Prague Treebanking for Everyone

Prague Arabic Dependency Treebank

Otakar Smrž

Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University in Prague

Vilem Mathesius Lecture Series 21

Prague, November 29, 2006
PADT is a project of linguistic annotation of Modern Written Arabic based on the theory of Functional Generative Description.
PADT is a project of linguistic annotation of Modern Written Arabic based on the theory of Functional Generative Description.

PADT 1.0 was published in 2004 by the Linguistic Data Consortium, and has been used by tens of academic and commercial institutions.
PADT is a project of linguistic annotation of Modern Written Arabic based on the theory of Functional Generative Description.

PADT 1.0 was published in 2004 by the Linguistic Data Consortium, and has been used by tens of academic and commercial institutions.

In PADT, which now consists of the morphological and the analytical levels of description of Arabic, the annotation of tectogrammatics and information structure is being established.
Outline

1. Introduction
Outline

1. Introduction
2. Morphology
   - MorphoTrees
   - ElixirFM
Outline

1. Introduction
2. Morphology
   - MorphoTrees
   - ElixirFM
3. Syntax and Beyond
   - Dependency Grammar
   - Analytical Syntax
   - Tectogrammatics
Outline

1. Introduction
2. Morphology
   - MorphoTrees
   - ElixirFM
3. Syntax and Beyond
   - Dependency Grammar
   - Analytical Syntax
   - Tectogrammatics
4. Software
   - TrEd Live
   - Encode Arabic
Outline

1. Introduction
2. Morphology
   - MorphoTrees
   - ElixirFM
3. Syntax and Beyond
   - Dependency Grammar
   - Analytical Syntax
   - Tectogrammatics
4. Software
   - TrEd Live
   - Encode Arabic
5. References
Outline

1. Introduction
2. Morphology
   - MorphoTrees
   - ElixirFM
3. Syntax and Beyond
   - Dependency Grammar
   - Analytical Syntax
   - Tectogrammatics
4. Software
   - TrEd Live
   - Encode Arabic
5. References
Arabic is a language of rich morphology, both derivational and inflectional, with highly ambiguous orthography.
Arabic is a language of rich morphology, both derivational and inflectional, with highly ambiguous orthography.

Boundaries of syntactic units, tokens, are obscure in writing—orthographical words, strings, consist of up to four lexemes.
Morphology Disambiguation

Arabic is a language of rich morphology, both derivational and inflectional, with highly ambiguous orthography.

Boundaries of syntactic units, tokens, are obscure in writing—orthographical words, strings, consist of up to four lexemes.

Disambiguation encompasses subproblems like tokenization.
Morphology Disambiguation

Arabic is a language of rich morphology, both derivational and inflectional, with highly ambiguous orthography.

Boundaries of syntactic units, tokens, are obscure in writing—orthographical words, strings, consist of up to four lexemes.

Disambiguation encompasses subproblems like tokenization, full morphological tagging or its simplified ‘part-of-speech’ versions.
Morphology Disambiguation

Arabic is a language of rich morphology, both derivational and inflectional, with highly ambiguous orthography.

Boundaries of syntactic units, tokens, are obscure in writing—orthographical words, strings, consist of up to four lexemes.

Disambiguation encompasses subproblems like tokenization, full morphological tagging or its simplified ‘part-of-speech’ versions, lemmatization.
Morphology Disambiguation

Arabic is a language of rich morphology, both derivational and inflectional, with highly ambiguous orthography.

Boundaries of syntactic units, tokens, are obscure in writing—orthographical words, strings, consist of up to four lexemes.

Disambiguation encompasses subproblems like tokenization, full morphological tagging or its simplified ‘part-of-speech’ versions, lemmatization, diacritization or restoration of the structural components of words.
Morphology Disambiguation

Arabic is a language of rich morphology, both derivational and inflectional, with highly ambiguous orthography.

Boundaries of syntactic units, tokens, are obscure in writing—orthographical words, strings, consist of up to four lexemes.

