Natural Language Engineering **4** (1): 41–55. Printed in the United Kingdom

© 1998 Cambridge University Press

Using semantic lexicons for full text message retrieval in a communication aid

STEFAN LANGER
CIS, University of Munich, München, Germany
e-mail: stef@cis.uni-muenchen.de

MARIANNE HICKEY
Applied Computing, University of Dundee, Dundee, Scotland
e-mail: mhickey@mic.dundee.ac.uk

(*Received 21 October 1997*)

Abstract

In this paper, we present results of a project that investigated the application of lexicon based text retrieval techniques to Alternative and Augmentative Communication (AAC). As a practical outcome of this research, a communication aid based on message retrieval by key words was designed, implemented and evaluated. The message retrieval module in the system uses a large semantic lexicon, derived from the WordNet database, for query expansion. Trials have been carried out with the device to evaluate whether the approach is suitable for AAC, and to determine the semantic relations that lead to efficient message retrieval. The first part of this paper describes the background of the project and highlights the retrieval requirements for a communication aid, which differ considerably from the requirements in standard text retrieval. We then present the overall design of the WordKeys communication aid and describe the tasks of its sub-modules. We summarise trials that have been carried out to determine the effect of semantic query expansion on the success of message retrieval. Evaluation results show that information about word frequency can solve problems that occurred in the semantic query expansion because of taxonomies that have too many intermediate steps between closely related words. Finally, a user evaluation with the improved system showed that full text retrieval is an effective approach to message access in a communication aid.

1 Full text retrieval for a communication aid

Computer based voice output systems have been designed to help non-speaking people, who often also have physical disabilities, to communicate independently. Although many voice output communication aids are available, there are major design problems to be overcome if they are to help non-speakers participate more in social interaction. A principal deficiency of the current generation of devices, which has received much research and development effort, is the low rate of communication that can be achieved by users. Rates of between 2 and 26 words per minute have been reported (Kraat 1985), which compares poorly to natural speech rates of

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

about 150 words per minute (Foulds 1980). The low communication rates do not encourage device users to create messages, or communication partners to maintain attention. Thus, one of the primary research aims in AAC is to find ways that allow non-speakers to achieve relatively high communication fluency. For this, there needs to be low cognitive load: AAC systems should take as much of the cognitive load as possible away from users, leaving them freer to participate more fully in conversations. It also implies high efficiency, incurring low physical load on device users.

Currently, there exist different types of communication aid for non-speaking people, technology and non-technology based, with some systems based on symbolic communication and others based on natural language. Among the systems using natural language, we can distinguish two different approaches. The communication strategy can be based on enhanced message composition (as with word or character prediction), or on the selection of pre-stored messages. Approaches of the latter type are based on the assumption that a lot of conversational material is re-used over time, in different conversations (Newell and Alm 1994). It is this type of communication aid that will be discussed further here.

Communication aids based on message selection that are in current use have several constraints:

- They rely on retrieval methods that put a high cognitive load on the user. The user must remember an access route, or in some cases a code, in order to speak a message. This effectively restricts the number of accessible messages, as the user of the system has to remember the links between symbols or codes and the messages. The high cognitive load and the resulting errors are likely to have an adverse affect on communication rate.
- The user, or a carer, has to enter message identifiers, or other knowledge, for each new message or has to decide on where to store a new message in the structure of the system.

The WordKeys approach to try to overcome these problems is the use of full text retrieval to access a message database. This eliminates the need for the user to remember any message numbers or another code. A non-speaker can select a conversational item from the database by entering one or several key words. Appropriate messages will be those related to these input words (Hickey and Page 1993).

At a first glance, the implementation of such a system for AAC users might seem straightforward as retrieval techniques for full text have been investigated in information retrieval research for decades. However, there are major differences between the requirements for a communication aid and the requirements for standard text retrieval. One of the differences is the document length. Most algorithms suggested in the literature for full text retrieval are designed for collections of larger documents and are not applicable to the short messages communicated by AAC users. A second difference is the purpose of the retrieval. Whereas in standard information retrieval documents are sought to gain access to information, messages in a communication aid are retrieved to be used in conversations. This means that, in

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

order for a non-speaker to participate in a timely fashion, it is crucial that his or her AAC device requires minimal input. In standard text retrieval, queries of 5–10 words are regarded as short queries (Hearst 1996). This is too long for message retrieval in a communication aid. Users of these devices typically have a very low typing rate and in a conversation they need to access the target message as quickly as possible. Therefore it is desirable that any message from the database can be retrieved by only one key word, without the need for query refinement. Under these circumstances sufficient retrieval of target messages can only be achieved with a high recall rate, which necessarily increases the number of irrelevant messages. Thus, appropriate ranking of the messages is a key requirement for an AAC system.

