Multimodal Building of Monolingual Dictionaries for Machine Translation by Non-Expert Users

Miquel Esplà-Gomis and **Víctor M. Sánchez-Cartagena** and **Juan Antonio Pérez-Ortiz**
Dep. de Llenguatges i Sistemes Informàtics
Universitat d'Alacant, E-03071 Alacant, Spain
{mespla,vmsanchez,japerez}@dlsi.ua.es

Abstract

This paper explores a new approach to help non-expert users with no background in linguistics to add new words to a monolingual dictionary in a rule-based machine translation system. Our method aims at obtaining the correct paradigm which explains not only the particular surface form introduced by the user, but also the rest of inflected forms of the word. An initial set of potential paradigms is automatically obtained and then interactively refined by the user with a novel graphical interface through active machine learning. We show the promising results of experiments performed with a Spanish monolingual dictionary.

1 Introduction

Rule-based machine translation (MT) systems heavily depend on explicit linguistic data such as monolingual dictionaries (MD), bilingual dictionaries (BD), grammars, and structural transfer rules (Hutchins and Somers, 1992). Although some automatic acquisition is possible, collecting these data usually requires in the end the intervention of domain experts (mainly, linguists) who master all the encoding and format details of the particular MT system. It could be interesting, however, to open the door to a broader group of non-expert users who could collaboratively enrich MT systems through the web.

In this paper we present a novel method for building or enlarging the MDs in rule-based MT systems by non-expert users. An automatic process is first run to collect as much linguistic information as possible about the new word to be added to the dictionary using a corpus-based technique. After that, the resulting set of potential hypothesis is filtered by eliciting additional knowledge from non-experts with no linguistic background through *active learning* (Olsson, 2009; Settles, 2010), that is, by interactively querying the user in order to efficiently reduce the search space. As these users do not possess the technical skills which are usually required to fill in the dictionaries, this elicitation is performed via the classification of a set of critical *inflected word forms* (IWFs) derived from the assignment of the word to a concrete paradigm. This technique only requires *speaker-level* understanding of the language and allows not only to incorporate to the dictionary the particular surface form introduced by the user (for example, *wants*), but it also discovers a suitable paradigm for the new word so that all the IWFs of the corresponding lexeme and their morphological information (such as *wanted, verb, past* or *wanting, verb, gerund*) are also inserted.

In a previous work (Esplà-Gomis et al., 2011), the interaction with the user was only possible through a text-based interface. In this paper, we introduce an advanced graphical drag-and-drop interface, which reduces the time needed by the user to supply the information necessary to assign a paradigm to the new entry and which is specially designed for devices equipped with touchscreens. The text-based interface still guarantees accessibility and compatibility with any kind of devices.

Monolingual Dictionaries (MDs). This work focuses on this kind of dictionaries, which basically have two types of data: *paradigms*, that group regularities in inflection, and *word entries*. The paradigm assigned to many common English verbs, for instance, indicates that by adding the ending *-ing*, the gerund is obtained. Paradigms make easier the management of dictionaries in two ways: by reducing the quantity of information that needs to be stored, and by simplifying revision and validation thanks to the explicit encoding of regularities in the dictionary. Once the most frequent paradigms in a dictionary are defined, entering a new IWF is generally limited to writing the stem and choosing an inflection paradigm. Our system helps to assign new words to one of the existing paradigms in a MD by efficiently interrogating the user.

In our experiments we used the free/open-source rule-based MT system Apertium (Forcada et al., 2011), which is being currently used to build MT systems for a variety of language pairs. Every word is assigned to a paradigm in Apertium's MDs, and specific paradigms are defined for words with irregular forms. In addition, all the lexical information is included in the paradigms; as a result, paradigms usually exist which only contain lexical information and do not add any suffix to the corresponding stem; the paradigm for the proper nouns is a good example of this. It is worth noting that stems in Apertium MD are as small as necessary to generate all the possible IWFs. For example, for word *teeth*, we have the stem *t*, which produces the IWF *teeth* and *tooth*.