Disambiguation encompasses subproblems like tokenization, full morphological tagging or its simplified ‘part-of-speech’ versions, lemmatization, diacritization or restoration of the structural components of words, plus combinations thereof.
He will notify them about that through SMS messages, the Internet, and other means.

<table>
<thead>
<tr>
<th>String</th>
<th>Token</th>
<th>Token Tag</th>
<th>Buckwalters’s M-Tags</th>
<th>Token Form</th>
<th>Token Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>سيخبرهم</td>
<td>F-----------</td>
<td>FUT</td>
<td></td>
<td>sa-</td>
<td>will</td>
</tr>
<tr>
<td>سيخبرهم</td>
<td>VIIA-3MS---</td>
<td>IV3MS+IV+IVSUFF_MOOD:I</td>
<td></td>
<td>yu-ḥbir-u</td>
<td>he-notify</td>
</tr>
<tr>
<td></td>
<td>S-----3MP4-</td>
<td>IVSUFF_DO:3MP</td>
<td></td>
<td>-hum</td>
<td>them</td>
</tr>
<tr>
<td>بذلك</td>
<td>P-----------</td>
<td>PREP</td>
<td></td>
<td>bi-</td>
<td>about/by</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>SD----MS--</td>
<td>DEM_PRON_MS</td>
<td></td>
<td>ﺩَلِيْکَا</td>
<td>that</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>P-----------</td>
<td>PREP</td>
<td></td>
<td>ِاْن</td>
<td>by/about</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>N--------2R</td>
<td>NOUN+CASE_DEF_GEN</td>
<td></td>
<td>ٌتَارِیق-ی</td>
<td>way-of</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>N--------2D</td>
<td>DET+NOUN+CASE_DEF_GEN</td>
<td></td>
<td>َذَرْـیل-ی</td>
<td>the-messages</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>A-------FS2D</td>
<td>DET+ADJ+NSUFF_FEM_SG+CASE_DEF_GEN</td>
<td></td>
<td>ِالـقَـیـر-اُت-ی</td>
<td>the-short</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>C-----------</td>
<td>CONJ</td>
<td></td>
<td>ِاَلـینتَرْنـت-ی</td>
<td>the-internet</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>Z--------2D</td>
<td>DET+NOUN_PROP+CASE_DEF_GEN</td>
<td></td>
<td>ِاـِنترنت-ی</td>
<td>the-internet</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>C-----------</td>
<td>CONJ</td>
<td></td>
<td>ِاَلـینتَرْنت-ی</td>
<td>the-internet</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>FN--------2R</td>
<td>NEG_PART+CASE_DEF_GEN</td>
<td></td>
<td>ِگاْر-ی</td>
<td>other/not-of</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>S-----3FS2-</td>
<td>POSS_PRON_3FS</td>
<td></td>
<td>-۷ا</td>
<td>them</td>
</tr>
</tbody>
</table>
He will notify them about that through SMS messages, the Internet, and other means.

