Procesamiento del Lenguaje Natural, Revista nº 66, marzo de 2021, pp. 89-98 recibido 29-11-2020 revisado 21-01-2021 aceptado 31-01-2021

Using Guarani Verbal Morphology on Guarani-Spanish Machine Translation Experiments

Uso de la Morfología Verbal del Guaraní en Experimentos de Traducción Automática Guaraní-Español

Yanina Borges, Florencia Mercant, Luis Chiruzzo
Universidad de la República
Montevideo, Uruguay
{yanina.borges.mijailovich, florencia.mercant, luischir}@fing.edu.uy

Abstract: This paper shows the results of a project for building computational tools and resources for processing the Guarani language, an under-researched language in the NLP community. We developed a baseline machine translation system for the Guarani-Spanish pair, and then performed a series of experiments trying to improve its quality using morphological information. In this work we focus on the analysis of verbs, which is the most complex part of speech in Guarani. We report the results of the different tools implemented for verbs analysis and detection in Guarani, as well as the experiments on machine translation carried on using different versions of the corpus augmented with morphological features.
Keywords: Guarani, Verbal Morphology, Spanish, Machine Translation.

Resumen: Este artículo muestra los resultados de un proyecto para construir herramientas y recursos computacionales para procesar el idioma guaraní, un idioma poco explorado por la comunidad de PLN. Se desarrolló una línea base de traducción automática para el par Guaraní-Español, y luego se realizaron una serie de experimentos para intentar mejorar la calidad de esta línea base utilizando información morfológica. Este trabajo se enfoca en el análisis de los verbos, los cuales componen la categoría gramatical más compleja en el idioma guaraní. Se reportan los resultados de las distintas herramientas implementadas para el análisis y detección de verbos en guaraní, así como los experimentos sobre traducción automática hechos sobre diferentes versiones del corpus aumentado con atributos morfológicos.
Palabras clave: Guaraní, Morfología Verbal, Español, Traducción Automática.

1 Introduction

Guarani is a language spoken by 12 million people in several Latin American countries, mainly in Paraguay, Argentina, Brazil and Bolivia. Paraguay is the country that speaks the Guarani language the most according to a 2002 census[1], and is one of the official languages in the country. In our case we will focus on the Jopara dialect, which is the current variant of Guarani spoken in Paraguay and includes many Spanish loanwords and neologisms (Estigarribia, 2015; Lustig, 2010).

Processing the Guarani language presents particular challenges because it is largely under-researched in the NLP community. There are very few tools or resources built for this language. Guarani is a morphologically rich language, it is agglutinative and polysynthetic (Estigarribia and Pinta, 2017), with words formed by combining prefixes and suffixes around a root, and often the roots or lemmas could be used to form different parts of speech.

In this work we focus on studying the grammar of verbs, which is the part of speech with the greatest complexity in Guarani. We also focus on building tools for the Guarani-Spanish language pair as Spanish is the second most spoken language in Paraguay and is comparatively a much more researched language in the NLP community. The main contributions of this work to the processing of Guarani can be summarized as follows:

- We propose a rule-based method for morphological analysis of verbs in Guarani.

[1] https://www.paho.org/English/DD/AIS/cp_600.htm

ISSN 1989-7553. DOI 10.26342/2021-66-7

- We implement two approaches for detecting verbs in Guarani: a rule-based method and a probabilistic system based on Hidden Markov Models.
- We create a baseline Guarani-Spanish machine translation system based on neural networks, and given the corpus we use is very small and consequently the translation quality is poor, we perform several experiments incorporating during training the linguistic knowledge extracted with the aforementioned methods to improve the results.

2 Related work

There are comparatively very few works that focus on applying NLP techniques to the Guarani language, and in particular few works on creating corpora and resources for this language. The corpus COREGUAPA (Secretaría de Políticas Lingüísticas del Paraguay, 2019) is a monolingual reference corpus of current Guarani, there is also a small corpus of Mbya Guarani sentences (which is a very different dialect from Jopara) annotated using the Universal Dependencies framework (Thomas, 2019; Dooley, 2006), and there have been attempts at creating bilingual Guarani-Spanish or Guarani-English corpora from Wikipedia, although the Guarani version of Wikipedia is itself very small.