Bilingual Dictionaries (BDs). Once a word and its corresponding translation have been added to the MDs of the source and target languages, respectively, of a MT system, the next step is to link both of them by adding the corresponding entry in the BD. How to adapt this task to non-experts is out of the scope of this paper and will be tackled in future works.

Social Translation. In spite of the vast amount of content uploaded to the web during the last years, linguistic barriers still pose a significant obstacle to universal collaboration as they lead to the creation of "islands" of content, only meaningful to speakers of a particular language. Until *fully-automatic high-quality* MT becomes a reality, massive online collaboration in translation may well be the only force capable of tearing down these barriers (Garcia, 2009) and produce large-scale availability of multilingual information.

The resulting scenario, which may be called *social translation*, will need efficient computer translation tools, such as reliable MT systems, friendly postediting interfaces, or shared translation memories. Remarkably, collaboration around MT should not only concern the postediting of raw machine translations, but also the creation and management of the linguistic resources needed by the MT systems; if properly done, this can lead to a significant improvement in the translation engines. Since as many hands as possible are necessary for the task, speakers that, in principle, do not have the level of technical know-how required to improve MT systems or manage linguistic resources must be involved, and, consequently, software that can make those tasks easier and elicit the knowledge of both experts and non-experts must be developed (Font-Llitjós, 2007; Sánchez-Cartagena and Pérez-Ortiz, 2010). Large-scale collaboration implies a change in the way linguistic resources are managed and a series of conditions should hold in order to fully accomplish these goals (Pérez-Ortiz, 2010).

Knowledge Elicitation and Active Learning. Two of the more prominent works related to the elicitation of knowledge for building or improving MT systems are those by Font-Llitjós (2007) and McShane et al. (2002). The former proposes a strategy for improving both transfer rules and dictionaries by analysing the postediting process performed by a non-expert through a special interface. McShane et al. (2002) design a complex framework to elicit linguistic knowledge from informants who are not trained linguists and use this information to build MT systems into English; their system provides users with a lot of information about different linguistic phenomena to ease the elicitation task. Ambati et al. (2010) show how to apply active learning (Olsson, 2009) to the configuration of a statistical MT system.

Automatic Extraction of Resources. Many approaches have been proposed to deal with the automatic acquisition of linguistic resources for MT, mainly, transfer rules and dictionaries, even for the specific case of the Apertium platform (Caseli et al., 2006; Sánchez-Martínez and Forcada, 2009). The automatic identification of morphological rules (a problem for which paradigm identification is a potential resolution strategy) has also been subject of many recent studies (Monson, 2009; Creutz and Lagus, 2007; Goldsmith, 2010; Walther and Nicolas, 2011).

Multimodal Machine Translation. Multimodality has been used to enhance MT technologies in many different ways. For instance, interactive MT systems, which involve human experts so that they help to obtain the translation through a standard interface controlled with a keyboard and a mouse (Koehn and Haddow, 2009; Ortiz-Martínez et al., 2010), can be enhanced with a touchscreen (Alabau et al., 2010).

Novelty. Our work introduces some novel elements compared to previous approaches:

1. Unlike the Avenue formalism used in the work by Font-Llitjós (2007), our MT system is a *pure* transfer-based one in the sense that a single translation is generated and no language model is used. Therefore, we are interested in the unique right answer and assume that an incorrect paradigm cannot be assigned to a new word.

2. Bartusková and Sedlácek (2002) also present a tool for semi-automatic assignment of words to declination patterns; their system is based on a decision tree with a question in every node. Their proposal, however, focuses on nouns and is aimed at experts because of the technical nature of the questions.