<table>
<thead>
<tr>
<th>String</th>
<th>Token</th>
<th>Token Tag</th>
<th>Buckwalters M-Tags</th>
<th>Token Form</th>
<th>Token Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>سيخبرهم</td>
<td>F-----------</td>
<td>FUT</td>
<td></td>
<td>sa-</td>
<td>will</td>
</tr>
<tr>
<td>بهذلك</td>
<td>VIIA-3MS---</td>
<td>IV3MS+IV+IVSUFF_MOOD:I</td>
<td>yu-ḥbir-u</td>
<td>he-notify</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>S-----3MP4-</td>
<td>IVSUFF_DO:3MP</td>
<td>-hum</td>
<td>them</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>P-----------</td>
<td>PREP</td>
<td>bi-</td>
<td>about/by</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>SD----MS--</td>
<td>DEM_PRON_MS</td>
<td>qālika</td>
<td>that</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>P-----------</td>
<td>PREP</td>
<td>ʿan</td>
<td>by/about</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>N--------2R</td>
<td>NOUN+CASE_DEF_GEN</td>
<td>ṭarīq-i</td>
<td>way-of</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>N--------2D</td>
<td>DET+NOUN+CASE_DEF_GEN</td>
<td>ar-rasā-il-i</td>
<td>the-messages</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>A--------FS2D</td>
<td>DET+ADJ+NSUFF_FEM_SG+ +CASE_DEF_GEN</td>
<td>al-qaṣīr-at-i</td>
<td>the-short</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>C-----------</td>
<td>CONJ</td>
<td>wa-</td>
<td>and</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>Z--------2D</td>
<td>DET+NOUN_PROP+ +CASE_DEF_GEN</td>
<td>al-ḥintar-i</td>
<td>the-internet</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>C-----------</td>
<td>CONJ</td>
<td>wa-</td>
<td>and</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>FN--------2R</td>
<td>NEG_PART+CASE_DEF_GEN</td>
<td>ḡayr-i</td>
<td>other/not-of</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>S-----3FS2-</td>
<td>POSS_PRON_3FS</td>
<td>-hā</td>
<td>them</td>
<td></td>
</tr>
</tbody>
</table>
He will notify them about that through SMS messages, the Internet, and other means.

<table>
<thead>
<tr>
<th>String</th>
<th>Token</th>
<th>Token Tag</th>
<th>Buckwalter’s M-Tags</th>
<th>Token Form</th>
<th>Token Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>سيخبرهم</td>
<td>F------------</td>
<td>FUT</td>
<td></td>
<td>sa-</td>
<td>will</td>
</tr>
<tr>
<td>بهذلك عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>VIIR-3MS--</td>
<td>IV3MS+IV+IVSUFF_MOOD:I</td>
<td>yu-ḥbir-u</td>
<td>he-notify</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S-----3MP4-</td>
<td>IVSUFF_DO:3MP</td>
<td>-hum</td>
<td>them</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>P------------</td>
<td>PREP</td>
<td>bi-</td>
<td>about/by</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD-----MS--</td>
<td>DEM_PRON_MS</td>
<td>dālika</td>
<td>that</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P------------</td>
<td>PREP</td>
<td>ʾan</td>
<td>by/about</td>
<td></td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>N----------2R</td>
<td>NOUN+CASE_DEF_GEN</td>
<td>ṭarīq-i</td>
<td>way-of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N----------2D</td>
<td>DET+NOUN+CASE_DEF_GEN</td>
<td>ar-rasā’il-i</td>
<td>the-messages</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A----------FS2D</td>
<td>DET+ADJ+NSUFF_FEM_SG+ +CASE_DEF_GEN</td>
<td>al-qaṣīr-at-i</td>
<td>the-short</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C------------</td>
<td>CONJ</td>
<td>wa-</td>
<td>and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Z----------2D</td>
<td>DET+NOUN_PROP+ +CASE_DEF_GEN</td>
<td>al-.internet-i</td>
<td>the-internet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C------------</td>
<td>CONJ</td>
<td>wa-</td>
<td>and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FN----------2R</td>
<td>NEG_PART+CASE_DEF_GEN</td>
<td>˙gayr-i</td>
<td>other/not-of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S----------3FS2-</td>
<td>POSS_PRON_3FS</td>
<td>-hā</td>
<td>them</td>
<td></td>
</tr>
</tbody>
</table>
He will notify them about that through SMS messages, the Internet, and other means.

