

# Enhancing the ElixirFM Lexicon with Verbal Valency Frames

Viktor Bielický

Institute of Near Eastern and African Studies  
Charles University in Prague  
viktor.bielicky@ff.cuni.cz

Otakar Smrž

Institute of Formal and Applied Linguistics  
Charles University in Prague  
otakar.smrz@mff.cuni.cz

## Abstract

This paper reports on the status of our ongoing project to extend, refine, and maintain the lexicon of the ElixirFM system. The latest release of ElixirFM has been enriched with valency frames for selected verbal lexemes. We briefly outline some other similar projects for Modern Standard Arabic and present crucial axioms of the approach we have adopted, which is Functional Generative Description. Our contribution focuses mainly on describing the annotation procedure, resources and tools, as well as on the representation of the valency phenomena in form of dependency trees in the lexicon. We also point at some intricate issues concerning reciprocal meaning associated with some verbs. Such a valency lexicon for Arabic can find broad use in various NLP tasks and applications, in addition to its reusability by the users of the language.

## 1. Introduction

Valency of a lexical unit, in particular a verb, represents its potential to bind specific numbers and types of other lexical units. Valency properties are lexicalized and cannot be inferred using grammatical rules only. Creating a valency lexicon of the most frequent verbs for Modern Standard Arabic (MSA) that will be reusable by both humans and computational systems is a desirable task. This kind of lexicon can find use in a wide variety of NLP problems including word-sense disambiguation, tagging, syntactic and semantic analysis, machine translation, and language generation.

The aim of this paper is to report on the status of our ongoing project to extend, refine, and maintain the lexicon of the ElixirFM system, which has been recently enhanced with valency frames for some chosen verbal lexemes. This is the result of the initial phase of our effort to build a verbal valency lexicon for MSA, the basic concept of which was presented in (Bielický and Smrž, 2008).

First, we describe the ElixirFM Lexicon and its web application ElixirFM Online Interface (Sec. 2), and we briefly outline some projects and resources closely related to ours (Sec. 3). Then we touch upon some crucial axioms of the theoretical framework we adopted (Sec. 4), discuss several topics concerning verbal valency frames and their annotation procedure, and examine closely reciprocal meaning associated with some verbs (Sec. 5).

## 2. ElixirFM

**ElixirFM** (Smrž, 2007a, b) is a high-level open-source implementation of Functional Arabic Morphology reusable for various NLP applications. Most of its lexical material was originally derived from the open-source Buckwalter lexicon (Buckwalter, 2002), however the latest release of ElixirFM (Smrž and Bielický, 2009) has undergone considerable computational and lexicographical corrections, refinements, and extensions.

The online implementation **ElixirFM Online Interface** (see Fig. 1 at the end) can serve as an example of the lexicon's computational potential.<sup>1</sup> This user interface

works in four different modes: 1) Resolve mode provides tokenization and morphological analysis of the inserted text (optionally in Unicode, Buckwalter or ArabTeX notations); 2) Inflect mode offers complete or partial paradigms of inflected forms for a given lexeme; 3) Derive mode gives verbonominal derivatives (verb, verbal noun, active and passive participles) for the particular class of lexemes; 4) Lookup mode enables to search for a desired lexeme and its possibly existing root-related derivatives, i.e. it serves as an online dictionary. It is even possible to browse the dictionary in a reverse direction by inserting expressions in English.

## 3. Similar Projects for Arabic

To the best of our knowledge, the only existing projects that are related to a systematic description of valency phenomena or semantic-syntactic properties of a considerable number of Arabic verbs are the Arabic Propbank (Palmer et al., 2008), the work by Loukil et al. (2008), and the printed study by al-Qahtani (2005). Let us shortly review their main characteristic features.

### 3.1. Arabic Propbank

The Pilot Arabic Propbank project, which is based on corpus data derived from the Penn Arabic Treebank and which is similar to the English and Chinese Propbanks, aims at adding a level of semantic annotation into a large Arabic corpus.

First, framesets of particular verbs are created by which individual senses for the given verbal lexeme are distinguished. A frameset contains all semantically required arguments for each sense of the verb (lexical unit). The arguments are labeled as ARG0-4 and are assigned specific semantic roles (experiencer, organizer, event, agent, theme, etc.). Consequently, the ARG0-4 labels must be seen as variables containing different semantic roles according to and depending on every single verb sense they are assigned to.

Second, the framesets are used as guidelines by annotators who match them to corpus instances of verbal predicates and their arguments. In addition to the above-mentioned 5 numbered arguments, 19 fixed adjunctive arguments are distinguished (time, location, direction, manner, etc.).

