Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 50 (2008) 1034–1046

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

On the impact of morphology in English to Spanish statistical MT

A. de Gispert *, J.B. Mariño

TALP Research Center, Universitat Politècnica de Catalunya (UPC), Campus Nord, c/Jordi Girona 1-3, 08034 Barcelona, Spain

Received 1 June 2007; received in revised form 24 March 2008; accepted 13 May 2008

Abstract

This paper presents a thorough study of the impact of morphology derivation on N-gram-based Statistical Machine Translation (SMT) models from English into a morphology-rich language such as Spanish. For this purpose, we define a framework under the assumption that a certain degree of morphology-related information is not only being ignored by current statistical translation models, but also has a negative impact on their estimation due to the data sparseness it causes. Moreover, we describe how this information can be decoupled from the standard bilingual N-gram models and introduced separately by means of a well-defined and better informed feature-based classification task.

Results are presented for the European Parliament Plenary Sessions (EPPS) English → Spanish task, showing oracle scores based on to what extent SMT models can benefit from simplifying Spanish morphological surface forms for each Part-Of-Speech category. We show that verb form morphological richness greatly weakens the standard statistical models, and we carry out a posterior morphology classification by defining a simple set of features and applying machine learning techniques.

In addition to that, we propose a simple technique to deal with Spanish enclitic pronouns. Both techniques are empirically evaluated and final translation results show improvements over the baseline by just dealing with Spanish morphology. In principle, the study is also valid for translation from English into any other Romance language (Portuguese, Catalan, French, Galician, Italian, etc.).

The proposed method can be applied to both monotonic and non-monotonic decoding scenarios, thus revealing the interaction between word-order decoding and the proposed morphology simplification techniques. Overall results achieve statistically significant improvement over baseline performance in this demanding task.

Keywords: Morphology generation; N-gram based translation; Statistical machine translation; Machine learning

1. Introduction

It is well known that the performance of statistical machine translation systems when translating from English into Spanish leaves room for improvement when compared with the opposite direction. The most reasonable explanation for this is that the Spanish language, having a richer morphology than English, tends to be represented by a larger vocabulary set, making decisions harder for SMT systems (i.e., models have higher perplexity).

Indeed, whereas from Spanish to English several input words may share the same (or very close) translation probability distributions and translate into the same target words, from English to Spanish a single input word may present a wider range of possible translations. In other words, while in the Spanish → English direction sparsity problems may arise in the source language (higher percentage of OOVs, fewer translation examples for each input word, etc.), in English → Spanish these problems arise in the target language (higher perplexity in translation and target language models, etc.). It is reasonable to expect that this would also hold for language pairs such as English and Portuguese, Catalan, Galician, Italian, French, or any

* Corresponding author. Present address: University of Cambridge, Department of Engineering, Trumpington Street, Cambridge CB2 1PZ, UK.

E-mail addresses: ad465@cam.ac.uk (A. de Gispert), canton@gps.tsc.upc.edu (J.B. Mariño).

0167-6393/$ - see front matter
doi:10.1016/j.specom.2008.05.003

Romance-family language, given their strong linguistic similarities.

Despite this fact, few research efforts have been devoted to dealing with the specific problems of translating from English into a richer morphology language such as Spanish, especially for large-data tasks (with notable exceptions mentioned in Section 1.2). The question of how much Spanish morphology derivation is affecting the statistical translation models needs to be addressed. In particular, how big an improvement would be achieved if morphology generation was flawless? What morphological relationships cannot be captured by current statistical models? Would it be possible to estimate these independently from the translation model?

This paper studies these questions in detail. For that, we work on two basic assumptions:

– A certain degree of morphology-related information is not only being ignored by current statistical translation models, but also has a negative impact on their estimation due to the data sparseness it causes.[1]
– This morphology information can be incorporated by means of an independent model.

In order to assess the validity of the first statement, we estimate simplified morphology models by removing various Spanish morphological features from the corpus and evaluate their oracle translation scores against simplified morphology references. By comparing these scores with the post-processing oracle scores obtained when simplifying Spanish morphology in the output of a standard SMT system, the (negative) impact of Spanish morphology derivation on the translation model can be measured.

With regards to the second statement, we propose considering Spanish morphology generation as a standard classification task. After defining a set of simple relevant features, we apply machine learning techniques to classify and evaluate this independent morphology generation task. Finally, we evaluate actual translation performance (against normal references) after morphology generation via boosting. This is done for both monotonic and non-monotonic decoding situations, and aims to reveal the interaction between morphology and word order when translating this language pair.

The paper is organized as follows. The rest of the present section introduces the SMT system used in this work and reviews previous publications in dealing with language-specific issues. Section 2 discusses the current limitations when generating Spanish morphology and proposes a system architecture to estimate simplified morphology translation models and morphology generation independently (Section 2.2). A way to measure the impact of morphology on translation modeling is addressed in Section 2.3. Section 3 studies the proposed morphology simplification models, including various alternative schemes and Spanish enclitic pronoun separation. It also analyses the oracle scores obtained when using simplified morphology reference translations.

Section 4 addresses Spanish morphology generation as a feature-based classification task which can be implemented via standard machine learning techniques, while Section 5 reports final translation scores after generation and studies the interaction of the presented morphology simplification techniques with word-order statistical models. Finally, Section 6 sums up the main conclusions drawn in the paper.

1.1. N-gram-based SMT

The SMT system used in this work follows the N-gram-based approach by Mariño et al. (2006a). It performs a log-linear combination of a translation model and additional feature functions. The translation model is estimated as a standard 4-gram model of a bilingual language expressed in *tuples* with Kneser–Ney smoothing (Kneser and Ney, 1995). This way it approximates the joint probability between source and target languages capturing bilingual context,[2] as described by the following equation:

$$p(s_1^J, t_1^I) = \prod_{i=1}^{K} p\big((s,t)_i | (s,t)_{i-N+1}, \ldots, (s,t)_{i-1}\big) \quad (1)$$

assuming a *unique* segmentation of the bilingual sentence into K tuples:

$$(s_1^J, t_1^I) = (s,t)_1, (s,t)_2, \ldots, (s,t)_K \quad (2)$$

Additional feature functions include a target language model, a word bonus model and two lexicon models. The *target language model* is estimated as a standard 5-gram over the target words with Kneser–Ney smoothing. This feature interacts with the *word bonus model* based on sentence length, as the latter compensates for the target language model preference for short sentences (in number of target words).

Finally, two lexicon feature functions account for the source-to-target and target-to-source IBM model 1 word translation probabilities to compute a lexical weight for each tuple.

For decoding we employ MARIE, a freely-available N-gram-based decoder (Crego et al., 2005) allowing both monotonic and non-monotonic search. The system has proved to be state-of-the-art in recent evaluations involving the English → Spanish language pair (Crego et al., 2006; Mariño et al., 2006b).

[1] If estimated at a word surface level, translation models are defined over each inflected form of the same base form. For morphologically rich languages, many forms will be sparsely represented in the training data, leading to uninformed model estimation.

[2] In (Casacuberta and Vidal (2004)) a Finite-State Transducer implementation of this model is used as full translation system, without any additional features.

1.2. Related work

Several authors have delved into language-specific problems and their impact on statistical translation. However, literature in this area is nearly entirely focused on translating into English.