There have been some attempts at creating machine translation or translation support systems that focus on the Guarani-Spanish pair, for example: (Gasser, 2018) describes a system for computer aided translation between Guarani and Spanish which uses morpho-syntactic rules to find translation candidates; similarly (Rudnick et al., 2014) presents a web system for collaborative translation between Guarani and Spanish with the aim of creating a parallel corpus; in (Abdelali et al., 2006) they describe a project for developing resources for a Guarani-English corpus using Spanish as a bridge language, and more recently (Alcaraz and Alcaraz, 2020) describes a web tool for analyzing Guarani sentences using a rule-based grammar approach and translating between Guarani and Spanish using an example-based method. However, none of these systems present an evaluation of results, and the systems that are readily available only perform simple translations at word level.

We differ from these works in that we collected a small corpus of parallel Guarani-Spanish documents from the web and tried to apply linguistically motivated techniques to improve the performance of a neural machine translation system over this corpus. As far as we know, this work is the first one to present a neural machine translation baseline for the Guarani-Spanish pair, and to try to improve it using morphological features. The use of morphological features to aid in machine translation for morphologically rich languages has been tried in the past for languages such as Turkish, Russian, Kazakh and Arabic (Bisazza and Federico, 2009; Myrzakhmetov and Makazhanov, 2016; El-Kahlout et al., 2019), being a helpful technique for low resourced languages.

3 Analyzing Guarani verbs

This section presents an introduction to Guarani verbal morphology and describes the tools we developed to approach the automated processing of Guarani verbs.

3.1 Introduction to Guarani verbal morphology

As mentioned before, Guarani is an agglutinative and polysynthetic language: its words con be generated by combining many different types of prefix and suffix morphemes around a root, and these morphemes generally act as independent units, i.e. they have their own meaning and do not change when they are attached to words (Academia de la Lengua Guaraní (ALG), 2018). For example, the inflected verb *"aguata"*, meaning "I walk", can be analyzed as the prefix *"a"* and the lemma *"guata"* which corresponds to the verb "walk". The prefix indicates the person who is performing the action, in this case the singular first person.

Verbs in Guarani can be classified into two groups based on the root word or lemma: proper verbs and verbalized lexical categories. Proper verbs have verbal roots, while verbalized lexical categories use a noun, adjective or adverb as root. The previous example *"aguata"* is a proper verb as it has the verbal root *"guata"* (to walk), while the verb *"amitã"* (I am a child) uses as root the noun *"mitã"* (child), so it is a verbalized noun.

The root of a verb gives meaning to the word, and other morphemes indicate the different verbal inflections. Guarani uses five main grammar categories for inflecting verbs, indicated by different verbal affixes: **number and person** (it works as a single prefix), **form**, **voice**, **mood** and **tense**.

Number and person: Guarani uses two grammatical numbers (singular and plural) and three grammatical persons (first, second and third), while the first person plural is further subdivided in two types whether the action of the verb includes the interlocutor or not (inclusive and exclusive). There are a total of 43 prefixes that are used to denote number and person.

Form: The verb form indicates if the action of the verb is affirmative, negative or a question. Interrogative forms do not need a particular symbol to be a question, for example, to ask “Do you walk?” we would say *“reguatapa”*, with the root *“guata”* (walk), the second person plural prefix *“re”*, and the interrogative suffix *“pa”*. There are four suffixes that could be used for indicating a question. Affirmative forms use the base form of the verb and do not add an additional morpheme. Negative forms require both a prefix and a suffix around the root. For example, the phrase “I don’t walk” would be *“ndaguatai”*, with the prefixes *“nd”* and *“a”* (for first person singular), then the root *“guata”* and the suffix *“i”*. The prefix *“nd”* with the suffix *“i”* indicate the negative form of the verb. There are eight combinations of prefix and suffix to indicate negative verbal forms.

Voice: Establishes the relationship between the subject and the action of the verb in a sentence, it could be either passive voice or active voice. For example, the verb *“oñembohérakuri”*, meaning “he was named”, is in the passive voice indicated by the prefix *“ñe”*. First, we find the prefix *“o”* that indicates the third person singular form, then the prefix *“ñe”*, after this the root *“mbohéra”* which means “name” and finally the suffix *“kuri”* indicates recent past. This verb is in the simple indicative mood because there is no morpheme between the root and the recent past morpheme. There are ten prefixes that could be used to indicate the voice category.

