EUSMT: Incorporating Linguistic Information into SMT for a Morphologically Rich Language.

Its use in SMT-RBMT-EBMT hybridization

PhD. Candidate: Gorka Labaka Intxauspe
Supervisors: Arantza Díaz de Ilarraza Sánchez
Kepa Sarasola Gabiola

Lengoaia eta Sistema Informatikoak/Lenguajes y Sistemas Informáticos
Euskal Herriko Unibertsitatea/Universidad del País Vasco

March 29, 2010

Basque Language

- Basque is a pre-Indo-European language [Trask, 1997] with no demonstrable genealogical relationship with other languages.
- There have been many unsuccessful attempts to relate Basque to other languages (Caucasian, Iberian, Berber).
- Most of the features present in Basque (agglutinative, ergative case system) are not unique, **but their combination makes Basque a real challenge for Human Language Technologies (HLT)**.

Sociological Status

- There are few fluent speakers of Basque.
- Basque speakers are distributed between Spain and France and it is in a diglossic situation in all its territories.
- There are not many linguistic resources for Basque:
  - Few corpora, both parallel and monolingual.
  - Syntactic and semantic processors are still on development.
  - But high quality morphological processors (analyzer and generator).
- **This mentioned lack of resources makes the application of HLT and Machine Translation even harder**.

Machine Translation for Basque

- Due to the co-official language status of Basque in some Spanish regions, many administrative texts have to be translated.
- Spanish-to-Basque translation is a real need.
- The Ixa group has already developed a Rule-Based Machine Translation system [Mayor, 2007], and attempts on EBMT have been also done [Alegria et al., 2008b].
- During the last years some SMT attempts have been developed by different authors [Sanchís and Casacuberta, 2007], [Pérez et al., 2008]. Most of them based on Stochastic finite-state transducers and synthetic corpora.
- Other RBMT systems: *Erderatu* [Ginestí-Rosell et al., 2009] or the system available in the website of the *Instituto Cervantes* (http://oesi.cervantes.es/traduccionAutomatica.html).

Objectives of this PhD thesis

Adaptation of SMT to Basque & First Hybridization Attempts

Objectives of this PhD thesis

Adaptation of SMT to Basque & First Hybridization Attempts

1. To deal with the agglutinative nature of Basque
   - [Agirre et al., 2006] -SEPLN 2006
   - [Labaka et al., 2007] - MT Summit 2007
   - [Labaka et al., 2008] - JTH 2008
   - [Díaz de Ilarraza et al., 2009] - EAMT 2008

Objectives of this PhD thesis

Adaptation of SMT to Basque & First Hybridization Attempts

1. To deal with the agglutinative nature of Basque
2. To implement different techniques to deal with word order differences in SMT
   - [Díaz de Ilarraza et al., 2009] - MT Summit 2009

Objectives of this PhD thesis

Adaptation of SMT to Basque & First Hybridization Attempts

1. To deal with the agglutinative nature of Basque
2. To implement different techniques to deal with word order differences in SMT
3. To combine by means of a Multi-Engine system SMT with previously developed RBMT and EBMT systems
   - [Alegria et al., 2008a] - MATMT 2008
   - [Alegria et al., 2008b] - AMTA 2008

Objectives of this PhD thesis

Adaptation of SMT to Basque & First Hybridization Attempts

1. To deal with the agglutinative nature of Basque
2. To implement different techniques to deal with word order differences in SMT
3. To combine by means of a Multi-Engine system SMT with previously developed RBMT and EBMT systems
4. To use of SMT for automatic post-edition of RBMT translations.
   - [Díaz de Ilarraza et al., 2008] - MATMT 2008

Objectives of this PhD thesis

Adaptation of SMT to Basque & First Hybridization Attempts

1. To deal with the agglutinative nature of Basque
2. To implement different techniques to deal with word order differences in SMT
3. To combine by means of a Multi-Engine system SMT with previously developed RBMT and EBMT systems
4. To use of SMT for automatic post-edition of RBMT translations.
5. To collect larger bilingual corpora and measure the impact of the size and nature of the corpora on the different techniques developed.

Objectives of this PhD thesis

Adaptation of SMT to Basque & First Hybridization Attempts

1. To deal with the agglutinative nature of Basque
2. To implement different techniques to deal with word order differences in SMT
3. To combine by means of a Multi-Engine system SMT with previously developed RBMT and EBMT systems
4. To use of SMT for automatic post-edition of RBMT translations.
5. To collect larger bilingual corpora and measure the impact of the size and nature of the corpora on the different techniques developed.
6. To carry out a final evaluation of the work done in this thesis.

Outline

1. General experimental setup

Outline

1. General experimental setup

2. Treatment of the morphological divergence between Spanish and Basque

Outline

1. General experimental setup
2. Treatment of the morphological divergence between Spanish and Basque
3. Treatment of the syntactic divergence between Spanish and Basque

Outline

1. General experimental setup
2. Treatment of the morphological divergence between Spanish and Basque
3. Treatment of the syntactic divergence between Spanish and Basque
4. Hybridization attempts

Outline

1. General experimental setup
2. Treatment of the morphological divergence between Spanish and Basque
3. Treatment of the syntactic divergence between Spanish and Basque
4. Hybridization attempts
5. Overall evaluation

Outline

1. General experimental setup
2. Treatment of the morphological divergence between Spanish and Basque
3. Treatment of the syntactic divergence between Spanish and Basque
4. Hybridization attempts
5. Overall evaluation
6. Contributions and Further Work

Outline

1. General experimental setup
2. Treatment of the morphological divergence between Spanish and Basque
3. Treatment of the syntactic divergence between Spanish and Basque
4. Hybridization attempts
5. Overall evaluation
6. Contributions and Further Work

Statistical Machine Translation

- We develop our systems using freely available tools (Moses, GIZA and SRILM)
- We use the same feature combination in all our experiments:
    - phrase translation probabilities (in both directions)
    - word-based translation probabilities (in both directions)
    - a phrase length penalty
    - a 4-gram target language model
    - lexicalized reordering (except on those cases where we specifically deactivate it)

Parallel corpus for Basque: Consumer

		sentence	tokens	vocabulary	singletons
training	Spanish	58,202	1,284,089	46,636	19,256
	Basque		1,010,545	87,763	46,929
development	Spanish	1,456	32,740	7,074	4,351
	Basque		25,778	9,030	6,339
test	Spanish	1,446	31,002	6,838	4,281
	Basque		24,372	8,695	6,077

Table: Some statistics of the corpus (Eroski Consumer).

- It is a collection of 1036 articles written in Spanish Consumer Eroski magazine, along with their Basque, Catalan and Galician translations.
- It contains more than 1,200,000 Spanish words and more than 1,000,000 Basque words.
- It was automatically aligned at sentence level [Alcázar, 2005].
- We have divided this corpus into three sets: training, development and test.

Evaluation of the machine translation

- In order to assess the quality of the systems developed in this thesis, we used metrics that compare the translation with human references.
- Accuracy metrics based on n-grams (higher values imply higher translation quality):
  - BLEU [Papineni et al., 2002]
  - NIST [Doddington, 2002]
- Error metrics (lower values imply higher translation quality).
  - Word Error Rate (WER) [Nießen et al., 2000]
  - Position-independent word Error Rate (PER) [Tillmann et al., 1997]
- Statistical Significance test by means of Paired Bootstrap Resampling [Koehn, 2004].

Outline

1. General experimental setup
2. Treatment of the morphological divergence between Spanish and Basque
   - Use of segmentation to adapt SMT to Basque
   - Different segmentation options
   - Experimental Results
3. Treatment of the syntactic divergence between Spanish and Basque
4. Hybridization attempts
5. Overall evaluation
6. Contributions and Further Work

Morphological divergences between Spanish and Basque

- Basque is agglutinative: words are formed by joining several morphemes together:
  - Each postpositional case has four different variants.
  - For a lemma more than 360 forms are possible.
  - In the case of ellipsis more than one suffix can be added to the same lemma, increasing the word forms that can be generated from a lemma.
- Postpositions are added to the last word of each phrase.