A primary work on the subject can be found in Nießen et al. (2000), where several transformations of the source string for a German → English task are proposed, leading to slightly increased translation performance. Transformations address issues such as compound word separation, reordering of separated verb prefixes (which are placed after the object in German), and word mapping to word plus POS to distinguish articles from pronouns, among others.

In Nießen and Ney (2004) hierarchical lexicon models including base form and POS information for translation from German into English are introduced, among other morphology-based data transformations. The same pair of languages is used in Corston-Oliver and Gamon (2004), where the inflectional normalization leads to improvements in the perplexity of IBM translation models and reduces alignment errors.

More recently, but still for the German → English pair, a sentence re-ordering as preprocessing is presented in Collins et al. (2005). Similarly to Nießen et al. (2000), German input strings are re-ordered into a more English-like sentence order, obtaining better translation quality.

Regarding Iberian languages, an approach to deal with inflected forms is presented in Ueffing and Ney (2003), tackling verbs in an English → Spanish task. The authors join personal pronouns and auxiliaries to form extended English units and do not transform the Spanish side, leading to an increased English vocabulary. Also in English → Spanish, Gispert et al. (2005) propose a statistical class-based model for full verb forms, which requires an additional instance model in order to decide the actual verb form given the base-form class at decoding time. In both publications, translation quality in a small-data task is improved.

In the opposite translation direction, Popovic et al. (2004) also transform a text in a more-inflected language (Catalan, Spanish and Serbian) to separate base forms and suffixes for verb forms, improving slightly the performance when translating into a less-inflected language (English). On the other hand, Gupta and Federico (2006) transform Spanish words into base forms and stems to obtain decreasing improvements as training size increases when translating into English.

As for translation from another highly-inflected language such as Czech into English, Al-Onaizan et al. (1999) and Goldwater and McClosky (2005) present a couple of techniques modifying Czech input words (substituting them for lemmas, POS tags or combinations of both) into a language more similar to English, again obtaining improvements in BLEU scores for a small-data task. More recently, in Talbot et al. (2006) vocabulary reduction is applied to the source language (Czech, French and Welsh) via automatic model clustering, by conflating those source words with similar translation distributions.

Table 1
Example of English → Spanish bilingual training sentence

I ask you and your party to give support for the release			
Les pido a usted y a su partido que	respalden	la	liberación
Part-Of-Speech information →	POS:VM M:sjve T:pres P:3rd N:pl	POS:DA G:fem N:sing	POS:NC G:fem N:sing
Base form →	[respaldar]	[el]	[liberación]

Regarding Arabic → English translation, Lee (2004) reports improved performance when automatically inducing Arabic word segmentation according to a word alignment to English material. A more recent and thorough study of the impact of Arabic word segmentation schemes for large-vocabulary translation into English is conducted in Habash and Sadat (2006). Similarly, Zollmann et al. (2006) investigate ways of mitigating the negative effect of Arabic inflection, showing improvements in a small-data Arabic → English task. In parallel to these research efforts, the Transtac project[3] addresses two-way English–Iraqi and English–Pashto speech translation, tackling inflectional issues in domain-limited tasks.

Finally, and aiming for a more general approach to dealing with language-specific challenges, Koehn and Hoang (2007) introduce factored translation models that can efficiently integrate morpho-syntactic information into phrase-based SMT with successful results for various language pairs.

2. Overall system and evaluation scheme

2.1. An illustrative example

Consider the bilingual training example shown in Table 1, where Part-Of-Speech (POS) information is included for the last Spanish words.[4] Additionally, the base form for these words is shown in the bottom row.

Regarding the Spanish verb form '`respalden`', certain considerations need be taken into account. On the one hand, the reason for it being third person plural (`P:3rd`, `N:pl`) is the necessary subject-verb agreement, where the subject of the relative clause is '`usted y su partido`'. It seems obvious that this specific dependency cannot be learned independently from lexical instances using a

[3] Visit <http://www.iraqcomm.com> for more information on this project.

[4] Note that 'POS' refers to Part-Of-Speech category ('VM', 'DA' and 'NC' meaning Main Verb, Determinant Article and Common Noun), 'M' refers to verb mode ('sjve' meaning subjunctive), 'T' to tense ('pres' meaning present), 'P' to person ('3rd' meaning third), 'N' to number ('sing' and 'pl' meaning singular and plural) and 'G' to gender ('fem' meaning feminine).

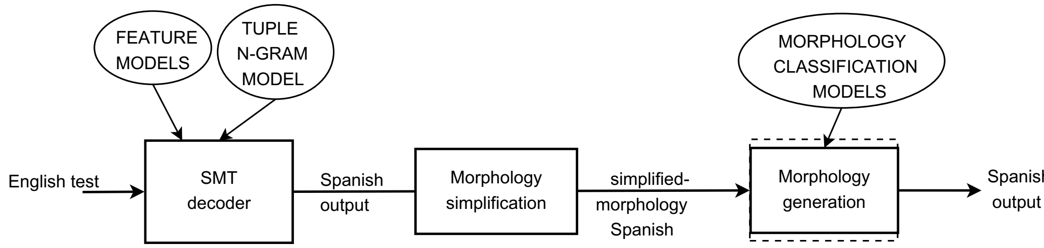

Fig. 1. Post-processing morphology in standard translation outputs.

standard N-gram-based or phrase-based translation model. In other words, unless the exact input sequence '`you and your party to give`' is found in a test set, current translation models will be unable to select the adequate morphology derivation.

On the other hand, the reason for the verb being present subjunctive (`M:sjve,T:pres`) is the structure '`pedir a alguien que HAGA algo`', or equivalently in English, '`to ask someone TO DO something`', where the capitalized Spanish word *must be* a verb in subjunctive mode, its tense depending on the tense of the preceding verb '`pedir`'. Again, it is evident that these complex dependencies cannot be captured by the bilingual N-gram model. During test decoding, a change in the person *being asked to do something* will lead to a very uninformed translation solution.

In addition, it is reasonable to expect that in training these morphology variations will cause such subjunctive examples to be different if the number and person information differs (as the Spanish verb form changes), weakening the chances of correctly translating this complex structure.

When it comes to the Spanish noun '`liberación`', its gender information is invariant, whereas it is reasonable to expect that its number information (`N:sing` or `N:pl`) will somehow depend on the English noun '`release`'.

And finally, regarding the Spanish article '`la`', its being feminine singular is solely due to its preceding a feminine singular Spanish noun ('`liberación`'). No relevant information for this gender and number decision can be extracted from either the English sentence or the preceding Spanish verb '`respalden`'. Therefore, assuming a certain tuple segmentation as in Table 2, it is clear that the trigram defined by tuples (T_3, T_4, T_5) is not useful to generate the article '`la`'. Only the bigram defined by (T_5, T_6) will be useful, or the less-sparse target language model if this particular bilingual bigram is not observed in training.

These kinds of morphology-related errors have been identified as one of the most important types of errors in this task (see Vilar et al. (2006), Popovic et al. (2006), Popovic et al. (2006) for human and automatic error analysis of this task). They have previously been studied in rule-based machine translation (Hutchins and Somers, 1992).