3. Our approach is addressed to non-experts, and, therefore, the answer to as few as possible simple questions is our main source of information (in addition to what an automated extraction method may deliver in a first step). Font-Llitjós (2007) already anticipated the advisability of incorporating an active learning mechanism in her transfer rule refinement system, asking the user to validate different translations deduced from the initial hypothesis. However, this active learning approach has not yet been undertaken. Unlike the work by McShane et al. (2002), we want to relieve users of acquiring linguistic skills.

4. Our work focuses on identifying the paradigm which could be assigned to a word, a task more restrictive than decompounding a word into morphemes. The technique defined by Monson (2009) tolerates some errors in the final output.

5. Our mid-term intention is to develop a system in line with the social translation principles which may be used to collaboratively build MT systems from scratch. This will also include the semi-automatic learning of the paradigms or the transfer rules which better serve the translation task, and which do not need necessarily correspond to the linguistically motivated ones.[1]

6. Our work uses a graphical interface for the *configuration* of an MT system. Although this kind of interfaces are commonly used in the process of translation (Alabau et al., 2010; Koehn and Haddow, 2009), they are rarely taken into account to do so.

2 Methodology

In this section, a short description of our method to assign new words to paradigms is presented; additional details may be found in the paper by Esplà-Gomis et al. (2011). Although our approach focuses on languages which generate IWFs by adding suffixes to the stems of words (as happens, for example, in Romance languages), it could be easily adapted to inflectional languages based on different ways of adding morphemes. Let $P = \{p_i\}$ be the set of paradigms in a MD. Each paradigm p_i defines a set of suffixes $F_i = \{f_{ij}\}$ which are appended to stems to build new IWFs. Given a *stem/paradigm* pair c composed of a stem t and a paradigm p_i, the *expansion* $I(t, p_i)$ is the set of possible IWFs resulting from appending each of the suffixes in p_i to t. For instance, an English dictionary may contain a paradigm p_i with suffixes $F_i = \{\epsilon$,-*s*, -*ed*, -*ing*$\}$ (ϵ denotes the empty string), and the stem *want* assigned to p_i; the expansion $I(\text{want}, p_i)$ consists of the set of IWFs *want*, *wants*, *wanted* and *wanting*. We also define a *candidate stem* t as an element of $\Pr(w)$, the set of possible prefixes of a particular IWF w. Note that here we are not referring to prefix as an affix which is placed before the stem, but to a sequence of characters at the beginning of the word.

Given a new IWF w to be added to a MD, our objective is to find both the candidate stem $t \in \Pr(w)$ and the paradigm p_i which expand to the largest possible set of morphologically correct IWFs. To that end, our method performs three tasks: obtaining the set of all stem/paradigm candidates which produce the IWF w when expanded; giving a *confidence score* to each of these stem/paradigm candidates so that the next step is made as short as possible; and, finally, asking the user about the correctness of some of the

[1] For example, a single inferred paradigm could group IWFs for verbs like *wait* (ϵ, -*s*, -*ed*, -*ing*) and nouns like *waiter* (ϵ, -*s*), whereas an expert would probably write two different paradigms in this case.

IWFs derived from each of the stem/paradigm candidates obtained in the first step. Next we describe the methods used for each of these tasks.

It is worth noting that in this work we assume that all the paradigms for the words in the dictionary are already included in it and all them are correct. The situation in which for a given word no suitable paradigm is available in the dictionary will be tackled in the future, possibly by following the ideas in related works (Monson, 2009).

2.1 Paradigm Detection

To detect the set of paradigms which may produce the IWF w we use a *generalised suffix tree* (GST) (McCreight, 1976) containing all the possible suffixes included in the paradigms in P. Each of the suffixes added to the GST is labelled with the name of the paradigms which contain it. In this way, the GST data structure allows to retrieve the paradigms compatible with w by efficiently searching for all the possible suffixes of w. Finally, a list L is built containing all the candidate stem/paradigm pairs. We will denote each of these candidates with c_n.

