

---

# Topologische Abhängigkeitsgrammatik fürs Arabisches

---

Vortrag von

Marwan Odeh

Betreut von Ralph Debusmann

Unter der Aufsicht von Prof. Gert Smolka

# Überblick

- Einführung
  - Phänomene
  - Modellierung in XDG
  - Abschluss und Demonstration
-

# Überblick

- Einführung
  - Phänomene
  - Modellierung in XDG
  - Abschluss und Demonstration
-

# Aufgabe

- Modellierung des *Arabischen* mit *Extensible Dependency Grammar* (XDG) (Debusmann, Duchier, Koller, Kuhlmann, Smolka, Thater 2004)
  - Phänomen-orientiert: viele Phänomene, kleines Lexikon
  - Proof-of-concept: XDG geeignet zur Behandlung des Arabischen?
-

# Arabisch

- Semitische Sprache (wie Hebräisch, Aramäisch)
  - Sprache des Korans
  - Eigenes Alphabet
  - Schreibweise von rechts nach links
  - Reiche Morphologie
  - Freie Wortstellung auf Satzebene
  - Einzigartige grammatische Kategorien
-

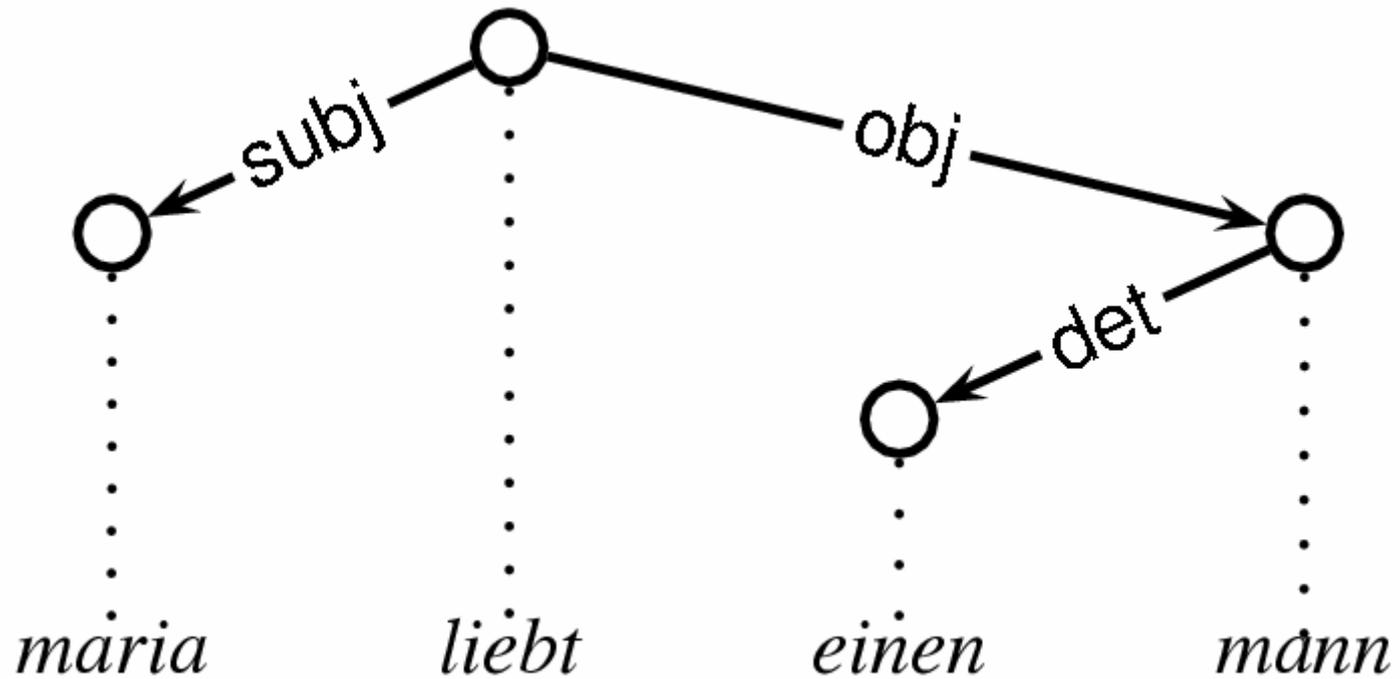
# Extensible Dependency Grammar (XDG)