Table 2
Example of tuple-segmented English → Spanish bilingual sentence

T_1	T_2	T_3	T_4	T_5	T_6
your	party	to	give support for	the	release
su	partido	que	respalden	la	liberación

Given the existence of these errors, it is reasonable to expect that an adequate post-processing stage could correct these errors in translation output. Indeed, such a process would involve first removing or simplifying Spanish morphological information (or parts of it), followed by a morphology generation module that would introduce the proper derivations into the Spanish text, as depicted in Fig. 1.

As we will see in the following section, we have implemented this generation module by means of a feature-based statistical classifier. For the simplification step, Part-Of-Speech tagging and lemmatization of the Spanish output is required.[5] Given this information, we can define different levels of morphology simplification, depending on which word categories are simplified.

2.2. System architecture

In order to investigate how much Spanish morphology is affecting the statistical translation model, we propose the framework defined by the architecture from Fig. 2.

After standard word alignment and tuple extraction, we substitute target language words (Spanish) with their simplified morphology forms. Then we estimate the bilingual N-gram translation model with these new tuples. Optionally, this morphology simplification can also be performed at the source-language side (English).

Several types of morphology simplification schemes can be applied, depending on which Part-Of-Speech category is modified (verbs, nouns, adjectives, etc.) or which of its morphology attributes. For instance, mode, tense, person or number information for Verbs, or person, number and gender for Adjectives. Combining these schemes can make the set of possible simplified morphology models grow endlessly. Section 3.3 presents all the simplified configurations investigated here.

The result is a standard bilingual model translating English into simplified morphology Spanish. This translation process may be independently evaluated if compatible sim-

[5] This linguistic processing will probably contain more errors than usual due to the fact of the sentence not being a correct Spanish but a machine translation hypothesis.

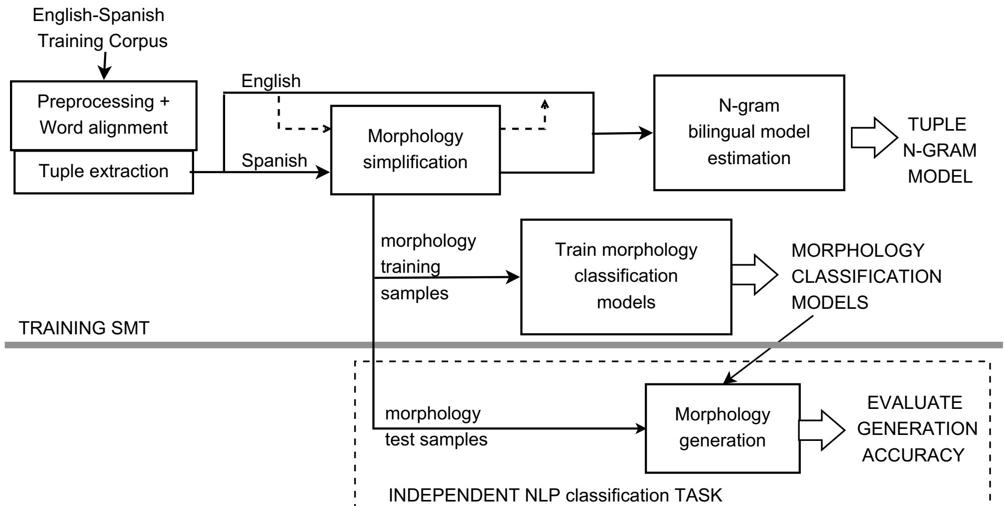

Fig. 2. Above, flow diagram of the training of simplified morphology translation models. Below, Spanish morphology generation as an independent classification task.

plified morphology references are provided, as discussed in Section 2.3.

Additionally, the morphology simplification module produces a set of samples whose correct morphology is known, as they belong to the training corpus. These samples can be used to train morphology classification (or generation) models. Any strategy capable of estimating the correct morphology class given the sample and its corresponding features (as defined in Section 4) can be implemented here. Interestingly for development purposes, this morphology generation task can be independently evaluated if a certain number of training samples is reserved as a development set, as illustrated below the gray line in Fig. 2.

To illustrate this process, let us consider again the example from Table 1 and assume that only Spanish verb person and number information is simplified. In this case, the verb form '`respalden`' is transformed into '`VMSPpn[respaldar]`', indicating simplified POS and base form. Under this simplification, the POS keeps information on word category ('VM' → main verb), mode and tense ('SP' → subjunctive, present), whereas 'p' and 'n' represent any person and number.

Furthermore, as the correct person and number for this verb is known beforehand, it also serves as a classification training sample. Assuming that the set of features describing this sample includes details on its subject ('`usted y su partido`'), accurate morphology classification models can be estimated.

During translation decoding, we follow the sequential approach depicted in Fig. 3, in which a single 1-best simplified morphology translation output is obtained, and posterior morphology generation is conducted independently.

In this approach, the translation system acts as a black box, all its models being estimated over simplified morphology parallel texts (including target language model and lexicon models). Apart from its simplicity, the advantage is that it allows for our study to be carried out with any statistical translation system through a simple black-box substitution.

2.3. Evaluation scheme

By evaluating a simplified morphology version of the translation output from Fig. 1 against compatible simplified morphology reference translations, we obtain an *oracle estimate for morphology post-processing*. This serves as an indicator of how much morphology can be corrected from baseline translation outputs. For example, if we remove all number and gender information from each Spanish noun in both the output and the references, the resulting BLEU score will theoretically reveal how much is to be gained by correcting that information in our current translation hypothesis.

On the other hand, if we evaluate the simplified morphology output of the translation system from Fig. 3 against the same compatible simplified morphology reference translations, we obtain an *oracle estimate for simplified morphology modeling*. The difference between this oracle estimate and the post-processing oracle estimate will represent how much gain we can expect to obtain by estimating simplified morphology translation models. In other words, the negative impact of sparsity caused by Spanish morphology will be empirically computed.

Finally, the same comparison needs to be carried out after morphology generation for both post-processing and simplified models. This will measure the actual gains in final translation quality (see Section 4).

3. Morphology simplification

3.1. Database and baseline

The following study was carried out on a large-data English → Spanish task, defined by a corpus containing the European Parliament Plenary Sessions (also known

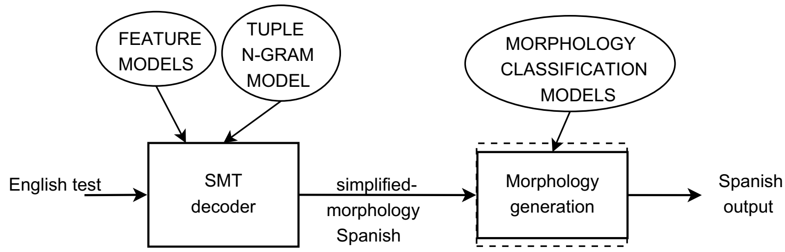

Fig. 3. Sequential translation architecture for simplified morphology translation models.

