Generating Text with Correct Verb Conjugation:
Proposal for a New Automatic Conjugator with NooJ

Héla Fehri, Sondes Dardour
MIRACL Laboratory, University of Sfax
`hela.fehri@yahoo.fr, dardour.sondes@yahoo.com`

Abstract

This paper describes a system that generates texts with correct verb conjugation. The proposed system integrates a conjugator developed using a linguistic approach. This latter is based on dictionaries and transducers built with the NooJ linguistic platform. The conjugator treats three languages: Arabic, French and English. It recognizes all verbs and allows their conjugation in different tenses. The results obtained are satisfactory and can easily be improved upon by processing other forms, such as the negative.

1 Introduction

Automatic Language Processing is an area of multidisciplinary research that permits the collaboration of linguists, computer scientists, logicians, psychologists, documentalists, lexicographers, and translators.

In this domain different conjugators are built and used (Rello and Basterrechea, 2010). The term conjugation is applied only to the inflection of verbs, and not to other parts of speech (inflection of nouns and adjectives is known as declension). The development of the conjugator is not an easy task and depends on the specificities of the processed language. Among existing conjuagators, for Arabic, we can cite *AlKanz*[1] and *qutrub*[2]. For the French language, we find *Le Figaro*[3] and *Reverso Conjugaison*[4]. And for English, we can cite The conjugator[5], conjugation.com and *Reverso Conjugaison*[6]. The difference between these conjugators lies in the number of languages, forms (negative, interrogative) and voices

processed. They can be in different forms such as a website or mobile application.

The aim of this paper is to *generate* a text with well-conjugated verbs. To reach this objective, we propose to develop a system that allows parsing a text, extracting different infinitive forms of verbs and conjugate them in the appropriate tense. This system integrates a conjugator, which makes it possible to conjugate Arabic, French, and English verbs in the desired tense. This conjugator should guarantee the correct conjugation of verbs without errors.

In this paper, after an introduction to the proposed method, we describe our resource construction and implementation using the NooJ linguistic platform (Silberztein and Tutin, 2005). Then, we give an idea of the experimentation and the results obtained and conclude with some future perspectives.

2 Proposed Method

As shown in Figure 1, the proposed method requires four steps or two phases: the identification, construction and compilation of resources phase and the conjugation phase in which the conjugator of verbs is integrated. In what follows, we will examine each phase in detail.

2.1 Identification, construction and compilation of resources

The step of constructing and compiling resources consists in identifying the lexical resources represented by dictionaries and building the syntactic grammars represented by transducers.

