Improving the Naturalness and Expressivity of Language Generation for Spanish

Cristina Barros
Department of Software
and Computing Systems
University of Alicante
Apdo. de Correos 99
E-03080, Alicante, Spain
cbarros@dlsi.ua.es

Dimitra Gkatzia
School of Computing
Edinburgh Napier University
Edinburgh, EH10 5DT, UK
d.gkatzia@napier.ac.uk

Elena Lloret
Department of Software
and Computing Systems
University of Alicante
Apdo. de Correos 99
E-03080, Alicante, Spain
elloret@dlsi.ua.es

Abstract

We present a flexible Natural Language Generation approach for Spanish, focused on the surface realisation stage, which integrates an inflection module in order to improve the naturalness and expressivity of the generated language. This inflection module inflects the verbs using an ensemble of trainable algorithms whereas the other types of words (e.g. nouns, determiners, etc) are inflected using hand-crafted rules. We show that our approach achieves 2% higher accuracy than two state-of-art inflection generation approaches. Furthermore, our proposed approach also predicts an extra feature: the inflection of the imperative mood, which was not taken into account by previous work. We also present a user evaluation, where we demonstrate that the proposed method significantly improves the perceived naturalness of the generated language.

1 Introduction

Improving the naturalness and expressivity of the generated language is key in the area of Natural Language Generation (NLG), which aims to automatically generate text from non textual inputs. Specifically, one way to address it is by enriching the language through its morphology. Existing NLG systems are usually applied to non-morphologically rich languages, such as English, where the *morphological realisation* (i.e. the production of well inflected sentences or words through the use of words' morpho-syntactic properties) of words during generation can be done using hand-written rules or existing libraries such as SimpleNLG (Gatt and Reiter, 2009). However, the use of this type of rules in morphologically rich languages, such as Spanish or German, can be expensive and lead to incorrect inflection of a word, thus generating ungrammatical or meaningless texts.

We propose a flexible and domain independent NLG approach for Spanish, focused on the surface realisation stage, which integrates an inflection module. This inflection module incorporates an ensemble of trainable algorithms to automatically inflect a sentence by learning the inflection of Spanish verbs in conjunction with some hand-crafted rules for inflecting others types of words.

Our contributions to the field are as follows: we propose a flexible NLG approach for Spanish, focused on the surface realisation stage, which includes a novel and efficient inflection module that tackles the challenge of inflection generation using an ensemble of algorithms together with some hand-crafted rules; we contribute a high-quality dataset which includes instances of Spanish verbs for all the grammatical moods (in contrast with the current inflection approaches which do not tackle the imperative mood); our inflection module achieves 2% higher accuracy than the state-of-the-art methods; and finally, the proposed method achieves significant improvement of the perceived naturalness of the generated language in terms of coherence, grammaticality and post-editing.

In the next section (Section 2), we refer to the related work on inflection generation in general and on inflection within NLG systems. In Section 3, we describe our overall surface realisation approach which consists of three modules, including the proposed inflection module for NLG. In Section 4, we

present the experimental setup for testing the inflection module both with automatic metrics against state-of-the-art approaches and with a user evaluation, and in Section 5, we discuss the results. Finally, in Section 6, directions for future work are discussed.

2 Related Work

An NLG system comprises a wide range of modules, commonly grouped into a pipeline of three broad stages: document planning, microplanning and surface realisation (Reiter and Dale, 2000). The latter is responsible for generating the output text in natural language, which includes the morphological realisation in order to make the generated language more natural and expressive.

Most of the existing NLG systems usually work with English where the morphological realisation does not represent a problem because of its simplicity. It can be addressed using existing libraries as in Khan et al. (2015) where the SimpleNLG software (Gatt and Reiter, 2009) is used to generate sentences from predicate argument structures. On the other hand, previous work has addressed the morphology employing information extracted from lexicons (Androutsopoulos et al., 2013) or has not included it (Ballesteros et al., 2015).

The grammatical richness of the Spanish language is a challenge for NLG. Existing methods to automatically learn/predict the inflection of the verbs for morphologically rich languages have used supervised or semi-supervised learning (Durrett and DeNero, 2013; Ahlberg et al., 2014; Nicolai et al., 2015; Faruqui et al., 2016) to learn morphological rules on word forms in order to inflect the desired words. Other approaches have relied on linguistic information, such as morphemes and phonology (Cotterell et al., 2016); morphosyntactic disambiguation rules (Suárez et al., 2005); and graphical models (Dreyer and Eisner, 2009).

