BUILDING A GRAMMAR WITH FOMA – ARAGONESE

MOTIVATION

The Aragonese language is one of the few endangered languages in Spain and one of many in Europe, which along with other languages such as the Astuarian-Leonese or Irish, to name but a few, is bound to disappear at any time soon. Aragonese is a Romance language spoken in several regions of the North of the shire of Huesca (Northeast of Spain) taking the form of several dialects. There are about 12,000 active speakers of the language and around 50,000 passive speakers, who can understand it native-like, but have no active command of it.

The motivation to take this language as the core of the assignment lies in the interest of the student in helping on the revitalization of the Aragonese by means of computation. By taking this task as a starting point, it might be possible to move to greener pastures in future works.

REFERENCE GRAMMAR USE

The Aragonese language is spoken taking the form of many local varieties depending on the area of the region and has not yet undergone the process of normalization and standardization, which other minority languages of the Spanish territory have. Considering those, it was complicated to stablish which variety was going to be taken as a reference for the present project. In the end, we decided to make use of the Verb Conjugator for the verb section, which provides a way of conjugating verbs following the advice from the DPGL of the languages of Aragon (General Council of Languages) and of the Wikipedia page for the noun-inflection section.

PHENOMENA COVERED

1- LEXC FILE

For carrying out the task, we first created a lexc file containing the lexicon to use and the different additions to form our grammar. We chose several verbs and nouns to cover some of the different aspects of the language, mostly regarding inflectional grammar. Those were the following:

VERBS			**NOUNS**		
charrar	saper	apetexer	flor	espinac	Falz
cremar	baxar	amanexer	voz	carnuz	bizino
creyer	debuxar	partir	peix	turcaz	dolor
ser	trayer	aber	aragonés	bel	

Regarding the conjugation of verbs, we decided to cover three tenses: infinitive, present tense and past tense. All of them where conjugated using the three personal forms that the language uses (1st person, 2nd person and 3rd person) in both singular and plural, giving as a result the following tags:

```
LEXICON VerbSpecial2
        [INF]:0 #;
        [PRESENTE]: VerbEndS2;
        [PASADO]: VerbEndS3;

LEXICON VerbEnd1
        [1PS]:+o #;
        [2PS]:+as #;
        [3PS]:+a #;
        [1PP]:+amos #;
        [2PP]:+az #;
        [3PP]:+an #;
```

Image 1: Tags used for the verbs

The verbs chosen for the task can be divided into regular and irregular (only 'aber' and 'ser'). Moreover, the regular ones can also be divided into groups according to the ending of the infinitive (conjugations) as it happens in many other Romance languages:

- 1st conjugation: verbs ending in –ar: charrar (to speak).
- 2nd conjugation: verbs ending in –er: creyer (to believe).
- 3rd conjugation: verbs ending in –ir: partir (to leave).

Therefore, the different endings for each division were described in the lexc file as 'VerbInfl', following these correspondences: VerbInfl1 -> 1st conjugation; VerbInfl2 -> 2nd conjugation; VerbInfl3 -> 3rd conjugation; VerbSpecial -> Verb 'aber'; VerbSpecial2 -> Verb 'ser'.

Each of these inflections would have a different ending depending on the tense, giving as a result five different endings for regular verbs and the past of the verb 'aber', and three special endings for the irregular verbs.

Regarding the nouns, we decided to cover the creation of plural nouns, choosing a few inflectional suffixes:

- When the word ends in -r, an -s is added to the noun.
- When the word ends in –c, -s, -z, -x, -ch; -es is added to the noun.
- When the words ends in -az or –oz; -os is added to the noun.

We described these three suffixes as 'NInfl' in the lexc file, giving as a result 'NInfl1', 'NInfl2' and 'NInfl3', which contained both the tag of singular (adding nothing to the noun) and plural (adding the corresponding suffix to the noun).

```
LEXICON Noun
        flor    NInfl1; !Flower
        voz     NInfl2; !Voice
        peix    NInfl2; !Fish
        aragonés NInfl2; !Aragonese
        espinac NInfl2; !Spinach
        carnuz  NInfl3; !Carrion
        turcaz  NInfl3; !Typical type of pigeon

LEXICON NInfl1
        [SG]:0 #;
        [PL]:+s #;
LEXICON NInfl2
        [SG]:0 #;
        [PL]:+es #;
LEXICON NInfl3
        [SG]:0 #;
        [PL]:+os #;
```

Image 2: Tags used for the nouns.