Mood: Indicates the way in which the action is performed. There are two main types: indicative or imperative. For example, the verb *“opu’ãva”*, meaning “he (usually) wakes up”, is in the usual indicative mood indicated by the suffix *“va”*. First, we find the prefix *“o”* that indicates the third person singular form, then the root *“pu’ã”* meaning “wake up” and finally the suffix *“va”*. This verb is in the simple indicative mood and simple present because there is no morpheme for these verbal categories. There are 46 possible suffixes for the indicative mood and eight for the imperative mood.

Tense: Indicates the time when the action of the verb is done. The basic tenses are present, past and future, but they have variants. There are 14 suffixes used to denote tense, no suffix is used for indicating present tense.

Verbal affixes always appear in the same order. To the left of the root we have the prefixes of form (only for negation), number and person, and voice. To the right of the root we have the suffixes of mood, form and tense. If the form is negative, the form prefix must match the number and person. Figure 1 shows this order graphically.

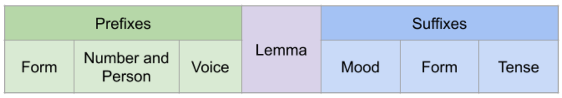

Figure 1: Canonical order of verbal affixes.

3.2 Manual annotation of the corpus

We used a small parallel corpus of around 14,500 sentence pairs extracted from the web. It has around 228,000 tokens in Guarani, corresponding to 336,000 tokens in Spanish, and it consists of articles (news, blog posts and stories) from Paraguayan websites. The corpus was aligned with a semi-automatic process: first aligning it automatically and then manually fixing the incorrect alignments (Chiruzzo et al., 2020). The news articles comprise the majority of the corpus, but they also present more noise in their translations, while the translation quality of the blog posts and stories is better. The corpus was divided into three sets for training, development and test of around 90 %-10 %-10 % keeping the same ratio between news, blog posts and stories for the three subsets.

We manually annotated a fraction of the corpus identifying all verbs and their morphological features. This was done for the blog posts and stories articles as the text was less noisy with a greater number of Guarani words and fewer Spanish loanwords or neologisms. There are 1,015 verbs in the training set, 412 in the development set, and 477 in the test set.

3.3 Verbal morphology analysis

We use the following rule-based heuristic for performing the morphological analysis of verbs: given a verb, we split it up as a concatenation of valid prefixes and suffixes in the possible order of appearance (see figure 1). We must consider some restrictions imposed by rules such as negation, where the use of a negation prefix restricts the number and person prefixes to be used.

The central part of the verb that is not recognized as a valid prefix or suffix is considered the root of the verb. This root is tested against a dictionary of valid verbs and words extracted from a web source[2].

When analyzing a word, we consider all possible combinations of prefixes, root and suffixes extracted in this way, and we sort the options based on the following priorities:

1. The resulting root or word is found in the dictionary and has a definition associated with a verb.
2. The resulting root is found in the dictionary and it is not a verb, but some grammatical rule of verbal affixes was applied.
3. The resulting root is found in the dictionary but it is not a verb and no grammatical rule of verbal affixes was applied.
4. The resulting root is not found in the dictionary, but some grammatical rule of verbal affixes was applied.
5. The resulting root is not found in the dictionary and no grammatical rule of verbal affixes was applied.

The process takes a verb and returns the analysis that has the highest priority in this list (where 1 is the highest).

We defined the following metrics to evaluate this method:

- Exact accuracy: Strict metric calculated as the number of the analyses that are exactly the same as expected in the gold standard, divided by the total number of verbs. The exact accuracy is defined in equation 1, with n verbs in the corpus, we define y_i with $i \in \{1, \ldots, n\}$ where $y_i = 1$ if the classification of the ith word is correct, $y_i = 0$ otherwise.

$$e\_accuracy = \frac{y_1 + y_2 + ... + y_n}{total\_number\_of\_verbs} \quad (1)$$

- Relaxed accuracy: It is a more relaxed metric that allows to evaluate the accuracy of the tag sequence found. To calculate this measure, we average number of hits in the tag sequence obtained for each word, that is, the number of correct tags divided by the number of tags in the sequence. The length of the tag sequence for each verb is fixed at six, since it corresponds to the five possible verbal affixes plus the lemma. The relaxed accuracy is defined in equation 2, given n verbs in the corpus, we define e_{ij} with $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, 6\}$, where $e_{ij} = 1$ if the jth tag of the ith verb was correctly classified, otherwise it is $e_{ij} = 0$.

$$r\_accuracy = \frac{e_{11}+...+e_{16}+...+e_{n1}+...+e_{n6}}{6 \cdot total\_number\_of\_verbs} \quad (2)$$

We considered two variants of these metrics: one of them (original lemma) is considering all the possible lemmas, and the other one (tagged lemma) is substituting the lemmas for a tag `LEMAESVERBO` (the lemma is a verb) or `LEMANOESVERBO` (the lemma is not a verb) whether the expected lemma of the verb is present as a verb in the dictionary or not. Table 1 shows the values for the metrics over the development and test sets.