<table>
<thead>
<tr>
<th>String</th>
<th>Token</th>
<th>Token Tag</th>
<th>Buckwalters's M-Tags</th>
<th>Token Form</th>
<th>Token Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>سيخبرهم</td>
<td>F------</td>
<td>FUT</td>
<td></td>
<td>sa-</td>
<td>will</td>
</tr>
<tr>
<td>بذلك</td>
<td>VI1A-3MS-</td>
<td>IV3MS+IV+IVSUFF_MOOD:I</td>
<td></td>
<td>yu-ϯbir-u</td>
<td>he-notify</td>
</tr>
<tr>
<td>عن طريق الرسائل القصيرة والإنترنت وغيرها</td>
<td>S-----3MP4-</td>
<td>IVSUFF_DO:3MP</td>
<td></td>
<td>-hum</td>
<td>them</td>
</tr>
<tr>
<td>عن طريق</td>
<td>P-------</td>
<td>PREP</td>
<td></td>
<td>bi-</td>
<td>about/by</td>
</tr>
<tr>
<td>الرسائل</td>
<td>SD-----MS--</td>
<td>DEM_PRON_MS</td>
<td></td>
<td>ḏālika</td>
<td>that</td>
</tr>
<tr>
<td>القصيرة والإنترنت</td>
<td>P-------</td>
<td>PREP</td>
<td></td>
<td>ʾan</td>
<td>by/about</td>
</tr>
<tr>
<td>S----3FS2-</td>
<td>POSS_PRON_3FS</td>
<td>DET+NOUN_PROP+CASE_DEF_GEN</td>
<td></td>
<td>ġayr-</td>
<td>other/not-of</td>
</tr>
<tr>
<td>S----3FS2-</td>
<td>POSS_PRON_3FS</td>
<td>DET+NOUN_PROP+CASE_DEF_GEN</td>
<td></td>
<td>-hā</td>
<td>them</td>
</tr>
</tbody>
</table>
Suppose you can list morphological analyses for a given input string . . .

<table>
<thead>
<tr>
<th>Morphs</th>
<th>Form</th>
<th>Token Tag</th>
<th>Lemma</th>
<th>Glosses per Morph</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>laY+(null)</td>
<td>ālā</td>
<td>VP-A-3MS--</td>
<td>ālā</td>
</tr>
<tr>
<td></td>
<td>liy~</td>
<td>āliy</td>
<td>A--------</td>
<td>āliy</td>
</tr>
<tr>
<td></td>
<td>liy~+u</td>
<td>āliy-u</td>
<td>A--------1R</td>
<td>āliy</td>
</tr>
<tr>
<td></td>
<td>liy~+i</td>
<td>āliy-i</td>
<td>A--------2R.</td>
<td>āliy</td>
</tr>
<tr>
<td></td>
<td>liy~+a</td>
<td>āliy-a</td>
<td>A--------4R</td>
<td>āliy</td>
</tr>
<tr>
<td></td>
<td>liy~+N</td>
<td>āliy-un</td>
<td>A--------1I</td>
<td>āliy</td>
</tr>
<tr>
<td></td>
<td>liy~+K</td>
<td>āliy-in</td>
<td>A--------2I</td>
<td>āliy</td>
</tr>
<tr>
<td>l+</td>
<td>āl</td>
<td>N--------R</td>
<td>āl</td>
<td>family/clan +</td>
</tr>
<tr>
<td>+ iy</td>
<td>-ī</td>
<td>S------1-S2-</td>
<td>ī</td>
<td>+ my</td>
</tr>
<tr>
<td>ilaY</td>
<td>īlā</td>
<td>P--------</td>
<td>īlā</td>
<td>to/towards</td>
</tr>
<tr>
<td>Iilay+</td>
<td>īlay</td>
<td>P--------</td>
<td>īlā</td>
<td>to/towards +</td>
</tr>
<tr>
<td>+ ya</td>
<td>-ya</td>
<td>S------1-S2-</td>
<td>ya</td>
<td>+ me</td>
</tr>
<tr>
<td>0a+liy+(null)</td>
<td>a-ī</td>
<td>VIIA-1-S--</td>
<td>waliya</td>
<td>I + follow/come after + [ind.]</td>
</tr>
<tr>
<td>0a+liy+a</td>
<td>a-liy-a</td>
<td>VISA-1-S--</td>
<td>waliya</td>
<td>I + follow/come after + [sub.]</td>
</tr>
</tbody>
</table>
Morphology