- Meta-Grammatikformalismus zur Beschreibung natürlicher Sprache
  - Basiert auf *Abhängigkeitsgrammatik* (Tesnière 1959, Melcuk 1988)
  - Grammatische Beschreibung auf *mehreren Dimensionen*
  - Hier benutzt: *Immediate Dominance* (ID) und *Linear Precedence* (LP) wie in TDG (Duchier und Debusmann 2001, Debusmann 2001)
-

# ID-Dimension (Immediate Dominance)

- Beschreibt grammatische Funktionen
  - Knoten entsprechen 1:1 Wörtern
  - Ungeordneter Baum
  - Kantenmarkierungen: Subjekt, Objekt...
-

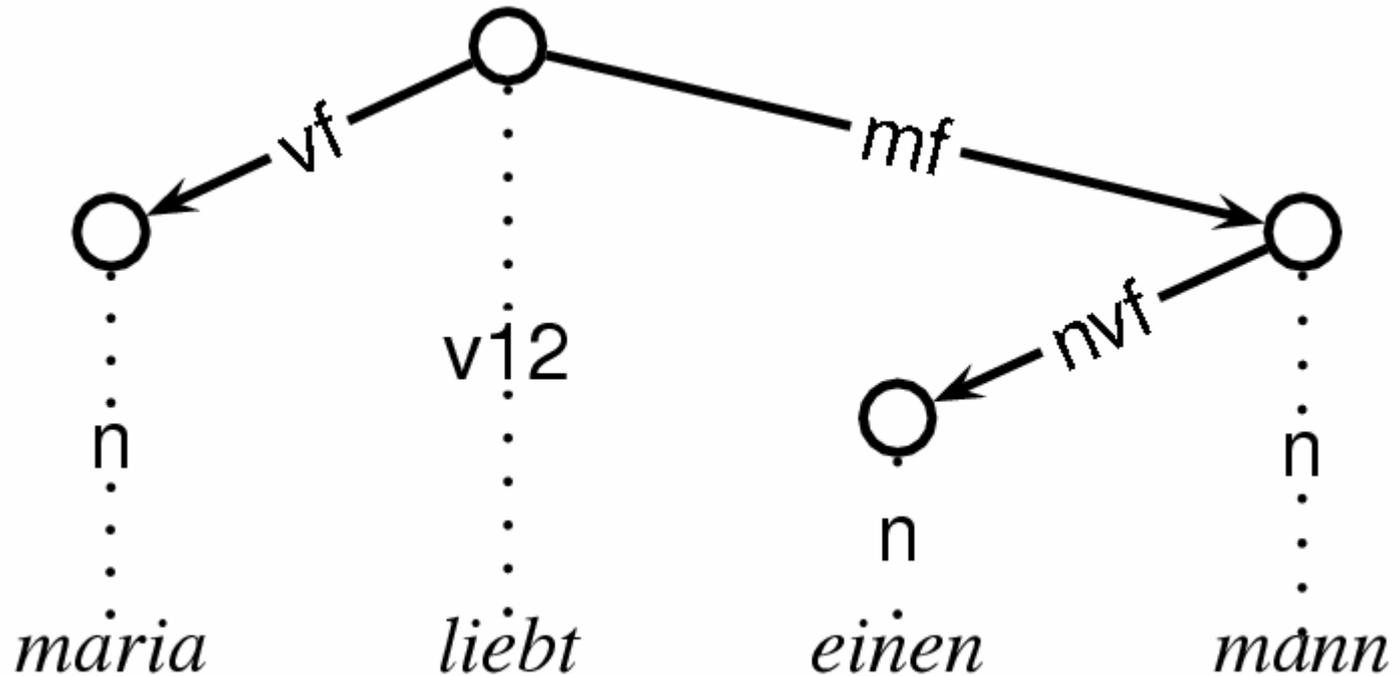
# ID-Beispielanalyse



# LP-Dimension (Linear Precedence)

- Beschreibt Wortstellung
  - Gleiche Knoten wie ID-Dimension
  - Geordneter, projektiver Baum
  - Kanten- und Knotenmarkierungen:  
topologische Felder
-

# LP-Beispielanalyse



# Überblick

- Einführung
  - Phänomene
  - Modellierung in XDG
  - Abschluss und Demonstration
-

# Phänomene

- Reiche Morphologie
  - Freie Wortstellung auf Satzebene
  - Mafool Motlak
  - Weitere Phänomene in der Ausarbeitung
-