Accuracy		Dev	Test
Exact	Original lemma	0.436	0.310
	Tagged lemma	0.386	0.304
Relaxed	Original lemma	0.751	0.615
	Tagged lemma	0.745	0.621

Table 1: Accuracy results for the rule-based method.

As expected, in all cases the relaxed accuracy is greater than the exact accuracy. The exact accuracy for both experiments (original lemma and tagged lemma) is around 0.3 for the test set, which is quite low, so using these

[2]http://descubrircorrientes.com.ar/2012/index.php/diccionario-guarani

Method	Precision		Recall		F1 Score		Accuracy	
	Dev	Test	Dev	Test	Dev	Test	Dev	Test
Rule-based method	0.574	0.611	0.838	0.716	0.681	0.660	0.874	0.862
HMM - original words	0.975	0.872	0.191	0.069	0.319	0.128	0.871	0.871
HMM - original lemma	0.975	0.891	0.191	0.083	0.319	0.153	0.871	0.871
HMM - tagged lemma	0.822	0.859	0.546	0.510	0.656	0.640	0.909	0.909

Table 2: Verb detection results.

methods for morphological analysis of verbs as the only tool would not be appropriate.

However, relaxed accuracy returns values greater than 0.6 in all cases, which indicates that in many cases, although the morphological analysis is not exactly the same as expected, several verbal affixes of the word are correctly labeled, which might indicate the method could be used to detect verbs as we will see in the following section.

Although this method has a lot of room for improvement, we consider it gives us starting point for Guarani verbs morphological analysis.

3.4 Verbs detection

We use two approaches for verb detection, first using the rule-based verb analysis method discussed above and then a machine learning method based on Hidden Markov Models (HMM).

3.4.1 Rule-based method

In order to know if a word is a verb or not, we take all the possible combinations found in the previous analysis and consider the word is a verb if at least an analysis of priority 1 or 2 is found. This means, we consider the word a verb if it can be split as a concatenation of valid prefixes, root and suffixes that honors the affixes combination rules and the root is found in the dictionary (either as a verb or another valid category).

We repeat this process for every word in the sentence, tagging all possible appearances of verbs.

3.4.2 Hidden Markov Models

We cast this problem as a sequence labeling problem, in which we take a sequence of words and we want to output a sequence of labels `VERB` or `NOT-VERB`. Notice that this can be seen as a simplified version of POS-tagging, in which we are only focusing on one part of speech (verbs). Because the size of the annotated corpus we have is rather small, we decided to use Hidden Markov Models, a statistical method that has proven to be very good at solving this kind of problems, and is comparatively less data intensive than more modern methods like Recurrent Neural Networks. We used the libraries `nltk` (Bird and Klein, 2009) and `sklearn` (Pedregosa et al., 2011) for implementing these methods. We trained the following variants:

HMM with original words: The standard HMM experiment was trained using all the original words from the sentences.

HMM with original lemma: In this experiment, instead of the words as they appear in the corpus, they are represented as a concatenation of labels representing their verbal affixes plus a label with the lemma. We used the morphological analysis described in 3.3 applied to each word.

HMM with tagged lemma: Finally, we carried out the same experiment but using a variant of the morphological analysis in which the verbal affixes are concatenated but instead of the original lemma we use the `LEMAESVERBO` or `LEMANOESVERBO` labels that represents whether or not the resulting lemma is a verb according to the dictionary. This representation might help reduce the data sparsity problem.

These last two experiments could be considered a hybrid model between the rule-based method and the Hidden Markov Model.

3.4.3 Results for verbs detection

All these experiments where trained over the training set and adjusted for the development set, then evaluated against the test set. Table 2 shows the Precision, Recall, F1 Score and Accuracy results obtained for the different verb detection methods.

Except for accuracy, the other metrics are calculated considering the positive class, that is, we calculate the performance for detecting verbs, where the majority of tokens in the corpus represent non-verb instances. Figure 2 shows a comparison of the different methods implemented for these metrics.