<sup>1</sup> <http://quest.ms.mff.cuni.cz/elixir/>

### 3.2. Arabic Syntactic Lexicon

As concerns the creation of a syntactic lexicon of Arabic verbs, the project announced by Loukil et al. (2008) focuses on building a machine-readable resource distinguishing syntactic (though surface) properties for each sense of a verbal lexeme, which are encoded into syntactic frames using the extensional Lexical Markup Framework formalism. Such a syntactic frame brings information on transitivity/intransitivity of a verb and a number and type of its syntactic complementations (its realization through the accusative case or after a preposition – if it is NP/PP, identification of its syntactic function (subject, object, etc.)). Each sense is illustrated with an example sentence. Unfortunately, the authors' standpoint seems to be affected by the traditional Arabic approach to syntactic structures, where classifying VSO and SVO word orders as different types of sentences (verbal vs. nominal) leads, in our opinion, to a lot of redundancy in the description.

### 3.3. Semantic Valence of Arabic Verbs

The printed study on semantic valence of Arabic verbs by al-Qahtani (2005), which has been briefly compared to our approach in (Bielický and Smrž, 2008), presents thorough examination of Arabic verbs in terms of the Case Grammar theory based on Cook's Matrix Model of verb classification (1979 and later improvements). It distinguishes three semantic types of verbs, i.e. state, process and action verbs, each of which is further subclassified into basic, experiential, benefactive and locative verbs according to a particular case role (semantic complementation) which it is characteristic for. These case roles are Agent, Experiencer, Benefactive, Object, Locative/Time. The obligatory Object is omnipresent with every verb and can occur more than once in a case frame. On the other hand, Experiencer, Benefactive and Locative are mutually exclusive. With some verbs, a particular case role is not realized on the surface ("covert case role"), i.e. it is either partially covert ("deletable") or totally covert ("coreferential" or "lexicalized"). The deletable case roles are simply those complementations which can be omitted on the surface as a result of the ellipsis. On the contrary, the coreferential and lexicalized case roles are always absent from the surface. The former coreferential roles denote instances where a single noun on the surface structure cumulates two case roles simultaneously, while the latter lexicalized roles include instances where a certain case role (usually Object) is incorporated in the semantics of a given verb. Let us look at some al-Qahtani's examples:

- (1) deleted case role from the surface  
*waṣafa Zaydun al-manẓara*  
 Zayd described the scene  
*waṣaf AEO/E-del* (Experiencer is deleted)
- (2) coreferential case roles  
*ġādara Zaydun al-bayta*  
 Zayd left the house  
*ġādar AOL/A=O* (Agent equals Object)
- (3) lexicalized case role  
*ʿamila Zaydun*  
 Zayd Worked = Zayd did some work  
*ʿamil AO/O-lex* (Object is lexicalized)

To conclude, all the aforementioned approaches suggest some interesting ideas which are worth noticing and can serve us as a source of inspiration.

## 4. Theoretical Framework

The theoretical framework that we adopted is the valency theory elaborated within the **Functional Generative Description** (FGD), a multi-stratal dependency-oriented language representation formalism (Panevová, 1974, 1975, 1994; Sgall et al., 1986). This approach has been successfully applied on two valency lexicons for Czech, namely VALLEX (Lopatková et al., 2006, 2008; Žabokrtský, 2005) and PDT-Vallex (Hajič et al., 2003).

Although the study of valency phenomena pertains mainly to the highest (underlying) tectogrammatical layer representing the linguistic meaning, it is also closely associated with the surface morphemic layer where deep-level entities emerge and are realized as morphemes.

According to FGD, it is possible to describe the semantic-syntactic properties of a verb or any other autosemantic lexical unit by its **valency frame** – a sequence of frame slots – where constraints on the specific type, number and possible morphemic realizations of the dependents of the lexical unit are encoded.

The valency frame of a verb is filled with a particular set of **functors** denoting a type of relation between the governor (a verb) and its dependent (verbal complementation/argument) as units of the tectogrammatical layer of language representation. Two kinds of verbal complementations are distinguished, i.e. **actants** (or inner participants – **ACTor**, **PATient**, **ADDRessee**, **ORIGin**, **EFFect**) and **free** (adverbial) **modifications** (or adjuncts, e.g. adverbials of direction, location, means, time, result, respect).<sup>2</sup> Both actants and free modifications are considered to be either obligatory or optional for each lexical unit (each sense of a verbal lexeme). The so-called "dialogue test" is introduced as the criterion of obligatoriness for both types of verbal complementations (e.g. Panevová, 1974, 1975). The valency frame in its narrow sense consists of both obligatory and optional actants and only obligatory free modifications (see Table 1).

	obligatory	optional
actants (inner participants)	+	+
free modifications (adjuncts)	+	-

Table 1: Members included in the valency frame

Thus, for instance, the valency frame of the basic meaning of the quadrilateral verb *tarġam* ("to translate") consists of two obligatory (ACT, PAT) and two optional (ORIG, EFF) actants (Fig. 2). If we intend to transform this sentence into the form of a dependency tree structure, its possible representation can be depicted as is shown in the following (Fig. 3).