Basque morphological generation

etxe	*/house/*			
	etxea	*/the house/*		
	etxeak	*/the houses/*		
	etxeok	*/these houses/*		
	[edozein] etxetara	*/to [any] house/*		
	etxera	*/to the house/*		
	etxeetara	*/to the houses/*		
	etxeotara	*/to these houses/*		
	etxeko	*/of the house/*		
			etxekoa	*/the one of the house/*
			etxekoak	*/the ones of the house/*
			...	
	etxeetako	*/of the houses/*		
			etxeetakoa	*/the one of the houses/*
			etxeetakoak	*/the ones of the houses/*
			...	
	etxeotako	*/of these houses/*		
			etxeotakoa	*/the one of these houses/*
	...			

Figure: Illustration of the Basque inflectional morphology.

Effect of morphology in the translation

- Sparseness (each Basque word appears few times in the corpus).
- Being Basque less-resourced, the sparseness problem is intensified.
- The agglutinative nature of Basque causes many 1:n alignments. Those alignments, even being allowed in the IBM models, harm the alignment quality.

	tokens	**vocabulary**	**singletons**
Spanish	1,284,089	46,636	19,256
Basque	1,010,545	87,763	46,929

Table: Figures on the Consumer training corpus.

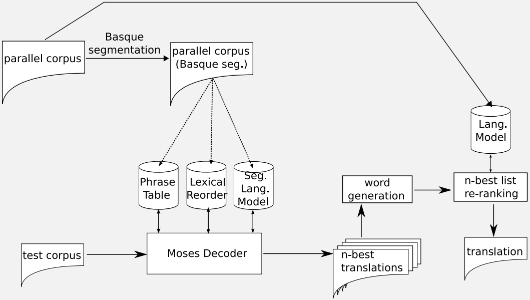
Different approaches for other highly inflected languages

- **Segmentation**. Words of the highly inflected languages are divided into several tokens [Goldwater and McClosky, 2005], [Oflazer and El-Kahlout, 2007], [Ramanathan et al., 2008].
- **Factored models**. Each word is tagged at different linguistic levels. Each level can be translated independently [Koehn and Hoang, 2007], [Bojar, 2007].
- **Morphology generation model**. The translation is carried out into target lemmas, and, then, their inflection is decided in a separated generation step [Minkov et al., 2007], [Toutanova et al., 2008], [Pérez et al., 2008].

Selected approach: Morphological segmentation

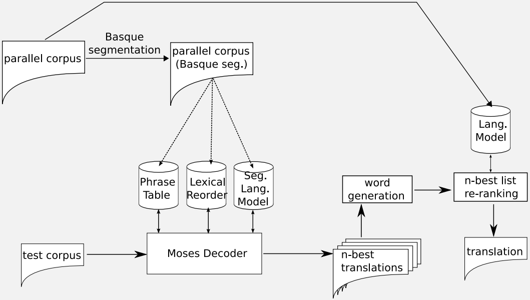
- Taking into account the work done for other highly inflected languages, we have chosen segmentation in order to adapt SMT to Basque.
  - High-precision morphological analyzer and generator are available for Basque.
  - The use of segmentation allows the generation of unseen words, unlike the factored model and the morphology generation model.
  - Complex translation steps make factored translation computationally unmanageable.
  - The biggest gains using factored models come from the incorporation of language models on different factors (lemmas, PoS or morphological information). This can also be combined with the segmentation.

Use of segmentation to adapt SMT to Basque



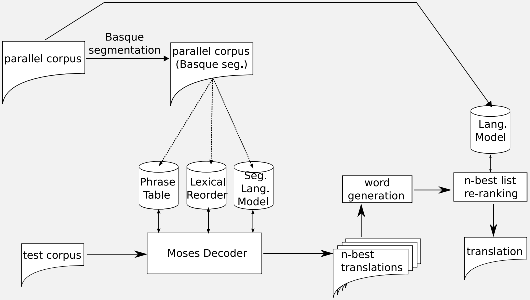
- Basque text is segmented before training, dividing each word into a set of tokens.

Use of segmentation to adapt SMT to Basque



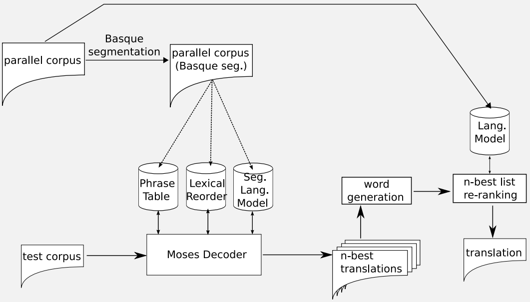
- Basque text is segmented before training, dividing each word into a set of tokens.
- An SMT system is trained over the segmented text.

Use of segmentation to adapt SMT to Basque



- Basque text is segmented before training, dividing each word into a set of tokens.
- An SMT system is trained over the segmented text.
- After translation, the final Basque word has to be generated. At generation, Basque morpho-phonologic rules have to be taken into account.

Use of segmentation to adapt SMT to Basque



- Basque text is segmented before training, dividing each word into a set of tokens.
- An SMT system is trained over the segmented text.
- After translation, the final Basque word has to be generated. At generation, Basque morpho-phonologic rules have to be taken into account.
- No word-level language model is used at decoding. It is incorporated by means of n-best lists.

Eustagger segmentation

- We based our segmentation of the analysis obtained by the Eustagger Basque lemmatizer [Aduriz and Díaz de Ilarraza, 2003].
- Straightforward segmentation, creating a new token for each morpheme recognized by Eustagger.
- We compare the performance of this segmentation with a baseline (out-of-the-box Moses trained on the tokenized corpus).

Eustagger segmentation

- We based our segmentation of the analysis obtained by the Eustagger Basque lemmatizer [Aduriz and Díaz de Ilarraza, 2003].
- Straightforward segmentation, creating a new token for each morpheme recognized by Eustagger.
- We compare the performance of this segmentation with a baseline (out-of-the-box Moses trained on the tokenized corpus).
- Automatic evaluation metrics did not show significant improvement. Worst BLEU scores, slightly better for the rest of the metrics.

	BLEU	NIST	WER	PER
Baseline	**10.78**	4.52	80.46	61.34
Eustagger segm.	10.52	**4.55**	**79.18**	**61.03**

Table: Evaluation of SMT systems.

Different segmentation options

- The lexicon of the Eustagger analyzer is too fine-grained.
- It defines morphemes according to the linguistic theories.
- This fine-grained morpheme definition does not agree with the functional usage.
- We conclude that, in case of using the segmentation, it is very important the way that the segmentation is carried out.

Different segmentation options

- We look for the best segmentation based on the analysis obtained by Eustagger.
- We define different ways to group the morphemes, giving rise to different segmentation options:
  1. **OneSuffix**: Groups all suffixes in a unique token.
  2. **AutoGrouping**: Groups those morpheme pairs scored over a threshold according to Pairwise Mutual Information.
  3. **ManualGrouping**: Morphemes are grouped according to hand-defined heuristics.

Different segmentation options

- We look for the best segmentation based on the analysis obtained by Eustagger.
- We define different ways to group the morphemes, giving rise to different segmentation options:
  1. **OneSuffix**: Groups all suffixes in a unique token.
  2. **AutoGrouping**: Groups those morpheme pairs scored over a threshold according to Pairwise Mutual Information.
  3. **ManualGrouping**: Morphemes are grouped according to hand-defined heuristics.

Original word:	aukeratzerakoan /*when choosing*/				
Analysis:	aukeratu+<adize>+<ala>+<gel>+<ine> aukeratu+tze +ra +ko +an				
Eustagger segm.:	aukeratu	+<adize>	+<ala>	+<gel>	+<ine>

Different segmentation options

- We look for the best segmentation based on the analysis obtained by Eustagger.
- We define different ways to group the morphemes, giving rise to different segmentation options:
  1. **OneSuffix**: Groups all suffixes in a unique token.
  2. **AutoGrouping**: Groups those morpheme pairs scored over a threshold according to Pairwise Mutual Information.
  3. **ManualGrouping**: Morphemes are grouped according to hand-defined heuristics.