Recently, the morphological inflection has been also addressed at SIGMORPHON 2016 Shared Task (Cotterell et al., 2016) where, given a lemma with its part-of-speech, a target inflected form had to be generated (Task 1). This task was addressed through several approaches, including align and transduce (Alegria and Etxeberria, 2016; Nicolai et al., 2016;

Liu and Mao, 2016); recurrent neural networks (Kann and Schütze, 2016; Aharoni et al., 2016; Östling, 2016); and linguistic-inspired heuristics approaches (Taji et al., 2016; Sorokin, 2016). Overall, recurrent neural networks approaches performed better, being (Kann and Schütze, 2016) the best performing system in the shared task, obtaining around 98%. For the purpose of this task, a dataset in 10 languages, including Spanish[1], was provided. This dataset consisted of examples of word forms with its corresponding morphosyntactic descriptions.

Finally, the work described here differs from existing statistical surface realisation methods which use phrase-based or n-grams learning (e.g., (Konstas and Lapata, 2012; Angeli et al., 2010)) since they do not include morphological inflection. In this respect, our work is more similar to (Dušek and Jurčíček, 2013), where the inflected word forms are learnt through multi-class logistic regression by predicting edit scripts; and (Bohnet et al., 2010) where a statistical morphology generator (evaluated for English, Spanish, German and Chinese) was employed as a part of a support-vector-machines based surface realiser from semantic structures.

3 Surface Realisation Approach

This section describes the overall surface realisation approach for Spanish. The approach is divided in three modules as shown in Figure 1: (1) vocabulary selection, (2) generation of related sentences and (3) inflection generation.

The vocabulary selection module chooses the vocabulary that is used for text generation. This vocabulary is then used to generate a set of related sentences in lemma forms with the chosen content words (i.e. all of the words contained in the sentences are lemmas). This content is then used to generate related sentences in lemma form that also contain terms included in the previous sentence. Finally, the inflection module inflects all the content of the sentences generated into inflected sentences, that will be the final output of the approach.

---

[1] The dataset also includes the imperative mood, however it does not include examples for all the tenses for a concrete verb. For instance, which corresponding, for the verb "abacorar" there are only 2 examples which corresponding to the first person plural of the present of indicative. Therefore, we opted to create a wide-coverage dataset for this work.
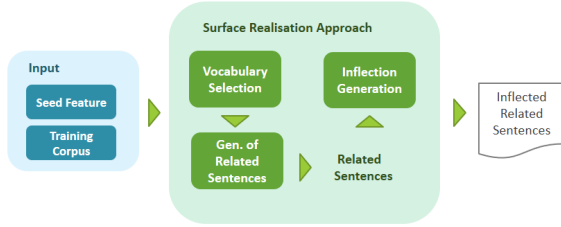
Figure 1: Diagram of the surface realisation approach divided in 3 modules: vocabulary selection, generation of related sentences and inflection generation.

3.1 Vocabulary Selection

As mentioned, the proposed approach mainly focuses on the surface realisation stage. Therefore, the content and the vocabulary needed to generate a sentence are determined by the input corpora and input *seed feature*. In general, a seed feature is an abstract object (which can be anything such as a topic, a sentiment, etc.) that is used to guide the generation process in relation to the most suitable vocabulary and content that the generated text will contain for a given domain (Barros and Lloret, 2015). The aim of the generated sentences is to meet the requirements expressed with this seed feature (e.g. to contain the maximum number of words for a specific phoneme, or to generate an opinionated sentence (Barros and Lloret, 2017)).

In this work, we selected phonemes (i.e. a small set of units different for each language, considered to be the basic distinctive units of speech by which morphemes, words, and sentences are represented) as the seed feature employed during the generation process. This approach will be useful in the context of assistive technologies for users with language impairments (e.g. dyslalia[2]) and the choice of the seed features is based on the end system. The generated sentences will contain the maximum number of words with the phoneme employed as the seed feature. These words are obtained from a part of the training corpus and stored into a bag of words that is used during the generation process. For example, for phoneme /d/, the bag of words could contain the words *delfín*-dolphin- or *dormir*-to sleep. The vocabulary contained in the bag of words is used to

---

[2]A disorder in phoneme articulation which implies the inability to correctly pronounce certain phonemes or groups of phonemes.

guide the sentences to be generated as it will be explained in the next section.

