NPFL094 – Morphological and syntactic analysis
ÚFAL, 24/25 adrirflorez@gmail.com

Homework 2:
Partial morphological analyzer with FOMA for Modern Greek

*This report is aimed at explaining the design and structural decisions of the **partial morphological analyzer** I developed for Modern Greek, based on my resulting lexicon from the previous homework assignment.*

1. Files

The submitted .zip file containing scripts and results consists of the following subdirectories:

- `foma/`, which contains the two files that construct my analyzer built with FOMA:
    - `el.foma`: defines the whole grammar and alternation rules (including phonotactical rules, deletion of word endings for base forms…).
    - `el.lexc`: definition of all supported inflectional classes and their respective endings, for each one of the supported lexicons (noun, adjective and verb). This file already contains one word example per each observed defined inflectional classification.
    - `analyzer.sh:` shell script that invokes the .foma file with the .lexc lexicon and that tests all input words given in `input-foma.txt`, and writes the results onto an `output-foma.txt` fule.
- `lexc/`, optional and extra Python program I developed to extend the base .lexc lexicon with all examples found in my resulting lexicon from the previous homework assignment.
    - `data/`: contains a base `el-empty.lexc` file with no known word examples per lexicon, only all the defined inflectional rules; and the full `el-lexicon.txt` file.
    - `lexc_builder.py`: adds all nouns, adjectives and verbs to their respective lexicon, provided that they belong to the supported inflectional classes that the morphological analyzer recognizes.
        - (!) You are welcome to test the .sh script with the updated el-full.lexc file, and by adding words from it to input-foma.txt

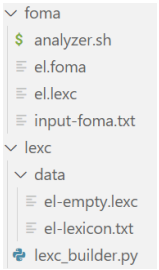

Figure 1: Directory and file structure of the partial morphological analyzer

2. Analyzer design

The resulting analyzer is a partial morphological analyzer for Modern Greek, which knows and recognizes the following parts of speech and inflectional classes as defined in `el.lexc`*:

** Each entry per lexicon consists of the form* `word:stem class ;`

Part of Speech *(LEXICON)*	Inflectional classes *(one sub-lexicon per each)*	Features *(multichar symbols)*

Noun (N)
 	
`LEXICON Noun`
`άνθρωπος:ανθρωπ NounMascOs ;`
`άντρας:αντρ NounMascAs ;`
`αφέντης:αφεντ NounMascIs ;`
`ίδρυμα:ιδρυ NounNeutMa ;`
`ψάρι:ψάρι NounNeutI ;`
`χαμόγελο:χαμογελ NounNeutO ;`
`χώρα:χωρ NounFemA ;`
`ύπαρξη:υπαρξ NounFemI ;`
	Masculine nouns
	- ending in -ός
	- ending in -ας
	- ending in -ης
	
	Neuter nouns
	- ending in -μα
	- ending in -ι
	- ending in -ο
	
	Feminine nouns
	- ending in -α
	- ending in -η
		Number:
		- Singular
		- Plural
		
		Cases:
		- Nominative
		- Accusative
		- Genitive
		- Vocative

Adjective (A)
 	
`LEXICON Adj`
`αγγλικός:αγγλικ AdjOs1;`
`άρρωστος:αρρωστ AdjOs2;`
`ακριβής:ακριβ AdjIs1;`
`ζηληαρις:ζηληαρ AdjIs2;`
`βαθύς:βαθ AdjUs;`
	Adjectives ending in -ός
	
	Adjectives ending in -ος
	
	Adjectives ending in -ης
	
	Adjectives ending in -ής
	
	Adjectives ending in -ύς
	
	And per each one of ending classes, a sub-lexicon per gender (masculine, feminine and neuter)
		Comparison degrees:
		- Comparative
		- Superlative

All of the defined sub-lexicons cover all of the noun and adjective inflectional classes that were established for the creation of the lexicon in homework 1.

➔ The only unsupported feature is that of negation, because in Greek this is done via: either the arbitrary addition of one prefix (α- or μη-), or by the negation of the phrase/sentence as a whole (with auxiliary particle δεν/δε).

Part of Speech *(LEXICON)*	Inflectional classes *(one sub-lexicon per each)*	Features *(multichar symbols)*
Verbs (V) LEXICON Verb πληρώνω:πληρων VerbA ; !! 3 syllables βελτιώνω:βελτιων VerbA ; διαβάζω:διαβαζ VerbA ; !! 2 syllables γράφω:γραφ VerbA ; αισθάνομαι:αισθαν VerbPath1 ; !! passive with θ λεγομαι:λεγ VerbPath1 ; αναλογίζομαι:αναλογιζ VerbPath2 ; !! pasive with τ συνεργάζομαι:συνεργα VerbPath2 ;	Active verbs from the A, first conjugation group, especifically with infinitive ending: - in - νω Passive verbs ending in -όμαι, especifically with past aorist ending: - in -*θηκα* - in -στηκα	Tenses: - Present - Past aorist - Past imperfect (continuous) - Perfect imperative (2nd persons) - Gerund - Past participle Number: - Singular - Plural Persons: - 1 first - 2 second - 3 third

All of the defined sub-lexicons cover conjugation alpha of verbs (active and passive), taking into account the respective phonotactical rules at morpheme boundaries when forming the different tenses (especifically past aorist).

These phonotactical rules are enforced to ensure that irregular endings and morpheme transformations are done correctly and systematically. However, we must know that even these phonotactical rules are irregular (i.e. ζ can arbitrarily change into either σ or ξ, φτ to σ or ξ…). Therefore, the examples given correspond to only one of these forms.

With respect to limitations and unsupported features:

➔ Aorist increment: regularly, for alpha conjugation, an έ- prefix is added at the start of 2-syllable verbs in the first, second and third person of singular, and third person of plural, for the past aorist in active voice. I simply do not know how to implement adding this prefix when the syllable count is exactly 2 for the verb.

- *For example, test with word **γράφψ**: past aorist for first person singular should be **έγραψα**, but I am only able to generate **γραψα**. :-(*

➔ Irregular verbs. Even though we have specific conjugations and rules to enforce, there are a lot of irregularities and exceptions for each verb (some of

them come from ancient Greek, others simply do not follow the rules… etc). It must be noted that the included phonotactical rules tackle some irregularities in past aorist formation.

➔ Compound tenses. Those formed with auxiliary particles (θα for future, conjugated forms of irregular verb έχω for pluscuamperfect and other compound forms…). These are usually formed in combination with the stem of the past aorist plus the ending morpheme of the present form. However, it all really depends on whether the verb is irregular and whether the action being described is momentaneous or continuous. So… pretty complicated feature to implement – it is already difficult for me to learn as a human trying to learn the language itself.

➔ Verbal adjective. These are usually formed with suffixes -μένος and the implementation should be straightforward, but I initially tried to follow the process of formation of adjectives like I had done with the degrees of comparison (+Comp and +Sup), but for some reason FOMA did not recognize this.

In general, many of the difficulties regarding verbs come with the fact that we have to fine-grain and deal with morpheme boundaries between verb stem and inflection morphemes very especifically. To generalize even better, I would have to improve the way the original lexicon is generated, especifically related to stem extraction for these verbs. For example, generated stem for verb πληρώνω is πληρω – and many times it is key to keep that last consonant, which is the one that will give us a hint towards which phonotactical rule to apply in each case, if any.