Original word:	aukeratzerakoan /*when choosing*/
Analysis:	aukeratu+<adize>+<ala>+<gel>+<ine> aukeratu+tze +ra +ko +an
OneSuffix:	aukeratu +<adize>+<ala>+<gel>+<ine>

Different segmentation options

- We look for the best segmentation based on the analysis obtained by Eustagger.
- We define different ways to group the morphemes, giving rise to different segmentation options:
  1. **OneSuffix**: Groups all suffixes in a unique token.
  2. **AutoGrouping**: Groups those morpheme pairs scored over a threshold according to Pairwise Mutual Information.
  3. **ManualGrouping**: Morphemes are grouped according to hand-defined heuristics.

Original word:	aukeratzerakoan /*when choosing*/
Analysis:	aukeratu+<adize>+<ala>+<gel>+<ine> aukeratu+tze +ra +ko +an
AutoGrouping:	aukeratu +<adize>+<ala> +<gel> +<ine>

Different segmentation options

- We look for the best segmentation based on the analysis obtained by Eustagger.
- We define different ways to group the morphemes, giving rise to different segmentation options:
  1. **OneSuffix**: Groups all suffixes in a unique token.
  2. **AutoGrouping**: Groups those morpheme pairs scored over a threshold according to Pairwise Mutual Information.
  3. **ManualGrouping**: Morphemes are grouped according to hand-defined heuristics.

Original word:	aukeratzerakoan /*when choosing*/				
Analysis:	aukeratu+<adize>	+<ala>	+<gel>	+<ine>	
	aukeratu+tze	+ra	+ko	+an	
ManualGrouping:	aukeratu+<adize>	+<ala>+<gel>+<ine>			

Experimental results: Different segmentations

	BLEU	NIST	WER	PER
Baseline	10.78	4.52	80.46	61.34
Eustagger segm.	10.52	4.55	79.18	61.03
OneSuffix segm.	11.24	**4.74**	**78.07**	**59.35**
AutoGrouping segm.	11.24	4.66	79.15	60.42
ManualGrouping segm.	**11.36**	4.67	78.92	60.23

Table: Evaluation of SMT systems with five different segmentation options.

- All the segmentations that group morphemes outperform both the baseline and the Eustagger segmentation.
- There are not big differences between grouping techniques, but according to BLEU the **improvement of the ManualGrouping segmentation is statistically significant over the others**.

Experimental results: Vocabulary size vs. BLEU score

Segmentation option	Running tokens	Vocabulary size	BLEU
Tokenized Spanish	1,284,089	46,636	-
Tokenized Basque	1,010,545	87,763	10.78
Eustagger segm.	1,699,988	35,316	10.52
AutoGrouping segm.	1,580,551	35,549	11.24
OneSuffix segm.	1,558,927	36,122	11.24
ManualGrouping segm.	1,546,304	40,288	**11.36**

Table: Correlation between token number in the training corpus and BLEU evaluation results

- There seems to be a correlation between the size of the vocabulary generated after segmentation and the BLEU score:
  - The closer the size of the vocabularies the bigger the obtained BLEU score.

Outline

Syntactic divergences between Spanish and Basque.

- The order of sentence constituents is very flexible, and mainly depends on the focus.
- Basque mainly follows the SOV sentence order.
- Spanish prepositions have to be translated into Basque postpositions (at the end of the phrase).
- Postpositional phrases attached to nouns are placed before nouns (instead of following them).

Effect of those divergences in the translation.

- SMT systems mainly follow a distance-based distortion method (both in word alignment and decoding).
- This method favour short-distance reordering, strongly penalize long-distance reordering.
- Spanish-to-Basque translation needs a high amount of long-distance reordering, and, as we will see, distance-based reordering produces worse translations.

Different approaches used in the literature

- **Lexicalized reordering**: reordering method integrated in Moses [Koehn et al., 2007].
- Methods based on pre-processing: they modify word order in source language to harmonize it with the target language's word order.
    - **Syntax-based**: based on source syntactic analysis and hand-defined reordering rules [Collins et al., 2005], [Popović and Ney, 2006], [Ramanathan et al., 2008].
    - **Statistical reordering**: based on word alignments and pure statistical information [Chen et al., 2006, Zhang et al., 2007, Sanchís and Casacuberta, 2007, Costa-Jussà and Fonollosa, 2006].

Moses' Lexicalized Reordering

- Reordering method implemented in Moses [Koehn et al., 2007].
- It adds new features to the log-linear framework.
- The orientation of each phrase occurrence is extracted at training, and their probability distribution is estimated.
- Those probability distributions are used to score each translation hypothesis at decoding.

Moses' Lexicalized Reordering: Possible Orientations

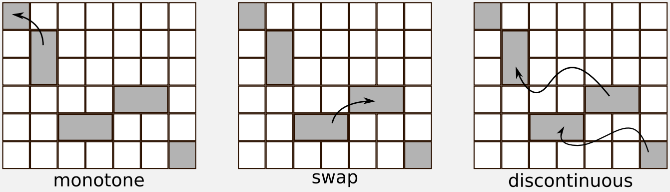
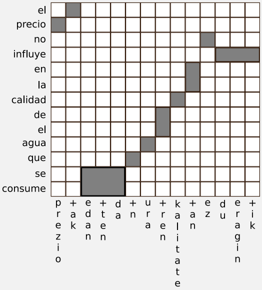


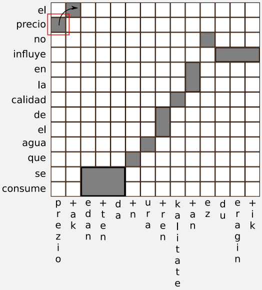
Figure: Possible orientations of phrases defined on the lexicalized reordering

- Three different orientations are defined:
  - **monotone**: continuous phrases occur in the same order in both languages. There is an alignment point to the top left.
  - **swap**: continuous phrases are swapped in the target language. There is an alignment point to the top right.
  - **discontinuous**: continuous phrases in the source language are not continuous in the target language. No alignment points to the top left or the top right.

Moses' Lexicalized Reordering: Training Example

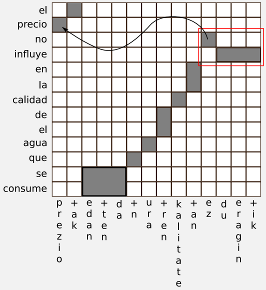


Moses' Lexicalized Reordering: Training Example



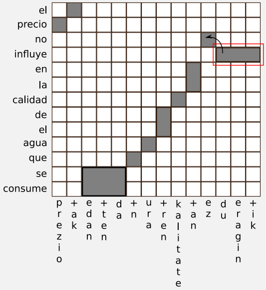
			mon.	swap	disc.
/prize/	precio	prezio	0.01	**0.79**	0.20

Moses' Lexicalized Reordering: Training Example



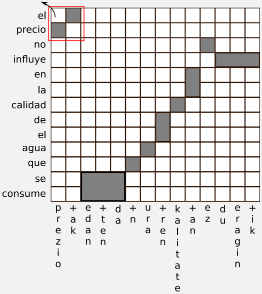
			mon.	swap	disc.
/prize/	precio	prezio	0.01	**0.79**	0.20
/does not influence/	no influye	ez du eragin +nik	0.20	0.20	**0.60**

Moses' Lexicalized Reordering: Training Example



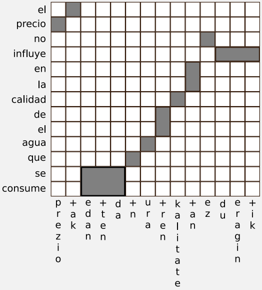
			mon.	swap	disc.
/prize/	precio	prezio	0.01	**0.79**	0.20
/does not influence/	no influye	ez du eragin +nik	0.20	0.20	**0.60**
/influence/	influye	du eragin +nik	**0.60**	0.20	0.20