The state of the art, and the special requirements for a retrieval module in an AAC device, suggest the use of enhanced full text retrieval using semantic expansion. It has been shown that it looks promising to further investigate the use of electronic semantic lexicons, both for query expansion and to overcome problems of word sense ambiguity (Richardson and Smeaton 1995). Especially relating to short messages, research on image caption retrieval has shown that the recall rate can be considerably higher if suitable methods are used to calculate semantic distances between query words and message words (Smeaton and Quigley 1996). Semantic query expansion is especially suited for communication aids, where minimal input and high recall are the most important factors. For large databases, precision also becomes a very important factor, but appropriate ranking of messages can help to alleviate the impact of the retrieval of unwanted messages.

The measurement of semantic distance can be based on the semantic relationship between words. Linguistic research in lexical semantics has, for many years, been investigating these semantic relations (Cruse 1986). Some of them are encoded in electronic thesauri and dictionaries. Relevant links for message retrieval are not confined to synonymy. In particular, hyponymic links should be considered for text retrieval purposes (Richardson and Smeaton 1995). Although there are other semantic relations that are interesting candidates for query expansion, such as meronymy, this paper will only discuss the integration of hyponymic relations in the text retrieval algorithm.

For semantic query expansion, a comprehensive electronic dictionary containing semantic information is needed. Research in electronic lexicography has been very intense during recent years, and many large dictionaries are being built for different languages. Some of them are based on printed dictionaries, others have been built from scratch as electronic dictionaries. Few of those dictionaries, however, are publicly available and, of those available, not all are suitable for the retrieval of unrestricted text. An electronic thesaurus, derived from the version of *Roget's Thesaurus* published in 1911, is available in the public domain but, despite some recent entries, the information contained there is partially outdated. The semantic database WordNet (Miller *et al.* 1990; Miller *et al.* 1993) has already been successfully used for information retrieval purposes (Guglielmo and Rowe 1996; Smeaton and Quigley 1996), and has been a source for the design of a lexical database for AAC systems (Zickus *et al.* 1995). WordNet (version 1.5) currently has 119,200 words in its database, which guarantees coverage of almost the complete English

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

vocabulary. The main information provided is about semantically related words for the different senses of each lemma in the database. As WordNet is a general lexicon of English lexemes, a query expansion algorithm based on this database is not restricted to any specific domain.

2 Design of the WordKeys communication aid

2.1 System overview

The WordKeys communication aid was designed to provide non-speaking people with an efficient means for message storage and retrieval, and to evaluate the performance of algorithms from intelligent message retrieval for use in AAC. There are two main functions of the system:

- In communication mode, WordKeys is used for verbal communication. It displays the search field, where the user can type in search words, the list of messages found in the search, the field containing the selected message and buttons for speaking, adding, modifying or deleting a message. Options such as search sensitivity can be set via the menu. A screen print of the interface is shown in Figure 1.
- To prepare a communication, the user or the carer types in new messages, which are then automatically indexed and integrated in the system.

There is strong emphasis in the design on reusability of the software, especially the lexicon, for other AAC systems. The internal structure is highly modular, and consists of the following components:

- Supervising module.
- Interface.
- Database for messages with index.
- Retrieval module.
- Main lexicon with syntactic category of words, frequency information and semantically related lexical items.
- Morphological analysis module, which makes use of a stop list, an affix lexicon and an exception list and relies on the main lexicon for lexical information.
- Interface to text to speech engine.

In the following we give a short overview of the morphological analysis module. We then discuss the role of the main lexicon for indexing and for the retrieval of messages.