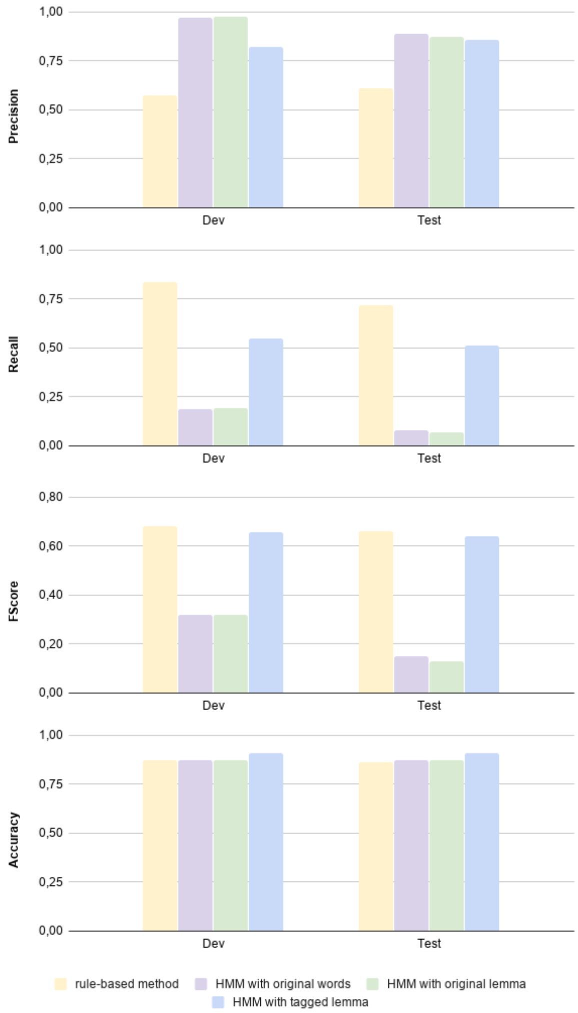

Figure 2: Performance for verbs detection.

From the figure we can see that the methods based on Hidden Markov Model with original lemma and with original words are the ones that return the best precision values. This indicates that these methods tend to get the result right when they classify a word as a verb.

For recall, the results of the rule-based method are much higher than the other methods. This indicates that it tends to classify more instances as verbs, while HMM-based methods tend to classify less (but more precisely). The recall results for the method based on HMM with original lemma and original words are much lower than the others, however with these methods the highest precision was obtained, this indicates that it is a method that tends to be correct when classifies a word as a verb, but given its recall it follows that it tends to classify very few words as verbs.

We consider that the F1 Score measure is the one that best reflects the performance for verbs detection, as it captures the trade-off between precision and recall. The values for the rule-based method and the HMM-based method with tagged lemma are similarly around 0.60. However, for HMM-based methods with original lemma and original words it gives a visibly lower result.

When evaluating the results of the methods with better F1 Score, we can see that with the rule-based method a higher recall is obtained, while with the method based on Hidden Markov Model with a tagged lemma, greater precision is obtained. These differences are balanced generating a similar F1 Score. Both methods seem to be good for verb detection. The difference is that the Hidden Markov Model-based method with tagged lemma has greater precision when classifying a word as a verb, but it tags fewer verbs. On the other hand, the rule-based method has less precision when classifying a word as a verb but classifies more instances as verbs.

In terms of accuracy, the values are very similar for all the methods. For both sets, the best result is the HMM-based method with tagged lemma. However, this measure is skewed by the high imbalance of the classes: there are many more instances of non-verbs than of verbs. For this reason, we consider this measure is not as significant for this study. The result of evaluating these metrics leads to the conclusion that the best methods for classifying verbs in the Guarani language, given the existing corpus, are the rule-based method, and the hybrid method based on Hidden Markov Model with labeled lemmas. Further research is needed to see if we can find a way of complementing both approaches in order to improve the results.

4 *Translation experiments*

We first performed an experiment to get a machine translation baseline for the Guarani-Spanish pair. We trained a sequence to sequence neural translation model composed by a RNN encoder with attention mechanism and a RNN decoder implemented in the OpenNMT (Klein et al., 2017) library. On our first experiments we used the default parameters for the OpenNMT model. This process ran for 100,000 iterations and took about three days to complete. We ran two variants of this experiment: the first one using all the words as they appear in the original corpus in Guarani, the second one transforming all the verbs in the Guarani corpus to a representation based on its morphological features. For the verbs detection and analysis we used the heuristics defined in sections 3.3 and 3.4.