<sup>2</sup> For a full list of verbal complementations and their description, detailed annotation guidelines for the tectogrammatical layer of the Prague Dependency Treebank for Czech (Mikulová et al., 2006) should be consulted.

تَرَجَّمَ أَكْثَرَ مِنْ خَمْسِينَ كِتَابًا مِنَ الْفَارِسِيَّةِ إِلَى الْعَرَبِيَّةِ  
*tarğama ʔaktara min ħamsīna kitāban min-a 'l-fārisīyati*  
*ʔilā 'l-ʔarabīyati*

he<sup>ACT</sup>-translated ʔa-more than fifty a-book<sup>PAT</sup> ʔfrom  
 Persian<sup>ORIG</sup> ʔinto Arabic<sup>EFF</sup>

he translated more than fifty books from Persian into  
 Arabic

Figure 2: A corpus instance with verb *tarğam* (“to translate”). The obligatory actants are highlighted in bold.

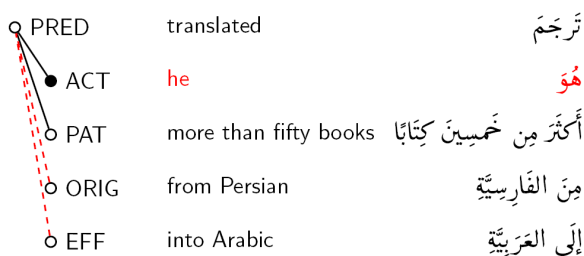


Figure 3: The sentence from Fig. 2 depicted in the form of a simplified dependency tree. The implicit Actor is restored with the additional node, since it is known from the context. Those actants marked with dashed edges (ORIG, EFF) are optional.

It is to be pointed out that the approach adopted by the FGD takes into account both syntactic and semantic criteria when assigning functors to verbal complementations. The first two actants, i.e. Actor and Patient, must be regarded in some cases as syntactic than pure semantic categories, since they undergo the process of “shifting of cognitive roles” (Panevová, 1974, 1975, 1994). That means that the first and the second actant of the given verb are always identified with Actor and Patient, whatever their actual semantics is. Consequently, the semantic Effect, Origin or Addressee may in fact shift to their position. As regards the free modifications, they cannot undergo such a “shift”.

In some languages, like English for instance, there are verbs whose valency frames can remain empty, i.e. no element on the tectogrammatical layer corresponds to the formal surface subject morpheme, and even the otherwise omnipresent Actor can be missing (4).

(4) *It is raining*

The valency frame is empty

As far as we know, an empty valency frame never occurs in Arabic. In the corresponding Arabic expression, the underlying Actor is always present in the valency frame, no matter if it is actually realized on the surface or elided. The implicit presence of the surface subject corresponding to the underlying Actor can be seen from the agreement between the verb and its elided subject (3rd pers. fem.).

(5) *tumṭiru* [ʔs-samāʔu]

it-rains [the-sky.ACT<sup>Obj</sup>]

approximately: [the sky] makes the rain fall

## 5. Valency frames in ElixirFM

### 5.1. Present coverage, tools and resources

The latest release of ElixirFM (Smrž and Bielický, 2009) has been enhanced with nearly 3.500 valency frames for selected verbal lexemes. Roughly 2.000 frames were created automatically exploiting the information on intransitivity from the Buckwalter Arabic Morphological Analyzer (Buckwalter, 2002). In addition to this, around 1.500 non-trivial valency frames have been created manually. All these frames are being subject to further additions, corrections, and refinements. In this initial phase of the annotation process, mostly basic meanings of verbal lexemes have been covered. In the following phases, we will focus more on figural meanings and on phraseology.