2.2 Message indexing and morphological analysis

At any stage of a conversation, and also when preparing for a conversation, the user of WordKeys can add a message to the database or modify an existing message. The system also offers the facility to import a text file. When a message is added to the database, the following actions are performed:

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

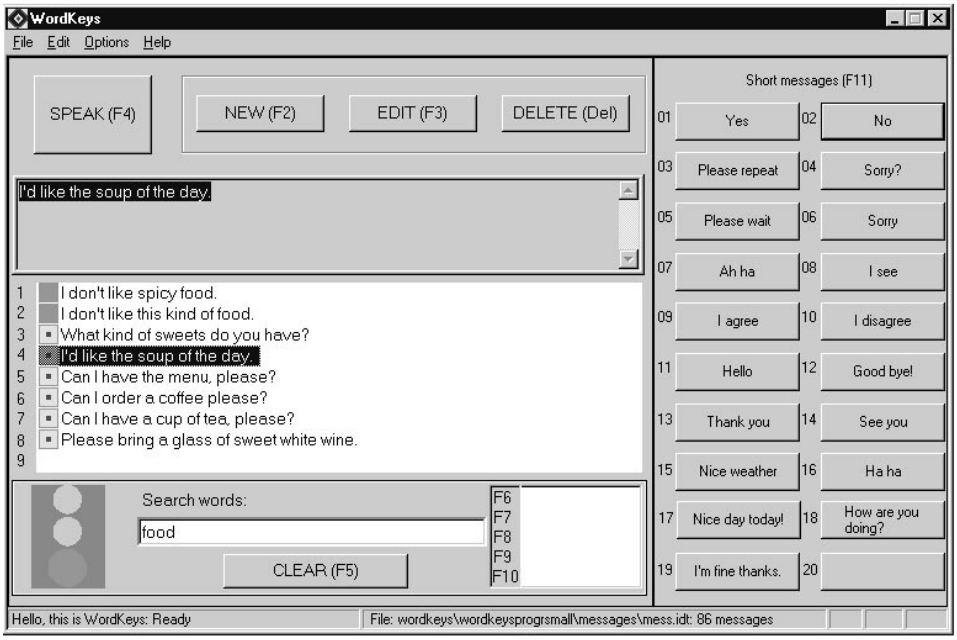


Fig. 1. The WordKeys interface.

- the message is parsed and indexed;
- the message with the index and optional user specified topic terms is added to the database.

The morphological module uses an affix list and is enhanced by an exception list and morpho-syntactic information from the WordNet database. The morphological analysis (Langer and Hickey 1997) of message words is carried out in two steps:

- lemmatisation;
- determination of the derivational root (only for semantically transparent derivation affixes).

The lemma of each word form in the message, apart from stop words, is always included in the index. A derivational analysis is carried out on the lemmatised word forms. Whereas all inflectional affixes are treated alike, derivational affixes are ranked according to the regularity of the semantic derivation; this influences the ranking of corresponding messages. During message access, the link between a word form and the lemma will be given a higher weight than links between morphologically complex words and their roots. The procedure of separating inflectional and derivational analysis is consistent with the findings reported in Hull (1996). He concludes that complex stemming algorithms can be slightly more effective than simple plural removal, and that the removal of derivational affixes is not always desirable. This is especially true for a retrieval system such as WordKeys, which uses semantic

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

relationships for retrieval and performs message ranking, since this approach can increase the impact of inaccuracies in the morphological analysis. Semantic relations between a lemma and some word form differ considerably from the semantic relations between derived words and their root. Correspondingly, we decided that the phases of stripping inflectional affixes – lemmatisation – and further morphological analysis have to be clearly separated.

To be able to determine semantically related words without loss of precision, information from the morphological analysis is also used to determine the morpho-syntactic categories of word forms and lemmas. The category can be clearly determined in the following cases:

- A word has one single entry in the main lexicon used for the analysis; this means the word is already a lemma.
- A word form has an inflectional or derivational affix, which only occurs with bases of one single morpho-syntactic category.

The removal of ambiguities concerning syntactic categories has an impact on the performance of the semantic expansion module. The less often words with inappropriate syntactic categories are included in the index, the higher the precision that will be achieved by the system. For many word forms in the messages, however, the category remains ambiguous. Currently, we are investigating the use of word frequency information, stochastic taggers and local grammars for determining syntactic information in these cases.

