Procesamiento del Lenguaje Natural, Revista nº 58, marzo de 2017, pp. 109-116 recibido 30-11-2016 revisado 13-01-2017 aceptado 13-02-2017

Spanish Morphological Generation with Wide-Coverage Lexicons and Decision Trees

Generación Morfológica del Español con Lexicones de Amplia Cobertura y Árboles de Decisión

Daniel Ferrés, Ahmed AbuRa'ed and Horacio Saggion
Large Scale Text Understanding Systems Lab
TALN - DTIC
Universitat Pompeu Fabra
C/ Tànger, 122-140
08018 Barcelona, Spain
{*daniel.ferres,ahmed.aburaed,horacio.saggion,*}*@upf.edu*

Abstract: Morphological Generation is the task of producing the appropiate inflected form of a lemma in a given textual context and according to some morphological features. This paper describes and evaluates wide-coverage morphological lexicons and a Decision Tree algorithm that perform Morphological Generation in Spanish at state-of-the art level. The Freeling, Leffe and Apertium Spanish lexicons, the J48 Decision Tree algorithm and the combination of J48 with Freeling and Leffe lexicons have been evaluated with the following datasets for Spanish: i) CoNLL2009 Shared Task dataset, ii) Durrett and DeNero dataset of Spanish Verbs (DDN), and iii) SIGMORPHON 2016 Shared Task (task-1) dataset. The results show that: i) the Freeling and Leffe lexicons achieve high coverage and precision over the DDN and SIGMORPHON 2016 datasets, ii) the J48 algorithm achieves state-of-the-art results in all of the three datasets, and iii) the combination of Freeling, Leffe and the J48 algorithm outperformed the results of our other approaches in the three evaluation datasets, improved slightly the results of the CoNLL2009 and SIGMORPHON 2016 reported in the state-of-the-art literature, and achieved results comparable to the ones reported in the state-of-the-art literature on the DDN dataset evaluation.
Keywords: Morphological generation, morphological lexicons, decision trees, natural language generation

Resumen: La Generación Morfológica es la tarea de producir la forma flexionada apropiada de un lemma en un determinado contexto textual y en concordancia con algunas características morfológicas. En este artículo se presentan y se evaluan algunos lexicones morfológicos de amplia cobertura y un algoritmo de árboles de decisión para la Generación Morfológica en español. Los lexicones para el español Freeling, Leffe y Apertium, el algoritmo de árboles de decisión J48 y la combinación de los lexicones Freeling y Leffe con el J48 han sido evaluados con los siguientes conjuntos de datos para el español: i) conjunto de datos de la CoNLL2009 Shared Task, ii) el conjunto de datos de verbos para el español de Durrett y DeNero (DDN), y iii) el conjunto de datos para el español de la evaluación SIGMORPHON 2016 Shared Task (task-1). Los resultados muestran que: i) los lexicones morfológicos consiguen alta cobertura y precisión en los conjuntos de datos DDN y SIGMORPHON 2016, ii) el algoritmo J48 por si sólo alcanza resultados en el estado del arte en los tres conjuntos de evaluación, y iii) que la combinación de predicciones de Freeling, Leffe y el algoritmo J48 mejora los resultados de nuestras otras implementaciones en los tres conjuntos de datos evaluados, que además mejoran ligeramente los resultados reportados en el estado del arte en los conjuntos de datos del CoNLL2009 y del SIGMORPHON 2016, y que consiguen resultados comparables con los reportados en el estado del arte de la evaluación del conjunto de datos DDN.
Palabras clave: Generador morfológico, lexicones morfológicos, árboles de decisión, generación de lenguaje natural

