Natural Language Engineering **23** (2): 217–244.

doi:10.1017/S135132491500039X

Translating text into pictographs

VINCENT VANDEGHINSTE, INEKE SCHUURMAN LEEN SEVENS and FRANK VAN EYNDE

Centre for Computational Linguistics, University of Leuven, Blijde Inkomststraat 21 - bus 3315 B-3000, Leuven, Belgium

e-mails: vincent@ccl.kuleuven.be, ineke@ccl.kuleuven.be, frank@ccl.kuleuven.be

(*Received 2 September 2014; revised 20 August 2015; accepted 24 August 2015; first published online 11 November 2015*)

Abstract

We describe and evaluate a text-to-pictograph translation system that is used in an online platform for Augmentative and Alternative Communication, which is intended for people who are not able to read and write, but who still want to communicate with the outside world. The system is set up to translate from Dutch into Sclera and Beta, two publicly available pictograph sets consisting of several thousands of pictographs each. We have linked large amounts of these pictographs to synsets or combinations of synsets of Cornetto, a lexical-semantic database for Dutch similar to WordNet. In the translation system, the Dutch input text undergoes shallow linguistic analysis and the synsets of the content words are looked up. The system looks for the nearest pictographs in the lexical-semantic database and displays the message into pictographs. We evaluated the system and results showed a large improvement over the baseline system which consisted of straightforward string-matching between the input text and the filenames of the pictographs.

Our system provides a clear improvement in the communication possibilities of illiterate people. Nevertheless there is room for further improvement.

1 Introduction

The importance of the digital society in various aspects of our lives is undeniable. Allowing people with cognitive disabilities to independently use the Internet can increase their quality of life, by reducing social isolation (Newell and Gregor 2000; Davies, Stock and Wehmeyer 2001; Dawe 2006).

Augmentative and Alternative Communication (AAC) assists people with severe communication disabilities to be more socially active in interpersonal interaction, learning, education, community activities, employment, volunteering, and care management. Picture-based communication systems are a form of AAC technology based on the use of graphics, such as drawings, pictographs, and symbols.

We can distinguish three types of pictographic communication (Takasaki and Mori 2007): more or less universal pictographs, pictographs for people with disabilities, and emoticons. Good examples of universal pictographs are road signs, direction boards at airports, and the symbols of each sport played in the Olympic Games. The second category of pictographic languages includes systems for AAC. The third category includes pictographs that decorate text messages such as *smilies, kaomojis*[1] and *emojis*.[2]

There are estimates that between two and five million people in the European Union could benefit from symbols or symbol-related text as a means of written communication (Keskinen *et al.* 2012). Consequently, there is an acute need for such picture-based communication interfaces that enable social contact for people with cognitive disabilities, and these interfaces should be easy to use, configurable and flexible, to allow adapting them to different situations for users with different (dis)abilities (Keskinen *et al.* 2012). Pictographs enable communication with pre-literate or illiterate people (Medhi, Sagar and Toyama 2006).

The system described in this paper is used in the WAI-NOT communication platform.[3] WAI-NOT is a Belgian non-profit organisation that aims at enabling internet access for people with mental disabilities. A number of specific computer applications were built, such as a special website adjusted to different intellectual levels. Pictographs and auditory support are used on this website wherever possible. Furthermore, it is possible to chat and to send emails with the help of pictographs through an adjusted e-mail client. This platform is widely spread over special needs education schools in Flanders. It uses two sets of pictographs (Beta[4] and Sclera[5]) for communication between WAI-NOT users leading to translation problems if Beta users communicate with Sclera users and vice versa, or if literate Dutch-writing users want to communicate with Sclera or Beta users or vice versa.

The users of the WAI-NOT e-mail client have two input modes which they can combine while composing e-mail: they can select pictographs from their pictograph set (either Beta or Sclera) through a two-level category system,[6] or they can type text. Whenever possible, the typed text is augmented with pictographs. In the *baseline* system, this is done through straightforward string-matching. Each input word that coincides with the filename of a pictograph (without the `.png` extension) is augmented with that pictograph. Notice that due to homonymy this can be a wrong pictograph. In a later stage, shallow linguistic analysis was added (Vandeghinste 2012), such as lemmatisation and part-of-speech tagging, in order to improve the coverage and accuracy of the system, making the message more understandable for illiterate users by converting more words into pictographs more accurately. When a word cannot be converted, it is displayed as text without an associated pictograph.[7]

[1] A *kaomoji* is the Japanese version of an emoticon. For instance ^-^ for a happy face.

[2] An *emoji* is a pictograph originally used in Japanese electronic messages or webpages. Japanese operators provide a wider range of such standardised icons into handsets.

[3] `http://www.wai-not.org`

[4] `http://www.betasymbols.com`

[5] `http://www.sclera.be`

[6] The pictograph input method requires the users to first select the main category (such as *family, food*), before entering the second level in which the actual pictograph (such as *sister, carrots*) can be chosen. It is far from optimal and requires further research that should result in a more suitable pictograph input method.

[7] Depending on the pictograph set used, some types of words are never converted (such as articles in Sclera).

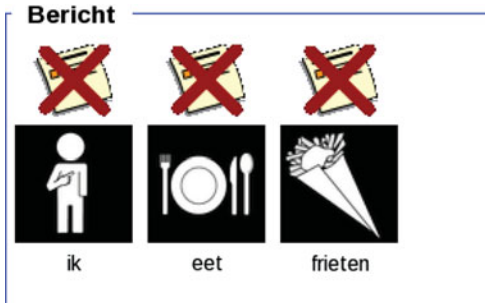

Fig. 1. (Colour online) An example e-mail message in Sclera pictographs as displayed in the composition interface. The message translates to English as *I eat fries*.

An example of a message augmented with pictographs, as in the WAI-NOT e-mail client, is displayed in Figure 1.

We have collected a corpus of 69,636 e-mail messages sent on the WAI-NOT platform, from the start of the platform on the 28th of April 2009 until the 23rd of May 2013. These e-mails have an average length of 7.7 words, and can be divided into several categories: some of the e-mails are clearly written by literate people, such as teachers and the WAI-NOT content providers. Those are the most standard e-mail messages, and perhaps the hardest to translate into pictographs, as they have a much broader vocabulary than the other e-mail categories. A second category consists of short messages sent by the intended users of the WAI-NOT platform, i.e. people with mental disabilities. These messages make up the largest part of the corpus, and in general contain only one sentence, no punctuation or capitalisation and several spelling errors. These messages are the main target of the whole translation exercise. Then, there is a substantial part of the messages which can be categorised as noise, as it is hard to see what the intended meaning of the message was. Some of these consist of apparently random keystrokes, others are most probably the result of repeatedly clicking on the same pictograph in the pictographic input interface.

From the two first categories, we selected two subsets: a *development set* of 186 messages, which were used as test cases during development of the system and an *evaluation set* of fifty messages. The average length of the emails in this evaluation set was about twenty words.

When presenting and demonstrating the tool[8] to remedial educationalists, i.e. people who are working on a daily basis with children with a cognitive impairment, they are, in general, very enthusiastic about the system. The tool allows *reading* messages which are originally text-only. As it concerns a form of *augmented* communication, adding pictographs to text can provide help in reading and understanding the text.

Additionally, such a system allows to convey information across language barriers (Mihalcea and Leong 2009), it helps in foreign language learning (Carney and Levin

[8] An online version of the tool is available at `http://picto.ccl.kuleuven.be`.

2002) and in language understanding for people with language disorders (Behrmann and Byng 1992; Alm *et al*. 2002).

The remainder of this paper describes our approach towards translating text into pictographs. Additional to a shallow source language analysis (involving processes such as part-of-speech tagging and lemmatisation), we use Cornetto (Vossen *et al*. 2008; van der Vliet *et al*. 2010), a lexical-semantic database for Dutch which is linked to Princeton WordNet.

In Section 2, we describe related work. Section 3, gives the detailed system description, whereas Section 4, presents our evaluation of the system, which shows a large improvement over the baseline system. Section 5, describes conclusions and the future work.

2 Related work

Pictographic communication has grown from local initiatives of which some have scaled up to larger communities. Across Europe, many pictograph systems are in place, such as Blissymbolics,[9] PCS,[10] Pictogram,[11] Beta, and Sclera. All these initiatives allow or provide aid with pictographic communication, but in the context of automatic conversion of text into pictographs they provide no or only a very limited amount of linguistic knowledge in order to appropriately disambiguate lexical ambiguities, which can lead to wrong conversions into pictographs (Vandeghinste 2012) or to the conversion into multiple pictographs per word, one for each sense of the word. An application of the latter can be seen on `www.widgit.com`.

Only few works related to the task of translating texts for pictograph-supported communication can be found in the literature. Mihalcea and Leong (2009) describe a system for the automatic construction of pictorial representations of the nouns and some verbs for simple sentences and show that the understanding, which can be achieved using visual descriptions, is similar to those of target-language texts obtained by means of machine translation. They automatically collected a picture set, which was validated through crowd sourcing in the PicNet project (Borman, Mihalcea and Tarau 2005). They used WordNet (Miller 1995) as a lexical resource, but it seems that they did not use the WordNet relations between concepts. The main other differences between their system and ours is the fact that our system tries to translate entire messages while theirs does not, and that our system is focused on AAC.