---

[1] `http://www.al-kanz.org/2007/06/26/conju-gaison-arabe/`
[2] `https://qutrub.arabeyes.org/`
[3] `http://leconjugueur.lefigaro.fr/`
[4] `http://conjugueur.reverso.net/conjugai-son-francais.html`
[5] `http://www.theconjugator.com/`
[6] `http://conjugueur.reverso.net/conjugai-son-anglais.html`
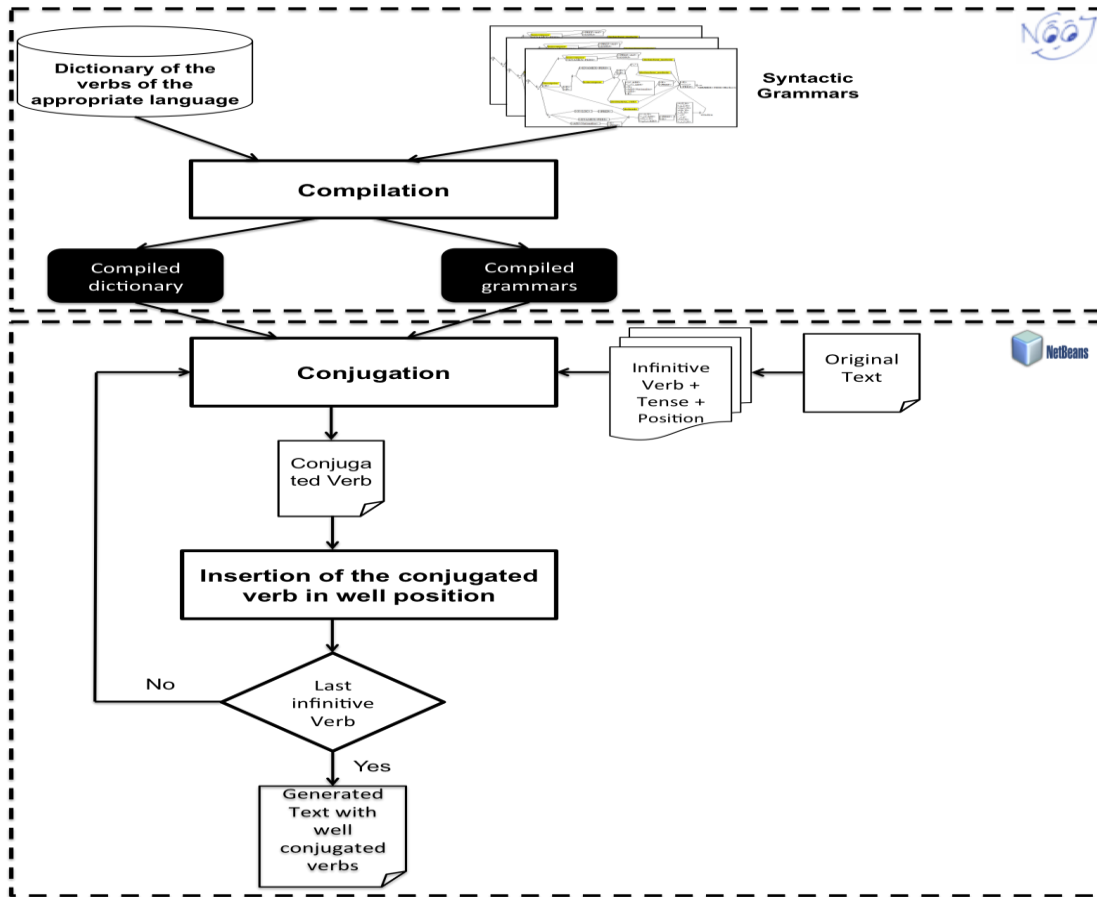
Figure 1: Proposed method

Identification of dictionaries: A NooJ dictionary is an electronic dictionary designed for use by computer systems. A NooJ dictionary contains different entries. The structure of an entry is specific to each dictionary, but contains at least the grammatical category of the entry (Name, Adjective, Verb, etc.).

In our work, we use three dictionaries: the dictionary of Arabic verbs (verbes arabes.nod) which contains 9,257 entries (Fehri et al., 2016), the dictionary of French verbs (_dm.nod) which contains 67,983 entries (Trouilleux, 2011) and the dictionary of English verbs (_sdic.nod) which contains 90,000 entries (Silberztein, 2003).

Each dictionary contains a derivation module to recognize the derived forms and a flexional module to recognize the inflected forms of the verb.

Construction of grammars: A grammar is a set of graphs. The number of grammars depends on the number of tenses treated to perform the conjugation. Note that each language has its proper tenses.

For Arabic, we have processed four tenses: the past tense (اَلْمَاضِي *al-māḍī*), the present tense (اَلْمُضَارِع *al-muḍāri‘*), the future tense and the imperative (اَلْأَمْر *al-amr*). Figure 2 represents the conjugation of the verbs in the future (F) with different Arabic pronouns.
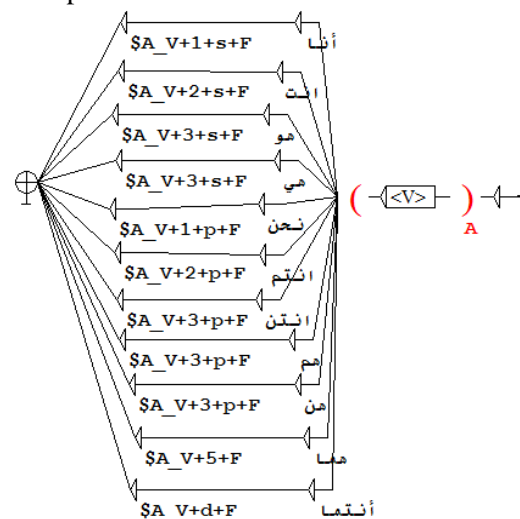