2.3 Message retrieval and message ranking

When the user has typed in one or several key words and decides to start the search, the following tasks are carried out:

- The content of the input field on the interface is parsed into word forms.
- Word forms are lemmatised.
- The word forms and lemmas are looked up in the message index. If they are found, the corresponding message numbers are added to the list of retrieved messages.
- The lemmas are looked up in the semantic lexicon to retrieve related words. The relations used are dependent on the semantic paths defined in the settings. The related words are re-applied in another query of the message database index.
- The messages that have been found are displayed on the screen in an order that corresponds to their scores.

The retrieved messages are ranked. Message that are considered less relevant for a query are put lower in the list or are excluded from the display. Ranking is performed according to semantic distance between key word and index word. Table 1 gives the figures for the message ranking criteria that apply in the case of one single key word. Where there is more than one key word, a combination of the semantic distances for the different key words is used. Any message being retrieved

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

Table 1. *Weight determination for message ranking*

Description	Ranking penalty	Comment	Example
Word in message is same word form as input word	0	Exact match, best rating	*key – key*
Word in message is lemmatized in index and matches input word	1	Less semantic distance than derivational analysis	*key – keys*
Word in message is reduced to root in index to match input word	2	Derivational analysis	*consider – consideration*
Synonym is looked up in lexicon	5	Synonymy considered closest semantic relation	*answer – reply*
Other semantically related word	≥ 10	Depends on semantic relation	*bird – seagull*

by more than one of the key words will be given an increased score. The semantic relations used, and their weightings, are defined in the preferences settings file.

The ranking procedure is best illustrated by an example. When typing in the search word *food* with the text database used for evaluation, the following messages were retrieved, in the given order:

(1) *What kinds of food do you like?*
(2) *I never eat low fat foods.*
(3) *Can I have onion soup, please?*

Given the ranking procedure described above, the reasons for this order are evident: *food* in message (1) is the same word form as the search word; in the second message, we have got another form of the search word; the third message is retrieved because of the synonym link between *food* and *soup*.

2.4 Main lexicon

One purpose of the main lexicon in WordKeys is to serve as a lexical database for the indexing module when performing morphological analysis. The main function of this lexicon, however, is to serve as a basis for the semantic query expansion. To choose the right lexicon, we had to bear in mind that WordKeys is a retrieval system for unrestricted text. It is not, as are many other full text retrieval systems, a domain specific system. This implies that it should be able to retrieve messages containing any word of the English language, apart from extremely domain specific vocabulary. Consequently, the ideal lexicon for the system should cover all English words. This is, of course, an unachievable aim as the complete lexicon of a language is not a stable structure. In practice, this means that the lexicon should be as large as possible and, to ensure efficient morphological analysis, should cover almost all morphologically simple English words and forms with derivational affixes.

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

We decided to use the semantic database WordNet for the following reasons:

- it is comprehensive, covering almost all English words;
- it contains most relevant links;
- the information contained in WordNet is stored in text files and can be easily converted to any other format.

To use the information in WordNet for our text retrieval algorithm, WordNet was converted to a more suitable format. We chose a format which was easily portable: a text file containing lemmas, together with their syntactic category and related words corresponding to the different senses. The original WordNet database is structured in so-called 'synsets', which are sets of words that share one sense, and it is subdivided into several sub-lexicons. The converted lexicon for WordKeys is organised in lemmas. The main reason for the conversion was that the lexical information used for WordKeys should be in a format that is relatively easy to achieve from any kind of lexical data. This makes it possible to use other lexicons (e.g. for other languages) and to include information from other sources, such as frequency counts (McRoy 1992). Mostly, such information is based on word forms or lemmas, not on senses.

There are several inconveniences with current comprehensive lexicons, especially concerning the semantic information that is contained in them:

- They include very rare words, including those that are homographs to more common words. When used for morphological analysis, this might imply reduction of a word form to an incorrect lemma.
- Semantic information includes rare and obsolete senses of a word.

Statistics on word frequencies, which are used in WordKeys, form one step towards solving these problems. The frequencies are retrieved from a large database of mainly written text, the British National Corpus (BNC). Although it can be argued that frequency counts from an independent corpus do not reflect frequencies as needed by an AAC system, we think that data from a balanced corpus, such as the BNC, can at least give a rough indication of the familiarity of most words in arbitrary collections of unrestricted text. The list used for WordKeys contains the most frequent 8000 words in this corpus. To achieve further improvement of the precision rate in retrieval, information on word sense frequencies would be useful but, unfortunately, this information is not easily available. Building a sense frequency counting lexicon from suitable corpora was beyond the scope of the current project phase.