Table 3
Spanish–English European Parliament Proceeding corpus statistics

EPPS corpus		Sent.	Words	vcb	OOVs	Avg. len.	References
	English		34.92 M	106.5 k	–	27.2	
Train	Spanish	1.28 M	36.58 M	153.1 k	0	28.5	1
Development	Spanish	504	15.4 k	2.7 k	19	30.6	3
Test	Spanish	840	22.8 k	4.1 k	28	27.1	2

as EPPS) from 1996 to May 2005, whose main statistics are shown in Table 3. Statistics include the number of sentences, running words, vocabulary size, out-of-vocabulary words, average sentence length and number of human references available. This task has been used in many previous research efforts, for both text and speech translation, especially within the framework of the European Project TC-STAR[6] and in three international shared tasks[7].

The development set was used to find the optimal model weights of the log-linear combination by optimizing the BLEU score. To obtain the required morphological analysis of each word, English was POS-tagged using the *TnT*[8] tagger by Brants (2000), which is claimed to have a 96.7% average accuracy on the English Penn Treebank. Spanish was tagged using the *FreeLing*[9] analysis toolkit by Carreras et al. (2004), which is claimed to have an accuracy of over 95% on general tasks.

For this task, the N-gram system with monotonic search (no re-ordering capabilities) gives a score of 47.85 BLEU. In this task, the margin of confidence at the 95% level is 0.5% for BLEU. Note that BLEU scores will be presented in the scale between 0 and 100 in this paper.

3.2. Enclitic pronoun separation

Spanish enclitic pronouns are attached to certain verb forms (mostly infinitive and imperative forms) as suffixes. Forms like ‘pedirle’ or ‘hágame’ cause vocabulary size to increase and contribute to data sparsity problems. To counteract this effect, we separate them according to the information provided by the above-mentioned FreeLing analysis toolkit.

While this would be a trivial strategy when translating from Spanish into English, this is not the case in the opposite direction, as we need to be able to regenerate full Spanish words after translating with separated enclitic pronouns. Therefore, we mark each enclitic pronoun with a ‘+’ symbol after POS tagging. For example, verb form ‘pedirle’ is converted into ‘pedir +le’.

We also force the translation decoder to produce the Part-Of-Speech of the generated Spanish output, so that enclitic pronouns can be linked back to their preceding word if that is an admissible verb form (mostly infinitive and imperative forms), as observed in training. Note that this process requires placing accents on vowels in non-infinitive cases, which can be detected according to the verb POS tag (as in ‘diga +nos’ → ‘díganos’).

Unfortunately, applying this technique to the baseline system does not cause any change in translation scores (47.88 in BLEU). Apparently, the monotonic decoding strategy seems insensitive to this linguistically-informed word segmentation. However, as we will see in Section 5.2, this behavior changes when non-monotonic search is conducted.

3.3. Morphology simplification schemes

The following morphology simplification schemes were considered: Table 4.

For each case, the original surface form (i.e., word) is transformed into a sequence of base form plus simplified POS tag. The word category and the level of simplification discriminate among studied schemes. For example, if we consider the S:*N* case, all words falling into the ‘noun’ category are transformed into their base form plus their simplified POS, which does not include the information on

[6] See <http://www.tc-star.org> for extensive list of publications.
[7] 2005 ACL Workshop on Building and Using Parallel Texts, 2006 HLT/NAACL Workshop on Statistical Machine Translation and 2007 ACL Workshop on Statistical Machine Translation.
[8] Available at <www.coli.uni-saarland.de/~thorsten/tnt>.
[9] Available at <http://garraf.epsevg.upc.es/freeling>.

Table 4
Morphology simplification schemes considered for Spanish (S) and English (E)

S:*D*	Determiners: gender and number
S:*A*	Adjectives: gender and number
S:*N*	Nouns: gender and number
S:*P*	Pronouns: person, number and gender
S:*V*	Verb: person, number
S:V_{MT}	Verb: person, number, mode and tense
S:V_{MT}+E:*VP*	S:V_{MT} + English pronouns and verbs: person, number
S:DAV_{MT}	S:V_{MT} + S:*D* + S:*A*
S:full	S:V_{MT} + S:*D* + S:*A* + S:*N* + S:*P*
S:full+E:full	S:full +E:VP + English nouns: number

gender and number as the normal POS would. Table 5 presents an example to illustrate these configurations. Note that all personal pronouns have a unique base form (indicated by '[]') and that English verb simplification only applies to third person singular present tense verbs.

On the other hand, the FreeLing package also provides a list of all pairs of base form and POS tag it can analyse. Accordingly, we only simplify a certain morphological attribute if there exist at least two pairs of this base form with identical POS except for that attribute. For example, as the Spanish noun '`liberación`' does not vary in gender (being always feminine), the process will not simplify the gender information (only number). In contrast, the noun '`perro`' is present in the list with different gender and number, and will thus simplify both attributes.

The post-processing oracle results for each morphology configuration are shown in Table 6. Note that corresponding simplified morphology references are used in each case.

Furthermore, we show for each case the number of output generated words (trgwrds) and the percentage of these being modified due to morphology simplification. For this post-processing case, the number of output words is obviously fixed. The number of simplified words represents the number of generation decisions that need to be taken in order to obtain the final Spanish text.

Table 5
Examples for each morphology simplification configuration

English	she	has	a	strong interest in		
POS	PRP	VBZ	DT	JJ	NN	IN
Spanish	ella	tiene	el	máximo interés en		
POS	PP3SF	VMIP3S	DAMS	AQMS	NCMS	SPS
S:*D*	ella tiene DAgn[el] máximo interés en					
S:*A*	ella tiene el AQgn[máximo] interés en					
S:*N*	ella tiene el máximo NCMn[interés] en					
S:*P*	PPpng[] tiene el máximo interés en					
S:*V*	ella VMIPpn[tener] el máximo interés en					
S:V_{MT}	ella VMmtpn[tener] el máximo interés en					
S:DAV_{MT}	ella VMmtpn[tener] DAgn[el] AQgn[máximo] interés en					
S:full	PPpng[] VMmtpn[tener] DAgn[el] AQgn[máximo] NCMn[interés] en					
E:VP	PRPpng[] VBPp[have] a very strong interest in					
E:full	PRPpng[] VBPp[have] a very strong Nn[interest] in					

Table 6
Post-processing oracle scores of each simplified morphology configuration

Baseline	BLEU	Trgwrds (% simp)
	47.85	25,733 (0)
S:*D*	**48.53**	25,733 (13.7)
S:*A*	48.23	25,733 (7.9)
S:*N*	47.89	25,733 (21.8)
S:*P*	47.93	25,733 (0.8)
S:*V*	48.33	25,733 (8.3)
S:V_{MT}	**49.16**	25,733 (8.3)
S:DAV_{MT}	50.42	25,733 (31.4)
S:full	**50.93**	25,733 (54.0)

As can be seen, the most promising oracles when simplifying one single class are determiners and verbs (when simplifying mode and tense as well). The bilingual model oracle for determiners ('S:*D*') improves BLEU by 0.8 points absolute over the baseline. For verbs, simplifying person, number, mode and tense ('S:V_{MT}') results in a better oracle score, growing up to 1.3 points. Simplifying the morphology of determiners affects ~14% of the generated words, whereas in the case of verbs this amount is only ~9%.

Adjectives ('S:*A*') present a smaller oracle gain of only 0.5 absolute BLEU, requiring morphology decisions on up to ~8% of the target words. In contrast, noun simplification involves modifying up to ~22% of the target words without a promising oracle result. Finally, pronouns involve very few words, with a greatly reduced impact on translation oracle scores.