Goldberg *et al*. (2008) show how to improve understanding of a sequence of pictographs by conveniently structuring its representation after identifying the different roles which the phrases in the original sentence play with respect to the verb (structured semantic role labelling is used for this).

Joshi, Wang and Li (2006) describe an unsupervised approach for automatically adding pictures to a story. They extract semantic keywords from a story and search an annotated image database. They do not try to translate the entire story.

[9] `http://www.blissymbolics.org`
[10] `http://www.mayer-johnson.com/category/symbols-and-photos`
[11] `http://www.pictogram.se/`

None of the previous works consider users with disabilities when designing the system.

Vandeghinste and Schuurman (2014) describe the linking of Sclera pictographs with synonym sets in a lexical-semantic database. Similar resources are PicNet (Borman *et al*. 2005) and ImageNet (Deng *et al*. 2009), both large-scale repositories of images linked to WordNet (Miller 1995), aiming to populate the majority of the WordNet synsets. These often contain photographs which might be less suitable for communication aids for the cognitively challenged, as they may lack clarity and contrast. The Sclera and Beta pictograph sets are specifically designed to facilitate communication with this user group.

Another area of work that is somewhat related is text-to-scene conversion, such as WordsEye (Coyne and Sproat 2001), a natural language interface for a 3D editor. It applies linguistic analysis with dependency parsing, and the parse is converted into a semantic representation, which is, in its turn converted into depictors representing 3D objects.

3 System description

In this section, we describe how we convert a textual Dutch message into a sequence of Sclera or Beta pictographs. The architecture of the system is displayed in Figure 2.

The first step in *translating* the source text into a sequence of pictographs consists of *shallow linguistic analysis*. This is described in Section 3.1. The conversion of a word into a pictograph can go along two different routes. The *direct route*, described in Section 3.2, uses specific rules for appropriately dealing with pronouns, and it uses a dictionary for parts-of-speech that are not present in Cornetto, which is limited to *verbs*, *nouns*, *adjectives*, and some *adverbs*, as well as for other words that cannot be appropriately linked to Cornetto synsets. The *semantic route* is only applied in the case of *content words*. It consists of *Semantic Analysis*, connecting the input text to Cornetto synsets, as described in Section 3.3, and of *Semantics to Pictos*, retrieving the pictographs linked to the synsets of the message, as described in Section 3.4. When words cannot be converted into pictographs by either route, we copy the textual input word to the output, only in cases of content words or words that are necessary to understand the message, such as, in some cases, prepositions. The final step consists of *choosing the optimal path*, choosing which pictographs are displayed. This is described in Section 3.5.

To make everything clear, we describe the translation of the sentence *Hij is genezen ‘He has recovered’* into Sclera as a running example. The resulting Sclera message is shown in Figure 3.

3.1 Shallow linguistic analysis

The incoming textual message undergoes shallow linguistic processing, which is a process consisting of several sub-processes. The result of this shallow analysis on the example sentence is shown in Figure 4.

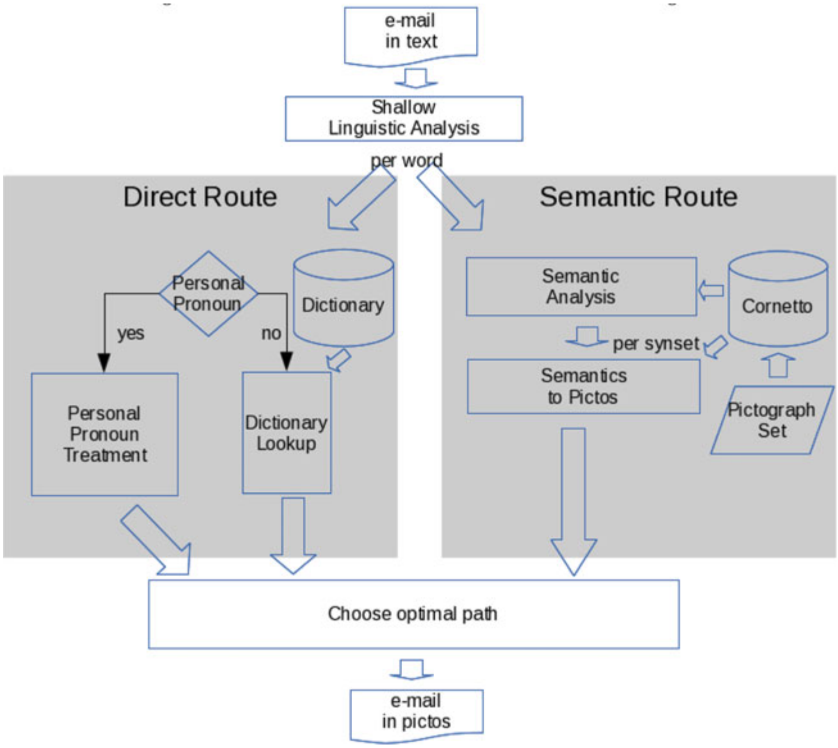

Fig. 2. (Colour online) Architecture of the Text2Picto translation engine.

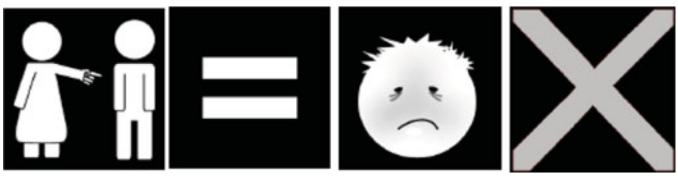

Fig. 3. Sclera translation of *Hij is genezen 'He has recovered'*.

The first sub-process that we apply is *tokenisation*, splitting of all the punctuation signs from the words, apart from the hyphen/dash and the apostrophe.

The next sub-process concerns *spelling correction*. As the users of the WAI-NOT platform have different levels of illiterateness we decided to perform at least some basic automatic spelling correction. An example message from the e-mail corpus reads *dag esra ik wes je en ge lukkig niew jaar adem* which supposedly should be corrected to (ignoring capitalisation rules) *dag esra, ik wens je een gelukkig nieuwjaar, adam 'hello esra, I wish you a happy new year, adam'*. As we did not find any tools available for automatic Dutch spelling correction, we decided to implement our own approach. We used the freely available and open lexicon from `www.opentaal.org`, which is the Dutch lexicon used in freely available spell checkers such as Hunspell,

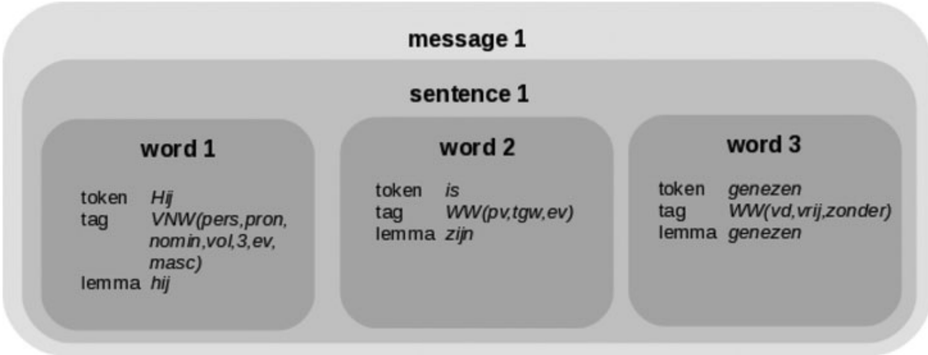

Fig. 4. The example sentence *Hij is genezen* after shallow linguistic analysis.

the open source spell checker for OpenOffice and MacOS and which received the hallmark for spelling of the Nederlandse Taalunie (Dutch Language Union),[12] the official international organisation responsible for the Dutch language policy. Besides the words in the lexicon, e-mail messages often contain first names. In order to avoid automatically and wrongly 'correcting' these names into proper Dutch words we have added all male and female first names that occurred more than fifty times in Belgium, last updated in 2009. This information was gathered from the Belgian National Institute of Statistics.[13] For every word that is not in this lexicon nor in the list of first names, we check all variants with one deletion, one insertion or one substitution. For all the variants present in the OpenTaal lexicon, the selection of the correct alternative is based upon a large frequency list (Vandeghinste 2002), containing roughly eighty million words of Belgian Dutch newspaper text. This is in fact a unigram model. In future versions we might consider higher order versions, or more complex approaches towards text normalisation, if deemed necessary.

Then, we apply *part-of-speech tagging*. We use HunPos (Halácsy, Kornai and Oravecz 2007), a trigram-based open source tagger similar to TnT (Brants 2000),[14] obtainable without requiring explicit permission, using the D-Coi tagset (Van Eynde 2005), which has become the de facto standard part-of-speech tagset since the release of the Corpus Gesproken Nederlands (CGN) (Spoken Dutch Corpus) (Oostdijk *et al.* 2002) and the Lassy-small corpus of written Dutch (van Noord *et al.* 2013). These corpora have nearly identical tagsets (Van Eynde 2005; Van den Bosch, Schuurman and Vandeghinste 2006). The tagger was trained on these manually corrected corpora. Both resources each contain about one million words with

[12] cf. `http://taalunieversum.org/inhoud/spelling-meer-hulpmiddelen/keurmerk`

[13] We used `http://statbel.fgov.be/nl/binaries/firstnamesall2009_tcm325-165200.xls` which is no longer available, but the updated version of 2013 can be found at `http://statbel.fgov.be/nl/binaries/firstnamesallages2013_nl_tcm325-237356.xls`.