Consequently, in each entry of the WordKeys lexicon, the following information is stored:

- main category of word (verb, noun, adjective or adverb), for morphological analysis and semantic links;
- frequency (0 if the word is not included in our frequency list);
- links to other words in the lexicon, specified according to the type of link.

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

An earlier version of WordKeys used synonym links only (Hickey 1995). In the current version, in order to retrieve a larger proportion of relevant messages, the coverage of semantic relations was widened to include synonyms and hyponyms/hypernyms. The information in the lexicon and the architecture of the retrieval module also allows us, by changing the preferences, to include meronyms and holonyms in the query expansion procedure. The benefits of these semantic relations for the system's performance have not yet been evaluated, and the following paragraphs only refer to query expansion with synonyms, hyponyms and hypernyms.

3 Evaluation of the semantic expansion module

An evaluation of the earlier prototype system was carried out to assess the effectiveness and efficiency of communication using message retrieval by key words. It was found that the kind of retrieval used in the WordKeys system is easy to understand by potential users and allows them to participate effectively in transactional and interactional encounters. Detailed results are reported in Hickey (1995).

The purpose of the evaluation reported in this section was to look at the benefits of semantic expansion in terms of retrieval success. To refine the system and to assess the suitability of the lexicon, we conducted formal assessments using a small, experimental database. Evaluation by a non-speaker took place once the system had been refined according to these formal trials.

To evaluate the semantic retrieval system, we used a database of short messages (average message length was 12 word forms), which contained approximately 100 messages. Sentences were selected from scenarios representing realistic conversations involving communication aid users. The performance of the system was tested using a set of possible key words for each message. We included key words that occurred in the message and morphologically related words, as well as key words that were related in meaning to the words in the message. Each search was restricted to one single key word. The key words chosen for each message were used to try to retrieve that message. The lexicon search for semantically related terms included hyponyms and hypernyms up to two steps up or down the hierarchy.

The evaluation results are given in Table 2, columns (1) and (2). These results only give a rough idea of the usefulness of semantic expansion for AAC. To obtain a more realistic impression of precision in a communication aid, we need to know how users access messages in practice. In future evaluation, we intend to use key words that have actually been used as search words to retrieve particular messages.

In Table 2 we present, in addition to the conventional precision results, a measurement that we label ranking imprecision. This is the average number of irrelevant messages presented higher in the message list than the target message. When the target message was not retrieved, all irrelevant messages in the list were counted. The ranking imprecision gives an indication of the cognitive load on the user when searching for the target message in the list. The precision for the search without semantic expansion is 1.0, and the ranking imprecision is 0.0. We assume that all

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

Table 2. *Results of tests of retrieval effectiveness on a database of 100 short messages for retrieval without semantic query expansion* (*column 1*), *with semantic query expansion* (*column 2*) *and with semantic query expansion and consideration of frequency information* (*column 3*)

	Without semantic expansion	With semantic expansion	With semantic expansion and frequency information
Recall	0·51	0·71	0·80
Precision	1·0	0·85	0·85
Ranking imprecision (number of messages)	0·00	0·17	0·20

messages containing a search word, or a morphological variation of it, are relevant to the user.

We think that the precision rate, and especially the ranking imprecision, obtained with semantic expansion are acceptable values for a communication aid. The number of retrieved irrelevant messages is not very high; the user will quickly be able to disregard these messages by visual scanning.

The performance of the system was not satisfying in terms of recall. The recall rate for message retrieval without semantic query expansion is very low. The messages could only be retrieved for half the key words we used as search words. The recall rate in column (2) shows the improvement contributed by the semantic module. The figures suggest that the average improvement in recall rate using the simple semantic search is approximately 40%; this still means that only 7 out of 10 messages are successfully retrieved. In order to improve the performance of the semantic expansion module, we identified the sources of the problems in the query expansion algorithm and in the lexicon structure.