Moses' Lexicalized Reordering: Training Example



			mon.	swap	disc.
/prize/	precio	prezio	0.01	**0.79**	0.20
/does not influence/	no influye	ez du eragin +nik	0.20	0.20	**0.60**
/influence/	influye	du eragin +nik	**0.60**	0.20	0.20
/the price/	el precio	prezio +ak	**0.17**	0.43	0.40

Moses' Lexicalized Reordering: Training Example



			mon.	swap	disc.
/prize/	precio	prezio	0.01	**0.79**	0.20
/does not influence/	no influye	ez du eragin +nik	0.20	0.20	**0.60**
/influence/	influye	du eragin +nik	**0.60**	0.20	0.20
/the price/	el precio	prezio +ak	**0.17**	0.43	0.40
/not/	no	ez	0.30	0.10	**0.60**
/does not influence in the/	no influye en la	+an ez du eraginik	0.08	**0.79**	0.13
/in the/	en la	+an	0.01	**0.83**	0.16
/in the quality/	en la calidad	kalitate +an	0.04	**0.56**	0.40
/in the quality of the/	en la calidad de el	+ren kalitate +an	0.14	**0.71**	0.15
/quality of the water/	calidad de el agua	ura +ren kalitate	0.01	**0.31**	0.68
/quality of the water that/	calidad de el agua que	+n ura +ren kalitate	0.03	**0.86**	0.11

Syntax-Based Reordering

- This method tries to reorder the source sentence before SMT translation, harmonizing the source word order to the target one.
- To reorder the source, we defined a set of rules that make use of syntactic analysis.
- Those rules have been defined to deal with the most important word order differences between both languages.
- They are divided into two sets: local reordering and long-range reordering

Syntax-Based Reordering: Local Reordering

- Deals with word order differences in phrases (Spanish noun and prepositional phrases).
- Uses Freeling [Carreras et al., 2004] to mark each word's PoS and phrase boundaries.
- Moves Spanish prepositions and articles to the end of the phrase, where Basque postpositions appear.

/the/ /price/	/no/	/has-influence/	/on/ /the/ /quality/	/of/ /the/ /water	/that/	/is/ /consumed/
El precio	no	influye	en la calidad	de el agua	que	se consume

Syntax-Based Reordering: Local Reordering

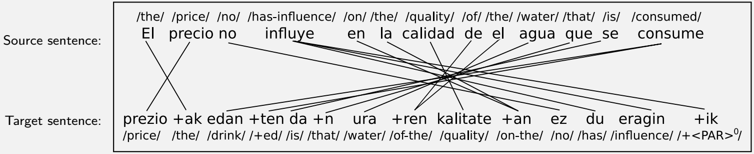
- Deals with word order differences in phrases (Spanish noun and prepositional phrases).
- Uses Freeling [Carreras et al., 2004] to mark each word's PoS and phrase boundaries.
- Moves Spanish prepositions and articles to the end of the phrase, where Basque postpositions appear.

/the/ /price/	/no/	/has-influence/	/on/ /the/ /quality/	/of/ /the/ /water	/that/	/is/ /consumed/
El precio	no	influye	en la calidad	de el agua	que	se consume
precio **El**	no	influye	calidad **la en**	agua **el de**	que	se consume

Syntax-Based Reordering: Long-range Reordering

- Based on the dependency tree of the source.
- Manually-defined rules move entire subtrees along the sentence.
- Allows longer reorderings which are the ones that most severely affect the translation.

Syntax-Based Reordering: Long-range Reordering



- We have defined four reordering rules which deal with the most important word order differences.

Syntax-Based Reordering: Long-range Reordering

Source sentence:
/price/ /the/ /no/ /has-influence/ /quality/ /the/ /on/ /water/ /the/ /of/ /that/ /is/ /consumed
precio el no influye calidad la en agua el de que se consum

(a)

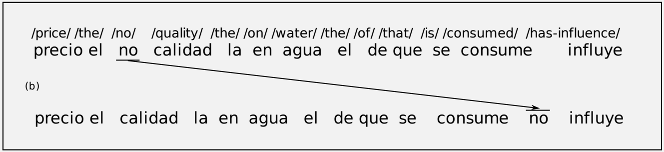
Reordered sent1:
precio el no calidad la en agua el de que se consume influye

- We have defined four reordering rules which deal with the most important word order differences.
  - (a) The verb is moved to the end of the clause, after all its modifiers.

Syntax-Based Reordering: Long-range Reordering

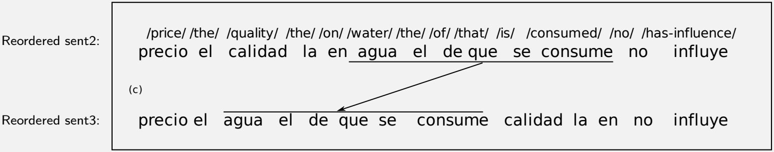
Reordered sent1:

Reordered sent2:



- We have defined four reordering rules which deal with the most important word order differences.
  - (a) The verb is moved to the end of the clause, after all its modifiers.
  - (b) In negative sentences, the particle 'no' is moved together with the verb to the end of the clause.

Syntax-Based Reordering: Long-range Reordering



- We have defined four reordering rules which deal with the most important word order differences.
  (a) The verb is moved to the end of the clause, after all its modifiers.
  (b) In negative sentences, the particle 'no' is moved together with the verb to the end of the clause.
  (c) Prepositional phrases and subordinate relative clauses which are attached to nouns are placed at the beginning of the whole noun phrase where they are included.

Syntax-Based Reordering: Long-range Reordering

Reordered sent3:
/price/ /the/ /water/ /the/ /of/ /that/ /is/ /consumed/ /quality/ /the/ /on/ /no/ /has-influence/
precio el agua el de que se consume calidad la en no influye

(c)

Reordered sent4:
precio el que se consume agua el de calidad la en no influye

- We have defined four reordering rules which deal with the most important word order differences.
  (a) The verb is moved to the end of the clause, after all its modifiers.
  (b) In negative sentences, the particle 'no' is moved together with the verb to the end of the clause.
  (c) Prepositional phrases and subordinate relative clauses which are attached to nouns are placed at the beginning of the whole noun phrase where they are included.

Syntax-Based Reordering: Long-range Reordering

Reordered sent4:

/price/ /the/ /that/ /is/ /consumed/ /water/ /the/ /of/ /quality/ /the/ /on/ /no/ /has-influence/
precio el <u>que</u> se consume agua el de calidad la en no influye

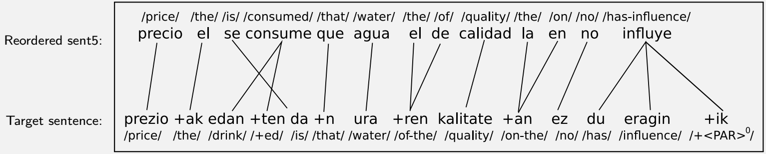
(d)

Reordered sent5:

precio el se consume <u>que</u> agua el de calidad la en no influye

- We have defined four reordering rules which deal with the most important word order differences.
  (a) The verb is moved to the end of the clause, after all its modifiers.
  (b) In negative sentences, the particle 'no' is moved together with the verb to the end of the clause.
  (c) Prepositional phrases and subordinate relative clauses which are attached to nouns are placed at the beginning of the whole noun phrase where they are included.
  (d) Conjunctions and relative pronouns placed at the beginning of Spanish subordinate (or relative) clauses are moved to the end of the clause, after the subordinate verb.