3.2 Generation of Related Sentences

This module generates a set of related sentences whose words are in lemma form by choosing the words of a sentence using over-generation and ranking techniques (Barros and Lloret, 2016). Starting from a training corpus, an input seed feature and a bag of words with the vocabulary, a factored language model is learnt over it. Factored language models (FLM) are an extension of language models, proposed by Bilmes and Kirchhoff (2003), where a word is viewed as a vector of k factors such that $w_t \equiv \{f_t^1, f_t^2, \ldots, f_t^K\}$. These factors can be anything, including lemmas, stems, words or any other lexical, syntactic or semantic features. Our approach uses the lemmas and Part-of-Speech (POS) tags as factors, due to the variability that they can bring to the generated sentences. The words chosen for generation will be in a lemma form, therefore, they are able to be further inflected to improve the naturalness and expressivity of the generated language. Furthermore, for the purpose of this research, a simple grammar (based on the basic structure that divides a sentence into subject, verb and object), shown in Figure 2, is used to guarantee the appearance of some elements in the generated sentence.

$$S \rightarrow NP\ VP$$
$$NP \rightarrow D\ N$$
$$VP \rightarrow V\ NP$$

Figure 2: Basic clause structure grammar.

In order to generate a set of related sentences with related content, we first generate independently a sentence by employing over-generation techniques, where a set of candidate sentences is generated based on the probabilities given by the FLM and the seed feature selected. The generation process prioritises the selection of words from the bags of words in order to ensure that the generated sentences will contain the maximum number of words related with the input seed feature.

These candidate sentences are subsequently ranked in order to select one, based on the sentence probability. The probability of a sentence is computed by the chain rule where the probability can be

calculated as the product of the probability of all the words:

$$P(w_1, w_2...w_n) = \prod_{i=1}^{n} P(w_i|w_1, w_2...w_{i-1}) \quad (1)$$

As suggested in Isard et al. (2006), the probability of a word is determined as the linear combination of FLMs, where a weight λ_i was assigned for each of them:

$$P(f_i|f_{i-2}^{i-1}) = \\ \lambda_1 P_1(f_i|f_{i-2}^{i-1})^{1/n} + \cdots + \lambda_n P_n(f_i|f_{i-2}^{i-1})^{1/n} \quad (2)$$

where f is the selected factors from the different FLMs and the total sum of the weights is 1.

After one sentence is generated, we perform pos-tagging, syntactic parsing and semantic parsing to identify different linguistic components of the sentence content (e.g. the subject, the object, name entities, etc.). To generate a sentence with related content, one of the identified linguistic components is chosen to influence the generation of the next sentence. This linguistic component is included in the sentence replacing the same type of linguistic component included in the next related sentence or it is used as a guide to select the content of the bag of words. For example, if the sentence generated by this module is "*mi padre tocar el suelo*"(my father to touch the ground), after performing the analysis of its content, this module would identify as linguistic components the subject ("*mi padre*"- my father) and the object ("*el suelo*" - the ground) of the sentence. The module would choose one of the two identified linguistic components, and would generate a sentence related to "*mi padre*"(my father) or a sentence related to "*el suelo*"(the ground).

Then, the remaining sentences are generated based on the probabilities estimated by the FLM, the input seed feature and the information extracted from the linguistic component.

3.3 Inflection Generation

After the set of related sentences are generated by the previous module, they are inflected by the last module integrated within the surface realisation approach (Barros et al., 2017). At this stage, we only address the inflection of Spanish verbs using supervised learning because of their complexity. The inflection of other simpler word types (e.g. determin-

ers, noun, adjectives, etc.) is done through a rule-based approach in order to ensure the gender and number concordance. In order to learn the inflection of Spanish verbs, we first created a dataset containing all the necessary information to inflect the verbs. The dataset was constructed by consulting the *Real Academia Española*[3] and the *Enciclopedia Libre Universal en Español*[4]. The dataset is composed of the following features: (1) *ending*, (2) *ending stem*, (3) *penSyl*, (4) *person*, (5) *number*, (6) *tense*, (7) *mood*, (8) *suff1*, (9) *suff2*, (10) *stemC1*, (11) *stemC2*, (12) *stemC3* (see also Table 1).