ISSN 1135-5948

1 Introduction

Morphological Generation is the task of producing the appropiate inflected form of a lemma in a given textual context and according to some morphological features. An example of morphological inflection in Spanish language is shown in Figure 1: the lemma *cantar* (sing) inflected with the verbal morphological features of *number* (plural), *person* (1st), *mode* (indicative), *tense* (imperfect) generates the inflected form *cantábamos* (sang). Morphological Generation is a crucial part of the Surface Realization phase of Natural Language Generation (NLG) systems. NLG for Spanish has been applied in complex applications such as Dialogue Systems (Amores, Pérez, and Portillo, 2006), Machine Translation (Forcada et al., 2011), and Textual Simplification (Bott et al., 2012) among others. Morphological Generation can be performed with the following resources: i) morphological lexicons (Molinero, Sagot, and Nicolas, 2009; Forcada et al., 2011; Padró and Stanilovsky, 2012), ii) hand-made or learned inflected rules or decision trees (Durrett and DeNero, 2013; Nicolai, Cherry, and Kondrak, 2015), and iii) other supervised learning systems (Bohnet et al., 2010; Dušek and Jurcicek, 2013; Ahlberg, Forsberg, and Hulden, 2015; Faruqui et al., 2016; Cotterell et al., 2016; Kann and Schütze, 2016). Morphological lexicons are handmade or semi-automatically generated dictionaries with inflected forms stored in the following way <*inflected form, lemma, morphological features*>. These lexicons can have wide coverage and achieve high precision but are not able to inflect new unseen lemmas.

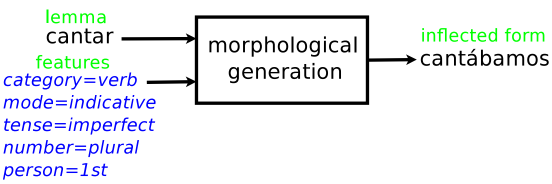

Figure 1: Example of Morphological Generation

Inflection rules can be generated manually or automatically by Rule Induction or Decision Tree learning algorithms. The advantage of rule induction systems over other supervised machine learning technologies is that the models (rules or trees) are human readable and interpretable. Thus these kind of models can be modified and extended with human supervision.

The main contribution of this paper is to present and evaluate a novel approach to Morphological Generation that combines predictions from the free-available and wide-coverage Spanish lexicons Freeling and Leffe with the ones from a human-interpretable J48 Decision Tree model. We will show that our approach achieves state-of-the-art results in coverage, precision and accuracy in Spanish Morphological Generation over several benchmarking datasets. The resource is available online.[1]

2 Related Work

This section describes morphological lexicons and supervised learning approaches to Morphological Generation for Spanish. COES is a morphological tool for Spanish (Rodríguez and Carretero, 1996) which is composed by a lexicon and a set of about 3,500 derivative rules that cover most of the morphological rules of the Spanish language. Freeling[2] (Padró and Stanilovsky, 2012) is an open source language analysis software that has a Spanish dictionary of about 650,000 inflected forms corresponding to 76,000 lemma to Part-of-Speech (PoS) combinations. This dictionary was obtained from the Spanish Resource Grammar (SRG) (Marimon, Seghezzi, and Bel, 2007). *Leffe*[3] (Molinero, Sagot, and Nicolas, 2009) is a wide coverage morphological and syntactic lexicon for Spanish that merged the following high quality existing lexicons for Spanish: Multext, USC, ADESSE, and SRG . Leffe has about 165,000 unique (lemma,PoS) pairs, which correspond to approximately 1,590,000 entries that associate a form with both morphological and syntactic information (approximately 680,000 unique (form,PoS) pairs). Apertium is a free/open-source platform for rule-based machine translation (Forcada et al., 2011) that has a morphological analyzer and generator based on finite state transducers. It also has a dictionary for Spanish[4] with over 46,000 lemmas and morphological inflectional rules that can cover more than 7.9 million inflected forms (of which most are verbal forms with enclitics).

[1] https://www.upf.edu/web/taln/resources

[2] http://nlp.lsi.upc.edu/freeling/

[3] https://gforge.inria.fr/projects/alexina/

[4] https://sourceforge.net/p/apertium/svn/HEAD/tree/languages/apertium-spa/apertium-spa.spa.dix

Unimorph[5] (Kirov et al., 2016) is a multilingual morphological resource extracted from Wiktionary that includes data in Spanish. Bohnet et al. (2010) presented a Support Vector Machine (SVM) based multilingual dependency oriented stochastic deep sentence realizer which has a morphological generator. They used the Levenshtein edit distance to map lemmas to word forms. The input to the classifier is the lemmata of a sentence, its dependency tree and the already ordered sentence. Dušek and Jurcicek, (2013) presented a morphological realizer that uses also edit scripts based on the Levenshtein distance and multi-class logistic regression classifiers. They used some generic features across all languages: lemma, PoS tag, morphological features (e.g. case, gender,...) and suffixes of the lemma up to 4 characters.