Figure 2: Conjugation of verbs in the future

For French, we have treated all tenses (present, past and future), aspects (perfective and imperfective), as well as all moods (indicative, imperative, subjunctive, conditional and gerundive). Figure 3 describes the conjugation of verbs in the present tense (PR) (tense: present, mood: indicative).
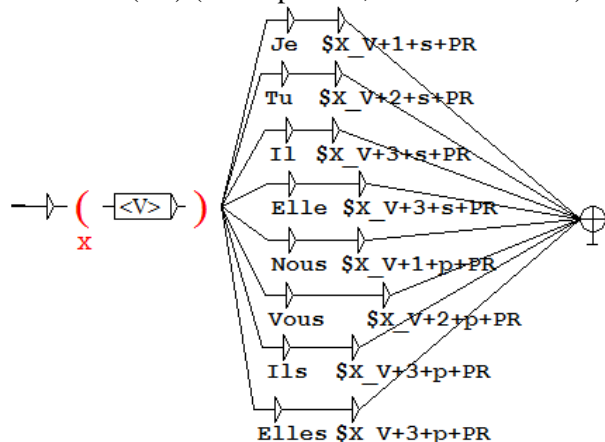

Figure 3: Conjugation of verbs in the present tense

For English, we have processed all possible combinations of tense, aspect and mood: present tenses (simple present and continuous present), past tenses (simple past and continuous past), present perfect tenses (present perfect (simple) and present perfect (continuous)), past perfect tenses (past perfect (simple) and past perfect (continuous)) and future tenses (simple future, continuous future, future perfect (simple) and future perfect (continuous)). Figure 4 describes the conjugation of verbs in the simple past (PT) with all pronouns.
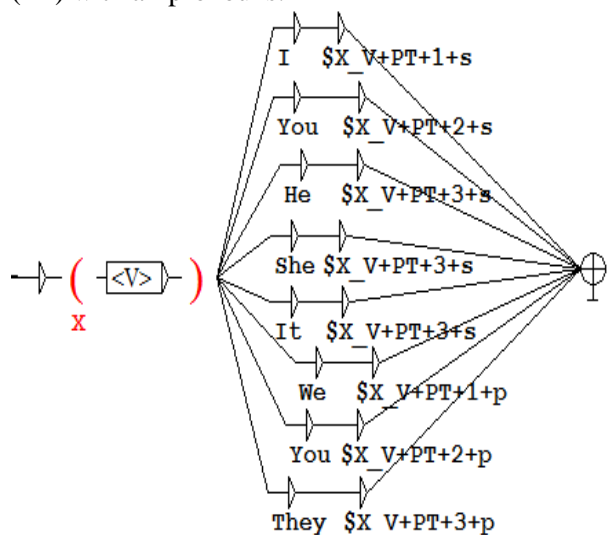

Figure 4: Conjugation of verbs in the past

Compilation of resources: The compilation phase consists of generating grammars and dictionaries in binary format that can be exploited in a later step.

2.2 Conjugation of verbs

The conjugation of verbs is done in three steps: parsing of the text, extraction of the infinitive form, its position and the desired tense, and conjugation of the extracted verb in the appropriate language. In our case, verbs to be conjugated and tenses are delimited by special characters such as parentheses.

To conjugate a verb, we use compiled resources described in section 2.1. Theses resources are used with command-line program *noojapply*, which is accessed from Java. Once the verb is conjugated, it will be inserted in the correct position in the generated text. The three steps mentioned will be repeated until processing of all verbs to be conjugated in the original text is complete.

3 Experimentation and evaluation

The experimentation of our system is done using NooJ and Java. As mentioned above, NooJ uses syntactic and morphological grammars already built. To evaluate our work, we have applied our resources to 300 texts in different languages: Arabic, French and English. Figure 5 represents an excerpt of results obtained when applying our system to an English text.

As shown in Figure 5, our system gives satisfactory results. However, some problems are related to the lack of standards for writing verbs (e.g., the hamza) in Arabic and the difficulties of dealing with some forms, such as the negative and interrogative forms and the passive voice.

Table 1 gives an idea about tenses and verbs processed by our system.

	Languages		
	Arabic	**French**	**English**
Number of tenses	4	17	12
Number of verbs	9 257	67 983	90 000

Table 1: Description of the conjugator.