The following example illustrates this stage of our method. Consider a simple dictionary with only three paradigms: p_1, with F_1=$\{f_{11}$=ϵ, f_{12}=*-s*$\}$; p_2, with F_2=$\{f_{21}$=*-y*, f_{22}=*-ies*$\}$; and p_3, with F_3=$\{f_{31}$=*-y*, f_{32}=*-ies*, f_{33}=*-ied*, f_{34}=*-ying*$\}$. Lets assume that a user wants to add the new word w=*policies* to the dictionary. The candidate stem/paradigm pairs which will be obtained after this stage are: c_1=*policies*/p_1, c_2=*policie*/p_1, c_3=*polic*/p_2, and c_4=*polic*/p_3.

2.2 Paradigm Scoring

Once L is obtained, a *confidence score* is computed for each stem/paradigm candidate $c_n \in L$ using a large monolingual corpus C. The score already introduced in our previous work (Esplà-Gomis et al., 2011) considers the frequency of occurrence in the corpus of each IWF in each candidate c_n. In this way, candidates producing a set of IWFs which are more likely to appear in the corpus get higher scores. Those IWFs which are very unusual (for example, some verbal tenses rarely used in Spanish) are discarded to obtain a more robust method.

Following our example, the IWFs for the different candidates would be: $I(c_1)$=$\{$*policies*, *policiess*$\}$, $I(c_2)$=$\{$*policie*, *policies*$\}$, $I(c_3)$=$\{$*policy*, *policies*$\}$, and $I(c_4)$=$\{$*policy*, *policies*, *policied*, *policying*$\}$. Using a large monolingual English corpus C, IWFs *policies* and *policy* will be easily found, and the rest of them (*policie*, *policiess*, *policied* and *policying*) will not. Therefore, c_3 would obtain the highest score.

2.3 Active Learning Through User Interaction

Finally, the best candidate is chosen from L by querying the user about a reduced set of the IWFs for some of the candidate paradigms $c_n \in L$. To do so, our system firstly sorts L in descending order using the confidence score previously computed. Then, users are asked (following the order in L) to confirm whether some of the IWFs in each expansion *exist* in the language; in this way, when an IWF w' is presented to the user

- if it is accepted, all $c_n \in L$ for which $w' \notin I(c_n)$ are removed from L;
- if it is rejected, all $c_n \in L$ for which $w' \in I(c_n)$ are removed from L.

We propose (Esplà-Gomis et al., 2011) an iterative algorithm to choose the IWF to be asked to the user each time. In a first stage (*confirmation stage*) a strategy is defined to find a possible solution between the candidates in L, i.e. a candidate c for which all the IWFs in $I(c)$ are correct. Then, in a second stage (*discarding stage*) some words produced by those candidates c' with $I(c) \subset I(c')$ are shown to the user to determine whether they are better than c or not. These two stages are repeated until only one candidate remains in L. Note that this method cannot distinguish between candidates which produce the same set $I(c_n)$: in the experiments, they are considered as a single candidate.

3 Two interfaces for user interaction

Our system provides two different user interfaces. Both of them are oriented to the classification of IWFs w into correct and incorrect forms, although they provide different levels of interaction to the user.

On the one hand, our system provides a very simple text-based interface, which asks the user about one single IWF at a time and gets a *yes*/*no* answer following the steps explained in Section 2.3. On the other hand, a more complex interface is provided to make easier the task of classifying the IWFs. In this drag-and-drop interface, a *word cloud* is shown to the user containing the most likely IWFs for a subset of $n = 10$ paradigms c_n in L. These word forms have

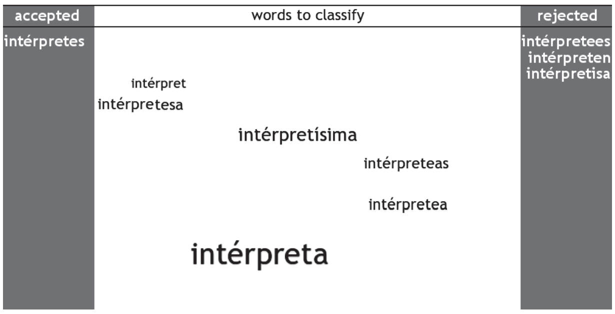

Figure 1: Screenshot of the drag-and-drop interface when the user wants to add the word *intérprete* to the MD.