Durrett and DeNero, (2013) presented a supervised approach to learn and predict morphological paradigms that automatically acquires the orthographic transformation rules of morphological paradigms from labeled examples, and then learns the contexts in which those transformations apply using a discriminative sequence model. Nicolai, Cherry, and Kondrak, (2015) used supervised inflection generation with discriminative string transduction and reranking. They transform character alignments into inflection rules and select them with a discriminative semi-Markov Model. Ahlberg, Forsberg, and Hulden, (2015) presented a system that learns morphological paradigms and is able to predict inflection tables from unseen lemmas. Their system is based on the longest common subsequence and SVMs. Faruqui et al. (2016) used character sequence to sequence learning for morphological inflection with encoder-decoder neural networks. The SIGMORPHON 2016 Shared Task (task-1)[6] on Morphological Reinflection (Cotterell et al., 2016) covered Morphological Inflection for 10 languages (including Spanish). The LMU system (Kann and Schütze, 2016) achieved the best accuracies for Spanish with 98.84% and 99.05%. They used a character-based sequence-to-sequence attention model called MED (Morphological EncoderDecoder) with an RNN encoder-decoder architecture (Bahdanau, Cho, and Bengio, 2014).

[5]`http://ckirov.github.io/UniMorph/`

[6]`http://ryancotterell.github.io/sigmorphon2016/`

3 *System Description*

Our Morphological Generation system can be executed in three ways: i) lexicon-based morphological inflection, ii) Decision Trees based predictions, and iii) a combination of lexicon-based generation and Decision Trees based predictions.

3.1 Lexicon-Based Inflection

The Lexicon-Based inflection simply uses the information existing in a morphological lexicon to generate a new inflected form. This methodology offers high quality predictions but does not perform predictions on unseen forms in the lexicon. The Morphological Lexicons used in this paper are: Freeling, Apertium (derived forms without enclitics), and Leffe (derived forms) (see in Table 1 some statistics about these lexicons).

3.2 Decision Trees

The Decision Trees algorithm used to predict is the J48 algorithm,[7] an open source Java implementation of the C4.5 algorithm (Quinlan, 1993) in the WEKA[8] data mining tool. The algorithm takes a lemma, PoS and PoS features as input and then generates the proper form according to the morphological features derived from the PoS and the features extracted from the lemma (see the description of these features in Tables 2 and 3).

Feature	Lexical categories
Case	Pronouns (i.e. ordinal, qualificative and possessive in case of Adjectives)
Gender	Adjectives, Determiners Nouns, Pronouns, Verbs (i.e. masculine, feminine, or common)
Mood	Verbs (i.e. indicative, infinitive, subjunctive, gerund, imperative and participle).
Number	Nouns, Verbs, Adjectives (i.e. singular, plural, and invariable)
Person	Determiners, Pronouns, Verbs (i.e. first, second and third person).
Possessornum	Determiners (i.e. singular, plural and invariable)
Polite	Pronouns (i.e. yes, no)
Tense	Verbs (i.e. past, present, future, imperfect and conditional).
Type	Adjectives, Adverbs, Determiners, Numerals (i.e. ordinal, qualificative and possessive in case of Adjectives).

Table 2: Morphological features associated to lexical categories

[7]J48 has been chosen because in our initial experiments it achieved better performance than other human-interpretable algorithms available in WEKA.