Syntax-Based Reordering: Long-range Reordering



- We have defined four reordering rules which deal with the most important word order differences.
  (a) The verb is moved to the end of the clause, after all its modifiers.
  (b) In negative sentences, the particle 'no' is moved together with the verb to the end of the clause.
  (c) Prepositional phrases and subordinate relative clauses which are attached to nouns are placed at the beginning of the whole noun phrase where they are included.
  (d) Conjunctions and relative pronouns placed at the beginning of Spanish subordinate (or relative) clauses are moved to the end of the clause, after the subordinate verb.

Statistical Reordering

- As syntax-based reordering, this method tries to reorder the source sentence before the SMT translation, harmonizing the source word order to the target one.
- It does not use any kind of syntactic information, it relies on pure statistical information.
- Translation process is divided in two steps, each of those steps is carried out by an SMT system:
  1. The first system is trained to reorder source words, without any kind of lexical transference.
  2. The second one carries out the lexical transference, as well as minor order movements.

Statistical reordering: Training process

1. Align source and target training corpora in both directions and combine word alignments to obtain many-to-many word alignments.
2. Modify the many-to-many word alignments to many-to-one (keeping for each source word only the alignment with a higher IBM-1 probability)
3. Reorder source words in order to obtain a monotonous alignment.
4. Train a state-of-the-art SMT system to translate from original source sentences into the reordered source
5. A second SMT system is necessary to carry out the lexical transference.

Experimental Results: Reordering techniques

- All the systems use the best segmentation option (*ManualGrouping*).
- In order to measure the impact of each reordering technique, we train and evaluate six different systems.
  - **Baseline**: a simplification of the system called *ManualGrouping* in segmentation experiments (deactivating the Moses' lexicalized reordering).
  - Individual techniques: **lexicalized reordering** (*ManualGrouping* in previous experiment), **syntax-based reordering** and **statistical reordering**.
  - Combination of methods: **Statistical+Lexicalized** and **Syntax-based+Lexicalized**.

Experimental Results: Reordering techniques

	BLEU	NIST	WER	PER
Baseline (ManualGrouping w/o Lexicalized reord.)	10.37	4.54	79.47	60.59
Lexicalized reord. (ManualGrouping)	11.36	4.67	78.92	60.23
Syntax-based reord.	11.03	4.60	78.79	61.35
Statistical reord.	11.13	**4.69**	78.21	**59.66**
Statistical+Lexicalized reord.	11.12	4.66	78.69	60.19
Syntax-based+Lexicalized reord.	**11.51**	**4.69**	**77.94**	60.45

Table: BLEU, NIST, WER and PER evaluation metrics.

- All individual reordering techniques outperform the baseline.
- Best results are obtained by the lexicalized reordering.
- System combinations have different behaviours.
- Syntax-based+Lexicalized combination statistically significantly outperforms the all single systems.

Outline

1. General experimental setup
2. Treatment of the morphological divergence between Spanish and Basque
3. Treatment of the syntactic divergence between Spanish and Basque
4. Hybridization attempts
    - Multi-Engine Combination
    - Statistical Post-Edition
    - Experimental Results
5. Overall evaluation
6. Contributions and Further Work

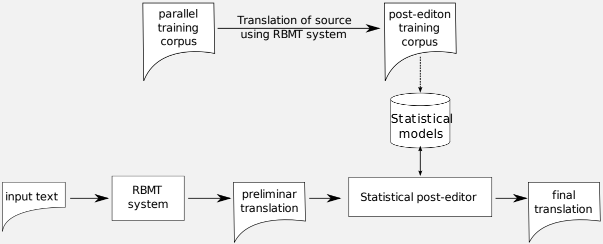
Hybridization

- After the development of a SMT system to translate from Spanish to Basque.
- Improve the translation by system combination:
  - SMT (this PhD thesis)
  - RBMT and EBMT (previously developed in Ixa)
- We experimented with two combination approaches:
  - Multi-Engine combination.
  - Statistical Post-Edition.

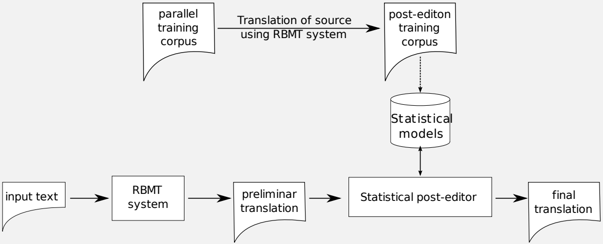
Multi-Engine combination

- We translate each sentence using the three engines.
- We select one of the possible translations, dealing with the following facts:
  - Precision of the EBMT approach is very high, but its coverage is low.
  - The SMT engine provides us a confidence score.
  - N-gram based techniques penalize the RBMT systems, although its translations are more adequate for human post-edition [Labaka et al., 2007]
- We use a simple hierarchical selection criterion:
  - If the EBMT engine covers the sentence, we choose its translation.
  - We only choose the SMT translation if its confidence score was higher than a threshold, defined on the development text set.
  - Otherwise, we choose the output from the RBMT engine.

General architecture of the Statistical Post-Edition

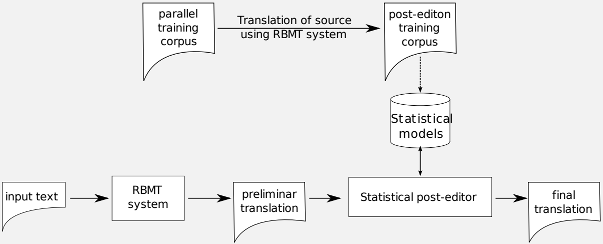


General architecture of the Statistical Post-Edition



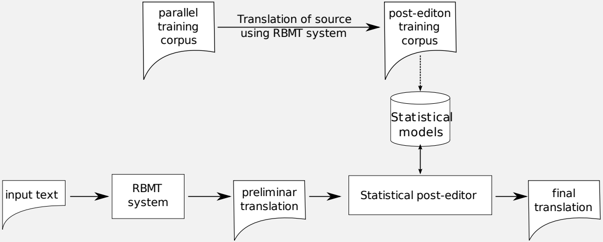
- It uses an SMT system to learn to post-edit the output of a RBMT system.

General architecture of the Statistical Post-Edition



- It uses an SMT system to learn to post-edit the output of a RBMT system.
- We do not have a real corpus of post-edited texts.

General architecture of the Statistical Post-Edition



- It uses an SMT system to learn to post-edit the output of a RBMT system.
- We do not have a real corpus of post-edited texts.
- We create a synthetic post-edition corpus from a parallel corpus.

Experimental Results: General domain (Consumer corpus)

	BLEU	NIST	WER	PER
Rule-Based (Matxin)	6.87	3.78	81.68	66.06
SMT-Segmentation+Reorder	**11.51**	**4.69**	**77.94**	**60.45**
EBMT system (0%)	-	-	-	-
Rule-Based + SPE	10.14	4.57	78.23	60.89
Multi-Engine	11.16	4.56	79.83	62.31

Table: Scores for the automatic metrics for systems trained on the Consumer corpus.

- For a general domain corpus, both hybridization techniques outperform the RBMT system.
- But they do not improve the results obtained by the SMT system.
- The bias of the automatic metrics against RBMT system can penalize the hybrid systems.
- A human evaluation would be necessary.

Labour Agreement corpus: Specific domain

Subset	Lang.	Doc.	Senten.	Words
Train	Basque	81	51,740	839,393
	Spanish	81		585,361
Development	Basque	5	2,366	41,408
	Spanish	5		28,189
Test	Basque	5	1,945	39,350
	Spanish	5		27,214

Table: Some statistics of the Labour Agreements Corpus

- We rerun the hybridization experiments on a specific domain corpus (Labour Agreement corpus).
- Administrative texts that contain many formal patterns that allow the EBMT system to extract them.

Experimental Results: Specific domain

	BLEU	NIST	WER	PER
Rule-Based (Matxin)	4.27	2.76	89.17	74.18
SMT-Segmentation+Reorder	12.27	4.63	77.44	58.17
EBMT system (64.92%)	32.42	5.76	60.02	54.75
Rule-Based + SPE	17.11	5.01	75.53	57.24
Multi-Engine	**37.24**	**7.17**	**56.84**	**45.27**

Table: Evaluation on domain specific corpus.