By simplifying all Spanish categories ('S:*full*') the BLEU oracle obtains an approximate 3 point increase and modifies more than half the target words (54.0%). However, by simplifying only determiners, adjectives and verbs the oracle scores are only 0.5 point lower, and only 31% of the generated words are simplified (see 'S:DAV_{MT}').

3.4. Analysis of oracle results

Let us now analyze the most promising oracles in detail, namely Determiners and Verbs. In order to do that, we study 100 random sentences whose oracle Word Error Rate is better than the baseline. The result is shown in Table 7, where four basic cases are distinguished. Firstly, an

Table 7
Morphology simplification for determiners. Post-processing oracle analysis

Frequency and type	Example
9% Adjacent disagreement error	nombrar a un Comisión Reference: una Comisión
39% Far disagreement error	las atroces situación Reference: la atroz situación
29% Wrongly Translated noun	las cautiverio Reference: los cautivos
23% Noun mismatch	las Naciones Unidas Reference: la ONU

adjacent or far disagreement error is denoted whenever the adequate Spanish noun is placed and its adjacent or nearby determiner does not present agreement. In most of the cases, the use of monotonic search limits the capability of the translation model.

On the other hand, whenever the adequate Spanish noun is wrongly produced or omitted (29% of the cases), the determiner contributes to an over-optimistic oracle, as there is no information to instantiate it correctly. Finally, whenever a different correct noun is produced (22.6% of the cases), the morphology simplification leads again to an over-optimistic oracle increase.

In contrast with these findings, a similar overview study of 100 sentences reveals that the oracle for verbs accounts for a true morphological error in 70% of the cases (see Table 8). The remaining cases include third person confusion and differences between the reference and correct translation.

We note that the English word 'you' is either translated as third person singular or plural in Spanish ('usted' or 'ustedes'), depending on the context. Unless the English sentence introduces the subject (as in '`mister President, as you know`'), then both singular and plural translations are valid, which causes mismatch with references in 14% of the cases. Together with the remaining mismatch cases, these situations do not represent a true morphology oracle.

Overall, the oracle gains described are small, especially taking into account the analysis in the previous section. It seems clear that morphology post-processing will not produce a very strong impact on translation quality, especially regarding determiners, nouns and adjectives. These findings correlate with the human error analysis in Vilar et al. (2006), where the authors identify an approximate 10% of word morphology errors for the same task.

3.5. Simplified morphology models

Let us now examine the possible gain from estimating simplified morphology models. As introduced in Section 3, we can obtain the same translation oracles for each simplification configuration by estimating a simplified morphology model. Results are presented in Table 9, where the S:N and S:P configurations have been omitted as they do not result in any significant change in post-processing oracles.

Table 8
Morphology simplification for verbs

Frequency and type	Example
69% Verb error	la Unión Europea, que legalizaron Reference: que legalizó
14% 3rd person confusion	como sabe (Input: as you know) Reference: como saben
17% Reference mismatch	el pueblo prefiere Reference: los ciudadanos prefieren

Post-processing oracle analysis

Table 9
Post-processing vs modeling oracle scores of each simplified morphology configuration

Baseline	BLEU 47.85	trgwrds (%simp) 25,733 (0)
S:D	48.40	25,697 (14.3)
S:A	48.19	25,833 (8.2)
S:V	48.88	26,733 (9.0)
S:V_{MT}	**49.72**	26,189 (9.5)
S:DAV_{MT}	50.60	25,433 (32.8)
S:full	**51.68**	26,028 (55.3)
S:full+E:full	51.43	26,209 (55.1)

Interestingly, translation oracles for determiners ('S:D') do not improve with respect to the post-processing case. This fact indicates that morphology simplification for determiners does not contribute to estimate a better translation model. The same conclusions can be drawn regarding adjectives ('S:A').

On the other hand, verb morphology simplification ('S:V_{MT}') achieves a 0.7 absolute BLEU higher oracle with respect to post-processing, showing the negative impact of verb morphology on the N-gram-based translation models. Verb simplification requires morphology generation decisions for $\sim$9% of the produced words.

Additionally, combined simplifications consistently yield 1 and 2 additional BLEU points ('S:DAV_{MT}' and 'S:*full*', respectively) over post-processing oracles from Table 6. However, they require simplifying around 32% and 55% of Spanish words, respectively, which demands stronger generation effort. Apart from that, according to the study in Section 3.3, we can expect these oracles to be overly optimistic as the morphology of determiners is also being simplified.

On the other hand, note that English morphology simplification ('+E:*full*') does not produce any relevant oracle improvement. Finally, we observe that morphology simplification also has an effect on target sentence length. In other words, in all experiments the simplified morphology models tend to produce longer output sentences than the post-processing case, or equivalently, use less tuples translating into NULL. This is possibly a by-product of the better estimation of the bilingual model since, when this is sparse, tuples translating into NULL usually get high unigram probability estimates, due to the Kneser–Ney smoothing, which assigns unigram estimates according to the number of previous contexts. Since these tuples originate from unaligned words, they do not tend to follow a pattern and typically occur after many different tuples, causing the output to be shorted when the model is sparse.

From this study, we can conclude that the main source of potential improvement lies in verb form morphology, which negatively affects the bilingual N-gram by 0.7 points absolute BLEU. In contrast to the baseline score, the simplified morphology oracle for verbs (S:V and S:V_{MT}) is nearly 2 points higher. This result is not surprising, since verbs are the morphological category that exhibit more derivation in Spanish and all Romance languages. In the following section, we generate the actual morphology for simplified verbs to obtain final translation scores.

4. Morphology generation

In this section we investigate whether it is feasible to implement morphology generation by means of an independent classification model which, making use of a set of relevant features for each simplified morphology word and its context, generates its appropriate morphology. The approach is evaluated for the two most promising oracles found in the previous section, namely those related to verb form morphology simplification (the S:V and S:V_{MT} schemes).

4.1. Classification of verb forms

For this purpose, we define a collection of context-dependent features and estimate a set of binary AdaBoost classifiers (Schapire and Singer, 1999) for each morphology derivation. In particular, to generate person and number information, we define six binary classifiers (the first one classifying whether the sample is 1st person singular or not, the second one whether the sample is 1st person plural, and so on) and select as output the class which obtains the best classification score according to its binary classifier. For mode and tense, we define eight binary classifiers (for present indicative, past perfect indicative, past imperfect indicative, conditional indicative, future indicative, present subjunctive, etc.).

In our setting, the AdaBoost algorithm combines several small fixed-depth decision trees as base rules (usually of depth 3). Each branch of a tree is, in fact, a conjunction of binary features, allowing the strong adaptative boosting classifier to work with complex and expressive rules. This particular implementation of the boosting technique introduced by Schapire and Singer (1999) was applied successfully to named entity extraction in Carreras et al. (2003).

For our purposes, any machine learning technique capable of inferring a classification strategy from a set of samples and features would have been equally appropriate.