[14] TnT requires a licence agreement to be faxed and a response from the author, and is only available for non-commercial non-profit research purposes, whereas HunPos is free and open source, directly available for download, even for commercial use.

manually corrected part-of-speech tags. We chose HunPos over the Frog-tagger (van den Bosch *et al.* 2007) as the latter proved, at the time of choosing a tagger, unstable in server mode, and is not easily trainable.

As the system is intended to translate e-mail messages for mentally challenged people, messages tend to be short, and mostly consist of merely one sentence. Nevertheless, some of the messages contain more than one sentence, so we apply *sentence detection*, as the translation engine works sentence based. This is done by rules based on punctuation signs.

A more language specific issue is the fact that Dutch contains *separable verbs*. These are verbs that have a lexical core and a separable particle. In some syntactic situations the core and the particle are written as one word, while in other situations they are written separately. Particles can have different part-of-speech tags, according to the tagset we use (Van Eynde 2005). The most frequent part-of-speech tags for particles are the final prepositions `VZ(fin)`,[15] such as in verbs like *afwerken* (*ik werk dit af → I/work/this/off 'I finish this'*). Other particles can be singular common nouns in standard case `N(soort,ev,stan)`, such as in verbs like *paardrijden* (*ik rij graag paard → I/ride/gladly/horse 'I like horse riding'*). Yet another set of particles can be tagged as uninflected adjectives `ADJ(vrij)`, such as in *vrijspreken* (*de rechter sprak hem vrij → the/judge/spoke/him/free 'the judge acquitted him'*). A final category of particles are the adverbs `BW`, such as in *bijeenbrengen* (*hij brengt geld bijeen → he/brings/money/together 'He collects money'*). Each of the words of a sentence that is tagged as a verb, be it in its finite, infinitive or past participle form, is combined with each of the words tagged with one of the potential tags for particles. The most likely combination according to the eighty million word corpus is selected, indicating what the most likely merger of a verb and its particle is. Additionally, we check whether the compound verb is more likely than the parts kept separately. This is done according to the methodology described in Vandeghinste (2002), which is based on the number of different particles a verb can have, the frequency of the compound and the frequency of the parts separately.

Then, we apply *lemmatisation*. We first look up whether the token and part-of-speech tag combination occurs in the already mentioned manually corrected corpora CGN and Lassy-small, and if this is not the case, apply lemmatisation rules implemented as regular expression substitutions conditioned on the part-of-speech label.

In the future versions, we might consider *dependency parsing*, as available for Dutch in the Alpino-parser (van Noord 2006) or the Frog-parser (van den Bosch *et al.* 2007), provided that these parsers are robust enough to deal with messages that contain many errors and dysfluencies.

[15] The tags consist of a prefix in uppercase which indicates the main word class, in this case `VZ`, and a suffix in round brackets indicating the value of the features of this word class, in this case `fin` which indicates that it is a *final* preposition that is used. *Final* prepositions are opposed to *initial* prepositions, where the latter appear before the noun phrase (hence *initial*), and the former appear after the noun phrase (hence *final*). Separable verb particles coincide with the final prepositions and receive the same part-of-speech tag. For full details see Van Eynde (2005).

3.2 The direct route

Not all words of a message can be analysed with Cornetto, and not all words should necessarily be converted into pictographs.

As *personal pronouns* are very frequent in the e-mail messages in our e-mail corpus and not included in the Cornetto lexical-semantic database, we have provided an explicit treatment for them, making sure that they are covered: the part-of-speech tag of the personal pronouns[16] contains a detailed list of features, of which the important ones in this case are the three last ones, indicating the values for person, number,[17] and gender.

Additionally, we have provided a translation mechanism that uses a dictionary. This mechanism allows to bypass the semantic analysis via Cornetto and provides a direct link between token/lemma/tag and the names of the pictographs, allowing for underspecification of any of the three input fields: the person creating the dictionary can choose whether to specify only the lemma, or to make it more specific, by also adding the part-of-speech tag and/or the token to which the word has to comply in order to allow translation into the pictograph. A detailed description of the dictionary is found in Vandeghinste and Schuurman (2014), and the effect of the dictionary is evaluated in Section 4.2. The dictionary entries are based on the most frequent words in the e-mail corpus which could not be appropriately converted through the semantic route. It concerns those words that occur at least fifty times in our email corpus (excluding our evaluation set).

3.3 Semantic analysis

As a first step in the semantic analysis of the source message, we detect words indicating a negative polarity, such as *niet* (*not*) and *geen* (*no*). When such a word is found, we look for its head. In the case of *niet*, we look for a verb within a window size of three, i.e. we look for a verb in the three preceding and the three following words. When a verb is found, we add the value `negative` to the verb's *polarity* feature.

As a second step in semantic analysis, we look up all the possible Cornetto-synsets connected to the lemma of each word. On its website Cornetto is described as

> ' ... a lexical semantic database for Dutch, covering 92K entries, including the most generic and central part of the language. The database combines the structure and content of WordNet and FrameNet-like data. It contains both vertical and horizontal semantic relations and combinatorial lexical constraints such as multiword expressions, idioms and collocations on the one hand, and lexical functions and frames on the other. The concepts are aligned with the English WordNet so that ontologies and domain labels were imported.' (`http://wordpress.let.vupr.nl/cornetto/`)

In our current system, we have limited ourselves to using the WordNet-like features of Cornetto, not (yet) using the FrameNet-like data. When developing and debugging the system, based on the development set, we made some adaptations in Cornetto,

[16] See the *tag* feature in the example sentence, as shown in Figures 4, 5, and 10.
[17] For the example sentence *ev* stands for *enkelvoud 'singular'*.

as it is rather biased towards Dutch as spoken in the Netherlands, while our system is mainly intended for the Dutch-speaking part of Belgium. Therefore, we added Belgian synonyms to the synsets from which they were lacking. Furthermore, we removed some links between lemmas and synsets because of the fact that Cornetto contains many synonyms for sexual concepts, such as e.g. genitals. The word *mossel 'mussel'*, for instance, is disabled as a synonym for the concept of female genitalia. An evaluator judged the appropriateness of each lemma belonging to sexual concepts: is the first, most intuitive sense of that lemma a sexual sense or not? If not, we disabled it. We also disabled meanings that were not prominent in our development set, for example the word *suiker 'sugar'* as referring to *suikerziekte 'diabetes'*, while it is much more frequent in the sugar-reading. Doing so, we improve the coverage of the language of the intended, cognitively challenged users. We discovered several such misconversions through extensive testing the system, not only with the messages from the development set, but also by test users[18] that were asked to report errors.

Cornetto is organised as such that each synset is linked to a number of lexical units (lexunits). Such a lexical unit consists of an identifier and a lemma. Linking words with synsets goes via the lexical units. Most synsets have a part-of-speech category encoded with them, which allows distinguishing between different synsets of the same lemma. We filter the synsets, keeping those where the part-of-speech of the synset agrees with the part-of-speech main category of the word, as labelled by the part-of-speech tagger. It should be noticed that the part-of-speech labels from Cornetto do not correspond in a one-to-one fashion with the labels from the part-of-speech tagger. Not only the labels are different, but so is the granularity. We manually encoded which Cornetto-labels are compatible with which part-of-speech labels according to Van Eynde (2005).

The result of the semantic analysis of the example sentence is shown in Figure 5. Note that by performing semantic analysis, we do not (yet) apply proper word-sense-disambiguation, but use the most common sense, based on DutchSemCor (Vossen *et al.* 2012).[19]

3.4 Retrieving the pictographs related to the semantic concepts

As stated in Mihalcea and Leong (2009), the use of pictographs has limitations. Complex concepts or combinations of concepts cannot always be depicted in a straightforward manner. There are certain properties in natural language which give it more power than pictograph languages. There are also many concepts that are difficult to depict, as their level of abstraction is too high. It is clear that pictograph symbols can be used to represent these concepts, but these concepts need to be learned by the users.

[18] We did not use the term *Beta users* as is common in software testing to avoid confusion with the users of the Beta pictograph set.

[19] In order to improve precision, in future versions we might implement this on the condition that the quality is good enough, because most of the messages are rather short, providing only little information on which to base the disambiguation.