- Both hybridization techniques entail important improvements.
- Statistical Post-Edition successfully corrects the RBMT output, outperforming the results of the SMT system.
- The higher contribution to the Multi-Engine system comes by the inclusion of EBMT systems.
- The inclusion of the RBMT engine causes a slightly negative effect (1% relative decrease for BLEU).

Outline

1. General experimental setup

2. Treatment of the morphological divergence between Spanish and Basque

3. Treatment of the syntactic divergence between Spanish and Basque

4. Hybridization attempts

5. Overall evaluation
    - Doubts about BLEU & evaluation alternatives
    - Systems selected to Human-targeted evaluation
    - Automatic Evaluation
    - Human-Targeted evaluation results

6. Contributions and Further Work

Overall Evaluation

- So far, we have evaluated each approach in isolation and by means of automatic metrics.
- But we only have one reference to calculate automatic metrics.
- The scores obtained in this situation could be biased.
- In order to corroborate the results obtained, we have carried out a final evaluation based on human-targeted metrics.

Doubts about BLEU measure

- In recent years many doubts have arisen about the validity of BLEU:
  - It is extremely difficult to interpret what is being expressed in *BLEU* [Melamed et al., 2003]
  - Improving *BLEU* does not guarantee an improvement in the translation quality [Callison-Burch et al., 2006]
  - It does not offer as much correlation with human judgement as was believed [Koehn and Monz, 2006]
- Those problems are intensified since we only have one reference per sentence.

Overall Evaluation: Linguistic similarity

- Recent researches have present new metrics that computes the similarity according to linguistic features [Liu and Gildea, 2007], [Albrecht and Hwa, 2007], [Padó et al., 2007], [Giménez and Màrquez, 2008]
- Two main reasons have led us to reject the use of metrics based on linguistic similarity:
  - The applicability of these deep evaluation techniques are strongly conditioned by the accessibility to the linguistic processors required and their accuracy.
  - Just like BLEU does, these metrics compare the automatic translations with human-defined references, and the evaluation is not so precise when we have only one reference.

Overall Evaluation: Human-Targeted evaluation

- Human-targeted metrics compare the automatic hypothesis with the closest human post-edited references.
- We can use the post-edited references to calculate metrics, such as BLEU, NIST or TER, giving rise to human-targeted metrics such as HBLEU, HNIST or HTER.
- HTER metric is particularly interesting, since TER (Translation Error Rate) measures the number of post-editions done by the human translator.

Overall Evaluation: Human-Targeted evaluation

- This method requires human post-edited references, and its high cost prevented us from evaluating many systems using this method.
- We have chosen the 5 systems we consider the most representative ones:
  - Rule-Based (Matxin)
  - SMT baseline
  - SMT systems that use segmentation and reordering
  - Multi-Engine combination
  - Statistical Post-Edition
- In order to evaluate all the systems properly we incorporate two variations:
  - A bigger corpus for training.
  - Matrex instead of Moses.

Training corpora used for the final evaluation

		tokens	vocabulary	singletons
Initial Bilingual	Spanish	1,284,089	46,636	19,256
	Basque	1,010,545	87,763	46,929
Initial Monolingual	Basque	1,010,545	87,763	46,929
Final Bilingual	Spanish	9,167,987	219,472	97,576
	Basque	6,928,907	438,491	236,238
Final Monolingual	Basque	27,950,113	1,057,237	580,477

Table: Statistics on the final training corpora.

- 7 times larger bilingual corpus.
- 27 times larger monolingual corpus.
- Heterogeneous corpora that cover different topics and styles:
  - News
  - Administrative texts
  - Popular science texts
  - ...

Matrex

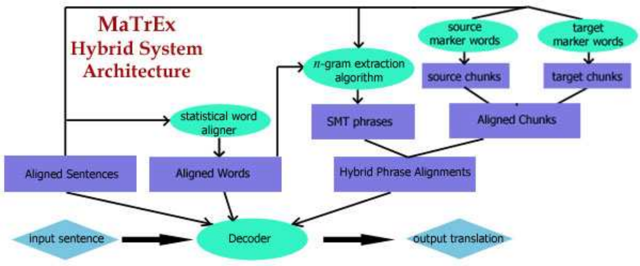


Figure: General design of the Matrex system [Stroppa and Way, 2006].

- MaTrEx is a data-driven MT system which combines both EBMT and SMT techniques.
- It aligns linguistic chunks using EBMT techniques and incorporates them into the SMT phrase table.
- The translation is carried out by a phrase-based decoder (Moses).

Automatic Evaluation: Reminder of previous evaluation

	BLEU	**NIST**	**WER**	**PER**
Matxin (RBMT)	6.87	3.78	81.68	66.06
SMT-baseline	10.78	4.52	80.46	61.34
SMT-Segmented	11.36	4.67	78.92	**60.23**
SMT-Segmented+Reorder	**11.51**	**4.69**	**77.94**	60.45
Multi-Engine	11.16	4.56	79.83	62.31
Statistical Post-Edition	10.14	4.57	78.23	60.89

Table: Scores for the automatic metrics for systems trained on the Consumer corpus.

Automatic Evaluation: larger training corpus

	BLEU	NIST	WER	PER
Matxin (RBMT)	6.87 (=)	3.78 (=)	81.68 (=)	66.06 (=)
SMT-baseline	11.12 (+0.34)	4.71 (+0.19)	78.13 (-2.33)	59.48 (-1.86)
SMT-Segmented	**11.56**(+0.20)	**4.83**(+0.16)	77.83 (-1.09)	**58.94**(-1.29)
SMT-Segmented+Reorder	11.19 (-0.32)	4.69 (=)	77.44 (-0.50)	60.09 (-0.36)
Multi-Engine	11.29 (+0.13)	4.73 (+0.17)	**76.99**(-2.84)	59.63 (-2.68)
Statistical Post-Edition	10.85 (+0.71)	4.67 (+0.10)	77.45 (-0.78)	60.42 (-0.47)

Table: Scores for the automatic metrics for all systems trained on the larger training corpus.

- Increasing the training corpus.
  - RBMT does not change, since it does not use the corpora for training.
  - All systems improve their scores, except the one we consider the best one (SMT-Segmented+Reorder).
  - The contribution of Syntax-based reordering is questioned.

Automatic Evaluation: MaTrEx vs. SMT

	BLEU	NIST	WER	PER
Matxin (RBMT)	6.87 (=)	3.78 (=)	81.68 (=)	66.06 (=)
MaTrEx-baseline	11.23 (+0.11)	4.75 (+0.04)	78.21 (+0.08)	59.66 (+0.18)
MaTrEx-Segmented	**11.71**(+0.15)	**4.82**(-0.01)	77.69 (-0.14)	**58.99**(+0.04)
MaTrEx-Segmented+Reorder	11.52 (+0.33)	**4.82**(+0.13)	**76.35**(-1.09)	**58.94**(-1.15)
Multi-Engine Hybridization	11.29 (=)	4.73 (=)	76.99 (=)	59.63 (=)
Statistical Post-Edition	10.85 (=)	4.67 (=)	77.45 (=)	60.42 (=)

Table: Scores for the automatic metrics for Matrex systems trained on the larger training corpus.

- The incorporation of EBMT phrases to SMT phrase-table consistently improves the results of the three SMT systems.

