Treatment Automatic:
Semantics Analysis and Traduction of the Frozen Expression with Nooj

Salma Tace¹, Said Nouh², Mohamed Azzouazi³
¹,²,³ Departement of Information Technology and Modeling Laboratory Hassan II University,

ABSTRACT
The purpose of this article is to define more closely the notion of freezing which, despite numerous publications in the field, remains a vague concept. The reflections are the result of a research project that aims to build an almost exhaustive database of fixed verbal expressions in English. After a review of the essential properties of the expressions, it is indicated why each of these criteria is problematic. One of the major problems lies in the existence of related phenomena: from a semantic point of view, fixed expressions participate in the general phenomenon of polysemy, lexical solidarity brings them closer to collocations and finally, morphosyntactic fixity is present in many fixed sentences that are conversational routines and even partly in so-called free syntax.

Keywords: Natural language, Nooj, Frozen expressions, Automatic language processing, Semantic analysis.

Corresponding Author:
Mohamed AZZOUAZI,
Departement of Information Technology and Modeling Laboratory Hassan II University, Casablanca, Morocco
Email: azouazii@gmail.com

1. Introduction

The automatic processing of natural language (TALN) or languages (TAL) is a discipline at the border of linguistics, data processing and artificial intelligence, which concerns the application programs and techniques of data processing to all aspects of human language. The TAL or TALN is sometimes called linguistic engineering.

Automatic Language Processing (T.A.L) is a multidisciplinary field of research that belongs to the field of artificial intelligence (I.A). It makes collaborate linguists, computer scientists, logicians, psychologists, documentalists, lexicographers or translators.

T.A.L. is also a socio-economic reality, with companies and specialized products like spelling correctors, translation software, voice dictation ...

With the increasing computerization of activities, the fields of application of T.A.L. are more and more numerous; even if its share varies according to the activities: low for cars "talking" or information synthesized, average for word processing programs with built-in corrector, important for "smarter" applications such as translation or summary automatic.

NooJ, presents features specifically designed for language teaching and linguistics. NooJ takes over...
some features of INTEx [Silberztein99], improving them and adapting them to the development of educational activities.

2. **Keywords**

2.1. Frozen expressions

The congestion is the process who gives us the possibility to designated a group of words whose elements are indissociable.

The congestion is characterized by loss proper sense of the elements constituting the group of words, which appears like:
- A new lexical unit.
- Autonomous and full-sense.

Regardless of its components.

The studies on congestion "have made it possible to highlight the fact that we are in the presence of one of the most important phenomena of the natural languages, which presents, for the automatic treatment, an obstacle as important as the polysemy, for example Nooj.

NooJ is a development environment used to construct large-coverage formalized descriptions of natural languages, and apply them to large corpora, in real time. The descriptions of natural languages are formalized as electronic dictionaries, as grammars represented by organized sets of graphs.

NooJ supplies tools to describe inflectional and derivational morphology, terminological and spelling variations, vocabulary (simple words, multi-word units and frozen expressions), semi-frozen phenomena (local grammars), syntax (grammars for phrases and full sentences) and semantics (named entity recognition, transformational analysis). In fact, Nooj allows linguists to combine in one unified framework Finite-State descriptions such as in XFST, Context-Free grammars such as in GPSG, Context-Sensitive grammars such as in LFG and unrestricted grammars such as the ones developed in HPSG.

NooJ is also used as a corpus processing system: it allows users to process sets of (thousands of) text files. Typical operations include indexing morpho-syntactic patterns, frozen or semi-frozen expressions (e.g. technical expressions), lemmatized concordances and performing various statistical studies of the results.

3. **Research goal**

The objective of this work is to traduction the frozen expressions with Nooj « arabe-english », he is an important subject of research.

4. **Experimentation and results**

The use of fixed expressions is far from being a marginal phenomenon in languages: Mr Gross (1982 & 19884) has listed more than 40,000 in French alone, which contradicts the traditional conception that considers them as exceptions - "ignoring these constructions means ignoring a large part of the language" (Danlos 1981). But because of this frequency, a translator or learner may remain speechless: fixed expressions "cause many difficulties when learning a foreign language or when translating (whether manual or automatic). Nevertheless, they are only the subject of a few bilingual dictionaries without scientific pretensions, and they are not at the centre of comparative linguistics studies. It would therefore be of the utmost interest that bilingual glossaries of fixed expressions be developed" (Danlos 1988). To our knowledge, there are no accessible studies that have given rise to exhaustive bilingual glossaries of fixed expressions.

First to translate the fixed expressions, it will be necessary to create a dictionary after a graph then to call them with the nooj platform.
Tableau 1: COMPARATIVE TABLE OF MORPHOLOGICAL ANALYZERS

<table>
<thead>
<tr>
<th>Technologie à états finis</th>
<th>Génération</th>
<th>Voyellation</th>
<th>Agglutination</th>
<th>Désambiguisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Cohen</td>
<td>x</td>
<td>v</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>AraMorph</td>
<td>x</td>
<td>v</td>
<td>v</td>
<td>±</td>
</tr>
<tr>
<td>Xerox</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>±</td>
</tr>
<tr>
<td>AraParse</td>
<td>x</td>
<td>v</td>
<td>v</td>
<td>±</td>
</tr>
<tr>
<td>AMSAAR</td>
<td>x</td>
<td>v</td>
<td>v</td>
<td>±</td>
</tr>
<tr>
<td>Sebawai</td>
<td>x</td>
<td>±</td>
<td>±</td>
<td>±</td>
</tr>
<tr>
<td>Nooj</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
</tr>
</tbody>
</table>

THE TABLE below explains why choosing the nooj platform as you see it, the nooj platform is perfect for different analyses and processing like de-ambiguating against xerox and david cohen.

Figure 1: CREATE A DICTIONARY OF NOOJ

Figure 2: MORPHOLOGICAL GRAPH OF NOOJ.
For this experiment, WE SELECTED to analyze the text we call the present graphs to finally have the number of sentences containing the times.

5. Conclusion

In view of all the above, and if we accept what has long been recognised by linguistic tradition, namely that the phraseological phenomenon is to be taken in the broad sense, it seems to me that we must reasonably admit that it is illusory to find a totally "watertight" definition of the freeze. In other words, we must be satisfied with a very general definition according to which a fixed expression is a phraseological unit composed of several words, contiguous or not, which have a certain degree of semantic freezing, a certain degree of lexical freezing and a certain degree of morphosyntactic fixity. Fortunately, the case of fixed expressions is not an isolated case.
References


Web sites:

1. www.nooj4nlp.net