

Multiple Use of Content in a Web-Based Language Learning System

Judith Knapp
European Academy of
Bolzano
jknapp@eurac.edu

Johann Gamper
Free University of Bolzano
jgamper@unibz.it

Peter Brusilovsky
University of Pittsburgh
peterb+@pitt.edu

Abstract

Content creation for a language learning system is a time-consuming, laborious, and expensive task. An important part of content in a language learning system is the illustrative content, namely example sentences. We discuss a solution to reuse illustrative content in a Web-based language learning system for the German and Italian languages. Our approach explores fully the potential of hypertext, which allows building complex networks of small pieces of learning material. A first evaluation shows promising results.

1. Introduction

At the European Academy of Bolzano an interdisciplinary research team is currently developing an innovative Web-based language learning system for the German and Italian languages, called ELDIT (<http://www.eurac.edu/eldit>). Its core module consists of a German and an Italian learners' dictionary, each one containing approximately 3,000 word entries.

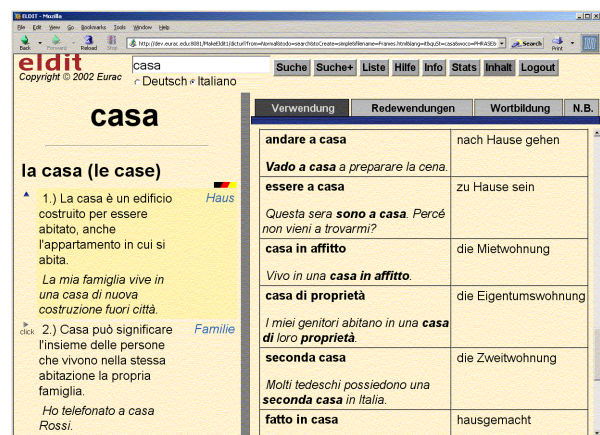


Figure 1: Dictionary entry of the Italian word "casa" (house) in ELDIT

Figure 1 shows a screenshot of the Italian dictionary entry "casa" (house). In the left-hand frame the lemma (casa) together with morphological information and different word meanings are shown. Each word meaning is described by a definition, a translation and an example sentence. In the right-hand frame additional information regarding the correct usage of the word is shown. A tab metaphor is used to illustrate typical usage patterns, word relations, etc. The tab "Verwendung" (usage), which is selected in Figure 1, shows collocations and word combinations, which are described by a pattern on the left, a translation equivalent on the right, and a lexicographic example below. More details can be found in [1, 5].

2. Illustrative Content in ELDIT

Psycholinguistic findings suggest that words should not be learned in an isolated way (e.g. as lists of words and respective translations), but in relation to each others and in a context, i.e. applied in example sentences and explored in texts [2, 7]. Therefore, ELDIT contains a large number of text sentences and lexicographic examples which are used to illustrate language in use, i.e. how the rules can be applied to form correct and meaningful sentences. To overcome problem with new words in illustrative sentences, each word is linked to the corresponding entry in the dictionary [5]. We call this example sentences *illustrative content*.

However, there is still a need for even more examples. Especially when learners have a specific context or sentence in mind, they might want to see an example sentence which is rather close to their specific needs. This is an important finding resulting from a first evaluation of ELDIT.

The authoring of didactic material is difficult and time-consuming. Modern hypermedia technologies, however, make it possible to use the existing content in different parts of the system. Hence we can neglect traditional paper based approaches which apply a 1:1 relation between information and illustration. In electronic systems the definition of an n:m relation between information and illustration is possible, which magnifies the amount of example sentences that can be provided for each piece of information.

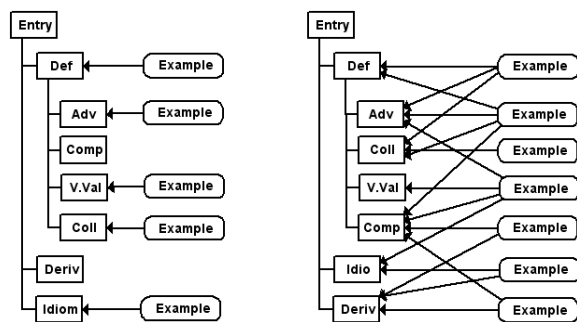


Figure 2: Traditional and new way of providing information and example sentences in ELDIT.

We have programmed a module that allows reusing the existing illustrative content in ELDIT at several places. Now for each piece of information not only the main example added by the linguists, but also additional

examples created for other contexts but showing the same information can be retrieved. The additional examples can be accessed via a button placed next to the standard example.

3. Reusing the Illustrative Content

The basic idea is to retrieve all example sentences which are useful to further illustrate the concept under consideration.

3.1. Problems

The retrieval of additional example sentences is not trivial. A simple search in our database does not lead to the desired results. The following problems occur:

1. The patterns under consideration might be unstructured, for instance they may contain Meta information such as slashes to indicate variations. For instance, the pattern *gli occhi, la bocca, il viso, ..., belli/bella/bello* indicates several patterns: “gli occhi belli (beautiful eyes)”, “la bocca bella (a beautiful mouth)”, and “il viso bello (a beautiful face)”. All these patterns should be considered when searching additional examples.
2. Words occur in declined or conjugated form both in the patterns and in the example sentences. For example, the collocation *to go home* occurs in the sentence “Yesterday I went home very late”, and therefore this sentence should be matched.
3. It is not sufficient that all words of a pattern occur in a lexicographic example, but they must occur as a collocation. For instance, the word combination *to go home* occurs as a collocation in the sentence “I went home very late”, but not in the sentence “I went out and came home very late”.
4. Words usually have several meanings. For instance, the word *house* may be a building but also a dynasty. Hence, the sentence “The royal house of Norway is a branch of the princely family of Glücksburg” is not a good illustration for the definition “A house is a place to live and to work”.