For the features, we consider the following set (in lowercase text only):

– Bilingual model features: current tuple; $n-1$ previous tuples (n being the order of the bilingual model).
– Target language features: target side of the tuple; whether there is a Spanish personal pronoun preceding the verb form, and which one; target verb form without 'tense' information; target verb form without 'mode' information; whether the form precedes the auxiliary verb 'haber' (to detect past participles, as in 'ha decidido').
– Source language features: if there is an English full verb form (including pronoun) in the region of the current tuple, whether it starts and ends in this tuple, or whether it is divided into various tuples; if there is an English verb form, itself, its Part-Of-Speech tag, base form and preceding personal pronoun (if any), whether it is in active or passive voice, whether the verb is followed by a personal pronoun (such as 'gave us'); the last preceding noun in the sentence (if any) and its POS tag (as it can presumably act as subject); and the last preceding verb base form (if any).

All these features can be extracted by a simple set of rules based on word, POS tag and base form information for both languages. Note that we define a broader set of features in the source language (i.e., based on the English text) compared to the target language. This is because during translation, the Spanish output will contain translation errors, which in turn will cause extracted features to be less reliable than those based on the fixed source sentence.

Overall, the average number of active features per verb form is 12. Next, details on the accuracy of the classifiers trained to generate verb form morphology are presented. In all cases, classification models were learned in less than 6 h in a single 3 GHz CPU with insignificant memory usage.

4.2. Classification of person and number

In order to generate morphology for these cases, we distinguish two subcategories. Firstly, those verb forms whose person and number information is missing, and secondly, those for which number and gender is missing (past participle, such as 'decidido', 'decidida', 'decididos', 'decididas', which can also be regarded as adjective).

In the first case, we extract features for all such verb forms in the training data (nearly 1 M samples), and reserve 10 k samples for classification testing. The resulting accuracy depends on the number of training samples considered, as shown in Table 10, which shows the number of training samples, the size of the feature vocabulary (number of different features observed) and the accuracy. Clearly, from 500 k samples, the learning curve is nearing saturation with this feature set.

For the second case, after reserving 10 k samples for classification testing, we find only about 300 k such verb forms in training. By using them all, the feature vocabulary size is 433 k and accuracy is 89.25%. Clearly, the feature that triggers when the Spanish verb form is preceded by the auxiliary verb 'haber' helps distinguish nearly all past perfect cases (as in 'hemos decidido', where the second verb

Table 10
Classification accuracy on Spanish verb person and number morphology information

Train samples	Feature vcb	Accuracy (%)
300 k	437 k	56.96
500 k	648 k	87.36
1 M	1.1 M	87.67

Table 11
Classification accuracy on Spanish verb mode and tense morphology information

Train samples	Feature vcb	Accuracy (%)
300 k	347 k	81.24
500 k	515 k	82.32
1 M	924 k	82.51

is always masculine singular) from adjectival cases (as in'una decisión tomada antes', where 'tomada' depends on the noun 'decisión').

Overall, these classification accuracies are quite high, especially taking into account that there is an inherent ambiguity in translating English verbs with 'you' (such as 'you said'), as there is no possible way to determine whether the subject is singular or plural from a sentence-limited view.

4.3. Classification of mode and tense

Regarding the S:V_{MT} simplification scheme, apart from the previous classifiers, we need to decide mode and tense. In this case, we extract features for all such verb forms in the training data (nearly 1 M samples), and again reserve 10 k samples for classification testing. Again, Table 11 presents the size of the feature vocabulary and the obtained accuracy given the number of training samples.

These results show that learning the adequate mode and tense for a Spanish verb form is a more difficult task, possibly asking for the use of much more complex features. Aiming at improving this accuracy, separate classifiers for only mode (88.93% accuracy) and only tense (90.2% for the five indicative mode tenses, 91.00% for the three subjunctive mode tenses) have also been trained, but the combined final accuracy is just 80.23%.

In the next section we apply these classifiers to generate morphology for the simplified morphology Spanish outputs for the S:V and S:V_{MT} configurations from Section 3.3.

5. Final system

5.1. Translation with morphology generation

Once morphological information has been obtained via Adaboost-based classification, we generate the actual Spanish verb form word by means of a straightforward substitution process. Given the list of possible words and their corresponding POS-tags, we only observe a very few cases that need be disambiguated (related to the Spanish subjunctive imperfect tense, as in 'cantara' or 'cantase'), and yet without any semantic difference. For simplicity, we always pick the first option in the list.

Table 12 shows in bold the final translation results obtained after morphology generation for both post-processing and simplified morphology modeling in the S:V (person and number) and S:V_{MT} (mode, tense, person and number) configurations. For comparison, oracle scores are also shown for each case.

Table 12
Post-processing vs modeling oracle and final translation BLEU scores after morphology generation in the two studied verb simplification configurations

Baseline		Post-processing	Modeling
		47.85	47.85
	oracle	48.33	48.88
S:V	**gen** V	**47.97**	**48.44**
	oracle	49.16	49.72
S:V_{MT}	**gen**V_{MT}	**48.13**	**48.31**

As can be seen, post-processing does not achieve any significant improvement in scores. However, slightly better results than the baseline serve as a sanity check, ensuring that the feature-based morphology generation strategy is not harming current translations.

As for the simplified modeling results, these achieve bigger improvements over the baseline. Interestingly, the result when simplifying mode and tense (S:V_{MT}) does not improve that of S:V, possibly due to the accuracy drop caused by the mode and tense classifiers, as shown in the previous section. The best obtained score is 48.44, a significant 0.6 BLEU increase over baseline and quite close to the 48.88 oracle.

To sum up, we can conclude that it is feasible to generate Spanish verb form morphology as a classification task, which can be carried out by machine learning techniques. The task is well-defined, in the sense that a correlation between classification accuracy on held-out data and final translation scores is observed. Therefore, we can expect that if a more accurate classification technique is used to replace the simple set of statistical binary classifiers used here, results will approximate the oracle scores.

5.2. Interaction with non-monotonic decoding

Recently, it has been shown that word order errors in the English → Spanish task, which were also shown to be important by the manual error analysis in Vilar et al. (2006), can be successfully addressed by means of non-monotonic search (Crego and Mariño, 2006, 2007). This can be done efficiently by extending the monotonic search graph with non-monotonic paths according to the POS-tag sequences found in the input.

In detail, a set of relevant re-ordering rules defined over source-language POS sequences are extracted from word alignments. Then, for each input test sentence, whenever a POS sequence matches one re-ordering rule, a new unweighed non-monotonic path is added to the input graph. Then, the input graph is translated via standard monotonic beam-search decoding, as the non-monotonic paths are already explicitly included in the graph. More

Table 13
Effect of Spanish enclitic pronoun separation on reordering baseline system

Baseline	49.18
+enclitic sep.	49.69

Obtained BLEU scores.

Table 14
Post-processing vs modeling oracle and final translation BLEU scores after morphology generation when non-monotonic decoding is applied

Baseline		Post-processing	Modeling
+ enclitic sep.		49.69	49.69
	oracle	50.29	50.78
S:*V*	**gen*V***	**49.81**	**50.21**

details on this can be found in (Crego and Mariño, 2006, 2007). By applying these techniques to the N-gram-based system, our baseline scores increase to 49.69 BLEU.