different font sizes to emphasise those which are supposed to be more common. The user can drag these words to two columns in the left and right side of the application window: one for the correct forms and the other one for the incorrect forms. When a form is classified, the words in the cloud change by removing or adding the new possible IWFs of the remaining candidates. In addition, the font size of the words in the cloud changes depending on the frequency of the new words.

An example of the drag-and-drop interface can be seen in Figure 1. This screenshot shows the interface when the noun *intérprete* is being added to the dictionary. The user has classified the word *intérpretes* as correct, which is the plural of *intérprete*, and discarded *intérpretees*, *intérpreten* and *intérpretisa*. Among the remaining IWFs we can find, for example, *intérpreta*, which is the result of adding the suffix *-a* to the stem *intérpret* by expanding an existing paradigm which inflects, for the substantives which end in *-e*, the IWFs of masculine singular (*-e*) and plural (*-es*), and feminine singular (*-a*) and plural (*-as*). Another example of incorrect form is *intérpretísima*, which is the result of adding the superlative-feminine suffix *-ísima* to the stem *intérpret* by expanding a paradigm which produces the same IWFs than the one which produced *intérperta* but in this case for adjectives; therefore, it includes the superlative inflection suffixes masculine (*-ísimo*) and feminine (*-ísima*).

This second interface was implemented using HTML5 canvas[2] and JavaScript, so it can run on most modern web browsers and devices. It provides a set of advantages to the user:

- if the most suitable paradigm is not the first in L, but it is in the top n paradigms of the list, it is shown to the user from the first iteration, so the number of interactions required to choose the paradigm is lower;
- users are oriented to choose the most usual IWFs thanks to the font size of the word forms in the cloud;
- users may ignore a particular IWF if they are not sure about whether it is correct or not; if the information provided by later decisions is enough for obtaining the right paradigm, this IWF will have remained unclassified;
- all the decisions taken so far are visible and amending one of them is as easy as dragging a word from one column to the other.

The graphical interface is much better adaptable than the text-based one to modern portable devices (such as smartphones or tablets) which are usually equipped with a touchscreen. In these devices is where the drag-and-drop interface could achieve its maximum potential.

Our textual interface can be useful mainly to guarantee the accessibility to users with disabilities, since it can be easily used with screen readers.

4 Experiments

The aim of the experiments is to asses, in a realistic scenario, whether our semi-automatic methodology is valid to find out, for a given word, its most suitable paradigm. Therefore, a group of evaluators has been told to add a set of words to a MD using the two previously described interfaces. For this task, we chose the Apertium Spanish MD from the Spanish–Catalan language pair. The dictionary was filtered to remove word entries belonging to a closed part-of-speech category, since they are so frequent that we may assume that they have been included at the moment of the creation of a dictionary, and multi-word units and IWFs which are obtained by adding prefixes to the stem, which are out of the scope of this paper. After that, a test set was created with words extracted from the filtered dictionary as follows. First, a stem assigned to each of the paradigms p_i was added to the test set; to build a more realistic test set, we chose one more stem from those common paradigms having more than 10 stems assigned. Then, we obtained, for

[2] http://www.w3.org/TR/html5/

each pair stem/paradigm, all the possible IWFs and included the most common ones into the test set. In this way, we obtained 226 words: 106 extracted from the first group of paradigms and 120 from the second one. Obviously, the stems from which we obtained the words included in the test set were removed from the dictionary.