For example, the verb *"omoheñóiva'ekue"* ("he/she/they produced") is transformed into the tag sequence *"TPRETPLUSCUAMPERFECTO INDEFFORM 3SINPLU VACTSIMPLE MINDSIMPLE moheñói"*' (meaning the verb *"moñehói"* conjugated in the pluperfect tense, indefinite form, third person plural or singular, simple active voice, simple indicative mood).

We evaluated the results of the machine translation experiments using the BLEU (Papineni et al., 2002) measure, which compares the quality between a candidate translation and one or more gold reference translations based on n-grams similarity, trying to capture at the same time notions of fluency and fidelity of the translation.

Figure 3 shows the BLEU results over the development corpus during the 100,000 iterations of these baseline experiments (with original and tagged verbs).

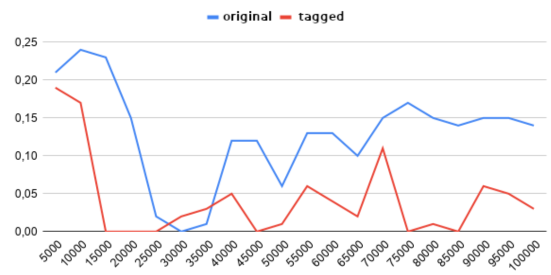

Figure 3: BLEU results for the baseline OpenNMT experiments during 100,000 iterations over the development set.

The best results seem to be achieved in the first iterations (less than 15,000 iterations) with BLEU values up to 0.237 for the original version and around 0.170 for the tagged version. After iteration 20,000 the values seem to decline and start to oscillate, but never achieving higher results. As can be seen in figure 3 after iteration 20,000 the values obtained are lower, and the model with original text performed almost always better than the tagged version.

We manually analyzed some of the predictions made by the model, and found out that they present a very low level of fluency and fidelity, which might be consistent with the

Original	
gn	*He'i pe reunión Gobierno representante-kuéra ndive omoîva reunión pyahu, pero nodefiníri mba'eve hikuái.*
Gold translation	
es	*Dijo que en la reunión representantes del Gobierno propusieron una nueva reunión, pero no se definió fecha para ello.*
en	*He said that at the Government representatives meeting they proposed a new meeting, but no date was set for it.*
Candidate translation	
es	*Dijo que la COMPRA en la COMPRA de la COMPRA de la COMPRA de la COMPRA de la COMPRA*
en	*He said the BUY in the BUY the BUY the BUY the BUY the BUY*

Table 3: Example of translation achieved with baseline experiment.

low BLEU values. Table 3 shows a translation example using the model trained with original text. As we can see, the translation candidate is not semantically correct (low fidelity), and it is not fluent either since it repeats the same phrase over and over again, which seems to be a common situation in this type of neural architecture (Holtzman et al., 2019).

Since the results of these first experiments indicated that the best models were generated in the first iterations, we decided to reduce the number of iterations to 20,000 and add more validation checkpoints. Also, for the baseline experiments the morphological information did not seem to improve the translation quality, so we tried to perform different variations in the input data in order to test if these features could be used during training in an advantageous way. For this second round of experiments, we used the following variations of representation for verbs:

1. **Original:** This means using the original text as in the baseline experiment, without morphological information. In this case the verb *"omoheñóiva'ekue"* would not be changed.

2. **Separation in verbal affixes:** Each verb was converted into consecutive labels describing its affixes, and the resulting lemma was added at the end, similarly to what was done in the tagged baseline experiment. In this case, the word *"omoheñóiva'ekue"* would be transformed into the sequence *"TPRETPLUSCUAMPERFECTO INDEFFORM*

3SINPLU VACTSIMPLE MINDSIMPLE mohenõi".

3. **Separation in verbal affixes without default labels:** The third experiment is very similar to the previous one but the labels of verbal inflections that do not add affixes to the lemma (the default values) are excluded. We also decided to remove the plural/singular third person label it is the most used label by default. For example, the word *"omoheñóiva'ekue"*, was transformed into the sequence *"TPRETPLUSCUAMPERFECTO moheñói"*.