Finally, in this section we study whether the above morphology simplification techniques interact with reordering strategies. Given the experimental architecture defined in Section 3, we just replace the previous translation system by its non-monotonic version and conduct the same morphology simplification experiment.

5.2.1. Enclitic pronoun separation

Table 13 compares the resulting translation scores when applying the enclitic pronouns separation strategy proposed in Section 3.2 to the reordering baseline described above.

Interestingly, even though this strategy did not produce large gains in a monotonic decoding framework, this time it yields a significant 0.5 BLEU increase, even though only approximately 0.4% of the output words actually have an enclitic pronoun. It seems clear that there is a positive interaction between this technique and the non-monotonic search, which ends up generating better statistical models and possibly outputting new Spanish words via joining verb forms with subsequent pronouns that were never observed as a training sequence before.

5.2.2. Morphology simplification and posterior classification

Finally, we would like to evaluate simplified morphology translation models in a non-monotonic decoding framework. For this experiment, we work on the S:*V* configuration (person and number for verbs), which provided the best results in the monotonic case (see Section 4). Even though this does not necessarily mean that it will also achieve the best result when compared to other simplification schemes in non-monotonic decoding, we believe it is reasonable to choose it if we take into account that verbs are the major source of morphology variation in Spanish. Therefore, their impact on this experiment will be crucial to determine the potential of the approach.

We apply this morphology simplification to the previous system with non-monotonic search and enclitic pronoun separation. In this case, the translation model is based on a bilingual N-gram of unfolded tuples with reordered source, which modifies the features used in verb morphology classification. However, that does not affect the proposed architecture for morphology estimation and generation via machine learning.

Table 14 shows in bold the final translation results with re-ordering, comparing both post-processing and simplified morphology modeling in the S:*V* (person and number) configuration. Again, the oracle scores are also shown for each case.

As can be seen, the approximate 0.5 BLEU difference in score between post-processing and modeling oracles persists. Clearly, this indicates that successfully addressing word order problems does not necessarily imply an improvement in morphology generation, a challenge which has a different cause (and therefore requires different solutions). Undoubtedly, Spanish morphology equally harms both monotonic and non-monotonic statistical models.

Nevertheless, the final translation scores obtained through machine-learning morphology generation continue to provide a statistically significant BLEU improvement over baseline scores (even when these incorporate enclitic pronoun separation). Therefore, we can conclude that the proposed simplified morphology modeling is independent from monotonic or non-monotonic search, yielding the same performance boost. This is because it tackles the specific problems related to Spanish morphology when translating from English, namely data sparseness and morphology generation.

In summary, the proposed techniques for Spanish enclitic pronoun separation and verb morphology simplification coupled with Adaboost-based generation provides an overall 1-point BLEU improvement in this English → Spanish large-vocabulary task, over a state-of-the-art statistical machine translation system with reordering capabilities.

Finally, Table 15 shows examples of two sentences in our test set, including the input English sentence, the baseline and translations and gen*V* for both monotonic and non-monotonic decoding, and the first Spanish reference translation. Translation errors are marked in italics, whilst improvements in gen*V* with respect to the baseline are marked in bold face. As we can see, various Spanish verb forms correct their agreement with the subject of the sentence, although that does not necessarily match the reference translation. Additionally, as the translation output is longer, some omissions are also corrected.

6. Conclusion

This paper is devoted to studying the impact of the target morphology when statistically translating from English into Spanish. We prove that a certain degree of morphology-related information is not only being ignored by current statistical translation models, but also affects them negatively at training time by creating sparseness.

Table 15
Two translation examples comparing baseline vs. gen*V* in monotonic and non-monotonic decoding

English input		Mr President, I ask you to give this your utmost priority, because, as Annetta Flanigan said, they only went there to help the people of Afghanistan
monotonic	baseline	Señor Presidente, le pido que su máxima prioridad, porque, como Annetta Flanigan, solo *ha* ido allí para ayudar al pueblo de Afganistán
	gen*V*	Señor Presidente, le pido que dé su máxima prioridad, porque, como Annetta Flanigan, sólo **han** ido allí para ayudar al pueblo de Afganistán
non-monotonic	baseline	Señor Presidente, quiero pedirle que su máxima prioridad, porque, como ha dicho, sólo *ha* ido allí para ayudar a la población de Afganistán
	gen*V*	Señor Presidente, le ruego que dé su máxima prioridad, porque, como decía **Annetta Flanigan**, sólo **han** ido allí para ayudar al pueblo de Afganistán
Reference		Sr. Presidente, le pido que ésta sea su mayor prioridad, porque, como dijo Annetta Flanigan, sólo fueron allí para ayudar al pueblo de Afganistán
English input		This is brutality for which there is huge political responsibility both on the part of the European Union, which legalised the war, and the occupying forces, and the political groups, which approved the resolution
monotonic	baseline	Esto es brutalidad *que la* gran responsabilidad política por parte de la UE, que *legalizaron* la guerra, y las fuerzas de ocupación, y los Grupos políticos, que *aprobó* la resolución
	gen*V*	Esta es *la* brutalidad de *los* que hay gran responsabilidad política por parte de la UE, que **legaliz**ó la guerra, y las fuerzas de ocupación, y los Grupos políticos, que **aprobaron** la resolución
non-monotonic	baseline	Esta es *la* brutalidad de gran responsabilidad política tanto de la UE, que *legalizaron* la guerra, *y* las fuerzas de ocupación, y los Grupos políticos, que *aprobó* la resolución
	gen*V*	Esta es *la* brutalidad de *los* que hay una gran responsabilidad política tanto de la UE, que **legaliz**ó la guerra, *y* las fuerzas de ocupación, y los Grupos políticos, que **aprobaron** la resolución
Reference		Esto es una crueldad porque hay una enorme responsabilidad política, tanto por parte de la UE, que legalizó la guerra, como por parte de las fuerzas de ocupación y los partidos políticos, que aprobaron la resolución

In fact, in this paper this impact has been quantified, and the main Spanish morphological features causing it have been pointed out (verb form derivation). This is done by comparing the obtained post-processing oracle scores with those obtained from simplified morphology translation models for a set of morphology simplification schemes.

On the other hand, we show that some Spanish morphology information can be introduced into simplified morphology translation hypotheses by means of an independent model. For this, we defined a set of relevant features for each Spanish verb base form, and trained statistical classifiers based on machine learning techniques. This classification task has been evaluated independently and high accuracy scores have been obtained when generating Spanish verb person, number and gender information. This has resulted in statistically significant improvements in final translation scores.

Additionally, we tackled the issue of Spanish enclitic pronouns which are attached to verbs as suffixes, and propose a simple separation technique that can be undone when the Spanish output is produced.

Finally, the last part of the paper was devoted to investigating the interaction between the proposed morphology-based techniques and word order models, proving that overall gains due to correct morphology processing are higher with non-monotonic search. This is due to enclitic separation, as morphology simplification seems independent from word re-ordering.

The findings of this work set two clear paths for further study. Firstly, classification accuracy needs to be improved in order to better generate Spanish morphology when translating from English, nearing final scores which are closer to the obtained oracles. This could be done both by using more complex features or by applying alternative methods for classification, such as support vector machines.