3.2. Implementation

With regards to the previously mentioned problems the retrieval of these additional examples is a four-step process:

1. Extracting “clean” patterns
2. Retrieving all example sentences
3. Recognition of collocations
4. Disambiguation of meaning

3.2.1. Extracting “clean” patterns The first step is to construct new “clean” patterns which can be passed to our search engine. In some cases it is easy to get such a “clean” pattern, since the given pattern consists only of one word, it occurs in the citation form, and different meanings are not considered. Hence, the word can directly be passed to the search engine. In other cases we have multiple word expressions which are given in an unstructured form. In these cases we unfold meta-symbols and remove general words. In some cases even different word meanings have to be considered. In these cases we build our search expression by adding some specific words from the main lexicographic example, from which we assume that they describe the meaning of the word in consideration.

3.2.2. Retrieving examples The second step is to retrieve all possible examples from the ELDIT dictionary by passing the search pattern obtained in step 1 to the search engine. Our search engine applies NLP techniques both on the search pattern and on the learning material in order to retrieve also results that might occur in inflected form [4].

3.2.3. Recognizing collocations The third step is to recognize collocations. A sophisticated concordance tool is required to check whether a word combination forms a collocation in a sentence or not. Currently, we are using a rather simple approach based on a set of rules to identify real collocations. A more sophisticated disambiguation could be done by including a tool such as “Phrase Manager” [8]. The inclusion of this system into ELDIT is part of our future work.

3.2.4. Meaning disambiguation The last and most difficult step is the disambiguation of word meanings. Currently this step is compiled into the search patterns which include nouns and verbs from the original example sentences.

4. Performance

We analyzed how many additional examples are found and how many of these items are useful/valid from the didactic point of view.

In many cases finding examples works with a precision of 100%, since only one word has to be searched, and different meanings are not considered in the presentation. Hence, in our analysis we got a precision of 100%. This result is very important, since we have no manually created examples for these words.

In cases of multiple word expressions some problems with invalid results arise (precision is approximately 90%) since our rule system is not a very precise concordance tool.

In the cases where we used the original example sentence to perform meaning disambiguation generally a lot of (mostly valid) examples are found for common, frequently used meanings, whereas only very few (and mostly invalid) examples are found for the less frequently used meanings and patterns.

5. Related Work

Searching text corpora with concordance tools is not new in language learning. For instance, the aim of the PET 2000 [3] and Lexica [6] projects, was to help students to acquire the necessary vocabulary for the English language. Students can search a text corpus using concordance tools. In this way they can explore words in context and build their own vocabulary database consisting of words with definitions and examples.

ELDIT takes a different approach by combining hand-crafted examples with automatically retrieved examples. The detailed data model, the hypertext nature of our system, and the application of computational linguistics technologies allow searching and reusing

these pieces on different places, hence we can offer a large number of carefully prepared text pieces on many different places which magnifies the amount of information provided in the system. Moreover, in ELDIT all words are linked to the corresponding dictionary entry, hence unknown words found in the example sentences can be checked by a simple mouse click.

6. Conclusions

We presented a solution to content reuse in the ELDIT language learning system for German and Italian. Since the authoring of learning material is very time-consuming, it is desirable to reuse it for different learning situations. Our approach to reuse illustrative example sentences explores the potential of hypertext and NLP technologies, which allows building complex networks of small pieces of learning material. A first evaluation shows promising results of our approach and the didactic usefulness of this additional information for the learner. Still, the automatic generation of the hypertext links using shallow reasoning and pattern matching techniques can and has to be improved.

7. References

- [1] A. Abel and V. Weber. ELDIT, prototype of an innovative dictionary. In *Proceedings of the 9th EURALEX International Congress on Lexicography (EURALEX'00)*, Stuttgart, Germany, 2000.
- [2] J. Aitchison. *Words in the Mind: An Introduction to the Mental Lexicon*. Blackwell Publishers Ltd, Oxford, UK, 2nd edition, 1994.
- [3] T. Cobb. Breath and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12(4):345–360, Oct. 1999.
- [4] M. Domenig and P. ten Hacken. *Word Manager: A System for Morphological Dictionaries*, volume 1 of *Informatik und Sprache*. Olms Verlag, Hildesheim, 1992.
- [5] J. Gamper and J. Knapp. A data model and its implementation for a Web-based language learning system. In *Proceedings of the 12th International World Wide Web Conference (WWW2003)*, pages 217–225, May 2003.
- [6] R. Goodfellow. Evaluating performance, approach and outcome in the design of CALL. In *CALL: Media, Design & Applications*, pages 109–140. Swets & Zeitlinger, The Netherlands, 1999.
- [7] B. Kielhöfer. Psycholinguistische Grundlagen der Wortschatzarbeit. *Babylonia*, 1996.
- [8] S. Pedrazzini. *Phrase Manager: A System for Phrasal and Idiomatic Dictionaries*, volume 3 of *Informatik und Sprache*. Olms Verlag, Hildesheim, 1994.