is doubtful if this number of unknown tokens (2-5%) falls within the total of an individual learner's unknown vocabulary. Are they really unknown words that are glossed? This is the question that only further research can answer.

The number of words to be defined so that a reader can understand a text is hard to arrive at (Nation 2001: 83, 8; Stoller – Grabe 1993: 32-33). What for one is enough, for another may be too little. According to Granger (ed.) (1998: 160) 'a dictionary is only of use to the student if it includes the word or phrase the student wants'. Otherwise this may lead to incomprehension (Nation 2001: 177, Vetulani et al. 1998: 73). Therefore, additionally to the target words highlighted on every screen page, there ought to be lexicographic assistance for all the words to enable the learner to autonomously (within some reasonable limits) choose what words to look up. And this goal is manageable in the light of the state-of-the-art lexicography.

3 Interaction-Compensation

'Forms and methods of presentation have not been recognized as inherent constituents in lexicography (...) And the question that has been neglected but what is central to lexicography is whether there are many intrinsic interdependencies between data given and the methods used to present them' (Stein 1984:124).

Here a step towards humanizing the dictionary is postulated and an analysis of the meaning of word 'interaction', as well as incorporating learning in which the dictionary becomes a partner in learning (cf Granström 2004) to allow for a variety of interactive styles and the levelling of the language used in the dictionary. The word 'interaction' and 'interactive' are still very ambiguous. Only when the learner can actively search for meaning, should we speak about interactive reading and interactive language learning. Interaction also ought to involve playing with the language allowing both the computer dictionary and the learner to have fun exchanging information and arriving not only at the proper meaning of a word but also at its shades and blurred senses.

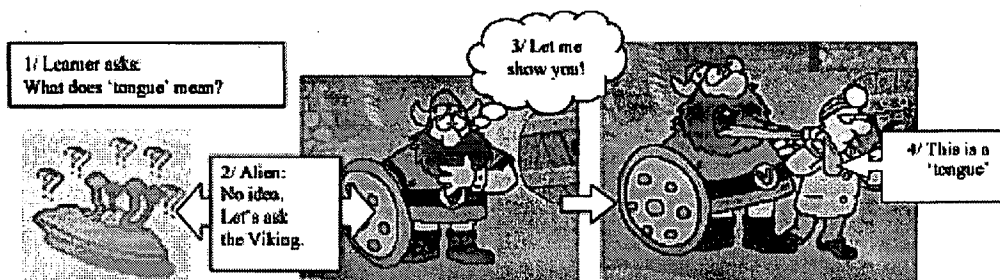


Figure 2. Tandem H-C Learning (Ferlacka 2005a).
Graphics taken from 'The Interactive Dictionary', 1995, Corel.

3.1 Human-Computer Tandem Learning

'Tandem learning' describes learning in which one part of the dictionary takes the place of a human and becomes a partner in learning (cf. Mindplay 2005, Kaliski in Roach (ed.) 1992: 105, Berleant et al. 1997) so that the learner may have the feeling that his communication with the dictionary extends beyond the interaction with an unfeeling reference tool (cf. Bickmore et al. 2005, Cole 2003, Massaro 2004, Brophy 1999, Ma – Kevitt 2003, Nakano et al. 2005, Johnson – Rickel – Lester 2000: 47). If mouse clicks, keyboard strokes or speech are to create a semblance of Human-to-Human Interaction that is believable and encouraging to learn, it should be both the reader AND the computer to decide which modality to choose for interaction (Kopp – Wachsmuth 2004, Johnson – Rickel – Lester 2000, Pangburn 2002).

4 Conclusion

A model may itself be conceptual or physical. The project I have presented here is still at its conceptual phase. A cooperation of computer scientists and lexicographers as well as scholars from many other related fields is crucial if the concept of this dictionary is to take a physical form. Although more research is needed in order to make such a dictionary a reality, this day is believed not to be that distant.

References

A. Dictionaries

- 'The Interactive Dictionary' on CD-ROM (1995). Corel.
- 'WordNet 2.1' at <http://poets.notredame.ac.jp/cgi-bin/wn?cmd=wn&word=bachelor> 2006-03-23..
- 'Cambridge Learner's Dictionary Online' at http://dictionary.cambridge.org/define.asp?key=HW*11027&dict=CLD2 2006-03-23.
- 'Onelook Dictionary Online' at <http://www.onelook.com/?w=bachelor&ls=a> 2006-03-23.

B. Other literature

- Berleant, D. et al. (1997), 'LEARN: Software for Foreign Language Vocabulary Acquisition from English Unrestricted Text'. *Computer Assisted Language Learning* 10(2). At <http://class.ee.ias-tate.edu/berleant/home/me/cv/papers/LEARN97.html>
- Biber, D., Conrad, S., Reppen, R. (1998), *Corpus Linguistics. Investigating Language Structure and Use*, Cambridge, CUP.