Then, the test set was split into 10 subsets and two human evaluation steps were carried out, one for each interface described in Section 3. An heterogeneous group of 10 non-expert evaluators was chosen. For each interface, each subset was assigned to an evaluator, who was asked to introduce each of its words in the Apertium dictionary using our system. Experiments were run using the filtered dictionary and a word list obtained from the Spanish Wikipedia dump[3] as the monolingual corpus C.

The different evaluation metrics obtained from the human evaluation process are:

- *success rate*: percentage of words from the test set that have been tagged with the paradigm assigned to them in the original MD;
- *average precision and recall*: precision (P) and recall (R) were computed as
$$P(c, c') = |I(c) \cap I(c')| \cdot |I(c)|^{-1},$$
$$R(c, c') = |I(c) \cap I(c')| \cdot |I(c')|^{-1},$$
where c is the stem/paradigm pair chosen by our system and c' is the pair originally in the dictionary; confidence intervals were estimated with 99% statistical confidence with a *t-test*;
- *average number of questions*: average number of questions made by our system for each word in the test set;
- *average number of initial paradigms*: the average number of compatible paradigms initially found as possible solutions in the first stage of our method.

Finally, an alternative approach without user interaction was designed as a baseline to better evaluate the impact of active learning. The baseline consists of directly choosing the first element in the list L as the most suitable candidate. The average position of the right candidate in L has also been computed.

[3] `http://dumps.wikimedia.org/eswiki/20110114/eswiki-20110114-pages-articles.xml.bz2`

5 Results and Discussion

We evaluated our approach and computed the results following the metrics depicted in Section 4. The average number of initial candidates was 56.4; it was specially high for verbs, whereas it was much lower for nouns and adjectives. Figure 2 shows an histogram representing the position of the right candidate in the initial list L for each word in the test set. In addition, it is worth mentioning that the average position of the right candidate in L was 9.1.

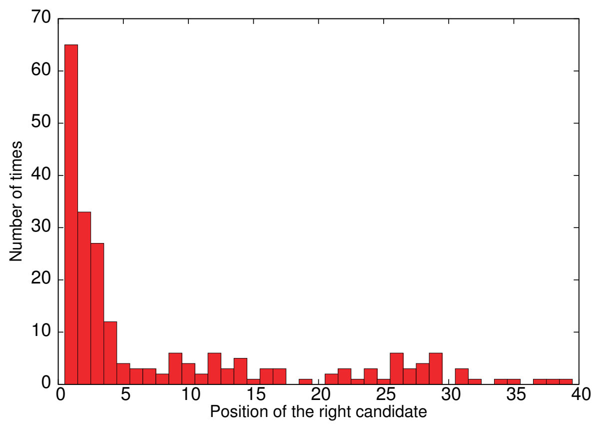

Figure 2: Distribution of the position of the right candidate in the initial list of candidates L for each word in the test set.

Success rate, precision and recall for the two active learning interfaces and the baseline system are presented in Table 1. It also shows the average number of actions users have to perform to find the paradigm. The difference between the success rate and the precision and recall of the active learning approaches stress the fact that those words which were assigned to incorrect paradigms, were assigned to paradigms producing similar IWFs. These results are clearly better than those obtained by the baseline approach and confirm the improvement provided by the intervention of the users, even when they are not experts.

Results show that there are not statistically significant differences between the results obtained by the two interfaces. Despite this fact, in general, users expressed that they prefer the graphical interface since it is more dynamic and allows them to freely choose the forms to classify. In addition, it is worth noting that evaluators needed less interactions to finish the task using the drag-and-drop interface, probably because they could choose the words which they doubtlessly knew. We observed that, in average, users needed

System	% success	P	R	Actions
base	$29\% \pm 8$	$70\% \pm 6$	$63\% \pm 7$	-
text	$73\% \pm 8$	$87\% \pm 5$	$87\% \pm 4$	5.2 ± 1
graph	$71\% \pm 8$	$87\% \pm 5$	$91\% \pm 4$	4.2 ± 1

Table 1: Success rate, precision and recall obtained by the non-interactive baseline system, the text-based interface, and the drag-and-drop graphical interface, and average number of operations carried out by the evaluators using each interface.

around 30 seconds in average to find the paradigm of each word in the test set.