We considered that a word can be divided into three parts: (1) the *ending* (in Spanish the verbs are classified by their ending); (2) *ending stem* (i.e. the closest consonant to the ending); and (3) *penSyl* (i.e. the previous syllable of the ending formed by the whole syllable or its dominant vowel is extracted), as shown in Figure 3.

Suff1 and *suff2* are the inflection predicted for the suffix of the verb form; and *stemC1*, *stemC2* and *stemC3*, refer to the inflection predicted for the stem of the verb form.
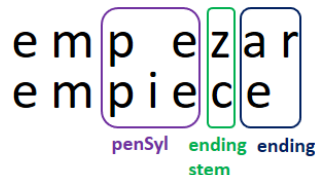

Figure 3: Division of the Spanish verb *to begin* and its inflection for the first singular person of the present tense and in the subjunctive mood.

We trained an ensemble of individual models for each of the features with a potential inflection value. We used the WEKA (Frank et al., 2016) implementation of the Random Forest algorithm to train the models for the *stemC3* and *stemC2* features, and the Random Tree algorithm to train the models for the *suff1*, *suff2* and *stemC1* features. We then predicted all the possible inflections given a verb in its base form, i.e., all the tenses for each mood in Spanish. For accomplishing this task, we first analysed the base form to extract the necessary features for the

Feature	Description
(1) *ending*	ending of the verb that can be "-ar", "-er" and "-ir", used to classify the verbs in 1st, 2nd, and 3rd conjugation respectively.
(2) *ending stem*	the closest consonant or group of letters to the ending, being part of the same syllable of the ending
(3) *penSyl*	the previous syllable of the ending, consisting of the whole syllable or the dominant vowel
(4) *person*	grammatical distinction between references to participants in an event, which can be 1st (the speaker), 2nd (the addressee) and 3rd (others) person
(5) *number*	grammatical category that expresses count distinctions, which can be singular (one) or plural (more than one)
(6) *tense*	category that expresses time reference, in Spanish there are 17 different verb tenses
(7) *mood*	grammatical features of the verbs used for denoting modality (statement of facts, of desire, of commands, etc.), in Spanish there are three different moods
(8) *suff1*	one of the possible inflections for the ending
(9) *suff2*	one of the possible inflections for the ending
(10) *stemC1*	one of the possible inflections for the stem
(11) *stemC2*	one of the possible inflections for the stem
(12) *stemC3*	one of the possible inflections for the stem

Table 1: Detailed description of features. A specific verb tense in Spanish can have more than one valid inflection, being necessary to predict each variant of the tense.

inflection, and then we predicted its predicted inflection using the models. Finally, the predicted inflections were employed to replace the features previously identified in the base form, leading to the reconstruction of the base form into the desired inflection, as it can be seen in Figure 4.
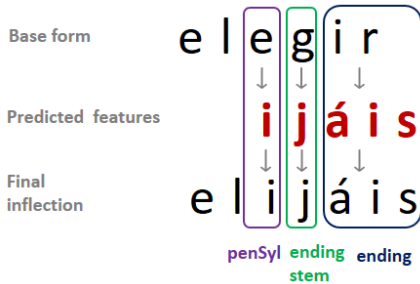

Figure 4: Reconstruction of the verb *elegir* (to choose) with the features predicted by the models.

4 Experiments

We performed two experiments: first, we tested the inflection module by comparing it against the state-of-the-art in order to ensure the accuracy for this task. Secondly, we generated and inflected the sentences using the whole surface realisation approach in order to test whether the quality of the generated sentences improved.

4.1 Experiments on Inflection Generation

For the first experiment, we compared our inflection module (RandFT) with two very competitive baselines by Durret13 (Durrett and DeNero, 2013) and Ahlberg14 (Ahlberg et al., 2014), by measuring the accuracy of their output for Spanish verb inflections under the same conditions. This experiment was done to validate the performing of the inflection module.

In order to compare our system with both baselines, we employed the test set of examples (200 different verbs) which was made available by Durrett and DeNero (2013), since this test-set included verbs with both irregular and regular forms. This test set does not include any of the entries used within our training dataset. For the experiments, we generated all the verb inflections for the 200 base forms.

Furthermore, the aforementioned baselines do not predict all the grammatical moods that exist in the Spanish language. Both baselines are only able to predict the indicative and subjunctive mood, but not the *imperative one*, which is complex especially for irregular forms. To tackle this, we used an additional test-set to evaluate this grammatical mood. We created the additional test-set by employing information from the Freeling's lexicon for the imperative forms of these 200 verbs (Padró and Stanilovsky, 2012).