[8]`http://www.cs.waikato.ac.nz/~ml/weka/`

dataset	size (#tokens)								
	nouns		verbs		adjectives		total (all lexical categories)		
	lemmas	forms	lemmas	forms	lemmas	forms	lemmas	forms	CRAE top10k(%)
Apertium	16,668	36,013	3,927	255,253	6,279	23,093	36,253	323,846	92.02
Freeling	49,528	107,638	7,660	497,801	18,618	62,979	76,335	669,216	91.48
Leffe	70,944	154,106	8,359	530,309	28,509	97,136	162,394	852,347	93.51

Table 1: Statistics of the Morphological Lexicons used in the evaluation

Features Set	Description	example
Length	number of characters	6
Last characters	last, penultimate, and antepenultimate characters	r,a,t
Last n-grams	last bigram last trigram	ar,tar
1-grams	all lemma unigrams	c,a,n,t,a,r
1-grams order	all lemma unigrams with associated order position	c1,a2,n3,t4,a5,r6
1-grams reverse order	all lemma unigrams with associated reverse order	c6,a5,n4,t3,a2,r1
2-grams	all lemma bigrams	ca,an,nt,ta,ar
Last 3-gram with a skip	the last three chars skipping the penultimate one	t_r
Penultimate 3-gram with a skip	last penultimate three chars skipping the antepenultimate one	n_a
Phonetics	consonant and vowels in the same way of the character and n-grams previous features	c,v,c, vc,cvc, c_c,c_v

Table 3: Features associated to the lemma. Includes the example of features extracted for the lemma *cantar* (sing)

The system is based on the Levenshtein edit distance algorithm between the lemma and the target word form. The edit distance algorithm calculates how many character operations are needed to transform and edit one string (e.g. lemma) to another (e.g. target word form). The possible operations are: Insert(index, character), implemented by inserting the presented character into the index position of the lemma; Replace(index, character), implemented by replacing the character in the index position of the lemma with the presented character; and Delete(index), implemented by deleting the character at the index position of the lemma (see an example of some edit scripts in Figure 2).

lemma	inflected form	edit script
rehusar	rehusemos	R(0,s) R(1,o) I(2,m) I(2,e)
enviar	envié	R(0,é) D(1)
millón	millones	I(0,s) I(0,e) R(1,o)
joven	jóvenes	I(0,s) I(0,e) R(3,ó)
hermoso	hermosísimo	I(1,m) I(1,i) I(2,í) I(2,s)
averiar	averiado	R(0,o) I(1,d)

Figure 2: Examples of edit scripts

The index starts from the last character of the lemma to the first (i.e. 0 indicates the last character, 1 indicates the last-1, etc...). For example: in order to obtain the verb form *envié* (sent) the operations *R(0,é)* and *D(1)* are supposed to be applied on the lemma *enviar* (send). The implementation starts with D(1) deleting the letter *'a'* then R(0,é) will replace the last letter with *'é'*. The J48 algorithm constructs a decision tree for each lexical category. The lexical categories that can be used by the system are common nouns, verbs, adjectives, adverbs, pronouns, determiners, and numerals. The decision tree built will learn a mapping index which refers to a sequence of operations to transform a lemma to an inflected form according to the lemma based and PoS based features.[9] After training, a J48 model is learnt in which the system takes a lemma, PoS and PoS features as input and then generates a test instance based on the training features from the input data, the model will predict a mapping index which refers to a sequence of operations forming the edit script as shown at Figure 2. The operations are implemented from right to left order in the sequence of edit scripts after applying the operations sequence on the lemma the final form will be obtained.

3.3 Lexicon and Decision Trees

This configuration combines lexicon-based and Decision Trees based predictions. This configuration gives priority to predictions that can be obtained with the lexicon; thus trying to ensure a high precision because of the wide-coverage of frequent forms by the lexicon. This wide-coverage is shown in Table 1: a significant number of inflected forms in the lexicons (more than 852,000 in the case Leffe lexicon) and the high coverage (about 92%-93%) of the *Corpus de Referencia del Español Actual (CREA)* corpus top 10,000

[9]Some specific features for verbs contained in the SIGMORPHON dataset are not described here. These features are *Alt*, *Aspect*, *Polar*, and *Polite*.

frequent words[10] in Spanish. The system works in the following way: i) firstly the system seeks if there are inflected forms associated to the lemma and the morphological features in the lexicon(s) and selects them[11], ii) otherwise, if no inflected forms are found in the lexicon, then the system will execute the J48 Decision Tree classifier associated to the lexical category of the current lemma to be inflected with the morphological features and the lemma based features as input data to predict the edit operations necessary to generate the new inflected form. The combinations of lexicons and J48 used are the following ones: i) Freeling executed in combination with the J48 algorithm (Freeling + J48), ii) Leffe executed in combination with the J48 algorithm (Leffe+J48), and iii) Freeling executed in combination with Leffe and the J48 algorithm (F+L+J48). The last combination uses firstly the Freeling lexicon to find the inflected form, otherwise uses the Leffe (if Freeling fails to retrieve an inflected form), and otherwise uses J48 (if both Freeling and Leffe fail).