MorphoTrees

... organize the analyses into a hierarchy with the **string** as its **root**
MorphoTrees

... organize the analyses into a hierarchy with the string as its root and the full tokens as the leaves
... organize the analyses into a hierarchy with the **string** as its **root** and the **full tokens** as the **leaves**, grouped by their **lemmas**.
MorphoTrees

... organize the analyses into a hierarchy with the **string** as its **root** and the **full tokens** as the **leaves**, grouped by their **lemmas**, canonical forms
MorphoTrees

... organize the analyses into a hierarchy with the string as its root and the full tokens as the leaves, grouped by their lemmas, canonical forms and partitionings of the string into such forms:
and, so to be ready, intend concern, interest they to understand understanding to make understand
ElixirFM is a high-level implementation of Functional Arabic Morphology.
ElixirFM is a high-level implementation of Functional Arabic Morphology.

ElixirFM uses the Functional Morphology library for Haskell and extends it.
ElixirFM is a high-level implementation of Functional Arabic Morphology.

ElixirFM uses the Functional Morphology library for Haskell and extends it.

Morphology is modeled in terms of paradigms, grammatical categories, lexemes and word classes. The computation of analysis or generation is conceptually distinguished from the general-purpose linguistic model.
ElixirFM is a high-level implementation of Functional Arabic Morphology.

ElixirFM uses the Functional Morphology library for Haskell and extends it.

Morphology is modeled in terms of paradigms, grammatical categories, lexemes and word classes. The computation of analysis or generation is conceptually distinguished from the general-purpose linguistic model.

The lexicon of ElixirFM is derived from the open-source Buckwalter lexicon and from the PADT annotations. It is redesigned in important respects.
Outline

1. Introduction
2. Morphology
   - MorphoTrees
   - ElixirFM
3. Syntax and Beyond
   - Dependency Grammar
   - Analytical Syntax
   - Tectограмmatics
4. Software
   - TrEd Live
   - Encode Arabic
5. References
Dependency vs. Linearity

... by providing the basic necessities of life to its people, including medical care ...

**Arabic:**

bi-tawfīri ḍarūrīyātí al-ḥayātí al-asāsīyātí li-šābihā

by-giving-of necessities-of the-life the-basic to-people-of-it

wa-min baynihā ṛ-rī'āyātu aṭ-ṭibbīyātu

and-from between-of-them the-care the-medical
Dependency vs. Linearity

... by providing the basic necessities of life to its people, including medical care ...

bi-tawfīrī darūrīyātī al-ḥayātī al-asāsīyātī li-šaʿbiḥā
by-giving-of necessities-of the-life the-basic to-people-of-it

wa-min bayniḥā ar-rīʿāyātū at-ṭibbiyyātū
and-from between-of-them the-care the-medical
Dependency vs. Linearity

... by providing the basic necessities of life to its people, including medical care ...

by-giving-of necessities-of the-life the-basic to-people-of-it

and-from between-of-them the-care the-medical
In the section on literature, the magazine presented the issue of the Arabic language and the dangers that threaten it.

وفي ملف الأدب طرحت ...
In the section on literature, the magazine presented the issue of the Arabic language and the dangers that threaten it.

وَفِي مَلِفْ الأَدَبِّ طَرَحَتِ...
In the section on literature, the magazine presented the issue of the Arabic language and the dangers that threaten it.

وَقَيَّمَ مَفْلِظَّةَ الْأَدَّابِ طَرْحَتٍ . .
In the section on literature, the magazine presented the issue of the Arabic language and the dangers that threaten it.

وَفي مَلفُ الأَدَّبِ تَرَاحَتْ ...
In the section on literature, the magazine presented the issue of the Arabic language and the dangers that threaten it.

وَفِي مِلْفِ الأَدْبِ طَرَّحَتِ...
In the section on literature, the magazine presented the issue of the Arabic language and the dangers that threaten it.