The ElixirFM lexicon is maintained and annotated in the environment of the open-source tree editor TrEd.<sup>3</sup>

The main resource for extracting the valency frames is the **Prague Arabic Dependency Treebank** (PADT) (Hajič et al., 2004a; Smrž et al., 2008) the second release of which containing a million word tokens is in preparation. This linguistically annotated corpus of Arabic newswire texts consists of three layers of annotations, i.e. functional morphology, surface analytical syntax, and tectogrammatics. The ElixirFM lexicon is already used in the MorphoTrees annotations of PADT, and the valency frames are being linked to the nodes in the tectogrammatical as well as analytical syntactic structures. PADT also serves us as a source for the study of frequency of verbs. Besides PADT, CLARA (Zemánek, 2001) and Arabic Gigaword are also exploited. We further consult major contemporary printed dictionaries (Baalbaki, 2000; Hoogland et al. 2003; Wehr, 1979).

### 5.2. Dependency-tree representation

Each annotated verbal lexeme, which is nested under its trilateral or quadrilateral root, is represented by a citation form, a lemma. The lexeme associates all its inflected forms and may include links to its deverbal nouns. Apart from these formal properties, a verbal lexeme comprises a set of lexical units, i.e. its individual senses (at least one), the semantic-syntactic properties of which are defined by means of their valency frame.

The valency frame, which is represented by a tree structure of alternative dependency subtrees, not only encodes the valency properties of a verb in terms of the FGD functors, but in most cases it also provides constraints on the possible surface morphemic realizations of a particular verbal complementation.

All valency frames in the ElixirFM lexicon include both obligatory and optional actants and only obligatory free modifications (see Table 1 above), but in some cases even optional free modifications are added, since they can be regarded as typical complementations.

To demonstrate the representation of all these phenomena in ElixirFM, the previously quoted sentence with verb *tarğam* (Fig. 2 and 3) can serve us as a suitable example (see Fig. 4 at the end).

<sup>3</sup> <http://ufal.mff.cuni.cz/~pajas/tred/>

### 5.3. Morphemic constraints

As is the case in VALLEX (Žabokrtský, 2005; pp. 63-66), two types of morphemic constraints can be distinguished and modified for our Arabic valency lexicon as well. These are:

*Explicit morphemic constraints* – one or more possible surface realizations for the given valency slot (alterantive subtree) is explicitly declared in the frame.

- Inflectional case ending (section of the morphological tag): 1- (nominative), 2- (genitive), 4- (accusative);
- Indefiniteness/definiteness (section of the morphological tag): -I-D;
- Constructions with adjective (section of the morphological tag): A-;
- i) Prepositions:<sup>4</sup> *bi-*, *li-*, *fī*, *ʿan*, *ʿalā*, *ʿilā*, *min*, *maʿa*, *ḥattā*, *munḍu*;
- ii) Prepositionals (yet incomplete list): *bayna*, *dūna*, *ḥawla*, *taḥta*;
- Subordinate content clauses with a conjunction (yet incomplete list): *inna*, *anna*, *an*, *mā* *iqā*;
- Dependent part of a phraseme: If the lexical choice of a particular verbal complement is very limited due to a very restricted collocability of such a lexical unit, the lexical choice should be listed in the lexicon, such as for example (6).<sup>5</sup>

(6) *inšaraha la-hu šadr-ī*  
it-was-delighted because-of-him.CAUS chest-of-me.ACT  
I was delighted by him  
valency frame: **ACT**<sup>Obl</sup> (*šadr*l*qalb*h*āḥīr*) **CAUS**<sup>Opt</sup> (*li-*)

*Implicit morphemic constraints* – the surface realization of the valency slot results from the semantics of the particular functor (usually a free modification), and thus it does not have to be specified in the valency frame, since it is registered in the list of possible forms (yet incomplete) gathered during the annotation process (e.g. LOC (locative) – *fī*, *bi-*, *ḥawla*, *bayna*, *ʿalā*, *taḥta*, *fawqa*, *ʿamāma*, *warāʿa*, *ḥalfā*, *dāḥila*, *ḥāriḡa*...; DIR1 (direction from) – *min*, *min* *ʿalā*, *min* *taḥti*, *min* *warāʿi*, *min* *fawqi*...; CAUS (cause) – *li-*, *bi-sababi*, *bi-faḍli*, *min* *ʿaḡli*, *naṯḡatan* *li-*, *bi-ḥukmi*, *li-ʿanna*...).

### 5.4. Reciprocity

When we try to capture the semantic-syntactic properties of a verb in its valency frame, we might face some difficulties, if the inherent reciprocal meaning of form VI verbs (and even of some form VIII verbs) is involved. This reciprocal character originates from the composition

<sup>4</sup> This distinction between prepositions and prepositionals is made according to Badawi et al. (2004).