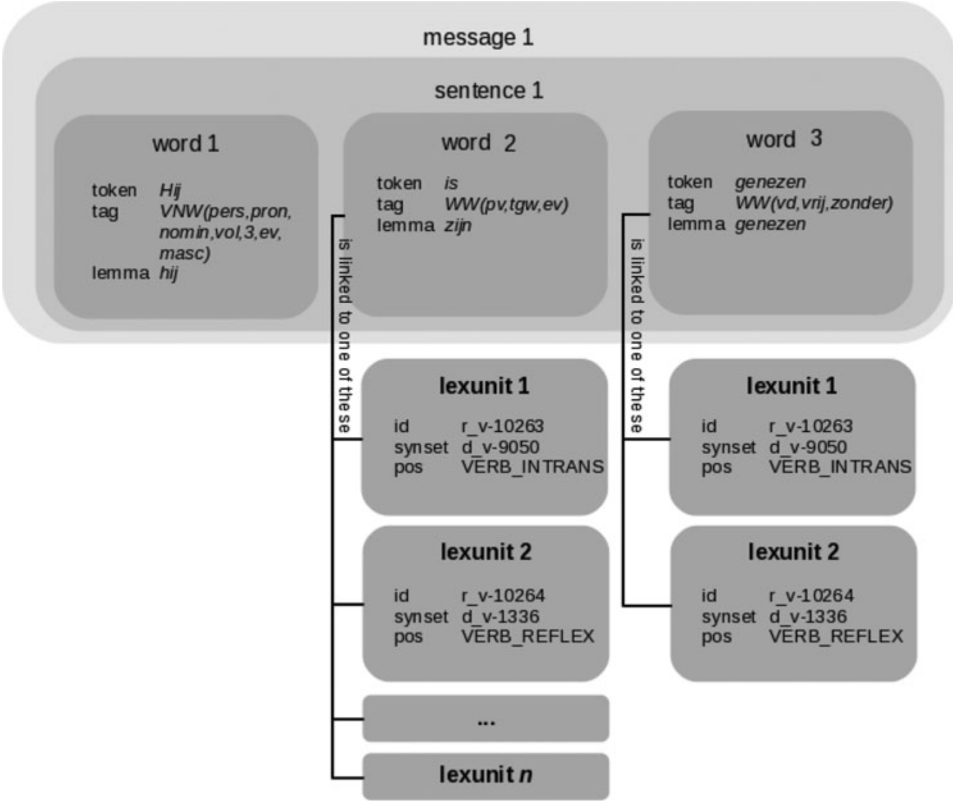

Fig. 5. The example sentence *Hij is genezen* after semantic analysis.

Some characteristics of natural language may not be present in certain pictographic languages, such as articles, inflection, or tenses. Only few auxiliaries are used. In general, no distinction is made between singular or plural. In some cases this is due to the fact that the concepts involved are hard to put into pictographs (like determiners, inflection of a verb), or because the pictographs mainly express the more abstract concept, as expressed by the lemma. In some exceptions the pictograph represents a more specific concept, such as a token in its plural form.

The Text2Picto translator works with two different *target languages*: Sclera, described in Section 3.4.1, and Beta, described in Section 3.4.2. Section 3.4.3 describes how the pictographs have been linked to the synsets and Section 3.4.4 describes how the pictographs are used in the translation engine.

3.4.1 The Sclera pictograph set

Sclera[20] is a large set of mainly black-and-white pictographs. Originally these were used as directives (*feed the dog, brush your teeth*), just like the pictographs we are confronted with in everyday life.

[20] Freely available under Creative Commons License 2.0.

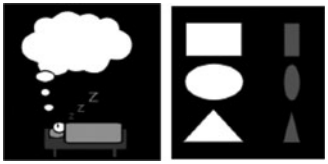

Fig. 6. Simplex Sclera pictographs for *dream* and *large*.

Fig. 7. Verb-object (Sclera) pictograph for *eat a sandwich* and *feed the dog*.

There are currently over 13,000 Sclera pictographs and new pictographs are created every month upon user request. These pictographs are freely available as `.png` files with a filename indicating their meaning in Dutch, English, French, and Spanish. As shown in Figure 6, they can represent simplex concepts corresponding to single Dutch words, but often they represent more complex concepts corresponding, for instance, to a verb and its objects (Figure 7), to two or more nouns or to nouns and prepositional phrases. Most pictographs are for content words and there are hardly any pictographs for prepositions or adverbs.

Although Sclera mainly contains black-and-white pictographs, some of them are green (indicating that something is permitted or approved) while others are red (indicating a ban or disapproval). In some others another colour is used for contrast or to indicate the colour itself. A ban or disapproval may also be expressed by a (red) cross through the pictograph.

As mentioned above, Sclera was originally used as a means to communicate directives to its users (pupils, residents) with as few pictographs as possible. However, the last decade more and more attention is being paid to the communicative needs of people with cognitive disabilities, focusing on *social inclusion*.

3.4.2 The Beta pictograph set

The Beta pictograph set used in the WAI-NOT environment consists of more than 3,000 colour pictographs. A licence for this pictograph set can be obtained at reasonable prices. Their black-and-white equivalents are available for free.

Figure 8 shows some Beta pictographs. Easy recognition of the pictographs is one of the main objectives. The pictograph set is built based on a set of rules, mainly determined by colour, form and position. A number of specific figures appear throughout the different pictographs, consistently playing the same role. The colouring in the drawings is done logically, such that the intended concept stands out from its context. Similarly, arrows are used to point to the object, whereas the rest of the pictograph serves as context to disambiguate the pictograph. An empty white arrow represents a change in space or time. Small dashes are used (like in comic strips) to indicate verbs. Figures in verb pictographs are depersonalised, i.e.

Fig. 8. (Colour online) Beta translation of *eet een boterham 'eat a sandwich'*.

the head is a simple circle, without any facial features in order to allow ignoring natural gender. A more detailed description of the rules to which Beta adheres can be found on `http://www.betasymbols.com/en/images`.

This approach is not free from criticism,[21] but understanding pictographs should be seen as a learning process, with regular exposure to the pictographs.

3.4.3 Linking the pictographs to synsets

We have linked the pictographs to the synsets in the Cornetto database for several reasons. The first reason is that by doing this we greatly improve the lexical coverage of the system, as not only the literal filename of the pictograph, but all its synonyms that are included in the database are now covered. A second reason also concerns the coverage. If for a certain word the synset is not covered by a pictograph, we can use the links between the synsets to look for an alternative pictograph with a similar meaning.

We have tried to minimise ambiguity in the process of linking the pictographs to the synsets. The links were created in two phases. First, we started from the pictographs and linked them with the synsets they were representing.[22] Second, we used the most frequent content words of the corpus, (with a threshold frequency of fifty) excluding our evaluation set, and made sure that links between these words and pictographs were established via the intended Cornetto-synset. For many words, the other possible synsets were never the intended sense of the words. One pictograph is linked to one synset, but note that it is theoretically possible to link different pictographs to the same synset, when they depict the same concept.[23] In those cases we selected a prototypical pictograph in order to provide consistent translations and not to confuse our users.

We have manually linked a subset of 5,710 Sclera and 2,760 Beta pictographs to Cornetto synsets (Vandeghinste and Schuurman 2014), and provide a detailed account of how this was done. As these pictographs sometimes depict complex concepts, they can be linked to one or to more synsets indicating that their meaning combines the meanings of the synsets. In these cases one of the synsets was identified as the head synset, indicating that the other linked synsets are in some kind of dependency relation with the head synset. Table 1 presents how many pictographs

[21] As stated on the Beta website.

[22] Excluding the linking of pictographs that were considered *too specific* for language usage, such as a pictograph depicting a specific command or prohibition, like *klas-kledij-frutselen-rood* which supposedly means *do not tinker with your clothing in class!*

[23] In Sclera, there are e.g. twelve pictographs for *August* and five for *lice*.

Table 1. *Distribution of the number of synsets per pictograph*

Nr of synsets	Distribution for Sclera	Distribution for Beta
1	2,689	2,690
2	2,416	16
3	559	
4	42	
5	3	
6	1	
Total	5,710	2,706

are linked to how many synsets. In comparison with Sclera, the Beta pictographs are much more word-based. They rarely do not form a translation of a single Dutch word, so they are more easily attributed to a single Cornetto synset.

In cases, where the pictograph meaning was not reflected by one or more synsets, they were often linked to the synset of its hyperonym. Note that, these links between Sclera and Beta pictographs and Cornetto-synsets are freely available to anyone upon simple request.

3.4.4 *Using the links between synsets and pictographs*

For every synset of every word we distinguish the *simplex* pictographs from the *complex* pictographs. The simplex pictographs are linked to one single synset. In the complex pictographs, we make a distinction between the *head* pictographs, which are linked to the head synset of that pictograph and the *as_dependent* pictographs, which are linked to the synsets of the dependents of the pictographs.

We decided to further extend the coverage of the pictographs in the system, by using the Cornetto relations between synsets. The `HAS_HYPERONYM` relation indicates the link between a subcategory and a super-category. Figure 9 shows the synset of the concept *vrouw* (woman) in the centre. As shown, there is a pictograph attached to that synset. When we have to translate the word *buurvrouw* (female neighbour) to which no pictograph is attached, we will connect the synset to that word through the `HAS_HYPERONYM` relation with its hyperonym synset *vrouw* and display that pictograph.

Other relations which allows improving the coverage of the system are the `XPOS_SYNONYM` or `XPOS_NEAR_SYNONYM` relations, which indicate a link between similar concepts but with a different part-of-speech. In Figure 9, this is shown by the link between the synsets of the words *vrouw* and *vrouwelijk* (feminine).