4 Evaluation

The evaluation of the systems presented was performed using separately the following datasets for Spanish (described below): i) CoNLL2009 Shared task Dataset for Spanish, ii) Durrett and DeNero datasets for Spanish Verbs, iii) SIGMORPHON 2016 Shared Task task-1 dataset for Spanish. The size in tokens of the training, development and evaluation splits of the datasets is reported in Table 4. The training split is used to train the Decision Trees models, the evaluation split is used for evaluation, and the development split is not used.

The experiments presented in this evaluation are designed to evaluate the following sets of measures: i) the *coverage*, *precision* and *accuracy* of the morphological lexicons over the datasets, and ii) the *accuracy* of the J48 algorithm applied over the datasets predicting alone or in combination with the morphological lexicons. The *coverage* measure of the morphological lexicons tell us about which percentage of the evaluation dataset can be automatically predicted with the lexicon and without performing prediction based on supervised learning. On the other hand, the *precision* measure indicates the percentage of this coverage that is correctly inflected. It is supposed that because lexicons have been produced by human experts, the inflections derived from a lexicon will have more confidence compared with the ones predicted by a supervised learning algorithm. The coverage and precision measures will be measured only over the DDN and SIGMORPHON 2016 datasets, because the CoNLL2009 includes the numeric lexical category not present in the lexicons evaluated. The accuracy measures will be calculated in all three datasets and with respect to some specific lexical categories in the case of CoNLL2009 (7 categories) and SIGMORPHON 2016 (3 categories). The CoNLL2009 evaluation will include the accuracy of some token subsets: tokens excluding the punctuation, only inflected forms, and only unseen forms in training set.

4.1 CoNLL2009 Dataset

The CoNLL2009 Shared Task[12] (Hajič et al., 2009) is to predict syntactic and semantic dependencies and their labeling. The Spanish datasets were generated from the AnCoraES[13] corpora (Taulé, Martí, and Recasens, 2008), a multilevel annotated corpora for Spanish (mainly news). It has about 528,000 tokens annotated manually, semiautomatically, or fully automatically. The data size of the training, test and development datasets for Spanish is 427,442, 50,630 and 50,368 tokens respectively. The lexical categories appearing in this dataset are: nouns, verbs, adjectives, adverbs, pronouns, determiners, conjunctions, adpositions, interjections, dates, numerals and punctuation.

4.2 Durrett and DeNero (DDN)

Durrett and DeNero, (2013) evaluated their approach to Morphological Paradigm prediction with full morphology tables extracted from Wiktionary.[14] For Spanish they extracted morphology tables of verbs (231,135 total items). They used 208,335 tokens to train, 11,400 tokens for development, and 11,400 tokens for test.

[10] `http://corpus.rae.es/lfrecuencias.html`

[11] Note that there are some special cases in which two or more inflected forms can be inflected (e.g. the verbal forms *cantara* and *cantase* for a set of morphological features of the verb *cantar* (sang)).

[12] `http://ufal.mff.cuni.cz/conll2009-st/`

[13] `http://clic.ub.edu/corpus/ancora`

[14] `http://cs.utexas.edu/~gdurrett`

	size (#tokens)			eval dataset information					
dataset	train.	dev.	eval.	-pun %	infl. % -	unk.%	#noun	#verb	#adj
ConNLL2009	427,442	50,368	50,630	85.42	29.96	6.16	11,500	5,941	3,431
DDN (ES-V)	208,335	11,400	11,400	100	98.25	100	-	11,400	-
SIGMORPHON	12,575	1,596	23,229	100	90.31	100	2,914	18,739	1,576

Table 4: Statistics of the Spanish evaluation datasets. (-pun) = indicates % excluding punctuation, (infl.) = % of only forms that differ from the lemma, (unk) = % of forms unseen in the training set

4.3 SIGMORPHON 2016 Dataset

The SIGMORPHON 2016 Shared Task (task-1) data came mainly from the English edition of Wiktionary. The data extraction process is described in (Kirov et al., 2016). The Spanish dataset has 1,596 instances for development, 12,575 for training, and 23,229 for testing. The lexical categories present in the dataset are nouns, verbs and adjectives.