4.2 Experiments on End-to-end NLG

For the second experiment, we integrated the inflection unit with the surface realisation approach described in Section 3 in order to test if the quality of the generated sentences improved. For this purpose, we generated a set of three related sentences for each Spanish phoneme (i.e. there are a total of 27 phonemes in Spanish). These sentences have related topics that will appear within the set so that the direct object of a sentence is used as the subject of the following sentence, obtaining a preliminary set of related sentences. We compared our realisation approach against a random baseline, where a random verb tense was assigned for each of the sentences forming the set. We set our proposed approach to a fixed tense, either present or indicative. These sentences were ranked according to the approach described in section 3.2 being the linear combination of FLM as follows: $P(w_i) = \lambda_1 P(f_i|f_{i-2}, f_{i-1}) + \lambda_2 P(f_i|p_{i-2}, p_{i-1}) + \lambda_3 P(p_i|f_{i-2}, f_{i-1})$, where f refers to a lemma, p refers to a POS tag, and λ_i are set $\lambda_1 = 0.25$, $\lambda_2 = 0.25$ and $\lambda_3 = 0.5$. These values were empirically determined by testing different values and comparing the results obtained.

For this experiment, we used a collection of Hans Christian Andersen tales, automatically gathered from *Ciudad Seva*[5], as a corpus. In order to train the FLM, used during the generation, we employed SRILM (Stolcke, 2002), a software that allows to build and apply statistical language models, which includes and implementation of FLM. In addition, we use Freeling language analyser (Padró and Stanilovsky, 2012) to tag the corpus with lexical information as well as to perform the analysis of the generated sentences.

5 Evaluation and Results

This section describes the results obtained with the experimentation carried out. First, the results obtained from the comparison of our inflection module in order to validate its performance are shown. Then, the results obtained from the integration of this module within the end-to-end NLG approach are described.

5.1 Results for the Inflection Module

The results obtained are shown in Table 2, where we compared the inflection of the same verb tenses as Durret and Ahlberg using the test-set described in the previous section. Our inflection module, which includes an ensemble of classifiers (RandFT), trained with our generalised dataset for Spanish, obtained a higher overall accuracy (but not significantly) with respect to the state-of-the-art baselines systems.

Approach	Correctly predicted verb tables	Correctly predicted verb forms
RandFT	**99**%	**99.98**%
Durret13	97%	99.76%
Ahlberg14	96%	99.52%

Table 2: Accuracy of predicting inflection of verb tables and individual verb forms given only the base form, evaluated with an unseen test set of 200 verbs. For the imperative mood, our system achieves 100% accuracy, however the baselines do not predict the imperative form.

Base form–*Inflected form*
contar–*cuenta*; **errar**–*yerra*; **haber**–*he*; **hacer**–*haz*; **oler**–*huele*; **ir**–*ve*; **oír**–*oye*; **decir**–*di*

Table 3: Variability of inflection in the imperative mood for the 2nd person singular of the present.

In addition, our model can correctly perform the inflection of the imperative mood, which was not included in the baseline systems. This grammatical mood, which forms commands or requests, contains unique imperative forms among the irregular Spanish verbs, as shown in Table 3. For this experiment, our system achieves 100% accuracy when evaluated on the additional test-set.

5.2 Results for the Generated Text

We also performed a user evaluation with three evaluators in order to discern if the inclusion of the inflection module improved the naturalness and expressivity of the language.