Yet another relation between synsets that we use is the `ANTONYM` relation, which indicates that synsets are the opposite of each other. In Figure 9, this is demonstrated by the link between the synsets for *vrouw* and *man* (man). Of course, when finding the pictograph of the antonym, this is marked by giving the *antonym* feature of the word the value `TRUE`, as shown in Figure 10, which will be displayed by the negation pictograph, as shown in Figure 3. The translation of the message *Hij is genezen* (E:

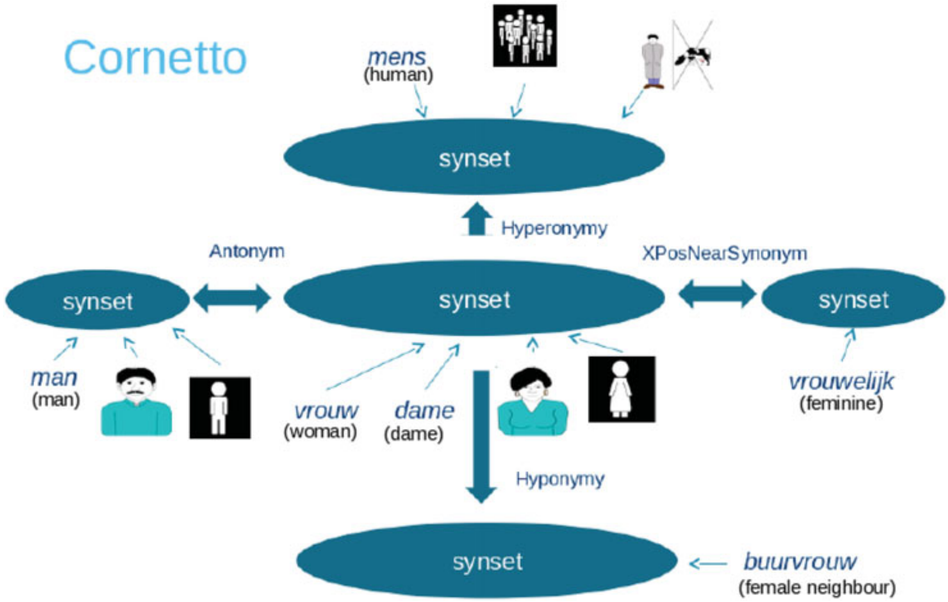

Fig. 9. (Colour online) The synset of *woman* and its links to closely related synsets.

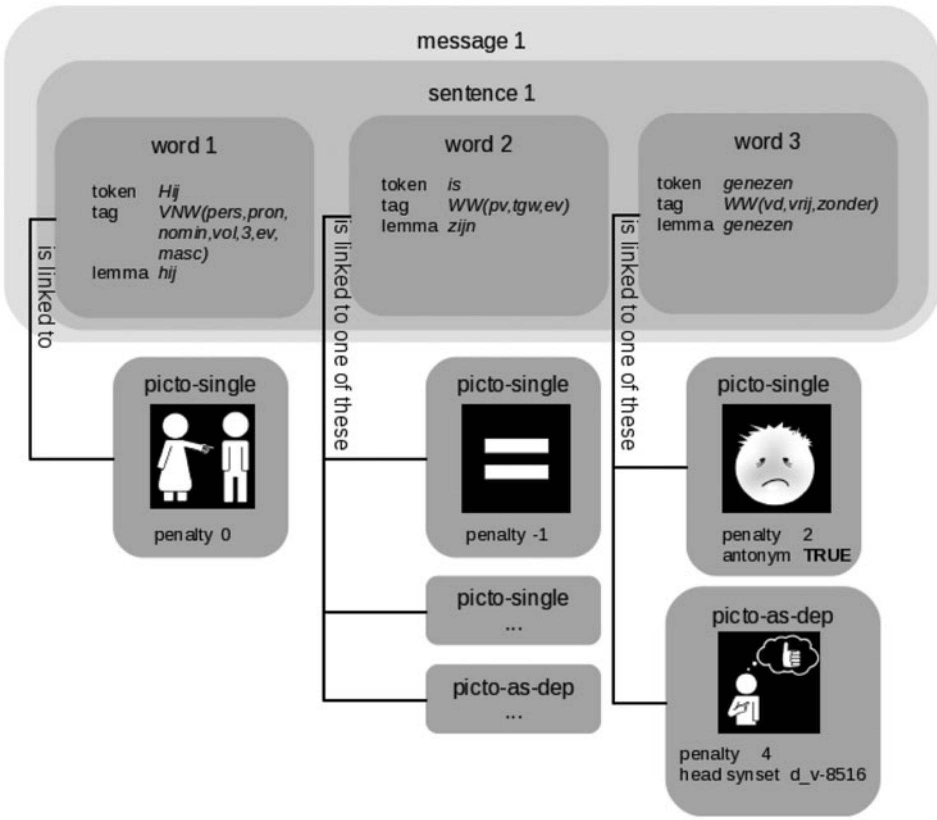

Fig. 10. The example sentence *Hij is genezen* with linked pictographs.

He has recovered) consists of the following sequence of pictographs (verbatim): *'Hij' (He) 'is' (is) 'ziek' (ill) 'niet' (not)*.

Using pictographs through *synset propagation*, as this is called, is controlled by parameters, for which we describe the tuning in Section 4.1. The penalties for using WordNet relations affect how the system selects the pictographs, as described in Section 3.5.

3.5 Finding the optimal path

Up to this point, we have, for every word in the sentence, looked up whether we can find one or more pictographs for it. We have connected these pictographs, together with their *penalty*, depending on the number and kind of synset relations we had to go through to connect them to the words.

The final step consists of finding the optimal path, and takes as its input the message annotated with the connected pictographs, as shown for the example sentence in Figure 10. We define the weight or cost of a path in Equation (1).

$$W(P) = \sum_{i=0}^{n} \sum_{j=0}^{q} W(s_{j_{w_i}}) \qquad (1)$$

where $W(P)$ is the weight of a path P, n is the number of words, w_i is the ith word, q is the number of pictographs associated with that word, and $s_{j_{w_i}}$ is the jth pictograph associated with w_i. If there are no pictographs associated with word w_i, then we set $q = 1$ and $s_{1_{w_i}} = w_i$. The weight of a pictograph $W(s_{j_{w_i}})$ is determined by the relations linking the synset associated with w_i with the synset of the pictograph $s_{j_{w_i}}$. In Algorithm 1 (cf. infra), which is an A* search algorithm (Hart, Nilsson and Raphael 1968) we show how we search for the best path in more detail. We used A* as it is well-known for its performance and accuracy (Dechter and Pearl 1985).

We first show the **findBestPath** subroutine, which takes as input the queue Q containing path P_0. Let P_0 be the path with all words left to process. We shift the first, currently best scoring path (P) from the queue and *extend* it, as explained in the subroutine **extend**. This subroutine returns a number of paths which are added to the queue Q. We remove the double paths, i.e. paths leading to the same solution, keeping only the variant with the lowest *estimated cost*.[24] We sort the queue Q by lowest estimated cost, and repeat until the first queue element Q_0 no longer has any words left to process.

The **extend** subroutine works as follows. The input path has an array of words to process W_P. We shift the first word w from W_P and check whether it has complex pictographs C_w attached to it. If this is the case, we check for each of these complex pictographs $c \in C_w$ whether each of the list S_c of other synsets that are linked to C_w but not to w are linked through the synsets of the remaining words to process W_P. If this is the case, we call the word to which they are linked w_x with x being the index of this word in W_P. Then, we copy the current path P to a new path P', and

[24] Through the WordNet links there are several paths possible between two synsets.

Algorithm 1 Find the best path

sub findBestPath
$Q = (P_0)$
repeat
  P = shift Q
  push(Q, P →extend)
  Q →removeDoubles
  Q →sort
until Q_0 →wordsToProcess = () (Q_0 is the first element of Q)

sub extend
The current path P has words to process $W_P = (w_0, \ldots, w_n)$
w =shift W_P
if $C_w = w$ →complexPictos **then**
  for all $c \in C_w$ **do**
    s_{c_w} is the synset of the picto c which is linked to w
    S_c is the list of synsets of the picto c which are not linked to w
    for all $s'_c \in S_c$ **do**
      if s'_c is linked to a word $w_x \in W_P$ **then**
        $P' = P$ →copy
        push(P', c)
        splice($W'_P, x, 1$)
        push(NewPaths,P')
      end if
    end for
  end for
end if
if $C_w = w$ →simplexPictos **then**
  for all $s \in C_w$ **do**
    $P' = P$ → copy
    push(P', s)
    push(NewPaths,P')
  end for
end if
if NewPaths = () **then**
  $P' = P$ →copy
  push(P', w)
  push(NewPaths,P')
end if
return NewPaths

add the complex pictographs to the list of already matched pictographs in P', and we remove (splice) the word w_x at position x that linked to this synset s_c from the list of words to process W_P. We add the new path P' to the list of new paths. We have implemented a similar, but simpler, treatment for simplex pictographs. If the word w has simplex pictographs C_w, we create a new path P' for each s of C_w by copying P and adding the pictograph s to the list of already matched pictographs. In the case where no matching pictographs were found and the list of new paths is empty, instead of adding a pictograph to the path, we copy the current path P to a new path P', as in the two previous cases, but this time we add the word form as it is to P' instead of a pictograph (as no matching pictographs were found). P' is added to the (empty) list of new paths. The **extend** subroutine returns the list of new paths.

When encountering words that have their *antonym* feature set to `TRUE`, we insert the negation-pictograph, as shown in Figure 3, to communicate the correct polarity of the message.

4 Tuning and evaluation

In this section, we describe how we tuned the parameters and performed an evaluation of the tool, comparing it to the baseline when no language technology was used at all.