Note that the number of verbs indicated in Table 1 represents only the lemmas that exist in our dictionary. The derived forms are also recognized by our system thanks to morphological grammars.
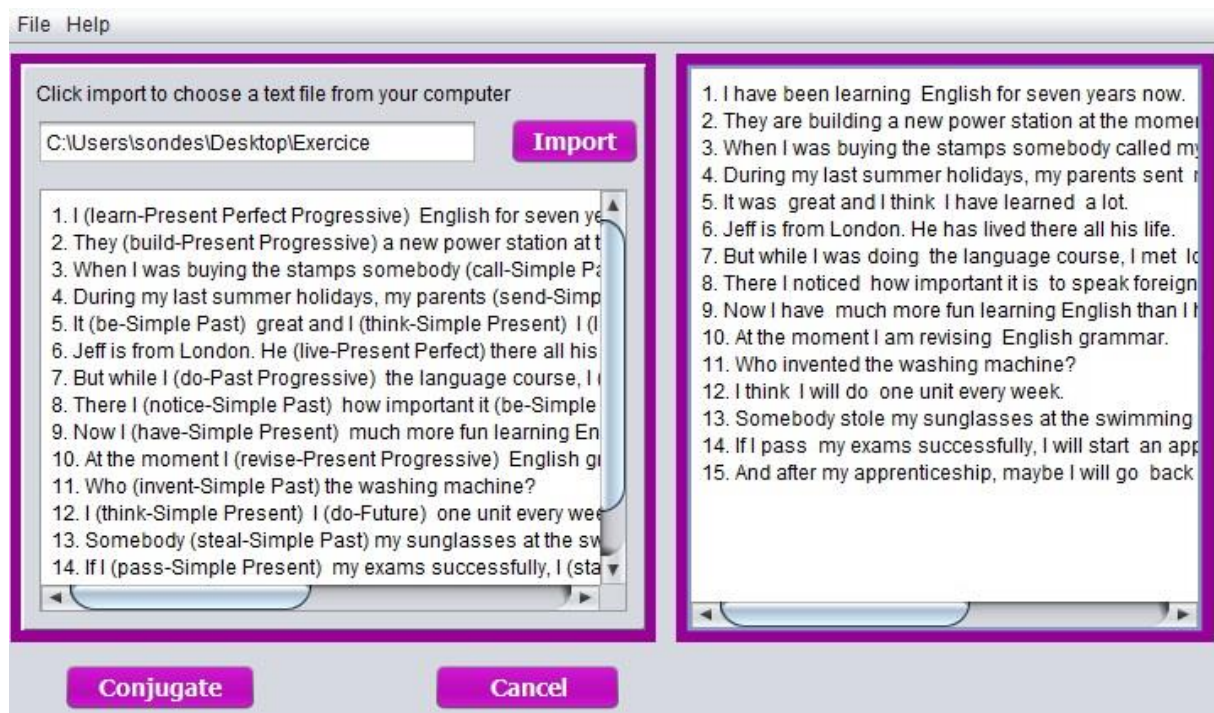
Figure 5: Excerpt of results obtained

4 Conclusion

The system we developed helps to learn how to conjugate a verb correctly. It can be used as a teaching tool for learning conjugation. It gives also sufficient results.

In the future, we aim to improve the conjugator by processing other forms (interrogative and negative) and the passive voice. Furthermore, we want to add other concepts and rules in order to know the tense of the verb without indicating it. This is possible by examining the context of the sentence.

References

Hela Fehri, Mohamed Zaidi and Kamel Boudhina. 2016. *Création d'un dictionnaire des verbes NooJ open source pour la langue arabe*. Report of the end of studies project. Higher Institute of Management of Gabes

Luz Rello and Eduardo Basterrechea. 2010. Automatic conjugation and identification of regular and irregular verb neologisms in Spanish, *Proceedings of the NAACL HLT 2010 Second Workshop on Computational Approaches to Linguistic Creativity*, pp 1–5, Los Angeles, California, June 2010.

Max Silberztein. 2003. *NooJ Manual* Available for download at: www.nooj4nlp.net.

Max Silberztein and Agnès Tutin. 2005. NooJ, un outil TAL pour l'enseignement des langues. *Application pour l'étude de la morphologie lexicale en FLE*. Special Atala, 8(2), pp 123-34.

François Trouilleux. 2011. *Le DM*, A French dictionary for NooJ. In *Automatic Processing of Various Levels of Linguistic Phenomena: Selected Papers from the NooJ 2011 International Conference* (eds. K. Vučković, B. Bekavac and M. Silberztein). Cambridge Scholars Publishing, pp 16-28.