

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,

Article Info

Received Jun 17, 2018

Keyword:

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

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.

Corresponding Author:

Mohamed AZZOUAZI,
Departement of Information Technology and Modeling Laboratory Hassan II University, Casablanca,
Morocco
Email: azzouazii@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. Keyzords

2.1. Frowen 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 composents

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 effet, 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 developped 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 a important subject of reserch.

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

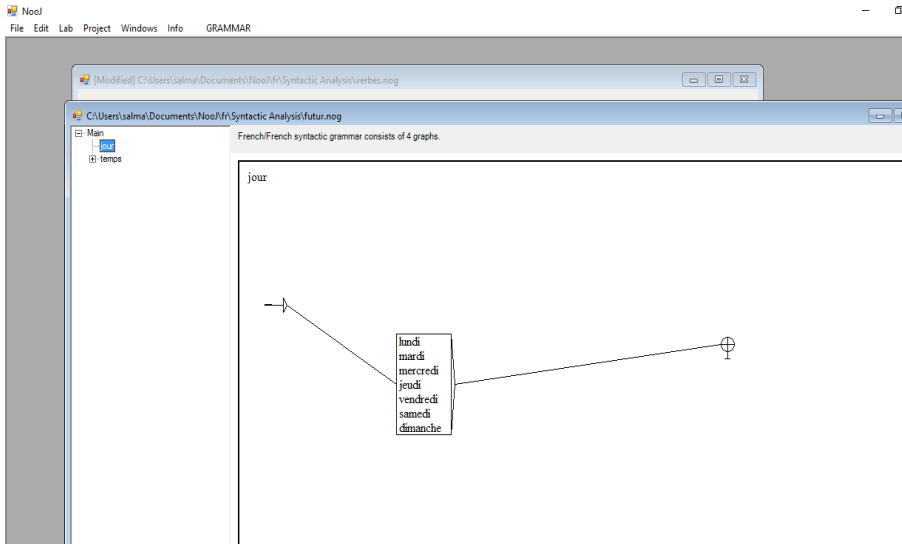


Figure 3 : GRAPH OF DATE

The screenshot shows a concordance analysis window in Nool. The window title is 'Nool - [Concordance for Text, C:\Users\salm\Documents\Nool\Projects\La femme de trente ans.net]'. The window contains a table with columns 'Text', 'Before', 'Seq.', and 'After'. The 'Seq.' column contains the word 'jour'. The 'Before' and 'After' columns contain the surrounding text from the source document.

Text	Before	Seq.	After
est heureuse aujourd'hui, le	sera	jour	elle toujours ? Car les
bourdonnant comme un essaim. 'Cela	sera	jour	donc bien beau ? demanda Mlle
du Carrousel des régiments qui	vont manœuvrer	jour	- Non, mon enfant, toutes les
à a bon cœur, f	entraînera	jour	peut-être à donner sa
fait colonel. - Ma chère, Victor	rentra	jour	colonel toute sa vie. Je
mais ! Comment espérer que vous	pourrez	jour	vous entendre tous deux avec
volontés différentes dont les tyrannies	seront	jour	inconciliables ? Tu seras ou victime
les tyrannies seront inconciliables ? Tu	seras	jour	ou victime ou tyran. L.
si douce et modeste, tu	plieras	jour	d'abord. Enfin tu as
une grâce de sentiment qui	sera	jour	astucieuse, et alors. ! Il n
seprit il après une pause.	blessera	jour	les naïves qualités de ta
Elle, de qui tu n	entendras	jour	bienôt plus la voix si
ma fille. Un jour tu	déploreras	jour	amèrement sa nullité, son défaut
mille autres chapitres qui te	viendront	jour	par lui. Alors, souviens-toi
'ange... Mais elle se se	gênera	jour	pas lui, j'espère, de
plus embarrassée que curieuse. 'Nous	allons donc faire	jour	connaissance, mon cher cœur ? repris
peut-elle, mon neveu	subira	jour	bienôt les inconvénients du mariage
silence, aujourd'hui tu en	deviendras	jour	peut-être la raison en
apprenais les mystères que je	vois traître	jour	Je les aurais à jamais
de tes plus poignants regrets	viendra	jour	du souvenir de ce que
peignaient si délicieuses. Cette soirée	fera	jour	ton désespoir, Louise. Dans ce
si non heureuse, un mari te	rendra	jour	, en peu de jours, ce
Et même comment te le	dirai	jour	je ? je ne me souviens
Ma sœur, je vous le	formai	jour	Je metrai fin au triste
je vous le formai. Le	metrai	jour	fin au triste désaccord, assez
venir ou bon lui semble ?	Demain	jour	nous ne mangerons plus dans
lui semble ? Demain nous ne	mangerons	jour	plus dans cette salle, quand
quand il ne nous y	verra	jour	plus, le jeune gentilhomme discontinuera
verra plus, le jeune gentilhomme	discontinuera	jour	de vous aimer par la
les Autrichiens et les Autais	vont faire	jour	leur attention à l'État ou