<sup>5</sup> It might be possible to include among the explicit morphemic constraints a possibility that one of valency slots is realized by the imperfect form of a verb (e.g. for verbs as *ḡaʿal* or *tarak*), but it is not clear yet if this element can be regarded as an actant or circumstantial qualifier (Badawi et al., 2004). This issue must be subject to further examination.

of the associative meaning of the form III verb with the anaphoric prefix *ta-* (LeTourneau, 1998).

In order to emphasize the reciprocal meaning of such verbs, we propose that a special attribute “Rcp” can be attached to the valency frame where particular complementations, which enter the relation of reciprocity, will be listed. In this case, we follow the practice in Vallex (Lopatková et al. 2006, 2008).

For example, if the surface subject of the verb *tabāḥaṭ* VI (“to discuss, to confer with each other”) is in plural or dual (or as two or more coordinated elements), in fact, two different actants (ACT, ADDR) are mixed up into a single surface realization (7). On the other hand, the very same verbal complementations can be realized as two surface entities, if the preposition *maʿa* is used. Thus, the valency frame created by us for this verb contains **ACT**<sup>Obl</sup> **ADDR**<sup>Obl</sup> (*maʿa*) **PAT**<sup>Opt</sup> (*fī*l*ḥawla*) and will receive the attribute Rcp: ACT-ADDR depicting that ACT and ADDR are in reciprocal relation (see (8) and Fig. 5 at the end).

(7) *tabāḥaṭa ʿz-zaʿīmāni fī ʿāḥiri ʿl-mustaḡiddāti*  
he-discussed the-two-leaders.ACT-ADDR last-the-news.PAT  
The two leaders discussed latest news

(8) *tabāḥaṭa Netanyahu maʿa wuzarāʿihi*  
he-discussed Netanyahu.ACT with-ministers-of-him.ADDR  
Netanyahu discussed with his ministers

With some other verbs than those of form VI another type of reciprocal relation can be found, if some other actants than Actor, e.g. PAT and EFF, enter the reciprocal usage. It is the case when the verb *qāran* III (“to compare”) is used with preposition *bayna* (10). Its standard valency frame can be determined as **ACT**<sup>Obl</sup> **PAT**<sup>Obl</sup> (4-) **EFF**<sup>Obl</sup> (*bi-*) and is illustrated on (10). For the reciprocal usage, this verb receives the attribute Rcp: PAT-EFF.

(9) *sa-yuqārinu ʿAristū bayna ʿl-ʿaḡnāsi*  
he-will-compare Aristotle.ACT between-the-genres.PAT-EFF  
Aristotle will compare the genres

(10) *wa-yuqārinu ʿl-ʿadaba bi-ʿl-funūni*  
and-he.ACT-compares the-literature.PAT with-the-arts.EFF  
and he compares the literature with arts

## 6. Conclusions and Future Work

In our contribution, we have overviewed the current status of our ongoing work on the valency lexicon of Arabic verbs, which is an integral part of the ElixirFM lexicon.

We have sketched the main characteristics of the latest release of ElixirFM including its web interface and we have reviewed the projects and sources for MSA that are related to our work. We have also outlined our theoretical background (Functional Generative Description), the present coverage of the valency frames of ElixirFM, and the representation of various valency phenomena in form of dependency trees. Last but not least, we have discussed issues about depicting the reciprocal meaning of verbs.

In the forthcoming phase of our work, we would like to focus on annotating less usual, figurative and idiomatic meanings of verbs as well as on solving some more intricate issues, such as the valency status of some

auxiliary, modal, impersonal and other defective verbs. We also intend to enrich the valency frames with additional information on potential passivization, correlation between secondary derived verb forms with their original counterparts, or semantic-syntactic class. We would like to link the entries with the instances of occurrence in PADT, as well as with other corpus examples and Czech glosses. We would like to continue building this complex language resource and make it most reusable for various NLP tasks and applications.

### Acknowledgements

We would like to thank Petr Zemánek for his valuable comments and suggestions. This work has been funded by the Ministry of Education of the Czech Republic, projects MSM0021620823 and MSM0021620838.