Each evaluator was shown 27 sets of sentences with different kinds of inflections (i.e. without inflecting the sentences, with a fixed inflection and with a random inflection, as described in Section
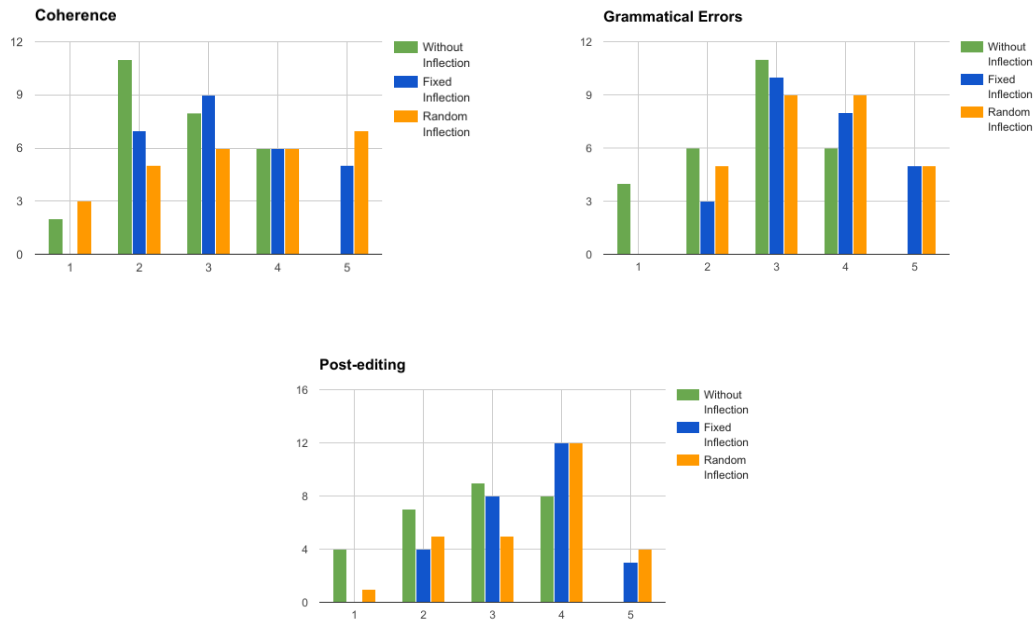
Figure 5: Number of sentences scored for each rating of the 5-pt Likert scale regarding the *coherence*, the *grammatical errors* and the *ease of correction*. The minimum values for the *coherence* indicate a lack of meaning of the sentences whereas the maximum values indicate a correct full meaning for a sentence. For the *grammatical errors* ratings, the minimum values represent a high number of errors in the sentences and the higher values indicate a lack of errors in the sentences. Finally, the minimum *ease of correction* values refers to a huge number of changes required to correct or improve the sentences while the maximum values indicates otherwise.

4.2) and had to overall rate each set using a 5-pt Likert scale, in terms of *coherence*, the *grammatical errors* an the *post-editing*. The *coherence*, which is very difficult to determine automatically being its analysis performed manually, refers to the meaning of the generated sentence, so that a sentence with no meaning would be rated with a 1 and a sentence with a full meaning would be rated with a 5. In contrast, the *grammatical errors* indicate the amount of errors in the sentence (i.e. fewer errors indicate a better sentence). The *post-editing* (ease of correction) refers to the amount of changes necessary to convert a sentence with many errors into one with no errors. In this sense, the lower values for the post-tagging indicates the need to make a lot of changes to the sentence whereas the higher values refers to not perform changes to the sentence. All the sentences contained in the sets were different since they were generated with each of the Spanish phonemes.

A summary of the results obtained can be seen in the Table 4. The results of inflecting a sentence in

Inflection Type	Coherence		Grammar errors		Post-editing	
	Mean	Mode	Mean	Mode	Mean	Mode
Without	2.65*	2	2.73*	3	2.75*	3
Fixed	**3.36***	3	**3.57***	3	**3.54***	4
Random	3.31*	5	3.51*	4	3.48*	4

Table 4: Results of the means and the modes of the 5-pt Likert scale with respect to the *coherence*, *grammatical errors* and *ease of correction*, of the inflected generated sentences. * denotes significance with $p < 0.01$.

contrast to not inflecting it are better, indicating that the quality of the generated sentences improved.

Figure 5 summarises the number of set of sentences (i.e. one set per phoneme) derived from the evaluation of the ratings mentioned before. As can be seen in the figure, the sentences without inflection are less coherent than the inflected sentences (both fixed and random inflection). In terms of grammaticality and ease of correction (post-tagging), the non-inflected sentences score lower than the inflected

sentences. These ratings in concordance to the results given in Table 4 demonstrate the improvement of the quality obtained after applying inflection to the generated sentences. In contrast, the ratings obtained from the fixed inflection and random inflection are quite similar, standing out the ratings of this last one in *coherence*. This is due to the fact that the inflection of the verb is the only thing that can be random or fixed in the sentence; however a sentence can be meaningful with more than one verb tense. For instance, consider the following two sentences: "I am in the ground" and "I was on the ground". Both are meaningful, and grammatically correct.