Section 4.1 describes how we tuned the parameters that are currently used in our system. Section 4.2 shows the procedures we applied for automatic evaluation whereas in Section 4.3 we describe a manual evaluation, comparing the results of translating into Sclera with the results of translating into Beta. In Section 4.4, we focus on some quantitative and qualitative aspects of the extrinsic evaluation of the system.

4.1 Parameter tuning

We can distinguish between three types of parameters that influence system behaviour. The first set consists of the parameters that tune the behaviour of using *Cornetto relations*, as described in Section 3.4.4. As using such relations increases the distance between the textual and the pictographic message, the use of these relations is subject to a series of penalty parameters that determine the cost of each of these relations. We keep track of how many of these synset links we apply and count penalties[25] for using these relations, which results in preferring the words closest to the original meaning. The maximum penalty that is allowed for finding pictographs for a synset can be set as the *threshold* parameter.

The second set of parameters tunes the behaviour of the system with respect to *features of the pictographs*. Some pictographs clearly depict a concept either in *plural* or in *singular*. For a subset of the Sclera pictographs,[26] we have manually checked whether they depict the concept either in plural or singular. We have introduced a parameter that determines the cost of using the *wrong number*. In cases where either the pictograph is not linked to a certain number, or the number of the word is underspecified (cannot be determined on the basis of the part-of-speech tag), we introduce the parameter for using *no number*.

A third set of parameters determines the behaviour *concerning what route to take*. The *out-of-vocabulary penalty* determines the cost of leaving a word untranslated,

[25] The extra cost of using related concepts instead of the proper concepts.

[26] For every pictograph with a name ending in `-en`, which is the most frequent ending for Dutch plurals. This could in future versions be extended to pictographs ending in `-s`, which is another frequent plural suffix.

Table 2. *Results of parameter tuning*

Parameter	Min	Max	Sclera	Beta
Cornetto relations				
Threshold	5	20	20	17
Hyperonym penalty	0	15	2	8
XPos penalty	0	15	7	7
Antonym penalty	0	15	8	7
Pictograph features				
Wrong number	0	10	8	8
No number	0	10	6	5
Route preference				
Out-Of-Vocabulary penalty	0	10	6	1
Direct route advantage	0	15	1	8

whereas the *direct route advantage* is a negative penalty (bonus) for using the direct route (Section 3.2) over the semantic route (Section 3.4).

We have tuned these parameters through an automated procedure. We built a local hill climber that varies the parameters between certain boundaries (*Min* and *Max*) and with certain granularity (size of parameter steps) when running the Text2Picto script. We ran this on fifty sentences selected from the development set, that were manually translated into Sclera and Beta, maximising the BLEU score (Papineni *et al.* 2002) of the Text2Picto script. BLEU is a precision metric often used in machine translation which compares the system output to one or more reference translations, by counting how many *n*-grams overlap (with *n* going from one to four), and correcting forbrevity.

Each trial of the tuning procedure consisted of a local hill climb with random initialisation of each of the parameters, until the BLEU score converged onto a fixed score. For each of the target pictograph sets, we ran five trials with different random initialisation and a granularity of one[27] in order to cover different areas of the search space. From these five trials we took the best scoring parameter values. These are presented in Table 2.

4.2 Automated evaluation

The *evaluation test set* of 50 Dutch messages that have been sent with the WAI-NOT e-mail system consists of 84 sentences (980 words). For each of these sentences, we created one reference translation in Sclera and one in Beta, translating, to the best of our ability the messages into the respective pictograph sets, focusing on the *content* of the message and how this content can best (most clearly and unambiguously) be expressed in pictographs, not performing a word-by-word translation, and not by post-editing the system output.

[27] The minimal variation steps in a parameter.

We have automatically evaluated different experimental conditions, progressively activating more features of the system. The first condition is the *Baseline* condition in which we only replaced words by pictographs if the pictographs have the same filename (without the `.png` extension) as the input words. In the next condition, we applied *lemmatisation*: the input sentences were tokenised, pos-tagged and lemmatised, and words that have a lemma that has the same filename as the pictograph are translated. Then, we added the *direct route*, as described in Section 3.2. This includes a specific treatment of the pronouns plus a set of dictionary entries per translation direction. The following condition uses the *synonym sets* of the Cornetto concepts: the pictographs are linked to synsets, and the input words are also linked to synsets, and whenever they share a synset, the words are translated into pictographs. The last condition also uses the *relations* between synsets, as described in the previous sections.

Table 3 shows the respective BLEU, NIST (Doddington 2002), Word Error Rate (WER) and Position-independent word Error Rate (PER) scores for the translation of messages into Sclera and into Beta. NIST is similar to BLEU but gives less credit to high-frequency non-informative *n*-grams. WER is often used in speech recognition and counts the number of words that are incorrect with respect to the reference translation(s). PER is like WER but treats the words as a bag and was included as there is no language model available for Sclera nor for Beta, so the position of the pictographs is not necessarily what we want to evaluate. For each condition we have included three variants. *No spelling correction* takes the input as it is and excludes the spelling correction process as described in Section 3.1. *Automated spelling correction* takes the input as it is and applies the spelling correction process as described in Section 3.1. In the *Manual spelling correction*-variant, we have manually corrected the spelling of the input to the best of our abilities, and sent this input to the Text2Picto engine. This was done to calculate an upper bound in order to estimate how much more room for improvement there is in the spelling correction process.

We have added significance levels for the BLEU and NIST scores, by comparing each condition with the condition on the previous line. Significance was calculated using bootstrap resampling (Koehn 2004). Although not presented in the tables, we have calculated significance levels between each pair of conditions. These are only used in the discussion of the results.

As is clear from Table 3 our tool improves translation quality over the baseline. The effect of the spelling corrector is negative or very small (and insignificant for BLEU) in the initial conditions, but it becomes significant for BLEU as well as NIST in the more advanced conditions that use Cornetto (Synonyms and Relations). There is clearly room for further improvement in the automated spelling correction process, as the scores for the upper bound in the advanced conditions are significantly better than the scores for the automated spelling correction process.

Adding lemmatisation results in a substantial improvement. For Sclera, there is significant improvement in comparing Baseline with Lemmatisation in all three variants, both for BLEU and NIST. In Beta, the rise in BLEU and NIST is smaller (because the Baseline has a better score), but still significant for NIST in all three variants, and significant for BLEU in the case of manual spelling correction. Adding

Table 3. *Automatic evaluation of Text2Picto conversion with one reference translation for the different experimental conditions*

	No spelling correction				Automated spelling correction				Manual spelling correction			
Condition	BLEU	NIST	WER	PER	BLEU	NIST	WER	PER	BLEU	NIST	WER	PER
Sclera												
Baseline	00.00	1.43	96.27	92.13	00.00	1.38	99.00	95.58	00.00	1.84	97.51	94.20
Lemmatis.	01.87*	1.68†	94.48	89.36	01.91*	1.73†	93.92	88.81	02.44*	2.29†	92.27	87.02
Direct	10.74†	2.93†	75.41	69.20	11.57†	3.05†	74.59	68.09	14.17†	3.68†	71.96	65.88
Synonyms	**12.02***	**3.32**†	**70.58**	**63.26**	**13.24***	3.41†	**70.03**	**62.43**	**16.55**†	*3.97*†	*67.54*	*60.50*
Relations	11.44	3.29	72.24	64.50	12.75	**3.42**	71.41	63.26	16.12	3.96	68.78	61.33
Beta												
Baseline	05.93	2.29	80.76	72.21	04.94	2.40	81.10	71.99	04.70	2.81	79.19	69.85
Lemmatis.	07.77	2.90†	77.05	66.93	08.15	3.01†	76.94	66.14	10.14†	3.53†	74.92	63.78
Direct	11.96†	3.65†	66.59	57.48	12.72†	3.76†	66.14	56.47	16.98†	4.43†	63.44	53.77
Synonyms	16.57†	4.12†	**56.24**	**46.91**	18.70†	4.28†	**55.12**	**46.01**	23.01†	5.00†	52.42	43.31
Relations	**18.56***	**4.22**†	56.47	47.24	**20.11***	**4.40**†	55.46	**46.01**	**25.91**†	**5.17**†	**51.29**	**42.07**