We found that most of the retrieval problems with semantically related words can be explained by properties of the semantic hierarchy in the lexical database. One of the reasons for the lack of satisfying query expansion terms for many words can be illustrated by a search for the hyponyms of *animal*: a search for hyponyms of this word first results in words for very general kinds of animals and then produces further subclasses of those kinds. To take an extreme example, the listing in Figure 3 shows that the link *animal – dog* has six intermediate steps. This large semantic distance does not correspond to the intuition of English language speakers. The first hypernym for *dog* that comes to mind would be *animal* or *pet*, and *dog* would certainly be among the more frequently named hyponyms of *animal*. Clearly, the underlying hierarchy is based on zoological classification rather than on actual language use. It might be argued that the intermediate levels are not very relevant, as long as we reach the right words in the end. However, it is not feasible to include such a large number of steps in the search algorithm for query expansion. For the majority of input words, a link more than three levels up or down in the hierarchy

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

```
dog
  → canine, canid
  → carnivore
   → placental mammal, eutherian, eutherian mammal
    → mammal
     → vertebrate, craniate
      → chordate
       → animal, animate being, beast, brute, creature, fauna
```

(source: WordNet)

Fig. 2. Hypernyms for *dog* in WordNet.

```
frump, dog
 → unpleasant woman, disagreeable woman

dog -- ("you lucky dog")
 → chap, fellow, lad, gent, fella, blighter
```

(source: WordNet)

Fig. 3. Sense 2 and 3 for *dog* in WordNet.

would result in too many, and imprecise, additional query words and, subsequently, to an explosion of search space.

In psycholinguistic research, it has been shown that certain words in most hierarchical structures would rarely be used when referring to objects in the world; others, designating what has been called 'basic level categories', are used more often (for a discussion of the properties of basic level categories – see Lakoff (1987, pp. 31–50)). These lexical items would be much more useful for retrieval purposes for unrestricted text than the former, which are often technical terms such as the ones shown in Figure 3. Later in the paper, we propose a method of determining those intermediate steps in the WordNet hierarchy that are too technical, in order to eliminate them.

A second problem for the semantic expansion module was caused by inappropriate word senses. As WordNet is not tailored for any special application and is meant to be as comprehensive as possible, it contains all senses of a word, even those that are very infrequent or obsolete. Here again, the information given for *dog* in WordNet is a very good example. Two senses for this word are shown in Figure 3. These senses are confined to special contexts and are only relevant for message retrieval in very few cases. Unfortunately, the lexical database does not offer appropriate information on sense frequencies or stylistic connotations, to prevent the unlikely query expansion from *dog* to *fellow* or the unwanted one resulting from the first sense given. Such expansions should be rated at a very low probability.

There are other reasons for defective message retrieval, such as inconsistencies in the database (e.g. *house* is not subsumed under *building*, whereas *apartment house*, *boathouse* are) or missing links in the database (e.g. *pet* is linked to *pet animals* such as *cats* or *dogs* only via the common hypernym *animal*). These are encoding mistakes

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

and, as they are specific to WordNet, we will not discuss them further here. However, the problems mentioned earlier – irrelevant intermediate levels and inappropriate word senses – will apply to every lexicon that tries to include a detailed semantic hierarchy.

4 Including word frequency

We have shown that it is necessary to offer the capability to detect technical/domain specific intermediate steps in semantic hierarchies. To improve the message retrieval algorithm by determining frequent words in the hierarchy, we used a lemmatised frequency list for the 8000 most frequent words in English, which was derived from the British National Corpus. The combination of information from WordNet and frequency information is not a new approach. Among other work, the approach has been taken in a lexical database for AAC systems presented in Zickus *et al.* (1995). However, the use of word frequency to create a reduced hyponymic hierarchy for WordNet that is then used for text retrieval has, to our knowledge, not been reported elsewhere.

The frequency list was used to compile a lexicon for WordKeys that allows us to avoid unproductive intermediate steps in the semantic hierarchy. Of course, all direct hyponyms and hypernyms were included – but for every lemma we also included the closest hypernyms that appeared both in more distant super/subordinated nodes and in the frequency list. We excluded the most frequent words (frequency more than 40,000); these were used to define a stoplist. This strategy improves the encoding for hypernyms of lemmas. Most words now have in their list the most intuitive superordinated word. For example, in the case of *dog*, the closest hypernym that appeared in the frequency list was *animal* and consequently this word was included in the hypernym list of *dog*, in addition to *canine* and *canid*. It can be argued that, going up in the taxonomy, the search for the next superordinated word that is frequent often leads to words designating basic level categories, or at least to a superordinated category that is not exclusively a technical category known by experts of a specific domain.