Phoneme: /n/
Without Inflection
Cuánto cosa tener nuestro pensamiento.
(How much thing to have our thinking.)
Cuánto pensamiento tener nuestro corazón.
(How much thought to have our heart.)
Cuánto corazón tener nuestro pensamiento.
(How much heart to have our thinking.)

Fixed Inflection
Cuánta cosa tiene nuestro pensamiento.
(How much thing our thinking has.)
Cuánto pensamiento tiene nuestro corazón.
(How much thought our heart has.)
Cuánto corazón tiene nuestro pensamiento.
(How much heart our thinking has.)

Random Inflection
Cuánta cosa tiene nuestro pensamiento.
(How much thing our thinking has.)
Cuánto pensamiento tuviere nuestro corazón.
(How much thought our heart had.)
Cuánto corazón tenga nuestro pensamiento.
(How much heart our thinking had.)

Figure 6: Example of a generated set of sentences for the phoneme /n/ without inflection, with a fixed inflection (e.g. in present of indicative) and with random inflection (first sentence in present of indicative, second sentence in present of subjunctive and third sentence in past imperfect of subjunctive).

Some examples of the inflection of an automatically generated set of sentences by the described approach are shown in Figure 6.

5.3 Discussion

With the experimentation carried out, on the one hand, the inflection module obtained almost 100% of accuracy, being able to inflect almost all the verbs in Spanish. On the other hand, the introduction of an inflection module in a surface realisation approach improves the generated language. This inflection approach could be further used in phrase-based NLG systems (i.e. systems trained to generated text based on n-grams rather than linguistic rules), in order to enhance the naturalness, grammaticality and coherence of the generated text. However, at this stage, while the NLG without the inflection module is language independent, the inflection module is only able to learn the inflection for Spanish verbs.

6 Conclusion and Future Work

This paper presented a flexible and domain independent NLG approach for Spanish focused on the surface realisation stage. Within the NLG approach, a robust light-weight supervised inflection module to obtain the inflected form of any Spanish verb for any of its moods (indicative, subjunctive and imperative) was integrated. This inflection module obtained accuracy close to 100%, outperforming existing state-of-the-art approaches. In addition, the integration of this inflection module within a surface realisation approach improves the quality of the generated sentences, adding naturalness and expressivity to the generated language. In the future, we plan to learn the inflection for other types of words (not only verbs), seeking for a whole sentence inflection model. Moreover, we will test this inflection approach to other languages.

References

Roee Aharoni, Yoav Goldberg, and Yonatan Belinkov. 2016. Improving sequence to sequence learning for morphological inflection generation: The biu-mit systems for the sigmorphon 2016 shared task for morphological reinflection. In *Proceedings of the 14th Annual SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 41–48. Association for Computational Linguistics.

Malin Ahlberg, Markus Forsberg, and Mans Hulden. 2014. Semi-supervised learning of morphological paradigms and lexicons. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 569–578.

Iñaki Alegria and Izaskun Etxeberria. 2016. Ehu at the sigmorphon 2016 shared task. a simple proposal: Grapheme-to-phoneme for inflection. In *Proceedings of the 14th Annual SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 27–30. Association for Computational Linguistics.

Ion Androutsopoulos, Gerasimos Lampouras, and Dimitrios Galanis. 2013. Generating natural language descriptions from owl ontologies: the naturalowl system. *Journal of Artificial Intelligence Research*, 48:671–715.

Gabor Angeli, Percy Liang, and Dan Klein. 2010. A simple domain-independent probabilistic approach to generation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 502–512, Cambridge, Massachusetts.

Miguel Ballesteros, Bernd Bohnet, Simon Mille, and Leo Wanner. 2015. Data-driven sentence generation with non-isomorphic trees. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 387–397. Association for Computational Linguistics.

Cristina Barros and Elena Lloret. 2015. Input seed features for guiding the generation process: A statistical approach for Spanish. In *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*, pages 9–17. Association for Computational Linguistics.

Cristina Barros and Elena Lloret. 2016. Generating sets of related sentences from input seed features. In *Proceedings of the 2nd International Workshop on Natural Language Generation and the Semantic Web (WebNLG 2016)*, pages 1–4. Association for Computational Linguistics (ACL).

Cristina Barros and Elena Lloret. 2017. A multilingual multi-domain data-to-text natural language generation approach. *Procesamiento del Lenguaje Natural*, 58(0):45–52.