5 *Results*

The coverage and precision measures of the lexicons over the DDN and the SIGMORPHON 2016 datasets are shown in Table 5: Freeling and Leffe achieve high coverage and precision with coverages of more than 88% and up to 92% and precisions over 99.4%.[15]

Lexicon	DDN	SIGMORPHON
Apertium	59.50 (99.95)	58.81 (99.23)
Freeling	91.08 (99.99)	88.87 (99.42)
Leffe	92.41 (99.99)	90.35 (99.41)

Table 5: Lexicon evaluation results: % of coverage and precision (between parentheses)

The accuracy measures of the lexicons and the J48 Decision Tree algorithm over all evaluation datasets are shown in Tables 6, 7, and 8. The J48 algorithm achieves state-of-the-art results in all of the three datasets, outperforming some statistical algorithms but with slightly inferior results with respect to other approaches existing in the literature. The combination of morphological lexicons and the J48 algorithm outperformed the results of our other approaches in the three evaluation datasets tested, improved slightly the results with respect to the CoNLL2009 and SIGMORPHON results reported in the state-of-the-art literature, and achieved results comparable to the ones reported in the state-of-the-art literature on the DDN dataset evaluation. The results of the DDN dataset evaluation experiments (see Table 6) show that the combination of J48 and these two lexicons (Freeling and Leffe) improve the results of Durrett and DeNero, (2013) and Nicolai, Cherry, and Kondrak, (2015) and equals the ones reported by Ahlberg, Forsberg, and Hulden, (2015) but are still slightly inferior to the ones obtained by Faruqui et al. (2016).

[15]The precision of the lexicons over the SIGMORPHON 2016 dataset could be increased if a manual revision detects annotation errors in this dataset.

Algorithm	Acc.(%)
Apertium	59.47
Freeling	91.07
Leffe	92.40
J48	99.57
Freeling+J48	99.89
Leffe+J48	99.85
Freeling+Leffe+J48	99.92
(Durrett and DeNero, 2013)	99.67
(Nicolai, Cherry, and Kondrak, 2015)	99.90
(Ahlberg, Forsberg, and Hulden, 2015)	99.92
(Faruqui et al., 2016)	**99.94**

Table 6: DDN evaluation results in accuracy

The results of the SIGMORPHON 2016 Shared Task task-1 dataset evaluation (see Table 7) show that the J48 algorithm achieves state-of-the art accuracy but slightly below the accuracies achieved by the best system at SIGMORPHON 2016 (Kann and Schütze, 2016). The combination of J48 with the lexicons outperforms the accuracies achieved by Kann and Schütze, (2016) and the other participants of SIGMORPHON 2016 (Cotterell et al., 2016).

Algorithm	Total	Noun	Verb	Adj.
Apertium	53.46	25.08	58.16	50.06
Freeling	88.36	76.38	91.72	70.49
Leffe (L)	89.82	81.91	91.59	83.37
J48	98.31	98.49	98.27	98.54
Freeling+J48	99.21	98.73	99.30	99.11
L+J48	99.19	98.73	99.26	**99.23**
Freeling+L+J48	**99.23**	**98.76**	**99.31**	**99.23**
Kann et al., 2016	98.94	-	-	-
Kann et al., 2016	99.05	-	-	-