4. **Tagged with original lemma:** The fourth experiment consists in using as representation of each verb a concatenation of its verbal affixes and the lemma joined with the "++" symbol. For example, the word *"omoheñóiva'ekue"* becomes the token *"TPRETPLUSCUAMPERFECTO++INDEFFORM ++3SINPLU++VACTSIMPLE++MINDSIMPLE ++moheñoi"*. Notice that in this case it is a single token instead of a sequence of tags.

5. **Tagged with tag lemma:** The fifth experiment is similar to the previous one with the difference that instead of concatenating the verb lemma at the end, we use the tag indicating if the lemma is a verb or not in the dictionary. For example, the word *"omoheñóiva'ekue"* was transformed into the token *"TPRETPLUSCUAMPERFECTO++INDEFFORM ++3SINPLU++VACTSIMPLE++MINDSIMPLE ++LEMAESVERBO"* for this experiment. With this representation, there could be different words that correspond to the same representation, so the rationale behind this is that it could alleviate the data sparsity.

Notice that experiments 2, 3, 4 and 5 could be considered hybrid models leveraging the rule-based methods and the neural machine translation method.

Table 4 shows the BLEU scores achieved for the different experiments over the development corpus.

Figure 4 shows a comparison of the results for these experiments over the development corpus. Notice that the best scores for all models are achieved around iteration 12,000.

Iter	Experiment				
	1	2	3	4	5
2000	0.11	0.20	0.18	0.23	0.20
4000	0.21	0.15	0.20	0.24	0.22
6000	0.20	0.19	0.26	0.21	0.24
8000	0.22	0.20	0.23	0.20	0.24
10000	0.22	0.23	0.24	0.24	0.23
12000	0.24	0.20	0.26	0.24	0.28
14000	0.15	0.11	0.25	0.16	0.26
16000	0.00	0.15	0.25	0.02	0.25
18000	0.08	0.05	0.23	0.05	0.06
20000	0.09	0.00	0.09	0.14	0.15

Table 4: BLEU results for the second round of experiments over the development corpus.

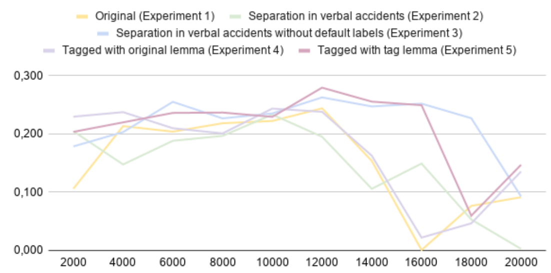

Figure 4: BLEU results over the development corpus.

The experiments with the best results were 3 and 5, reaching a BLEU score of 0.263 and 0.279 respectively for the 12,000 iteration. The experiment with the worst results in most cases is 2. On the other hand, experiments 1 and 4 behave in a similar way.

There seemed to be a certain improvement in BLEU scores using the model from experiment 5 with respect to the original model (experiment 1) over the development set. So we ran all models over the test set to check if this improvement still held on new data. Table 5 shows the results for the second round of experiments over the test set.

Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
0.174	0.138	0.174	0.157	0.203

Table 5: BLEU results for iteration over the test corpus.

Although the BLEU scores are lower in all cases, we can see that the results keep the same general behavior, that is, experiment 5 achieves the highest result, then experiments 3 and 1, then 4 and finally 2. This seems to indicate that incorporating morphological information during training might help improve

Original	
gn	*Temimbo'e o'yta haguã ojehechauka Ayolas-pe*
Gold translation	
es	*Estudiantes de natación realizan exhibición en Ayolas*
en	*Swimming students perform an exhibition in Ayolas*
Candidate translation	
es	*Estudiantes verifican circuitos turísticos de Ayolas*
en	*Students verify tourist circuits in Ayolas*

Table 6: Example of translation achieved with experiment 5.

the translation quality.

We also performed a manual revision of some results for experiment 5, which is the one with the best results. In this case, the translations in general were more fluent and there were no longer so many repetitions. Table 6 shows a translation example achieved with the result of experiment 5. The translation candidate is clearly more fluent in the target language, and it manages to pick up some of the concepts from the original sentences. However, its semantic is very different. Most of the analyzed predictions still have errors regarding their fidelity, although in some cases they transmit similar messages.

Since no BLEU value exceeded 0.3 in our experiments, we conclude that with the existing corpus, this method for automatic translation requires other complementary methods or tools to obtain better translations. We consider that, even with the modest improvements provided by training with morphological features, there is still a long way until good translation quality can be achieved.