Taking a closer look at the results, we observed some relevant causes for the errors which reduce the success rate. On the one hand, we detected human errors for words which should have been accepted but were rejected or vice-versa. These mistakes, caused by a lack of knowledge of the users (for example, about accentuation rules), should be taken into account in the future; they could be solved, for instance, by using *reinforcement questions* or combining the answers of different users for the same or similar words. Moreover, it could be possible to give a kind of *confidence score* to the paradigms in the dictionary based on how frequently words are incorrectly assigned to them.

We also observed that most of the words which were not assigned to the expected paradigm were verbs. Spanish morphological rules allow multiple concatenations of enclitic pronouns at the end of verbs. In many occasions, users rejected forms of verbs with too many enclitic pronouns or for which some concrete enclitics had no semantic sense. This happens because, in order to reduce the number of possible paradigms, Apertium's dictionaries can assign some words to existing paradigms which are a superset of the correct one; since the included semantically incorrect IWFs will never occur in a text to translate, this, in principle, may be safely done.

6 Limitations and Work Ahead

In this paper we have described a system for interactively enlarging dictionaries and selecting the most suitable paradigm for new words. Our preliminary experiments have brought to light several limitations of our method which will be tackled in the future.

Detection of Lexical Information. One of the most important limitations of our approach is that, as already commented in Section 2, candidate paradigms producing the same $I(c_n)$ set cannot be distinguished. This situation usually holds when the expansions of two different stem/paradigm pairs are equal but the lexical information in each paradigm is different. For example, in Spanish two different paradigms may contain the same suffixes F=$\{\epsilon$, -$s\}$ although one of them is assigned to substantives and the other one is assigned to adjectives.

We have started to explore a method to semi-automatically obtain this lexical information. A statistical part-of-speech tagger may be used to obtain initial hypothesis about the lexical properties of a word w; this information could then be refined by querying users with complete sentences in which w plays different lexical roles.

Lack of Suitable Paradigms. Our approach assumes that all the paradigms for a particular language are already included in the dictionary, but it could be interesting to have a method to also add new paradigms. The work by Monson (2009) could be a good start for the new method.

Improvement of the Graphical Interface. We plan to make some changes to make our drag-and-drop interface more user-friendly . Some evaluators detected that the changes in the cloud when a word is classified and the remaining IWFs change of position and size may be confusing; these changes should be introduced in a soft and gradual way.

Enlargement of BDs. To make a word in a MD available for the MT system, it must be also added to the BD. We plan to define interactive methods similar to the one in this paper to adapt this task to non-expert users.

Other Improvements. We plan to improve our approach by using simple statistical letter models of bigrams or trigrams to discard candidates producing morphologically unlikely IWFs, or by using additional information in the scoring stage, such as word context, number of occurrences, etc.

7 Conclusions

We have shown an active learning method for adding new entries to MDs. Our system allows non-expert users with no linguistic background to contribute to the improvement of rule-based MT systems by means of textual and graphical interfaces. The Java source code for the method described in Section 2 is published[4] under an open-source license.

Acknowledgements

This work has been partially funded by Spanish Ministerio de Ciencia e Innovación through project TIN2009-14009-C02-01 and by Generalitat Valenciana through grant ACIF/2010/174 from VALi+d programme.

References

V. Alabau, D. Ortiz-Martínez, A. Sanchis, and F. Casacuberta. 2010. Multimodal interactive machine translation. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10, pages 46:1–46:4, New York, NY, USA. ACM.

V. Ambati, S. Vogel, and J. Carbonell. 2010. Active learning and crowd-sourcing for machine translation. In *Proceedings of the Seventh conference on International Language Resources and Evaluation, LREC 2010*.