Automatic Evaluation: MaTrEx vs. SMT

	BLEU	**NIST**	**WER**	**PER**
Matxin (RBMT)*	6.87 (=)	3.78 (=)	81.68 (=)	66.06 (=)
MaTrEx-baseline*	11.23 (+0.11)	4.75 (+0.04)	78.21 (+0.08)	59.66 (+0.18)
MaTrEx-Segmented	**11.71**(+0.15)	**4.82**(-0.01)	77.69 (-0.14)	**58.99**(+0.04)
MaTrEx-Segmented+Reorder*	11.52 (+0.33)	**4.82**(+0.13)	**76.35**(-1.09)	**58.94**(-1.15)
Multi-Engine Hybridization*	11.29 (=)	4.73 (=)	76.99 (=)	59.63 (=)
Statistical Post-Edition*	10.85 (=)	4.67 (=)	77.45 (=)	60.42 (=)

Table: Scores for the automatic metrics for Matrex systems trained on the larger training corpus.

- The incorporation of EBMT phrases to SMT phrase-table consistently improves the results of the three SMT systems.
- The systems evaluated by means of human-targeted metrics are those marked with a *.

Automatic Evaluation: MaTrEx vs. SMT

	BLEU	NIST	WER	PER
Matxin (RBMT)*	6.87 (=)	3.78 (=)	81.68 (=)	66.06 (=)
MaTrEx-baseline*	11.23 (+0.11)	4.75 (+0.04)	78.21 (+0.08)	59.66 (+0.18)
MaTrEx-Segmented	**11.71**(+0.15)	**4.82**(-0.01)	77.69 (-0.14)	**58.99**(+0.04)
MaTrEx-Segmented+Reorder*	11.52 (+0.33)	**4.82**(+0.13)	**76.35**(-1.09)	**58.94**(-1.15)
Multi-Engine Hybridization*	11.29 (=)	4.73 (=)	76.99 (=)	59.63 (=)
Statistical Post-Edition*	10.85 (=)	4.67 (=)	77.45 (=)	60.42 (=)

Table: Scores for the automatic metrics for Matrex systems trained on the larger training corpus.

- The incorporation of EBMT phrases to SMT phrase-table consistently improves the results of the three SMT systems.
- The systems evaluated by means of human-targeted metrics are those marked with a *.
- As a consequence of the unexpected behaviour at increasing the training corpus, we have not evaluated the system that gets the highest BLEU score.

Human-Targeted evaluation results

	HTER	HBLEU	HNIST	HWER	HPER
Matxin	54.74	26.88	6.84	58.51	42.98
MaTrEx-baseline	53.59	27.86	7.23	58.48	40.23
MaTrEx-Segmented+Reorder	48.10	33.29	7.60	54.52	35.45
Multi-Engine	47.62	34.71	7.64	53.74	**35.27**
Statistical Post-Edition	**47.41**	**34.80**	**7.74**	**52.04**	36.05

Table: Scores for the human-targeted metrics for selected systems.

- The Matrex system that uses the improvements proposed in this PhD thesis outperform the Matrex baseline consistently.
- The two hybridization attempts obtain even better results, showing up as an interesting field in which to continue our investigation.
- All the differences between systems are statistically significant except those between Multi-Engine and Statistical Post-edition systems.

Human-Targeted evaluation results vs. BLEU

	HTER	HBLEU	HNIST	HWER	HPER	BLEU
Matxin	54.74	26.88	6.84	58.51	42.98	6.87
MaTrEx-baseline	53.59	27.86	7.23	58.48	40.23	11.23
MaTrEx-Segmented+Reorder	48.10	33.29	7.60	54.52	35.45	**11.52**
Multi-Engine	47.62	34.71	7.64	53.74	**35.27**	11.29
Statistical Post-Edition	**47.41**	**34.80**	**7.74**	**52.04**	36.05	10.85

Table: Scores for human-targeted metrics and BLEU.

- The automatic evaluation penalizes the RBMT system and the hybrid systems that use it.

Comparison with other systems

	BLEU	**NIST**	**WER**	**PER**
UPV-PRHLT	7.11	3.65	82.64	65.56
Avivavoz	**8.12**	3.90	81.60	64.22
EHU-IXA (MaTrEx-Segmented)	**8.10**	**3.98**	**78.70**	**62.25**

Table: Official results provided by the Albayzin evaluation organizers.

- We obtained the best results in Albayzin evaluation campaign:
  - Our system gets the best results by means of NIST, WER and PER.
  - The difference between our system and the *Avivavoz* system were not significant regarding BLEU.
- It was the only occasion that we could directly compare our work with other translation systems for Basque.
- The system we presented to the evaluation was the one called *MaTrEx-Segmented* in this thesis.

Outline

1. General experimental setup
2. Treatment of the morphological divergence between Spanish and Basque
3. Treatment of the syntactic divergence between Spanish and Basque
4. Hybridization attempts
5. Overall evaluation
6. Contributions and Further Work

Contributions: SMT to Basque

- Development of a state-of-the-art SMT system for Basque.
- Improvement of that baseline by means of segmentation.
  - Better scores in automatic evaluation for small and large corpora.
  - Definition of a hand-defined heuristic for morpheme-grouping that outperforms automatic segmentations.
- Combination of syntax-based reordering and lexicalized reordering.
  - Statistically significant improvement in 1M words corpus.
  - Those results are not corroborated at enlarging the training corpus.
- The combination of segmentation and syntax-based reordering clearly outperforms the baseline.
  - Statistically significant improvements in human-targeted evaluation.
  - 10% relative improvement in HTER and 16% in HBLEU.

Contributions: System combination

- Development of Multi-Engine and Statistical Post-Edition systems.
  - Both systems considerably outperform single systems in a specialized text like Labour Agreement corpus.
  - For a general domain corpus those gains are not perceived by automatic metrics.
  - But human-targeted evaluation shows statistically significant improvement.

Further work

- Investigate segmentation based on Bootstrapping and Word-Packing [Ma et al., 2007].
- Clarify, by means of human evaluation, the contribution of the syntax-based reordering method.
- Go deeper into Multi-Engine hybridization, creating new translation hypothesis combining phrases from the translation proposed by the different engines.
- Make use of factored machine translation implemented in Moses to integrate bilingual information at Statistical Post-Edition.
- Collect a real post-edition corpus to rerun post-edition experiments.
- Automatically learn post-editing rules to correct SMT translation, in the way Elming (2006) does.

Thanks for your Attention

Thank you!
Eskerrik asko!

EUSMT: Incorporating Linguistic Information into SMT for a Morphologically Rich Language.

Its use in SMT-RBMT-EBMT hybridization

PhD. Candidate: Gorka Labaka Intxauspe
Supervisors: Arantza Díaz de Ilarraza Sánchez
Kepa Sarasola Gabiola

Lengoaia eta Sistema Informatikoak/Lenguajes y Sistemas Informáticos
Euskal Herriko Unibertsitatea/Universidad del País Vasco

March 29, 2010

Outline

Aduriz, I. and Díaz de Ilarraza, A. (2003).
Morphosyntactic Disambiguation and Shallow Parsing in Computational Processing of Basque.
In *Inquiries into the lexicon-syntax relations in Basque. Bernarrd Oyharçabal (Ed.)*, Bilbao.

Agirre, E., Díaz de Ilarraza, A., Labaka, G., and Sarasola, K. (2006).
Uso de información morfológica en el alineamiento Español-Euskara.
Journal of the Spanish Association for Natural Language Processing, 37:257–265.

Albrecht, J. S. and Hwa, R. (2007).
A Re-examination of Machine Learning Approaches for Sentence-level MT Evaluation.
In *Annual Meeting of the Association for Computational Linguistics (ACL'07)*, pages 880–887, Prague, Czech Republic.

Alcázar, A. (2005).
Towards Linguistically Searchable Text.
In *Proceedings of BIDE (Bilbao-Deusto) Summer School of Linguistics*, Bilbao.

Alegria, I., Casillas, A., Díaz de Ilarraza, A., Igartua, J., Labaka, G., Lersundi, M., Mayor, A., and Sarasola, K. (2008a).
Mixing Approaches to MT for Basque: Selecting the Best Output from RBMT, EBMT and SMT.
In *Proceedings of the Mixing Approaches to Machine Translation workshop*, Donostia, Spain.