Figure 4 : TRAITEMENT ANALYSIS OF DATE

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

- [1] ANDRIENIERENANA, C.-L. et al., 2000, « Problème de traduction des expressions figées : un exemple bilingue français-malgache, malgache-français », Actes du Colloque international « Traduction humaine, Traduction automatique, Interprétation », Série Linguistique n°11.
- [2] ANSCOMBRE, J.-C., 2003, « Les proverbes sont-ils des expressions figées? », Cahiers de lexicologie, 1, 159-173.
- [3] BERCKER, F., 1994, Etude comparative anglais-français des expressions figées figurées comportant un ou plusieurs noms de parties du corps humain, Thèse de l'Université Paris III Sorbonne Nouvelle.
- [4] Blumenthal, Peter / Hausmann, Franz Josef (éds.), Collocations, corpus, dictionnaires (Langue française, 150). – Paris : Larousse, 2006.
- [5] Bolly, Catherine, Les unités phraséologiques : un phénomène linguistique complexe ?. Thèse de doctorat (non publiée). – Louvain-la-Neuve, 2008.
- [6] Silberztein M. 2015. Silberztein, M. (2014). Formaliser les langues : l'approche de NooJ. ISTE Ed.: Paris. (426 p.).
- [7] Silberztein Max, 2004. NooJ: A Cooperative, Object-Oriented Architecture for NLP. In INTEX pour la Linguistique et le traitement automatique des langues. Cahiers de la MSH Ledoux, Presses Universitaires de Franche-Comté.
- [8] Mesfar Slim, Silberztein Max, 2008. Transducer minimization and information compression for NooJ dictionaries. In Proceedings of the FSMNLP 2008 conference, Frontiers in Artificial Intelligence and Applications. Ed IOS Press, Netherlands.
- [9] Silberztein Max, 2007. An Alternative Approach to Tagging. Invited Paper In Proceedings of NLDB 2007. LNCS series, Springer-Verlag Eds, pp. 1-11.
- [10] Silberztein Max, 2005. NooJ's Dictionaries. In the Proceedings of LTC 2005, Poznan University.
- [11] Muller Claude, Royauté Jean, Silberztein Max Eds. 2004. INTEX pour la Linguistique et le Traitement Automatique des Langues. Proceedings of the 4th and 5th INTEX workshops, Bordeaux, May 2001, Marseille, May 2002: Presses Universitaires de Franche-Comté (400 p).
- [12] Gross, Gaston, Les expressions figées en français. Noms composés et autres locutions. – Paris : Ophrys, 1996.
- [13] Gross, Maurice, « Une classification des phrases < figées > du français », in : Revue Québécoise de Linguistique, 11 / 2 (1982), p. 151–185.
- [14] Gross, Maurice, « Les limites de la phrase figée », in : Langages, 90 (1988), p. 7–22.
- [15] Rey, Alain / Chantreau, Sophie, Dictionnaire des Expressions et Locutions. – Paris : Dictionnaires Le Robert, 1997.
- [16] Senellart, Jean, « Reconnaissance automatique des entrées du lexique-grammaire des phrases figées », in : Lamiroy, Béatrice (éd.), Le lexique-grammaire (Travaux de Linguistique, 37). – Louvain-la-Neuve : De Boeck, 1998, p. 109–127.
- [17] <http://pen.ius.edu.ba/index.php/pen/issue/view/17>
- [18] <http://pen.ius.edu.ba/index.php/pen/issue/view/18>
- [19] <http://pen.ius.edu.ba/index.php/pen/issue/view/19>

Web sites:

1. www.nooj4nlp.net