D. Bartusková and R. Sedlácek. 2002. Tools for semi-automatic assignment of czech nouns to declination patterns. In *Proceedings of the 5th International Conference on Text, Speech and Dialogue*, pages 159–164, London, UK. Springer-Verlag.

H. Caseli, M. Nunes, and M.L. Forcada. 2006. Automatic induction of bilingual resources from aligned parallel corpora: application to shallow-transfer machine translation. *Machine Translation*, 20:227–245.

M. Creutz and K. Lagus. 2007. Unsupervised models for morpheme segmentation and morphology learning. *ACM Trans. Speech Lang. Process*, 4.

M. Esplà-Gomis, V.M. Sánchez-Cartagena, and J.A. Pérez-Ortiz. 2011. Enlarging monolingual dictionaries for machine translation with active learning and non-expert users. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, Hissar, Bulgaria. Accepted for publication.

A. Font-Llitjós. 2007. *Automatic improvement of machine translation systems*. Ph.D. thesis, Carnegie Mellon University.

M.L. Forcada, M. Ginestí-Rosell, J. Nordfalk, J. O'Regan, S. Ortiz-Rojas, J.A. Pérez-Ortiz, F. Sánchez-Martínez, G. Ramírez-Sánchez, and F.M. Tyers. 2011. Apertium: a free/open-source platform for rule-based machine translation. *Machine Translation. Special Issue on Free/Open-Source Machine Translation*, in press.

[4]https://apertium.svn.sourceforge.net/svnroot/apertium/branches/apertium-dixtools-paradigmlearning

I. Garcia. 2009. Beyond translation memory: Computers and the professional. *The Journal of Specialised Translation*, 12:199–214.

J.A. Goldsmith, 2010. *The Handbook of Computational Linguistics and Natural Language Processing*, chapter Segmentation and morphology. Wiley-Blackwell.

W. J. Hutchins and H. L. Somers. 1992. *An introduction to machine translation*. Academic Press, London.

P. Koehn and B. Haddow. 2009. Interactive assistance to human translators using statistical machine translation methods.

E.M. McCreight. 1976. A space-economical suffix tree construction algorithm. *Journal of the Association for Computing Machinery*, 23:262–272, April.

M. McShane, S. Nirenburg, J. Cowie, and R. Zacharski. 2002. Embedding knowledge elicitation and MT systems within a single architecture. *Machine Translation*, 17:271–305.

C. Monson. 2009. *ParaMor: From Paradigm Structure to Natural Language Morphology Induction*. Ph.D. thesis, Carnegie Mellon University.

F. Olsson. 2009. A literature survey of active machine learning in the context of natural language processing. Technical report, School of Electronics and Computer Science, University of Southampton.

D. Ortiz-Martínez, I. García-Varea, and F. Casacuberta. 2010. Online Learning for Interactive Statistical Machine Translation. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT)*, pages 546–554, Los Angeles.

J.A. Pérez-Ortiz. 2010. Social Translation: How Massive Online Collaboration Could Take Machine Translation to the Next Level. In *Second European Language Resources and Technologies Forum: Language Resources of the Future, FlarenetNet Forum 2010*, pages 64–65, Barcelona, Spain.

V.M. Sánchez-Cartagena and J.A. Pérez-Ortiz. 2010. Tradubi: open-source social translation for the Apertium machine translation platform. In *Open Source Tools for Machine Translation, MT Marathon 2010*, pages 47–56.

F. Sánchez-Martínez and M.L. Forcada. 2009. Inferring shallow-transfer machine translation rules from small parallel corpora. *Journal of Artificial Intelligence Research*, 34:605–635.

B. Settles. 2010. Active learning literature survey. Technical report, Computer Sciences Technical Report 1648, University of Wisconsin–Madison.

G. Walther and L. Nicolas. 2011. Enriching morphological lexica through unsupervised derivational rule acquisition. In *Proceedings of the International Workshop on Lexical Resources*.