### References

- Al-Qahtani, D. M. (2005). *Semantic Valence of Arabic Verbs*. Beirut: Librairie du Liban Publishers.
- Baalbaki, R. (2000). *Al-Mawrid: A Modern Arabic English Dictionary*. 14<sup>th</sup> ed. Beirut: Dar El-Ilm Lilmalayin.
- Badawi, E., Carter, M.G. & Gully, A. (2004). *Modern Written Arabic: A Comprehensive Grammar*. London: Routledge.
- Bielický, V., & Smrž, O. (2008). Building the Valency Lexicon of Arabic Verbs. In Proceedings of the 6<sup>th</sup> Conference on Language Resources & Evaluation (LREC 2008). Marrakech, Morocco.
- Buckwalter, T. (2002). *Buckwalter Arabic Morphological Analyzer Version 1.0*. LDC catalog number LDC2002L49, ISBN 1-58563-257-0.
- Cook, W.A. (1979). *Case Grammar: Development of the Matrix Model (1970-1978)*. Washington D.C.: Georgetown University Press.
- Graff, D (2007). *Arabic Gigaword Third Edition*. LDC2007T40, 1-58563-460-3.
- Hajič, J., Panevová, J., Uřešová, Z., Bémová, A., Kolářová, V., & Pajas, P. (2003). PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Trebank Annotation. In Proceedings of the Second workshop on Treebanks and Linguistic Theories (pp. 57–68). Växjö: Sweden.
- Hajič, J., Smrž, O., Zemánek, P., Pajas, P., Šnaidauf, J., Beška, E., Kráčmar, J. & Hassanová, K. (2004a). *Prague Arabic Dependency Treebank 1.0*. LDC catalog number LDC2004T23.
- Hajič, J., Smrž, O., Zemánek, P., Šnaidauf, J. & Beška, E. (2004b). *Prague Arabic Dependency Treebank: Development in Data and Tools*. In Proceedings of the NEMLAR International Conference on Arabic Language Resources and Tools (pp. 110–117), Cairo, Egypt.
- Hoogland, J., Versteegh, K. & Woidich, M. (2003). *Woordenboek Arabisch-Nederlands*. Amsterdam: Bulaaq.
- LeTourneau, M.S. (1998). *Implicit Reciprocals in Standard Arabic*. In Benmamoun, E., Eid, M. & Haeri, N. (Eds.), *Perspectives on Arabic Linguistics XI* (pp. 91–113). Amsterdam – Philadelphia: John Benjamins Publ. Comp.
- Lopatková, M., Žabokrtský, Z. & Benešová, V. (2006). *Valency Lexicon of Czech Verbs VALLEX 2.0*. Technical Report TR-2006-34, Praha: ÚFAL MMF UK.
- Lopatková, M., Žabokrtský, Z. & Ketnerová, V. (2008). *Valenční slovník českých sloves [Valency lexicon of Czech verbs]*. Praha: Karolinum.
- Loukil, N., Haddar, K. & Ben Hammadou, A. (2008). *Towards a syntactic lexicon of Arabic verbs*. In HLT & NPL within the Arabic World : Arabic Language & Local Languages Processing – Status Updates & Prospects (LREC 2008). Marrakech, Morocco.
- Mikulová, M. et al. (2006). *A Manual for Tectogrammatical Layer of the Prague Dependency Treebank*. Technical Report 30, Praha: ÚFAL MFF UK.
- Palmer, M., Babko-Malaya, O., Bies, A., Diab, M., Maamouri, M., Mansouri, A. & Zaghouni, W. (2008). *A Pilot Arabic Propbank*. In Proceedings of the 6<sup>th</sup> Conference on Language Resources & Evaluation (LREC 2008). Marrakech, Morocco.
- Panevová, J. (1974). *On Verbal Frames in Functional Generative Description. Part I*. In Prague Bulletin of Mathematical Linguistics, 22, 3–40.
- Panevová, J. (1975). *On Verbal Frames in Functional Generative Description. Part II*. In Prague Bulletin of Mathematical Linguistics, 23, 17–52.
- Panevová, J. (1994). *Valency Frames and the Meaning of the Sentence*. In Luelsdorff, P.A. (Ed.), *The Prague School of Functional and Structural Linguistics* (pp. 223–243). Amsterdam – Philadelphia: Benjamins Publ. Comp.
- Sgall, P., Hajičová, E. & Panevová, J. (1986). *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Dordrecht: D. Reidel Publishing Company.
- Smrž, O. (2007a). *ElixirFM – Implementation of Functional Arabic Morphology*. In ACL 2007. Proceedings of the Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources (pp. 1v8). Prague, Czech Republic.
- Smrž, O. (2007b). *Functional Arabic Morphology: Formal System and Implementation*. PhD thesis, Charles University in Prague.
- Smrž, O. & Bielický, V. (2009). *ElixirFM. High-level implementation of Functional Arabic Morphology*. <http://sourceforge.net/projects/elixir-fm/>. 2009.
- Smrž, O., Bielický, V., Kouřilová, I., Kráčmar, J., Hajič, J. & Zemánek, P. (2008). *Prague Arabic Dependency Treebank : A Word on the Million Words*. In HLT & NPL within the Arabic World: Arabic Language & Local Languages Processing – Status Updates & Prospects (LREC 2008). Marrakech, Morocco.
- Wehr, H. (1979). *Arabic-English Dictionary: The Hans Wehr Dictionary of Modern Written Arabic*. Urbana, IL: Spoken Language Services.
- Zemánek, P. (2001). *CLARA (Corpus Linguae Arabicae): An Overview*. In: ACL 2001. Workshop Proceedings on Arabic Language Processing: Status and Prospects (pp. 111–112), Toulouse, France.
- Žabokrtský, Z. (2005). *Valency Lexicon of Czech Verbs*. PhD thesis. Charles University in Prague.