Table 7: SIGMORPHON 2016 Spanish datasets evaluation accuracy (%) results

Algorithm	Total	-pun	infl.	unk.	Noun	Verb	Adj	Adv	Pron	Det	Num
Apertium-spa	71.61	67.56	62.10	39.28	71.26	85.18	79.27	0	0.77	47.14	-
Freeling	84.04	81.77	86.34	43.94	72.60	95.03	76.21	72.89	43.29	87.69	-
Leffe	77.56	74.35	75.11	44.64	72.47	82.86	51.47	75.97	0	82.70	-
J48	99.05	98.92	**97.26**	92.81	99.80	94.32	99.24	**98.51**	**99.56**	99.95	**92.16**
Freeling+J48	**99.06**	**98.93**	97.11	**94.74**	**99.84**	**95.77**	**99.44**	96.19	98.00	**99.97**	**92.16**
Leffe+J48	98.13	97.87	94.02	94.70	**99.84**	95.70	99.41	96.19	**99.56**	92.91	**92.16**
Freeling+Leffe+J48	**99.06**	**98.93**	97.11	**94.74**	**99.84**	**95.77**	**99.44**	96.19	98.00	**99.97**	**92.16**
(Bohnet et al., 2010)	98.48	-	-	-	-	-	-	-	-	-	-
(Dušek and Jurcicek, 2013)	99.01	98.86	97.10	91.11	-	-	-	-	-	-	-

Table 8: CoNLL2009 Spanish results in accuracy (%)

Finally, the results of the CoNLL2009 Shared Task evaluation experiments (see Table 8) show that both J48 and J48 combined with the lexicons improved very slightly the results of the statistical learning approaches that evaluated the CoNLL2009 dataset for Spanish: the SVM approach (Bohnet et al., 2010) and the Logistic Regresion one (Dušek and Jurcicek, 2013).

6 Discussion

Morphological Generation is a crucial task in several advanced Language Technology applications that require so high precision that no errors should be passed to the output presented to final users; because a single wrongly inflected word could affect the end user's trustworthyness in the application. The system presented in this paper pretends to minimize inflection errors and improve the precision of Morphological Generation systems by incorporating the benefits of the lexicons to the ones obtained by supervised learning algorithms. The evaluation of the *coverage* and *precision* measures over two of these lexicons indicate that the lexicons can predict with a precision over 99.95% and 99.23% in the DDN and SIGMORPHON 2016 datasets respectively. In addition to the fact that the results of combining lexicons and Decisions Trees compare or outperform most of the state-of-the-art results, it has to be taken into account that the model generated by the J48 Decision Trees algorithm is human-interpretable and can be modified and extended in the same way as decision rules. On the other hand, common accuracy errors obtained in all three datasets were those produced by a wrong prediction of the edit scripts to inflect the form by the J48 model (e.g. given the lemma *endeudar* (indebt) generates *endieudas* instead of the correct form *endeudas*). But the most frequent errors were dataset-specific errors such as: i) verbal forms with enclitics without enough features in the dataset to be learnt by the model or recognized by the lexicon, phrasal verbs, and numerals (not present in the lexicons) for CoNLL2009 dataset, and ii) wrong morphological features present in the SIGMORPHON 2016 dataset that affect the lexicon-based predictions but not the J48 ones.

7 Conclusions and Further Work

This paper describes and evaluates free-available morphological lexicons and a Decision Tree algorithm that can perform Morphological Generation in Spanish at state-of-the art level. The Freeling, Leffe and Apertium Spanish lexicons, the J48 Decision Tree algorithm and the combination of J48 with Freeling and Leffe lexicons have been evaluated with the following datasets for Spanish: i) CoNLL2009 Shared task dataset, ii) Durrett and DeNero dataset of Spanish Verbs, iii) SIGMORPHON 2016 Shared Task (task-1) dataset. The results show that: i) the Freeling and Leffe lexicons achieve high coverage and precision over the DDN and SIGMORPHON 2016 datasets, ii) the J48 algorithm achieves state-of-the-art results in all of the three datasets, and iii) the combination of Freeling, Leffe and the J48 algorithm outperformed the results of our other approaches in the three evaluation datasets, improved slightly the results of the CoNLL2009 and SIGMORPHON 2016 reported in the state-of-the-art literature, and achieved results comparable to the ones reported in the state-of-the-art literature on the DDN dataset evaluation.

Further work includes: i) performing tuning of the J48 algorithm parameters using the development data, ii) the adaptation of the system to other Ibero-Romance languages such as: Catalan, Galician, and Portuguese, and iii) investigate data-driven methods to detect wrong predictions.