Alegria, I., Casillas, A., Díaz de Ilarraza, A., Igartua, J., Labaka, G., Lersundi, M., Mayor, A., and Sarasola, K. (2008b).
Spanish-to-Basque MultiEngine Machine Translation for a Restricted Domain.
In *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas*, Hawaii, USA.

Bojar, O. (2007).
English-to-Czech Factored Machine Translation.
In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 232–239, Prague, Czech Republic. Association for Computational Linguistics.

Callison-Burch, C., Osborne, M., and Koehn, P. (2006).

Re-evaluating the Role of BLEU in Machine Translation Research.
In *Proceedings of the International Conference of European Chapter of the Association for Computational Linguistics (EACL)*, pages 249–256.

Carreras, X., Chao, I., Padró, L., and Padró, M. (2004).
Freeling: an Open-Source Suite of Language Analyzers.
In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, pages 239–242.

Chen, B., Cettolo, M., and Federico, M. (2006).
Reordering Rules for Phrase-based Statistical Machine Translation.
In *IWSLT 2006*, pages 182–189.

Collins, M., Koehn, P., and Kucerova, I. (2005).
Clause Restructuring for Statistical Machine Translation.
In *ACL*, pages 531–540.

Costa-Jussà, M. R. and Fonollosa, J. A. R. (2006).
Statistical Machine Reordering.

In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 70–76, Sydney, Australia. Association for Computational Linguistics.

Díaz de Ilarraza, A., Labaka, G., and Sarasola, K. (2008).
Statistical Post-Editing: A Valuable Method in Domain Adaptation of RBMT Systems.
In *Proceedings of the Mixing Approaches to Machine Translation workshop*, Donostia, Spain.

Díaz de Ilarraza, A., Labaka, G., and Sarasola, K. (2009).
Relevance of Different Segmentation Options in Spanish-Basque SMT.
In *EAMT-2009: Proceedings of the 13th Annual Conference of the European Association for Machine Translation*. European Association for Machine Translation.

Díaz de Ilarraza, A., Labaka, G., and Sarasola, K. (2009).
Reordering in Spanish-Basque SMT.
In *Proceedings of the MT Summit 2009*, Ottawa, Canada. International Association for Machine Translation.

Doddington, G. (2002).
Automatic Evaluation of Machine Translation Quality using N-gram Co-Occurrence Statistics.
In *Proceedings of the Second International Conference on Human Language Technology Research*, pages 138–145, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Giménez, J. and Màrquez, L. (2008).
Heterogeneous Automatic MT Evaluation through Non-Parametric Metric Combinations.
In *Proceedings of the IJCNLP 2008: Third International Joint Conference on Natural Language Processing*, pages 319–326, Hyderabab, India.

Ginestí-Rosell, M., Ramírez-Sánchez, G., Ortiz-Rojas, S., Tyers, F. M., and Forcada, M. L. (2009).
Development of a Free Basque to Spanish Machine Translation System.
Journal of the Spanish Association for Natural Language Processing, 43:187–195.

Goldwater, S. and McClosky, D. (2005).
Improving Statistical MT through Morphological Analysis.
In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 676–683, Vancouver, Canada.

Koehn, P. (2004).
Statistical significance tests for machine translation evaluation.

Koehn, P. and Hoang, H. (2007).
Factored Translation Models.
In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Processing and Computational Natural Language Learning*, pages 868–876.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007).
Moses: Open Source Toolkit for Statistical Machine Translation.
In *Annual Meeting of the Association for Computational Linguistics (ACL)*, Prague, Czech Republic.

Koehn, P. and Monz, C. (2006).
Manual and Automatic Evaluation of Machine Translation between European Languages.
In *In Proceedings of NAACL 2006 Workshop on Statistical Machine Translation*, pages 102–121.

Labaka, G., Díaz de Ilarraza, A., and Sarasola, K. (2008).
Descripción de los sistemas presentados por IXA-EHU a la evaluación ALBAYCIN'08.
In *V Jornadas en tecnología del Habla*, Bilbao, Spain.

Labaka, G., Stroppa, N., Way, A., and Sarasola, K. (2007).
Comparing Rule-Based and Data-Driven Approaches to Spanish-to-Basque Machine Translation.
In *Proceedings of MT-Summit XI*, pages 297–304.

Liu, D. and Gildea, D. (2007).
Source-Language Features and Maximum Correlation Training for Machine Translation Evaluation.

In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT'07)*, pages 41–48.

Ma, Y., Stroppa, N., and Way, A. (2007).
Bootstrapping word alignment via word packing.
In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 304–311, Prague, Czech Republic.

Mayor, A. (2007).
Matxin: erregeletan oinarritutako itzulpen automatikoko sistema.
PhD thesis, Euskal Herriko Unibertsitatea.

Melamed, I. D., Green, R., and Turian, J. P. (2003).
Precision and Recall of Machine Translation.
In *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 61–63, Morristown, NJ, USA. Association for Computational Linguistics.

Minkov, E., Toutanova, K., and Suzuki, H. (2007).
Generating Complex Morphology for Machine Translation.

In *Proceedings of 45th Annual Meeting of the Association for Computational Linguistics (ACL'07)*, pages 128–135, Prague, Czech Republic.

Nießen, S., Och, F. J., Leusch, G., and Ney", H. (2000).
An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research.
In *Proceedings of LREC-2000: Second International Conference on Language Resources and Evaluation*, pages 39–45.

Oflazer, K. and El-Kahlout, I. D. (2007).
Exploring Different Representation Units in English-to-Turkish Statistical Machine Translation.
In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 25–32, Prague, Czech Republic.

Padó, S., Galley, M., Jurafsky, D., and Manning, C. (2007).
Robust Machine Translation Evaluation with Entailment Features.
In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP (ACL-IJCNLP-2009)*, pages 297–305, Prague, Czech Republic.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. (2002).
BLEU: A Method for Automatic Evaluation of Machine Translation.
In *Proceedings of 40th ACL*, pages 311–318, Philadelphia, PA.

Pérez, A., Inés Torres, M., and Casacuberta, F. (2008).
Joining linguistic and statistical methods for spanish-to-basque speech translation.
Speech Communication, 50(11-12):1021–1033.

Popović, M. and Ney, H. (2006).
POS-based Word Reorderings for Statistical Machine Translation.
In *International Conference on Language Resources and Evaluation*, pages 1278–1283, Genoa, Italy.

Ramanathan, A., Bhattacharya, P., Hegde, J., M.Shah, R., and M, S. (2008).
Simple Syntactic and Morphological Processing Can Help English-Hindi Statistical Machine Translation.
In *Third International Joint Conference on Natural Language Processing (JCNLP'08)*, pages 513–520, Hyderabad, India.

Sanchís, G. and Casacuberta, F. (2007).
Reordering via N-Best Lists for Spanish-Basque Translation.
In *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-07)*, pages 191–198, Skövde, Sweden.

Stroppa, N. and Way, A. (2006).
MaTrEx: DCU Machine Translation System for IWSLT 2006.
In *Proceedings of the International Workshop on Spoken Language Translation*, pages 31–36, Kyoto, Japan.

Tillmann, C., Vogel, S., and Zubiaga, A. (1997).
A DP Based Search Using Monotone Alignments in Statistical Translation.
In *Proceedings of the EACL-EACL-1997: 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 289–296.

Toutanova, K., Suzuki, H., and Ruopp, A. (2008).
Applying Morphology Generation Models to Machine Translation.

In *Proceedings of Human Language Technologies: The Annual Conference of the Association for Computational Linguistics (ACL-HLT'08)*, pages 514–522.

Trask (1997).
The History of Basque.
Routledge, London, England.

Zhang, Y., Zens, R., and Ney, H. (2007).
Chunk-Level Reordering of Source Language Sentences with Automatically Learned Rules for Statistical Machine Translation.
In *SSST '07: Proceedings of the NAACL-HLT 2007/AMTA Workshop on Syntax and Structure in Statistical Translation*, pages 1–8, Morristown, NJ, USA. Association for Computational Linguistics.