Including frequent hypernyms for all lemmas in the lexicon was feasible without an explosion in search space. However, constructing the lexicon in such a way that all words also have in their entries all frequent hyponyms, and using this information at run time, would have led to a much larger lexicon and an explosion in search space, and would have made the system too slow. So, instead of performing the query expansion at run time, we decided to index messages with the frequent and the direct hypernyms of the content words in the message. In this way, we are exploiting information about hypernyms in the lexicon in order to access messages with hyponyms of search words.

Despite the fact that we still had the problem of the lack of word sense frequencies, the evaluation results for the new lexicon were considerably better than those for the first lexicon (see Table 2, column (3), compared with column (2)). The new lexicon, as expected, resulted in a higher recall rate than the previous one: 80% of the target messages were now correctly retrieved, compared with about 70%

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

with the first lexicon. The precision did not change. The figures also show that the average number of irrelevant messages appearing higher in the list than the target message rose slightly, which increases the cognitive load on the user when searching for the target message. The increase in recall rate was so significant that this easily compensates for the slight increase in ranking imprecision.

5 User evaluation

The second version of WordKeys, with the improved lexicon, was evaluated in a study with a non-speaking user. We decided to have a long-term trial with one user, rather than several short studies with more users, for the following reasons:

- It takes several weeks for a user to become familiar with any AAC system.
- The resources of the project: each user needs the hardware and speech synthesis software; each user needs a personalised message database at the start of the evaluation, and support during the evaluation.
- A long-term user study should clearly show the benefits and problems of the system.

Our subject is female, between 30 and 40 years old, and had had, prior to our evaluation, several years of experience with communication aids based on pre-stored messages. She has Cerebral Palsy, uses a wheelchair and is unable to use spoken language. She can use her hands for input via the keyboard, although at a relatively slow rate. She also used another message based communication system in the evaluation period. The messages that were used for WordKeys were taken from message files that had been typed in for the other system. The size of the message file was approximately 1000 messages. The subject used WordKeys regularly for five consecutive months, in conversations with friends and relatives. Data on her use of the system was stored in log files. The researcher met with the subject every two weeks, where the subject used WordKeys to report her experiences with the system and to converse with the researcher about a variety of topics.

The results of this evaluation are summarised as follows:

- The user found the system useful and suitable for her needs. She made the suggestion to integrate the retrieval module into the system she normally used.
- The user had no problems in understanding the retrieval mechanism and the interface.
- The retrieval module was used efficiently and the user was able to find relevant messages quickly.
- The user required only one key word to retrieve each desired message.
- The semantic expansion module was seldom used. The user retrieved most messages by typing key words that occurred in the message. The main reason she gave for this was that the relevance of messages found by the semantic expansion was unpredictable – sometimes they were relevant and sometimes completely irrelevant to the query.

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

- The user frequently needed to produce unique messages, a task for which WordKeys was not designed. The input field for new message was used for that purpose.

The user evaluation shows that the overall idea of message access based on full text retrieval by key words offers good perspectives for AAC users. Further work is needed in order to improve the semantic expansion module. The retrieval module would be most useful if it was combined with other types of communication aids that offer the possibility for rapid composition of unique text.

6 Conclusions

In this paper, we have described the overall design of a communication aid for non-speakers based on ideas from text retrieval with semantic query expansion. Several NLP modules, all based on data from the electronic dictionary WordNet, were used to build WordKeys. Indexing is carried out using a morphological analysis tool based on lexical knowledge together with an exception list. The query expansion module uses different kinds of semantic links, together with frequency data, to find related terms.

We have shown that the use of a large semantic database for the purpose of returning pre-stored messages in an AAC system enhances recall rates. We have described difficulties associated with semantic query expansion. In particular, under-generation of relevant related words, because of too many intermediate links between closely related words, proved to be a serious problem. We were able to demonstrate that the inclusion of information on word frequency in the lexicon can partly overcome this problem. This approach considerably increased recall without lowering precision.