5 Conclusions

We presented the results of an ongoing research on developing tools for processing the Guarani language and the Guarani-Spanish pair. First we implemented a rule-based approach to Guarani verb morphology. This approach showed promising results, although it could further be improved in order to use it in larger scale experiments. For example, we left out of the scope of this project the analysis of some verbal inflections, like the ones that indicate degree. Furthermore, it would be interesting to extend this approach to other parts of speech such as nouns or adjectives, that can also be richly analyzed for Guarani.

We proposed two approaches for verbs detection: a rule-based approach and some HMM-based models. We found out that using a hybrid between the rule-base and HMM methods seems to be amongst the best performing models for detecting verbs in Guarani sentences.

Finally, we tried several experiments incorporating morphological information in order to improve the performance of a baseline neural machine translation system for the Guarani-Spanish pair. The best performing methods use some of the morphological features along with part of the original text. However, the translation results in general are still far from being perfect, so further research is needed in order to improve these systems. The corpus size was one of the major limitations, so we would like to expand the corpus and also improve the alignment quality of some pairs. Nonetheless, we consider that the augmentation of the training data using morphological features served as a proof of concept for the kind of improvements that could be done in this machine translation system without using more data.

References

Abdelali, A., J. Cowie, S. Helmreich, W. Jin, M. P. Milagros, B. Ogden, H. M. Rad, and R. Zacharski. 2006. Guarani: a case study in resource development for quick ramp-up mt. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas, "Visions for the Future of Machine Translation*, pages 1–9.

Academia de la Lengua Guaraní (ALG). 2018. *Gramática Guaraní.*

Alcaraz, N. A. and P. A. Alcaraz. 2020. Aplicación web de análisis y traducción automática guaraní–español/español–guaraní. *Revista Científica de la UCSA*, 7(2):41–69.

Bird, Steven, E. L. and E. Klein. 2009. Natural language processing with python.

Bisazza, A. and M. Federico. 2009. Morphological pre-processing for turkish to english statistical machine translation. In *nnnn.*

Chiruzzo, L., P. Amarilla, A. Ríos, and G. Giménez Lugo. 2020. Development of a Guarani - Spanish Parallel Corpus. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages

2629–2633, Marseille, France, 05. European Language Resources Association.

Dooley, R. A. 2006. Léxico guarani, dialeto mbyá com informações úteis para o ensino médio, a aprendizagem e a pesquisa lingüística. *Cuiabá, MT: Sociedade Internacional de Lingüística*, 143:206.

El-Kahlout, I. D., E. Bektaş, N. Ş. Erdem, and H. Kaya. 2019. Translating between morphologically rich languages: An arabic-to-turkish machine translation system. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pages 158–166.

Estigarribia, B. 2015. Guarani-spanish jopara mixing in a paraguayan novel: Does it reflect a third language, a language variety, or true codeswitching? *Journal of Language Contact*, 8(2):183–222.

Estigarribia, B. and J. Pinta. 2017. *Guarani linguistics in the 21st century*. Brill.

Gasser, M. 2018. Mainumby: un ayudante para la traducción castellano-guaraní. *arXiv preprint arXiv:1810.08603*.

Holtzman, A., J. Buys, L. Du, M. Forbes, and Y. Choi. 2019. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.

Klein, G., Y. Kim, Y. Deng, J. Senellart, and A. Rush. 2017. OpenNMT: Open-source toolkit for neural machine translation. In *Proceedings of ACL 2017, System Demonstrations*, pages 67–72, Vancouver, Canada, July. Association for Computational Linguistics.

Lustig, W. 2010. Mba'éichapa oiko la guarani? guaraní y jopara en el paraguay. *PAPIA-Revista Brasileira de Estudos do Contato Linguístico*, 4(2):19–43.

Myrzakhmetov, B. and A. Makazhanov. 2016. Initial experiments on russian to kazakh smt.

Papineni, K., S. Roukos, T. Ward, and W.-J. Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. volume 12, pages 2825–2830.

Rudnick, A., T. Skidmore, A. Samaniego, and M. Gasser. 2014. Guampa: a toolkit for collaborative translation. In *LREC*, pages 1659–1663.

Secretaría de Políticas Lingüísticas del Paraguay. 2019. Corpus de Referencia del Guaraní Paraguayo Actual – COREGUAPA. http://www.spl.gov.py. Accessed: 2019-11-01.

Thomas, G. 2019. Universal dependencies for mbyá guaraní. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 70–77.