## ElixirFM 1.1 Derive Online

[Home](#) [Resolve](#) [Inflect](#) [Derive](#) [Lookup](#)

ElixirFM lets you derive words of similar meaning but different grammatical category. You only need to tell the desired grammatical categories.

You can either enter natural language descriptions, or you can specify the parameters using the positional morphological tags.

### Your Request

verb noun adjective (7646,1)

### ElixirFM Reply

Point the mouse over the data to receive further information.

t r ġ m ترجم						Lookup
v	tarġam	ترجم	t r ġ m ترجم	KaRDaS	I	"translate", "interpret" <input type="button" value="Inflect"/> <input type="button" value="Derive"/> <input type="button" value="Lookup"/>
V-----	I	tarġam	ترجم	KaRDaS	verb	
N-----	I	tarġamat	ترجمة	KaRDaS-at	noun	
A--A-----	I	mutarġim	مُترجم	muKaRDiS	adjective, active	
A--P-----	I	mutarġam	مُترجم	muKaRDaS	adjective, passive	

Figure 1: Screenshot of the Derive mode of the ElixirFM Online Interface showing verbonominal derivatives (verbal noun, active and passive participles) for the verb *tarġam* (“to translate”).

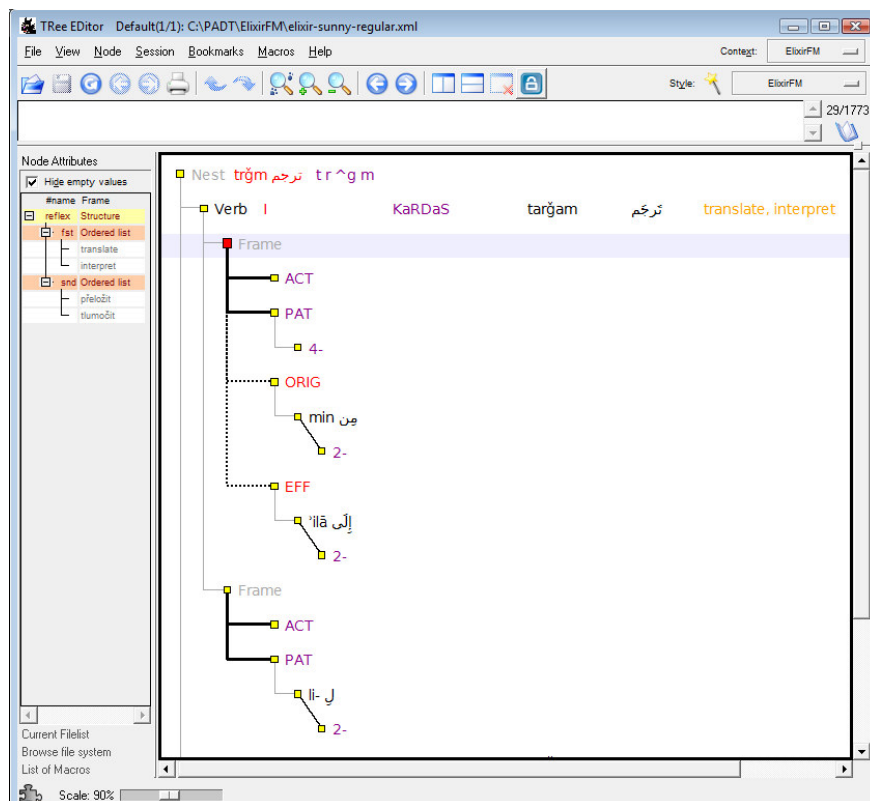


Figure 4: Screenshot of part of the lexical entries nested under the quadriliteral root *trġm*. The ElixirFM lexicon is displayed in the TrEd environment. The verbal lexeme *tarġam* contains two lexical units, since two different senses are distinguished (1. “to translate, interpret”; 2. “to write someone’s biography”). The first level of the valency frame (dependency tree) represents slots for various valency complements; those marked with dashed edges are considered to be optional. Their children represent all alternative morphemic representations of such complementations, which are expressed by their subtrees. Entries are encoded in form of morphophonemic patterns – the pattern *KaRDaS* corresponds to the quadriliteral strong verb that is traditionally denoted by the pattern *fa<sup>l</sup>al*.