According to our long term user evaluation, the performance of the retrieval module must be improved further in terms of recall and precision. An important area of future work is to reduce over-generation due to obsolete or infrequent words and word senses. Over-generation keeps the precision rate down, and this is only remedied in the current system by message ranking. A combination of strategies could be tried out to solve this and other problems, particularly the inclusion of usage information on words, or information about word sense frequencies achieved through the analysis of corpora. Improvement in long term use may also be achieved by adapting the lexicon according to user selections. This would mean that successful semantic links will get a higher weight than other ones. New links could be added based on knowledge of the user's message selections. Interactions where initial search words do not retrieve relevant messages could be recorded.

A further aim for the future is the integration of the WordKeys retrieval engine with other AAC systems. Although we have recently added the possibility for the direct selection of 20 short messages (such as *yes*, *good morning*, *How are you* etc.), that can be modified by the user, WordKeys lacks modules for rapid entry of unique text through word prediction, and the support for alternative input devices and for alternative selection methods.

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press

Acknowledgements

The WordKeys project has been supported by the European Union through a HCM/TMR-fellowship (January 1996–September 1997). We would also like to thank the reviewers for their helpful comments on the first version of this paper.

References

Cruse, D. A. (1986) *Lexical Semantics.* Cambridge University Press.

Foulds, R. (1980) Communication rates for non-speech expression as a function of manual tasks and linguistic constraints. *Proceedings of the International Conference on Rehabilitation Engineering*, pp. 83–87. Toronto, RESNA.

Guglielmo, E. J. and Rowe, N. C. (1996) Natural-language retrieval of images based on descriptive captions. *ACM Trans. Infor. Syst.* **14**(3): 237–267.

Hearst, M. (1996) Improving full-text precision on short queries using simple constraints. *Proceedings of the Fifth Annual Symposium on Document Analysis and Information Retrieval (SDAIR)*, Las Vegas, NV.

Hickey, M. (1995) Communication Enhancement in an Aid for Severely Disabled People. *PhD Thesis*, Coventry University, UK.

Hickey, M. and Page, C. J. (1993) Polyvox: Flexible message selection in a communication prosthesis for non-speakers. *Proceedings of the 2nd European Conference on the Advancement of Rehabilitation Technology*, pp. 89–91. Stockholm, Sweden.

Hull, D. (1996) Stemming algorithms: A case study for detailed evaluation. *J. Am. Soc. Infor. Sci.* **47**(1): 70–84.

Kraat, A. (1985) *Communication Interaction between Aided and Natural Speaker. A State of the Art Report*. IPCAS.

Lakoff, G. (1987) *Women, Fire and Dangerous Things.* Chicago University Press.

Langer, S. and Hickey, M. (1997) Automatic message indexing and full text retrieval for a communication aid. *Proceedings of the Workshop on Natural Language Processing for Communication Aids at EACL/ACL Madrid*, pp. 9–16.

McRoy, S. W. (1992) Using multiple knowledge sources for word sense discrimination. *Computational Linguistics* **18**(1): 1–30.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. J. (1990) Wordnet: An On-Line Lexical Database. *Int. J. Lexicography* **3**(4).

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. J. (1993) *An Introduction to WordNet, an On-Line Lexical Database.* Princeton University. (Electronic documentation for the WordNet program, available at: http://www.cogsci.princeton.edu/ wn/)

Newell, A. F. and Alm, N. (1994) Developing AAC technologies: a personal story and philosophy. *Euro. J. Disorders of Communication* **29**: 399–411.

Richardson, R. and Smeaton, A. (1995) Using WordNet in a Knowledge-Based Approach to Information Retrieval. Working paper CA-0395, School of Computer Applications, Trinity College Dublin.

Smeaton, A. and Quigley, I. (1996) Experiments on using semantic Distances between words in image caption retrieval. *Proceedings of the 19th International Conference on Research and Development in Information Retrieval*, pp. 174–180. Zürich, Switzerland.

Zickus, W. M., McCoy, K. F., Demasco, P. W. and Pennington, C. A. (1995) A lexical database for intelligent AAC systems. *Proceedings of the RESNA '95 Annual Conference*, pp. 124–126.

https://doi.org/10.1017/S1351324998001855 Published online by Cambridge University Press