$^*p < 0.05$, $^\dagger p < 0.01$

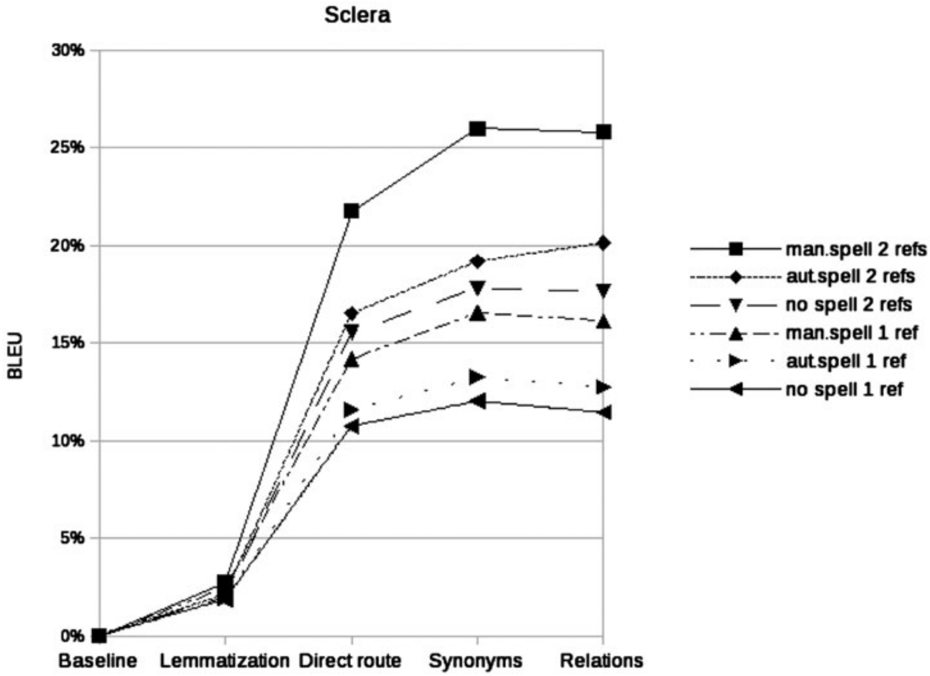

Fig. 11. Bleu scores for Sclera for all experimental conditions in the three variants.

the direct route provides a clear and significant improvement for Sclera and Beta in both BLEU and NIST in all three variants. This can be attributed to the proper treatment of high-frequency phenomena. Adding the synonyms results in another significant improvement for both target languages in BLEU and NIST in all three variants, as this step greatly improves the coverage of the systems, as now all synonyms in a synset can be converted to the pictograph associated with that synset. The effect of adding the relations is less straightforward. It does not seem to progress scores (significantly) in translations into Sclera, in any of the three variants in any of the metrics. For Beta, we see that there is a significant improvement for BLEU and NIST when the relations are added to the system, for all three variants.

Additionally, there is quite a large gap between the results for Sclera and the results for Beta. To find an explanation for this gap, we dug deeper and performed more experiments. The Sclera pictograph set consists of a much larger amount of pictographs than Beta, and it is therefore much more difficult to manually *translate* text into Sclera messages. Several different *paraphrasing* translations are possible, resulting in a less accurate measurement of translation quality by BLEU or NIST. In general WER and PER results are consistent with the results for BLEU and NIST.

To test this hypothesis, we have created a second reference translation for Sclera, based upon post-editing the system output. Figure 11 shows the effect this has on the BLEU score, comparing with the evaluation with only one reference, as presented in Table 3. For two references the effect of adding the synsets is significant in all three variants. The effect of adding the relations however remains insignificant for NIST

Table 4. *Manual evaluation of the Text2Picto translation engine*

		With proper names		Without proper names	
Condition	Precision	Recall	F-Score	Recall	F-Score
Sclera					
Baseline	77.60%	41.42%	54.01%	36.39%	49.55%
Text2Picto	89.24%	86.23%	87.71%	85.18%	87.16%
Rel. improv.	*15.00%*	*108.19%*	*62.39%*	*134.06%*	*75.92%*
Beta					
Baseline	82.73%	62.23%	71.03%	59.57%	69.27%
Text2Picto	85.91%	89.45%	87.64%	88.68%	87.27%
Rel.improv.	*3.84%*	*43.73%*	*23.38%*	*48.88%*	*26.00%*

and BLEU for all variants, apart from NIST on the automated spelling correction variant.

Although we would like to compare our work with the work of others, the only more or less comparable system that we found in literature is the system by Mihalcea and Leong (2009), but they provide a completely different type of evaluation, performing more *psycholinguistic* experiments with human subjects estimating the understandability of messages in which some words had been replaced by pictographs, which is not what we intend to measure. As they do not convert adjectives or adverbs at all, and only convert a selection of verbs, their conversion accuracy would result in a rather low recall compared to our system.

4.3 Manual evaluation

We have performed a manual annotation with one judge, who removed untranslated words that were considered not to contribute to the content. This allows calculating the recall. For each of the translated words, she judged whether the pictograph generated was the correct pictograph, in order to calculate precision. Results are presented in Table 4. These results differ from the results presented in Vandeghinste and Schuurman (2014) as we redid the manual evaluation, but now on the tuned system (cf. Section 4.1), and using a more systematic and objective approach to manual evaluation.

As proper names occur rather frequently in e-mail messages, we have calculated recall and F-score with and without proper names, in the latter case removing all proper names from the output. Precision remains the same in both conditions. In the case where proper names are included, they are not converted into pictographs, affecting recall negatively. In the WAI-NOT environment, proper names occurring in the contact lists of the users are converted into the pictures attached to these profiles, resulting in more *personalised* messages.

The improvements for Sclera are very large, especially the rise in recall, although we still have a substantial rise in precision as well. The Beta baseline system was already much better than the Sclera baseline system, so the improvements for Beta cannot be of the same magnitude. Nevertheless, there is still a substantial rise in recall, and a small rise in precision.

The F-scores for both target pictograph sets, while very different in the baseline, are now all around eighty-seven per cent. The difference in performance of the baseline systems led to the fact that Sclera was nearly unusable. This is now resolved, with both the translation of text into Beta and Sclera reaching similar levels of accuracy.

4.4 Extrinsic evaluation

To have an idea of the real-life effects of the improvements of the system requires an extrinsic evaluation. This can be a *quantitative evaluation*, using actual *usage numbers*, although changes in usage can never be exclusively attributed to the changes in the text-to-pictograph conversion process. Other possible reasons for change are differences in promotional activities, internal competition (other changes in the WAI-NOT environment allowing new possibilities to communicate, such as audio messages, or a WAI-NOT social network), and external competition (other communication environments and possibilities, outside of WAI-NOT).

We have collected the WAI-NOT statistics for over four years of usage, and if we look at usage number for the two years since the Text2Picto system's first version (as described in Vandeghinste (2012)) has been implemented and compare them with the two years before that, there is a relative rise of 38.55 per cent in number of e-mail messages that has been sent with the system.

Up till now we have not systematically collected qualitative feedback. Nevertheless we have received feedback that the biggest improvements of the system over the baseline are due to the translation of conjugated verbs instead of just infinitives, and of nouns, not only translating singular nouns, but also plurals and diminutives. The communication between Beta and Sclera users has improved due to the better conversions, and especially the big improvement in recall makes the WAI-NOT communication platform much more usable.

There have also been a number of improvements that are not inherent to the Text-to-Picto translation engine, but that were implemented at the same time, as the result of brainstorming sessions between the authors and the people of WAI-NOT, such as the use of actual photos from the address book in the case of proper nouns, making the messages much more personal.

5 Conclusions and future work

It is clear from the evaluation that our system provides an improvement in the communication possibilities of illiterate people, although further improvements are surely possible provided more research is done. The implementation of a proper word-sense-disambiguation mechanism should allow for an improvement in precision. The implementation of a higher order model for spelling correction should also result in further improvements, as the targeted user group easily makes spelling mistakes. The current system uses only shallow linguistic analysis. Using chunkers or syntactic parsers as source language analysis might further improve the system, as relations between words might provide further clues towards correct translation into pictographs.

Apart from these implementational issues, future work remains to be done concerning *in vivo* validation of the results. We will deploy the Text2Picto translation system into a Web 2.0 environment enabling the use of social media for illiterate people and assess the effect of the application on the quality of life of the target group.

In order to allow proper Web 2.0 communication we need to investigate how users can enter information through the use of pictographs. These pictograph messages will then be automatically converted into natural language text, again using the links with the Cornetto database. The current WAI-NOT environment employs a two-level menu system for pictograph selection. Users have to first select the category (e.g. *profession, animals*), and in a second click select the appropriate pictograph. We intend to investigate a more FrameNet-oriented approach (Baker, Fillmore and Lowe 1998) towards pictograph input, grouping together pictographs that are likely to appear in the same message (such as e.g. *baker* and *bread*).

Another direction of future work will consist of converting the system so that it translates from English and Spanish into Sclera and Beta, for which the first steps have already been set (Sevens, Vandeghinste and Van Eynde 2014). The Cornetto database lacks a number of twenty-first century concepts such as tablets, smart phones and other technological innovations, often discussed in e-mails by our target group, while these are present in Princeton WordNet, and in Sclera or Beta. We will improve the coverage of Cornetto by transferring the missing concepts from Princeton WordNet to Cornetto, including the relations that link the new concept to the existing Cornetto network. It is also worth mentioning that licences for Cornetto can no longer be obtained, and that Dutch Open WordNet is considered to be its open and free alternative, although the coverage of the latter is much lower. We are also working on a version that uses this resource for Dutch instead of Cornetto.

Acknowledgements

We would like to thank the Support Fund M.M. Delacroix for financing the Picto-project, in which we developed the tool described in this paper, the European Commission's Competitiveness and Innovation Programme for funding the Able-To-Include project, which allows *in vivo* implementations and the Flemish Agency for Innovation by Science and Technology (IWT) which allows further development and valorisation.

We also thank the reviewers for their valuable comments and suggestions, not only leading to improvements in the quality of the paper, but also in the quality of the Text2Picto engine.

We thank the people from WAI-NOT and K-POINT that came up with the idea of text-to-pictogram translation and asked us whether we could do something to improve the baseline.