وَفِي مَفْلِفَ الْأَدْبَبِ طَرْحَتْ . . .
Tectogrammatics

Description of *linguistic meaning* in its **semantic** and **pragmatic** aspects.

- **B** ملف *milaff* collection/file Masc.Sing.Def
- **C** أدب *adab* literature Masc.Sing.Def
- **B** طرح *ṭarah* to-present Ind.Ant.Act
- **B** مجلة *mağallah* magazine Fem.Sing.Def
- **N** قضية *qадdiyah* issue Fem.Sing.Def
- **N** لغة *luğah* language Fem.Sing.Def
- **N** عربي *arabīy* Arabic Adjective
- **N** و *wa* and Coordination
- **N** خطر *ḥatar* danger Masc.Plur.Def
- **N** هد *haddad* to-threaten Ind.Sim.Act
- **B** هي *hiya* it PersPronoun
- **B** هي *hiya* it PersPronoun
Tectogrammatics

Description of linguistic meaning in its semantic and pragmatic aspects.
Description of **linguistic meaning in its semantic and pragmatic aspects.**
Tectogrammatics

Description of **linguistic meaning** in its **semantic** and **pragmatic** aspects.

- **B** مِلَفٌ *milaff* collection/file Masc.Sing.Def
- **B** أدب *adab* literature Masc.Sing.Def
- **B** طرح *ṭarāḥ* to-present Ind.Ant.Act
- **B** مجلة *mağallah* magazine Fem.Sing.Def
- **N** قضَىَةٌ *qaḍīyah* issue Fem.Sing.Def
- **N** لغة *luğah* language Fem.Sing.Def
- **N** عَرَبِيّ *ḥarbiyy* Arabic Adjective
- **N** و *wa* and Coordination
- **N** خَطْرُ *ḥatār* danger Masc.Plur.Def
- **N** هَدَد *haddad* to-threaten Ind.Sim.Act
- **B** هُيَ *hiya* it PersPronoun
- **B** هُيَ *hiya* it PersPronoun
Outline

1. Introduction
2. Morphology
   - MorphoTrees
   - ElixirFM
3. Syntax and Beyond
   - Dependency Grammar
   - Analytical Syntax
   - Tectogrammatics
4. Software
   - TrEd Live
   - Encode Arabic
5. References
يُولد جميع الناس أَحْرَارًا متساويين في الكرامة والحقوق. وقد وُهَبوا عَقْلاً وضَحْمِيّاً وَعَلَّيْهِمْ أن يَعَامِلَ بَعْضُهُم بَعْضًا بِرُوحِ الْإِخَاهِ.

yuwladu jamiyEu {ln~aAsi OaHoraArFA mutasaAwynayn fiy {lokaraAmapi wa {loHuquwqi. waqado wuhibuwA EaqolAF waDamiyrFA waEalayohimo Oano yuEaAmila baEoDuhumo baEoDFA biruwHi {loIixaA’i.

يولد جميع الناس أَحْرَارًا متساويين في الكرامة والحقوق. وقد وُهَبوا عَقْلاً وضَحْمِيّاً وَعَلَّيْهِمْ أن يَعَامِلَ بَعْضُهُم بَعْضًا بِرُوحِ الْإِخَاهِ.

ywld jmyE AlnAs OHrArA mtsAwyn fy AlkrAmp wAlHqwq. wqd whbwA EqlA wDmyrA wElyhm On yEAm1 bEDhm bEDA brwH AlIxA’.
يُولد جميع الناس أَحْرَارًا متساويين في الكرامة والحقوق. وقد وُهِبوا عَقْلاً وضُميِّراً وعَلِيهم أن يُعامل بعضهم بعضاً بروح الإخاء.