2- RULES FILE

Having created the lexicon of the grammar in the lexc file, we created another txt. document containing the rewriting rules in order to apply some morphophonological changes and get our final results. The phenomena in this part can be summarised as follows:

- Removing the endings –ar, -er, -ir to get the root of the verbs for each conjugation.
- Changing 'y' for 'ig' in the first person of the verbs ending in –yer such as in 'creyer' -> 'creigo'.
- Removing all '+' symbols that were added in the lexc, which served as a fair reference point for creating the first rewriting rules.
- Erasing the root ab- in the present of the auxiliary verb 'aber' in order to get the correct forms: 'he', 'has', 'ha', 'emos', 'ez', 'han'.
- Erasing 's' when it is followed by an -y, -s, -f to form the past tense of 'ser': 'fue', 'fues', 'fue', 'fuemos', 'fuez', 'fueron'.
- Adding 'c' after 'x' for the first person of the present in verbs ending in –xer.
- Erasing 'e' after 'x' for the third person of the present in verbs ending in –xer, and adding back such 'e' when it is between an 'x' and a consonant (rather than 'c', such as in 'apetexco').
- Changing 'z' into 'c' when it is followed by -es, such as in 'voz' -> 'voces'.

- Erasing the accent of 'és' when it is followed by -es, such as in 'aragonés' -> 'aragoneses'.
- Changing 'c' into 'qu' when it is between an 'a' and an 'e', such as in 'espinac' -> 'espinaques'.

TESTING

Having finished our grammar, we ran it with foma, where we could observe the following result:

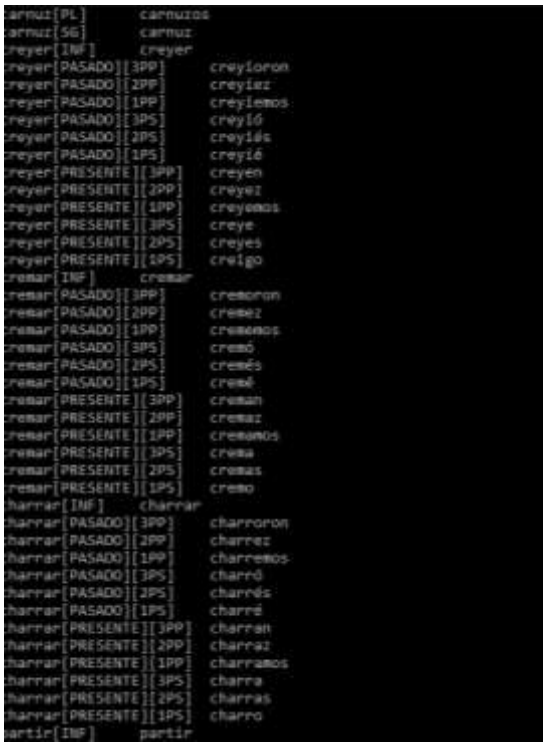

Image 3: Part of the final grammar.

Then, by using the commands 'up' and 'down', we tested that whichever word or morphological analysis searched would appear (taking special care in the irregular forms) and therefore, we could conclude that the grammar was working properly.

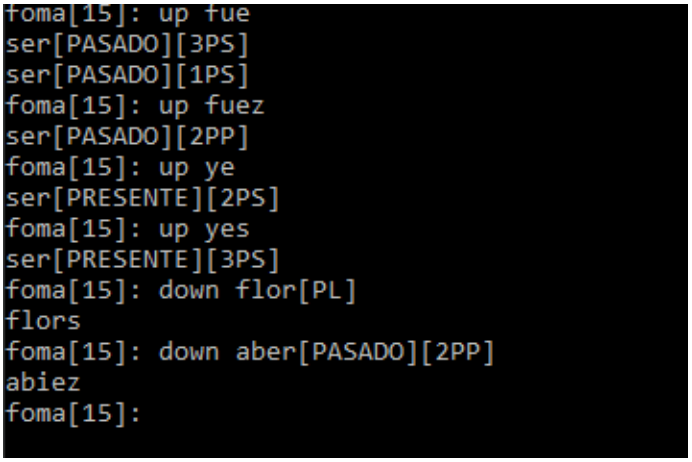

Image 4: Testing the grammar.