Secondly, further improvements could potentially be achieved through a tighter integration between the translation decoder and the morphology generation module. In the short term, this could be explored by generating morphology for a set of N-best translation hypotheses instead of the single 1-best solution, and performing a posterior hypothesis rescoring based on Spanish language models. Finally, a more ambitious line of research is to investigate ways of incorporating feature-based classification directly in translation decoding, including morphology as an additional feature into the search.

Acknowledgements

The authors would especially like to thank Lluís Márquez (UPC, Barcelona) for providing the machine learning toolkit implementing Adaboost that has been used for morphology generation experiments, and Josep Maria Crego (UPC, Barcelona) and Graeme Blackwood (University of Cambridge) for their most helpful contribution. We would also like to give credit to the anonymous reviewers of this paper, who helped improve the quality of the presentation with their insightful suggestions.

This work has been funded by the European Union under the integrated project TC-STAR – (IST-2002-FP6-5067-38) and by the Spanish Government under the project AVIVAVOZ – (TEC2006-13694-C03).

References

Al-Onaizan, Y., Curin, J., Jahr, M., Knight, K., Lafferty, J., Melamed, D., Och, F., Purdy, D., Smith, N., Yarowsky, D., 1999. Statistical machine translation: final report. Tech. Rep., Johns Hopkins University Summer Workshop, Baltimore, MD, USA.

Brants, T., 2000. TnT – a statistical part-of-speech tagger. In: Proceedings of the Sixth Applied Natural Language Processing (ANLP-2000). Seattle, WA.

Carreras, X., Chao, I., Padró, L., Padró, M., May 2004. Freeling: an open-source suite of language analyzers. In: 4th International Conference on Language Resources and Evaluation, LREC'04.

Carreras, X., Màrquez, L., Padró, L., 2003. A simple named entity extractor using adaboost. In: Proceedings of CoNLL-2003. Edmonton, Canada, pp. 152–155.

Casacuberta, F., Vidal, E., 2004. Machine translation with inferred stochastic finite-state transducers. Computational Linguistics 30 (2), 205–225.

Collins, M., Koehn, P., Kucerova, I., 2005. Clause restructuring for statistical machine translation. In: 43rd Annual Meeting of the Association for Computational Linguistics, pp. 531–540.

Corston-Oliver, S., Gamon, M., 2004. Normalizing German and English inflectional morphology to improve statistical word alignment. In: Proceedings of the 6th Conference of the Association for Machine Translation in the Americas, pp. 48–57.

Crego, J., de Gispert, A., Lambert, P., Costa-jussà, M., Khalilov, M., Banchs, R., Mari no, J., Fonollosa, J., 2006. N-gram-based SMT system enhanced with reordering patterns. In: Proceedings of the Workshop on Statistical Machine Translation. Association for Computational Linguistics, New York, pp. 162–165.

Crego, J., Mariño, J., 2006. Reordering experiments for N-gram-based SMT. In: 1st IEEE/ACL Workshop on Spoken Language Technology.

Crego, J., Mariño, J., 2007. Improving statistical MT by coupling reordering and decoding. In: Machine Translation, vol. 21. Springer.

Crego, J., Mariño, J., de Gispert, A., 2005. An Ngram-based statistical machine translation decoder. In: Proceedings of the 9th European Conference on Speech Communication and Technology, Interspeech'05, pp. 3193–3196.

de Gispert, A., Mariño, J., Crego, J., 2005. Improving statistical machine translation by classifying and generalizing inflected verb forms. In: Proceedings of the 9th European Conference on Speech Communication and Technology, Interspeech'05, pp. 107–114.

Goldwater, S., McClosky, D., 2005. Improving statistical MT through morphological analysis. In: Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing, HLT/EMNLP'05. Association for Computational Linguistics, Vancouver, Canada, pp. 676–683.

Gupta, D., Federico, M., 2006. Exploiting word transformation in statistical machine from spanish to english. In: Proceedings of 11th Annual Conference of the European Association for Machine Translation, EAMT'06.

Habash, N., Sadat, F., 2006. Arabic preprocessing schemes for statistical machine translation. In: Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers. Association for Computational Linguistics, New York City, USA, pp. 49–52.

Hutchins, W., Somers, H., 1992. An Introduction to Machine Translation. Academic Press, London, UK.

Kneser, R., Ney, H., 1995. Improved backing-off for m-gram language modeling. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing 1, pp. 181–184.

Koehn, P., Hoang, H., 2007. Factored translation models. In: Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP/CoNNL'07. Association for Computational Linguistics, Prague, Czech Republic, pp. 868–876.

Lee, Y., 2004. Morphological analysis for statistical machine translation. In: Susan Dumais, D.M., Roukos, S. (Eds.), HLT-NAACL 2004: Short Papers. Association for Computational Linguistics, Boston, Massachusetts, USA, pp. 57–60.

Mariño, J., Banchs, R., Crego, J., de Gispert, A., Lambert, P., Fonollosa, J., Costa-jussà, M., 2006a. N-gram based machine translation. Computational Linguistics 32 (4), 527–549.

Mariño, J., Banchs, R., Crego, J., de Gispert, A., Lambert, P., Fonollosa, J., Costa-jussà, M., Khalilov, M., 2006b. UPC's bilingual N-gram translation system. In: TC-STAR Workshop on Speech-to-Speech Translation. Barcelona, Spain, pp. 43–48.

Nießen, S., Ney, H., 2000. Improving SMT quality with morpho-syntactic analysis. In: Proceedings of the 18th International Conference on Computational Linguistics, COLING'00, pp. 1081–1085.

Nießen, S., Ney, H., 2004. Statistical machine translation with scarce resources using morpho-syntactic information. Computational Linguistics 30 (2), 181–204.

Popovic, M., de Gispert, A., Gupta, D., Lambert, P., Ney, H., Mariño, J., Federico, M., Banchs, R., 2006. Morpho-syntactic information for automatic error analysis of statistical machine translation output. In: Proceedings of the Workshop on Statistical Machine Translation. Association for Computational Linguistics, New York, pp. 1–6.

Popovic, M., Ney, H., 2004. Towards the use of word stems and suffixes for statistical machine translation. In: 4th International Conference on Language Resources and Evaluation, LREC'04, pp. 1585–1588.

Popovic, M., Ney, H., 2006. Error analysis of verb inflections in spanish translation output. In: TC-STAR Workshop on Speech-to-Speech Translation. Barcelona, Spain, pp. 99–103.

Schapire, R.E., Singer, Y., 1999. Improved boosting using confidence-rated predictions. Machine Learning 37 (3), 297–336.

Talbot, D., Osborne, M., 2006. Modelling lexical redundancy for machine translation. In: Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Sydney, Australia, pp. 969–976.

Ueffing, N., Ney, H., 2003. Using POS information for SMT into morphologically rich languages. In: 10th Conference of the European Chapter of the Association for Computational Linguistics, pp. 347–354.

Vilar, D., Xu, J., D'Haro, L., Ney, H., 2006. Error analysis of statistical machine translation output. In: 5th International Conference on Language Resources and Evaluation, LREC'06, pp. 697–702.

Zollmann, A., Venugopal, A., Vogel, S., 2006. Bridging the inflection morphology gap for arabic statistical machine translation. In: Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers. Association for Computational Linguistics, New York City, USA, pp. 201–204.