\cap yUladu ^gamI\'u an-nAsi 'a.hrAraN mutasAwIna fI al-karAmaTi wa-al-.huqUqi. \cap wa-qad wuhibU \cap \\cap wa-qad wuhibUA 'aqlaN wa-.damIraN wa-'alayhim 'an yu\'Amila ba\'.duhum ba\'.daN bi-rU.hi al-'i_hA'i.
Encode Arabic

biruwHi {loIixaA’i ← بِرَوحُ الْإِلَّهَاءِ ← bi-rU.hi al-’i_hA’i

Implemented in Perl and available on CPAN as Encode-Arabic:

$encoded = encode "buckwalter", decode "arabtex", $decoded
$encoded = encode("buckwalter", decode("arabtex", $decoded))

Implemented in Haskell and available along with ElixirFM:

encoded = encode Buckwalter $ decode ArabTeX decoded
encoded = encode Buckwalter (decode ArabTeX decoded)
encoded = (encode Buckwalter . decode ArabTeX) decoded

[cmd] decode ArabTeX < decode.d | encode Buckwalter > encode.d
Encode Arabic

biruwHi {loIixaA’i ← بَرُوحُ أَلَّاَخَاءِ ← bi-rU.hi al-’i_hA’i

Implemented in Perl and available on CPAN as Encode-Arabic:

$encoded = encode "buckwalter", decode "arabtex", $decoded
$encoded = encode("buckwalter", decode("arabtex", $decoded))

Implemented in Haskell and available along with ElixirFM:

encoded = encode Buckwalter $ decode ArabTeX decoded
encoded = encode Buckwalter (decode ArabTeX decoded)
encoded = (encode Buckwalter . decode ArabTeX) decoded

[cmd] decode ArabTeX < decode.d | encode Buckwalter > encode.d
Encode Arabic

biruwHi {loIixaA’i ← بهُوحُ آلِأخاء ← bi-rU.hi al-’i_hA’i

Implemented in Perl and available on CPAN as Encode-Arabic:

$encoded = encode "buckwalter", decode "arabtex", $decoded
$encoded = encode("buckwalter", decode("arabtex", $decoded))

Implemented in Haskell and available along with ElixirFM:

encoded = encode Buckwalter $ decode ArabTeX decoded
encoded = encode Buckwalter (decode ArabTeX decoded)
encoded = (encode Buckwalter . decode ArabTeX) decoded

[cmd] decode ArabTeX < decode.d | encode Buckwalter > encode.d
Encode Arabic

biruwHi {loIixaA’i ← بِرُوحَ الْأَلْحَاءِ ← bi-rU.hi al-’i_hA’i

Implemented in **Perl** and available on CPAN as **Encode-Arabic**:  

```perl
$encoded = encode "buckwalter", decode "arabtex", $decoded
$encoded = encode("buckwalter", decode("arabtex", $decoded))
```

Implemented in **Haskell** and available along with **ElixirFM**:  

```haskell
encoded = encode Buckwalter $ decode ArabTeX decoded
encoded = encode Buckwalter (decode ArabTeX decoded)
encoded = (encode Buckwalter . decode ArabTeX) decoded
```

[cmd] decode ArabTeX < decode.d | encode Buckwalter > encode.d
Encode Arabic

biruwHi {loIixaA’i ← بِروحَ أَلِيْخَايِ ← bi-rU hi al-’i_hA’i

Implemented in Perl and available on CPAN as Encode-Arabic:

$encoded = encode "buckwalter", decode "arabtex", $decoded
$encoded = encode("buckwalter", decode("arabtex", $decoded))

Implemented in Haskell and available along with ElixirFM:

encoded = encode Buckwalter $ decode ArabTeX decoded
encoded = encode Buckwalter (decode ArabTeX decoded)
encoded = (encode Buckwalter . decode ArabTeX) decoded

[cmd] decode ArabTeX < decode.d | encode Buckwalter > encode.d
Outline

1 Introduction
2 Morphology
   • MorphoTrees
   • ElixirFM
3 Syntax and Beyond
   • Dependency Grammar
   • Analytical Syntax
   • Tectogrammatics
4 Software
   • TrEd Live
   • Encode Arabic
5 References
References


PADT++ http://ufal.mff.cuni.cz/padt/online/