- Bickmore, T. W. et al. (1998), 'Animated Autonomous Personal Representatives', in *Proceedings Of The Second International Conference On Autonomous Agents*. New York: ACM Press. At <http://doi.acm.org/10.1145/280765.280768> 2006-03-23.
- Brophy, S. et al. (1999), 'Teachable Agents: Combining Insights From Learning Theory And Computer Science'. At <http://citeseer.ist.psu.edu/544982/10> 2006-03-23.
- Cole, R., van Vuuren, S., Pellom, B. et al. (2003), 'Perceptive Animated Interfaces: First Steps Toward a New Paradigm for Human Computer Interaction', in *Proceedings of the IEEE: Special Issue on Multimodal Human Computer Interface*. CSLR. At <http://cslr.Colorado.edu>
- Disney Interactive (1996), '101 Dalmatians – Animated Storybook' on CD-ROM.
- Disney Interactive (1996), 'The Hunchback Of Notre Name' on CD-ROM.
- DK Multimedia (1995), 'My First Amazing World Explorer' on CD-ROM.
- Granström, B. (2004), 'Towards A Virtual Language Tutor', in *Proceedings Of INSTIL/ICALL2004 – NLP And Speech Technologies In Advanced Language Learning Systems. Venice* (1-8). At <http://www.speech.kth.se/ctt/publications/#2004> 2006-03-23.
- Ferlacka, W. (2005a), 'The Place Of Dictionaries In Interactive CALL', in *Proceedings Of 36th Poznań Linguistic Meeting (PLM 2005)*. Abstract at <http://elex.amu.edu.pl/ifa/plm/2005/> 2006-03-23.
- Ferlacka, W. (2005b), 'How Can E-readers Stimulate Phonological Development?' (forthcoming), in *Proceedings Of Phonetics In FLT 5, Soczewka*. At <http://elex.amu.edu.pl/~sobkow/Soczewka05/Ferlacka.ppt> 2006-03-23.
- Ferlacka, W. (2006), 'The Conceptualization Of Immigration in Children's Narratives' (forthcoming), in *Proceedings Of 37th Poznań Linguistic Meeting (PLM 2006)*.
- Johnson, W. L., Rickel J. W., Lester, J. C. (2000), 'Animated Pedagogical Agents: Face-to-Face Interaction', in *Interactive Learning Environments. International Journal Of Artificial Intelligence In Education* 11: 47-78. At http://aied.inf.ed.ac.uk/abstracts/Vol_11/johnson.html 2006-03-23.
- Kaliski, T. (1992), 'Computer-Assisted Language Learning (CALL)', in Roach, P. (ed.). *Computing And Phonetics*, London – San Diego, Academic Press Ltd, pp. 97-110.
- Kopp, S., Wachsmuth I. (2004), 'Synthesizing Multimodal Utterances For Conversational Agents'. *Journal Of Visualization And Computer Animation* 15(1): 39-52. At <http://www3.interscience.wiley.com/cgi-bin/jissue/106562740> 2006-03-23.
- Kuczyński, M. (1997), 'Selection Of Words To Be Taught In the English Philology Department' (PhD diss.) Poznań: UAM.
- Minhua, M., Mc Kevitt, P. (2003), 'Building Character Animation For Intelligent Storytelling With The H-Anim Standard', in McNeill, M. (ed.) *Proc. Of Eurographics Ireland 2003 Workshop* (9-15), Faculty Of Informatics: University of Ulster, Coleraine, Northern Ireland.
- Massaro, D. W. (2004), 'Symbiotic Value Of An Embodied Agent In Language Learning', in *Proceedings Of The 37th Hawaii International Conference On Systems Sciences*. At <http://doi.ieeecomputersociety.org/10.1109/HICSS.2004.1265333> 2006-03-23.
- Mindplay* (2005). At <http://www.myreadingcoach.com> 2006-03-23.
- Nakano et al. (2005), 'Converting Text Into Agent Animations: Converting Gestures To Text', in *Proceedings Of The Human Language Technology Conference (HLT-NAACL04)*. At http://acl.ldc.upenn.edu/hlt-naacl2004/shorts/ps/272_Paper.ps 2005-05-28.
- Nation, I.S.P. (2001), 'Learning Vocabulary In Another Language'. CUP, Cambridge.
- Neubauer, F. (1984), 'The Language Of Explanation In Monolingual Dictionaries', in Hartmann, R.R.K. (ed.) *LEXeter '83 Proceedings* Tuebingen, Max Niemeyer Verlag, pp. 117-123.
- Nunan, D. (1991), *Language Teaching Methodology*. Englewood Cliffs, NJ: Prentice Hall.
- Pangburn, B. E. (2002), *Experience-Based Language Acquisition: A Computational Model Of Human Language Acquisition* (PhD diss.). Louisiana State University. At http://www.greatmindsworking.com/downloads/01_full_dissertation.pdf 2006-03-23.
- Quine, Willard Van Orman (1951), 'Two Dogmas of Empiricism'. *The Philosophical Review* 60: 20-43. At <http://www.ditext.com/quine/quine.html> 2006-03-23.
- Proquest Learning Page (2006), *Reading A-Z* at <http://www.readinga-z.com> 2006-03-23.

- Sobkowiak, Włodzimierz (2005), 'Phonolapsological Equivalence And Similarity In the English lexicon', in Kiefer, F., Kiss, G., Pajzs, J. (eds). *Papers In Computational Lexicography. Proceedings Of The 8th International Conference On Computational Lexicography (COMPLEX 2005)* (200-212). Budapest, Linguistics Institute (abstract).
- Stein, G. (1984), 'Towards A Theory Of Lexicography: Principles And/Vs. Practice In Modern English Dictionaries', in Hartmann, R.R.K. (ed.) *LEXeter '83 Proceedings*, Max Niemeyer Verlag, Tuebingen, pp. 124-130.
- Stoller, F. L., Grabe W. (1993), 'Implications For L2 Vocabulary Acquisition And Instruction From L1 Vocabulary Research', in Huckin, T., Haynes, M., Coady, J. (eds). *Second Language Reading And Vocabulary Learning* (24-45), Ablex Publishing Corporation.
- Vetulani, Z. et al. (1998), *Dictionary Based Methods And Tools For Language Engineering*. Wydawnictwo Naukowe UAM, Poznań.
- Wiegand, H. E. (1984), 'On The Structure And Contents Of A General Theory Of Lexicography', in Hartmann, R.R.K. *LEXeter '83 Proceedings*, Tuebingen, Max Niemeyer Verlag, pp. 13-30.