References

Alm, N., Iwabuchi. M., Andreasen, P., and Nakamura, K. 2002. A multi-lingual augmentative communication system. In *Univeral Access: Theoretical Perspectives, Practice and Experience*, pp. 398–408. Lecture Notes in Computer Science (LNCS), vol. 2615. Berlin: Springer.

Baker, C., Fillmore, C., and Lowe, J. 1998. The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (ACL/CoLing)*. Association for Compututational Linguistics, Montreal, Quebec, Canada, vol. 1, pp. 86–90.

Behrmann M., and Byng S. 1992. A cognitive approach to the neurorehabilitation of acquired language disorders. In D. Margolin (ed.), *Cognitive Neuropsychology in Clinical Practice*, pp. 327–50. Oxford, UK: Oxford University Press.

Borman, A., Mihalcea, R., and Tarau, P. 2005. PicNet: augmenting semantic resources with pictorial representations. In T. Chklovski, P. Domingos, H. Lieberman, R. Mihalcea, and P. Singh (eds.), *Technical Report SS-05-03. Proceedings of the AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors*, pp. 1–7. Menlo Park, California: The AAAI Press.

Brants, Th. 2000. A statistical part-of-speech tagger. In *Proceedings of the 6th Applied Natural Language Processing Conference (ANLP)*. Association for Computational Linguistics, Seattle, Washington, pp. 224–331.

Carney, R., and Levin, J. 2002. Pictorial illustration *Still* improve students' learning from text. *Educational Psychology Review* **14**(1): 5–26.

Coyne, B., and Sproat, R. 2001. WordsEye: an automatic text-to-scene conversion system. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, Association for Computing Machinery (ACM), New York, pp. 487–96.

Davies, D. K., Stock, S. E., and Wehmeyer, M. L. 2001. Enhancing independent internet access for individuals with mental retardation through use of a specialized web browser: a pilot study. *Education and Training in Mental Retardation and Developmental Disabilities* **36**(1): 107–13.

Dawe, M. 2006. Desperately seeking simplicity: how young adults with cognitive disabilities and their families adopt assistive technologies. In R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson (eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1143–52. New York, U.S.: Association for Computing Machinery (ACM).

Dechter, R., and Pearl, J. 1985. Generalized best-first search strategies and the optimality of A*. *Journal of the ACM* **32**(3): 505–36.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. 2009. ImageNet: a large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Institute of Electrical and Electronics Engineers, Miami, FL, pp. 248–55.

Doddington, G. 2002. Automatic evaluation of machine translation quality using N-gram co-occurrence statistics. In *Proceedings of the 2nd International Conference on Human Language Technology Research*, San Diego, California, pp. 138–45.

Goldberg, A., Zhu, X., Dyer, C. R., Eldawy, N., and Heng, L. 2008. Easy as ABC? Facilitating pictorial communication via semantically enhanced layout. In *Proceedings of the 12th Conference on Computational Natural Language Learning (CoNLL)*, Coling 2008 Organizing Committee, Manchester, England, pp. 119–26.

Halácsy, P., Kornai. A., and Oravecz, C. 2007. HunPos – an open source trigram tagger. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, Association for Computational Linguistics, Prague, Czech Republic, pp. 209–12.

Hart, P. E., Nilsson, N. J., and Raphael, B. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics SSC* **4**(2): 100–7.

Joshi, D., Wang, J., and Li, J. 2006. The story picturing engine — a system for automatic text illustration. *ACM Transactions on Multimedia Computing, Communications and Applications* **2**(1): 1–22.

Keskinen, T., Heimonen, T., Turunen, M., Rajaniemi, J. P., and Kauppinen, S. 2012. SymbolChat: a flexible picture-based communication platform for users with intellectual disabilities. *Interacting with Computers,* vol. **24**(5), pp. 374–86. Oxford, UK: Oxford University Press.

Koehn, P. 2004. Statistical significance tests for machine translation evaluation. In Lin, D., and Wu, D. (eds.) *Proceedings of 2004 Conference on Empirical Methods on Natural Language Processing (EMNLP 2004)*, pp. 388–95. Association for Computational Linguistics, Barcelona, Spain: Association for Computational Linguistics.

Medhi, I., Sagar, A., and Toyama, K. 2006. Text-free user interfaces for illiterate and semi-literate users. In *International Conference on Information and Communication Technologies and Development (ICTD)*, pp. 72–82. Berkeley, CA: Institute of Electrical and Electronics Engineers.

Mihalcea, R., and Leong, C. W. 2009. Toward communicating simple sentences using pictorial representations. *Machine Translation* **22**(3): 153–73.

Miller, G. A. 1995. Wordnet: A lexical database for english. *Communications of the ACM* **38**(11): 39–41.

Newell, A., and Gregor, P. 2000. 'User sensitive inclusive design' – in search of a new paradigm. In *Proceedings of the Conference on Universal Usability (CUU'00)*, Association for Computing Machinery (ACM), Arlington, VA, pp. 39–44.

Oostdijk, N., Goedertier, W., Van Eynde, F., Boves, L., Martens, J. P., Moortgat, M., and Baayen, H. 2002. Experiences from the spoken dutch corpus project. In M. Rodríguez, and C. Araujo (eds.), *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC)* , pp. 340–7. Las Palmas, Spain: European Language Resources Association.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL)*. Philadelphia, PA, pp. 311–8.

Sevens, L., Vandeghinste, V., and Van Eynde, F. 2014. Improving the precision of synset links between Cornetto and Princeton WordNet. In *Proceedings of the COLING Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014)*, Association for Computational Linguistics and Dublin City University, Dublin, Ireland, pp. 120–6.

Takasaki, T., and Mori, Y. 2007. Design and development of a pictogram communication system for children around the world. In T. Ishida, S. R. Fussell, and P. T. J. M. Vossen (eds.) *Intercultural Collaboration*, pp. 193–206. Berlin, Heidelberg: Springer.

Vandeghinste, V. 2002. Lexicon optimization: maximizing lexical coverage in Speech recognition through automated compounding. In M. Rodríguez and C. Araujo (eds.), *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC)*, pp. 1270–6. Las Palmas, Spain: European Language Resources Association.

Vandeghinste, V. 2012. Bridging the gap between pictographs and natural language. In *Proceedings of the W3C/WAI Research and Development Working Group (RDWG) Online Symposium: Easy-to-Read on the Web*. W3C Web Accessibility Initiative. http://www.w3.org/WAI/RD/2012/easy-to-read/paper14/

Vandeghinste, V., and Schuurman, I. 2014. Linking pictographs to synsets: Sclera2Cornetto. In N. Calzolari, K. Choukri, Th. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis (eds.), *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14),* pp. 3404–10, Reykjavik, Iceland: European Language Resources Association.

Van den Bosch, A., Busser, G. J., Daelemans, W., and Canisius, S. 2007. An efficient memory-based morphosyntactic tagger and parser for Dutch. In F. Van Eynde, P. Dirix, I. Schuurman, and V. Vandeghinste (eds.), *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting*, pp. 99–114, Utrecht: Landelijke Onderzoeksschool Taalkunde.

Van den Bosch, A., Schuurman, I., and Vandeghinste, V. 2006. Transferring PoS-tagging and lemmatization tools from spoken to written Dutch corpus development. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, and D. Tapias (eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy: European Language Resources Association.

van der Vliet, H., Maks, I., Vossen, P., and Segers, R. 2010. The Cornetto database: Semantic issues in linking lexical units and synsets. In A. Dijkstra, T. Schoonheim (eds.), *Proceedings of the 14th EURALEX 2010 International Congress*. pp. 477–83, July 6–10, 2010, Leeuwarden, the Netherlands: Fryske Akademy/De skriuwers.

Van Eynde F. 2005. Part-of-Speech tagging en lemmatisering van het D-Coi corpus. Centrum voor Computerlinguïstiek. University of Leuven, Belgium. p. 88.

van Noord, G. 2006. At last parsing is now operational. In P. Mertens, C. Fairon, A. Dister, and P. Watrin (eds.), *Verbum Ex Machina. Actes de la 13e conference sur le Traitement Automatique des Langues Naturelles (TALN06),* pp. 20–42. Belgium: Presses universitaires de Louvain, Louvain-la-Neuve.

van Noord G., Bouma, G., Van Eynde, F., de Kok, D., van der Linde, J., Schuurman, I., Tjong Kim Sang, E., and Vandeghinste, V. 2013. Large scale syntactic annotation of written Dutch: Lassy. In P. Spyns, and J. Odijk (eds.), *Essential Speech and Language Technology for Dutch: Resources, Tools and Applications*, pp. 147–64. Berlin Heidelberg: Springer.

Vossen, P., Görög, A., Izquierdo, R., and Van den Bosch, A. 2012. DutchSemCor: targeting the ideal sense-tagged corpus. In N. Calzolari, K. Choukri, T. Declerck, M. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk and S. Piperidis (eds.), *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, pp. 584–9, Istanbul, Turkey: European Language Resources Association.

Vossen, P., Maks, I., Segers, R., and van der Vliet, H. 2008. Integrating lexical units, synsets, and ontology in the Cornetto Database. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, and D. Tapias (eds.), *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*, pp. 1006–13, Marrakech, Morocco: European Language Resources Association.