Cristina Barros, Dimitra Gkatzia, and Elena Lloret. 2017. Inflection generation for spanish verbs using supervised learning. In *Proceedings of the 1st Workshop on Subword and Character LEvel Models in NLP (SCLeM)*. Association for Computational Linguistics.

Jeff A. Bilmes and Katrin Kirchhoff. 2003. Factored language models and generalized parallel backoff. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Companion Volume of the Proceedings of HLT-NAACL 2003–short Papers - Volume 2*, pages 4–6.

Bernd Bohnet, Leo Wanner, Simon Mille, and Alicia Burga. 2010. Broad coverage multilingual deep sentence generation with a stochastic multi-level realizer. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 98–106. Association for Computational Linguistics.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden. 2016. The sigmorphon 2016 shared task—morphological reinflection. In *Proceedings of the 2016 Meeting of SIGMORPHON*. Association for Computational Linguistics.

Markus Dreyer and Jason Eisner. 2009. Graphical models over multiple strings. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, pages 101–110. Association for Computational Linguistics.

Greg Durrett and John DeNero. 2013. Supervised learning of complete morphological paradigms. In *Proceedings of the North American Chapter of the Association for Computational Linguistics*, pages 1185–1195.

Ondřej Dušek and Filip Jurčíček. 2013. Robust multilingual statistical morphological generation models. In *51st Annual Meeting of the Association for Computational Linguistics: Proceedings of the Student Research Workshop*, pages 158–164. Association of Computational Linguistics.

Manaal Faruqui, Yulia Tsvetkov, Graham Neubig, and Chris Dyer. 2016. Morphological inflection generation using character sequence to sequence learning. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 634–643.

Eibe Frank, Mark A. Hall, and Ian H. Witten. 2016. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. Morgan Kaufmann, 4 edition.

Albert Gatt and Ehud Reiter. 2009. Simplenlg: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG)*.

Amy Isard, Carsten Brockmann, and Jon Oberlander. 2006. Individuality and alignment in generated dialogues. In *Proceedings of the Fourth International Natural Language Generation Conference*, pages 25–32. Association for Computational Linguistics.

Katharina Kann and Hinrich Schütze. 2016. Med: The lmu system for the sigmorphon 2016 shared task on morphological reinflection. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 62–70. Association for Computational Linguistics.

Atif Khan, Naomie Salim, and Yogan Jaya Kumar. 2015. A framework for multi-document abstractive summarization based on semantic role labelling. *Applied Soft Computing*, 30:737 – 747.

Ioannis Konstas and Mirella Lapata. 2012. Concept-to-text generation via discriminative reranking. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 369–378.

Ling Liu and Lingshuang Jack Mao. 2016. Morphological reinflection with conditional random fields and unsupervised features. In *Proceedings of the 14th Annual SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 36–40. Association for Computational Linguistics.

Garrett Nicolai, Colin Cherry, and Grzegorz Kondrak. 2015. Inflection generation as discriminative string transduction. In *NAACL HLT 2015, The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 922–931.

Garrett Nicolai, Bradley Hauer, Adam St. Arnaud, and Grzegorz Kondrak. 2016. Morphological reinflection via discriminative string transduction. In *Proceedings of the 14th Annual SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 31–35. Association for Computational Linguistics.

Robert Östling. 2016. Morphological reinflection with convolutional neural networks. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 23–26. Association for Computational Linguistics.

Lluís Padró and Evgeny Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Eight International Conference on Language Resources and Evaluation*.

Ehud Reiter and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.

Alexey Sorokin. 2016. Using longest common subsequence and character models to predict word forms. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 54–61. Association for Computational Linguistics.

Andreas Stolcke. 2002. Srilm - an extensible language modeling toolkit. In *Proceedings International Conference on Spoken Language Processing, vol 2.*, pages 901–904.

Octavio Santana Suárez, José Rafael Pérez Aguiar, Luis Javier Losada García, and Francisco Javier Carreras Riudavets. 2005. Spanish morphosyntactic disambiguator. In *Proceedings of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing*, pages 201–204.

Dima Taji, Ramy Eskander, Nizar Habash, and Owen Rambow. 2016. The columbia university - new york university abu dhabi sigmorphon 2016 morphological reinflection shared task submission. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 71–75. Association for Computational Linguistics.