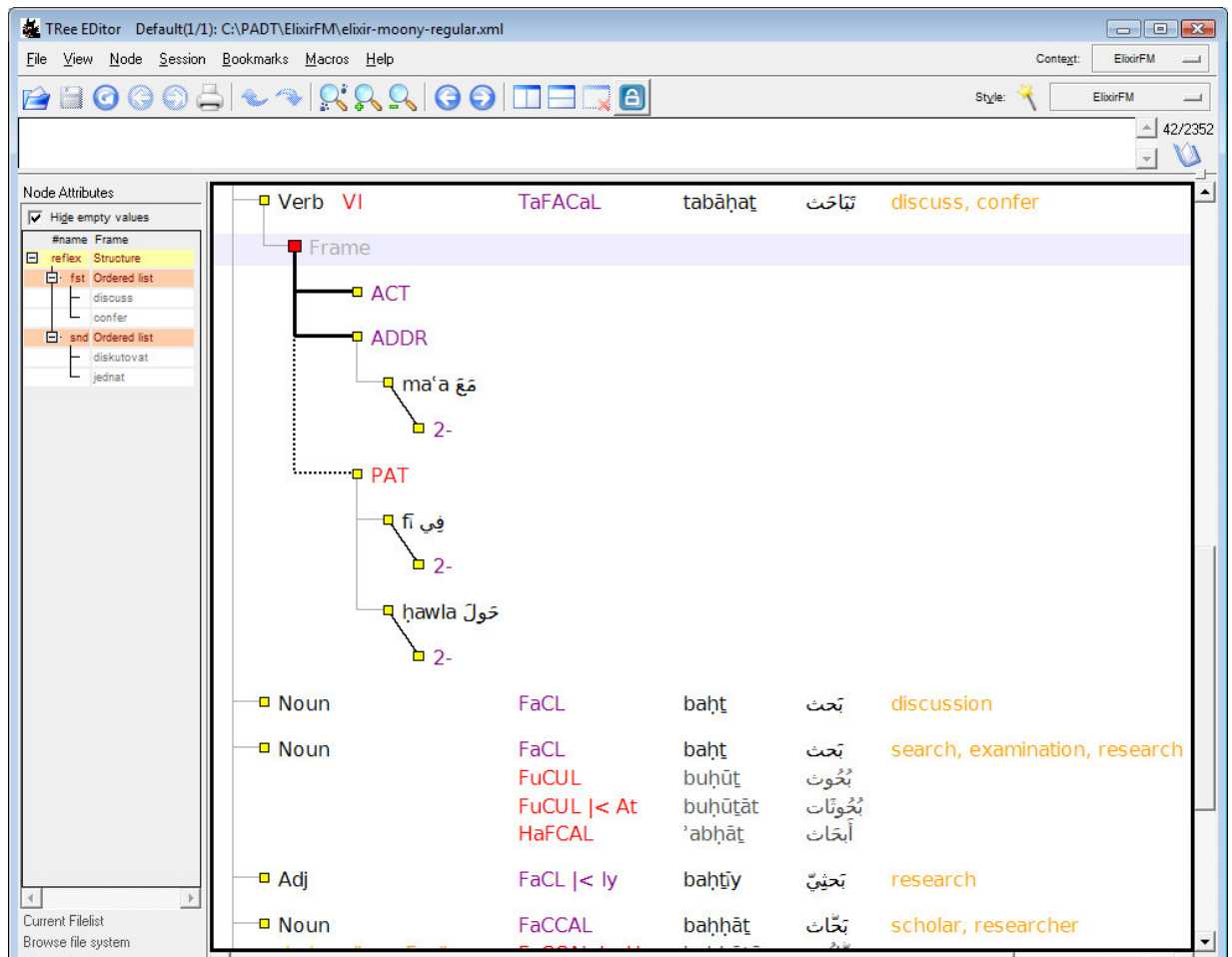


Figure 5: Screenshot of the valency frame for the reciprocal verb *tabāḥaṭ* (“to discuss, to confer”) with obligatory Actor, and Addressee. Optional Patient allows two possible surface morphemic realizations with prepositions *fī* and *ḥawla*. The reciprocal attribute Rcp: ACT-ADDR should be attached to this valency frame in order to emphasize the implicit reciprocal meaning of this verb.