Acknowledgements

This work was partly funded by the ABLE-TO-INCLUDE project (European Commission CIP Grant No. 621055), the TUNER project (TIN2015-65308-C5-5-R, MINECO/FEDER, UE), and the Spanish MINECO Ministry (MDM-2015-0502).

References

Ahlberg, M., M. Forsberg, and M. Hulden. 2015. Paradigm Classification in Supervised Learning of Morphology. In *Proceedings of NAACL 2015*.

Amores, G., G. Pérez, and P. M. Portillo. 2006. Reusing MT Components in Natural Language Generation for Dialogue Systems. *Procesamiento del Lenguaje Natural*, (37):215–224.

Bahdanau, D., K. Cho, and Y. Bengio. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. *CoRR*, abs/1409.0473.

Bohnet, B., L. Wanner, S. Mille, and A. Burga. 2010. Broad Coverage Multilingual Deep Sentence Generation with a Stochastic Multilevel Realizer. In *Proceedings of COLING 2010*, pages 98–106.

Bott, S., L. Rello, B. Drndarevic, and H. Saggion. 2012. Can Spanish Be Simpler? LexSiS: Lexical Simplification for Spanish. In *Proc. of COLING 2012*, pages 357–374.

Cotterell, R., C. Kirov, J. Sylak-Glassman, D. Yarowsky, J. Eisner, and M. Hulden. 2016. The SIGMORPHON 2016 Shared Task—Morphological Reinflection. In *Proceedings of SIGMORPHON 2016*.

Durrett, G. and J. DeNero. 2013. Supervised Learning of Complete Morphological Paradigms. In *Proc. of NAACL 2013*.

Dušek, O. and F. Jurcicek. 2013. Robust Multilingual Statistical Morphological Generation Models. In *Proceedings of ACL 2013 Student Research Workshop*, pages 158–164.

Faruqui, M., Y. Tsvetkov, G. Neubig, and C. Dyer. 2016. Morphological Inflection Generation Using Character Sequence to Sequence Learning. In *Proceedings of NAACL 2016*.

Forcada, M. L., M. Ginestí-Rosell, J. Nordfalk, J. O'Regan, S. Ortiz-Rojas, J. A. Pérez-Ortiz, F. Sánchez-Martínez, G. Ramírez-Sánchez, and F. M. Tyers. 2011. Apertium: a Free/Open-Source Platform for Rule-Based Machine Translation. *Machine Translation*, 25(2):127–144.

Hajič, J., M. Ciaramita, R. Johansson, D. Kawahara, M. A. Martí, L. Màrquez, A. Meyers, J. Nivre, S. Padó, J. Štěpánek, et al. 2009. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In *Proceedings of the CoNLL-2009: Shared Task*, pages 1–18.

Kann, K. and H. Schütze. 2016. MED: The LMU System for the SIGMORPHON 2016 Shared Task on Morphological Reinflection. In *Proceedings of SIGMORPHON 2016*.

Kirov, C., J. Sylak-Glassman, R. Que, and D. Yarowsky. 2016. Very-large Scale Parsing and Normalization of Wiktionary Morphological Paradigms. In *Proceedings of LREC 2016*.

Marimon, M., N. Seghezzi, and N. Bel. 2007. An Open-Source Lexicon for Spanish. *Procesamiento del Lenguaje Natural*, 39.

Molinero, M. A., B. Sagot, and L. Nicolas. 2009. Building a Morphological and Syntactic Lexicon by Merging Various Linguistic Resources. In *Proceedings of NODALIDA 2009*.

Nicolai, G., C. Cherry, and G. Kondrak. 2015. Inflection Generation as Discriminative String Transduction. In *Proceedings of NAACL 2015*.

Padró, L. and E. Stanilovsky. 2012. FreeLing 3.0: Towards Wider Multilinguality. In *Proceedings of LREC 2012*.

Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Rodríguez, S. and J. Carretero. 1996. A Formal Approach to Spanish Morphology: the COES Tools. *Procesamiento del Lenguaje Natural*, 19:119.

Taulé, M., M. A. Martí, and M. Recasens. 2008. AnCora: Multilevel Annotated Corpora for Catalan and Spanish. In *LREC 2008*.