# Reiche Morphologie

- Finites Verb bestimmt sogar Genus des Subjekts, z.B. yadhaku: Subjekt maskulin, tadhaku: Subjekt feminin:

yadhak-u	mohammed
يُضْحِكُ	محمد
lacht	Mohammed

# Reiche Morphologie

- Verbform bestimmt Genus des Subjekts, z.B. yadhaku: Subjekt maskulin, tadhaku: Subjekt feminin:

yadhak-u	mohammed
يُضْحِكُ	محمد
lacht	Mohammed

tadhak-u	mohammed
تُضْحِكُ	محمد
lacht	Mohammed

# Reiche Morphologie

- Verbform bestimmt Genus des Subjekts, z.B. yadhaku: Subjekt maskulin, tadhaku: Subjekt feminin:

yadhak-u	mohammed
يُضْحِكُ	محمد
lacht	Mohammed

tadhak-u	mohammed
* تُضْحِكُ	محمد
lacht	Mohammed

# Freie Wortstellung auf Satzebene

- Ermöglicht durch reiche Morphologie
  - Grammatische Funktion der NPs kann damit eindeutig bestimmt werden
-

# Beispiele

mohammed-a	fatat-u	al	tuhib-u
محمد	فتاة	ال	تحب
Mohammed	Frau	die	liebt

# Beispiele

mohammed-a	fatat-u	al	tuhib-u
محمد	فتاة	ال	تحب
Mohammed	Frau	die	liebt

mohammed-a tuhib-u fatat al

---

# Beispiele

mohammed-a	fatat-u	al	tuhib-u
محمد	فتاة	ال	تحب
Mohammed	Frau	die	liebt

mohammed-a tuhib-u fatat al

fatat al mohammed-a tuhib-u

---

# Beispiele

mohammed-a	fatat-u	al	tuhib-u
محمد	فتاة	ال	تحب
Mohammed	Frau	die	liebt

mohammed-a tuhib-u fatat al

fatat al mohammed-a tuhib-u

fatat al tuhib-u mohammed-a

---

# Beispiele

mohammed-a	fatat-u	al	tuhib-u
محمد	فتاة	ال	تحب
Mohammed	Frau	die	liebt

mohammed-a tuhib-u fatat al

fatat al mohammed-a tuhib-u

fatat al tuhib-u mohammed-a

tuhib-u fatat al mohammed-a

---

# Beispiele

mohammed-a	fatat-u	al	tuhib-u
محمد	فتاة	ال	تحب
Mohammed	Frau	die	liebt

mohammed-a tuhib-u fatat al

fatat al mohammed-a tuhib-u

fatat al tuhib-u mohammed-a

tuhib-u fatat al mohammed-a

tuhib-u mohammed-a fatat al

---

# Mafool Motlak

- Einzigartige grammatische Kategorie
  - Spezielle Verbform
  - Verstärkt das Verb
  - Steht am Satzende
  - Wiederholung: Verstärkung der Verstärkung
-

# Mafool Motlak: Beispiele

laib-an	walad-u al	laib-a
لعباً	الولدُ	لعبَ
spielen	Kind das	spielte

# Mafool Motlak: Beispiele

laib-an	walad-u al	laib-a
لعباً	الولدُ	لعبَ
spielen	Kind das	spielte

laib-an laib-an walad-u al laib-a

---

# Mafool Motlak: Beispiele

laib-an	walad-u al	laib-a
لعباً	الولدُ	لعبَ
spielen	Kind das	spielte

laib-an laib-an walad-u al laib-a

laib-an laib-an laib-an walad-u al laib-a

---

# Überblick

- Einführung
  - Phänomene
  - Modellierung in XDG
  - Abschluss und Demonstration
-

# Modellierung in XDG

- Verwendung des *XDG Development Kits* (XDK) (Debusmann, Duchier, Niehren 2004)
  - XDG-Instanz: wie in TDG (Duchier und Debusmann 2001, Debusmann 2001) fürs Deutsche
-

# Reiche Morphologie

- Lexikon: jedes Wort beschreibt Menge von *Agreement*-Tupeln aus Genus, Numerus und Kasus:

$mohammed - u = [agrs : \$masc \& sg \& nom]$

$mohammed - a = [agrs : \$masc \& sg \& acc]$

$mohammed = [agrs : \$masc \& sg]$

- Analyse: jeder Knoten wählt ein Tupel aus:

$\forall v \in V : agr(v) \in agrs(v)$

---

# Reiche Morphologie weiter

- Verben bestimmen Agreement ihrer Töchter

$$\forall h-l \rightarrow d : agr(d) \in outagrs(h)(l)$$

- Beispiele:

$$tadhaku = [outagrs : [subj : \$fem \& nom]]$$

$$yadhaku = [outagrs : [subj : \$masc \& nom]]$$

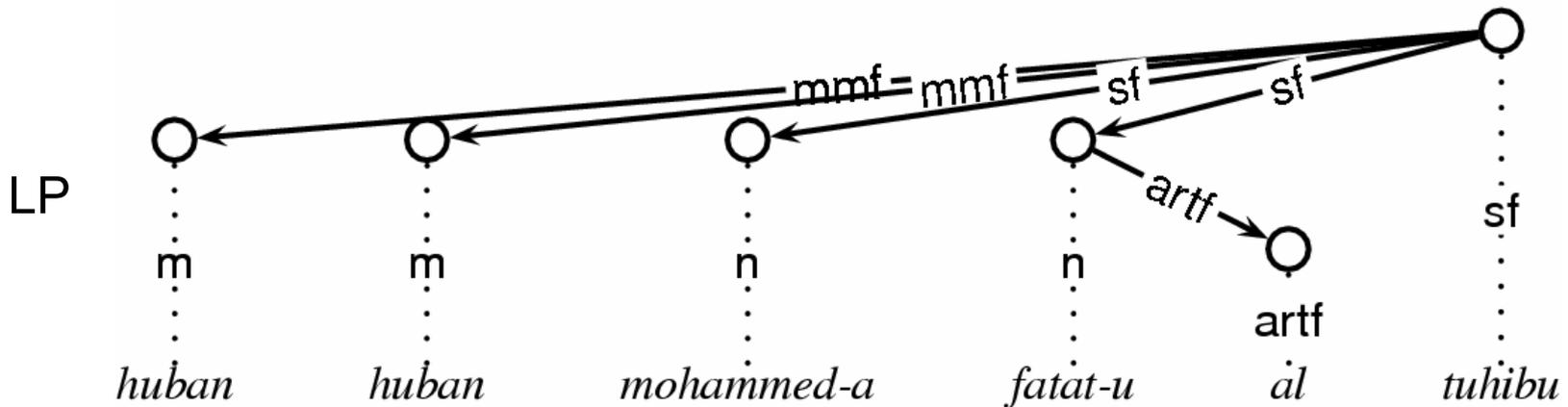
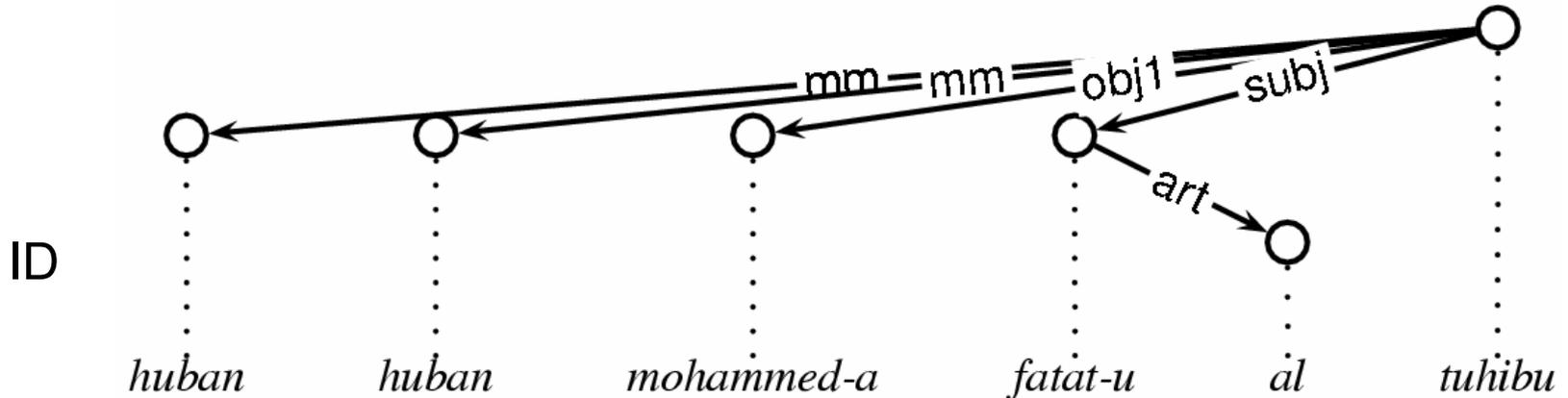
---

# Freie Wortstellung auf Satzebene

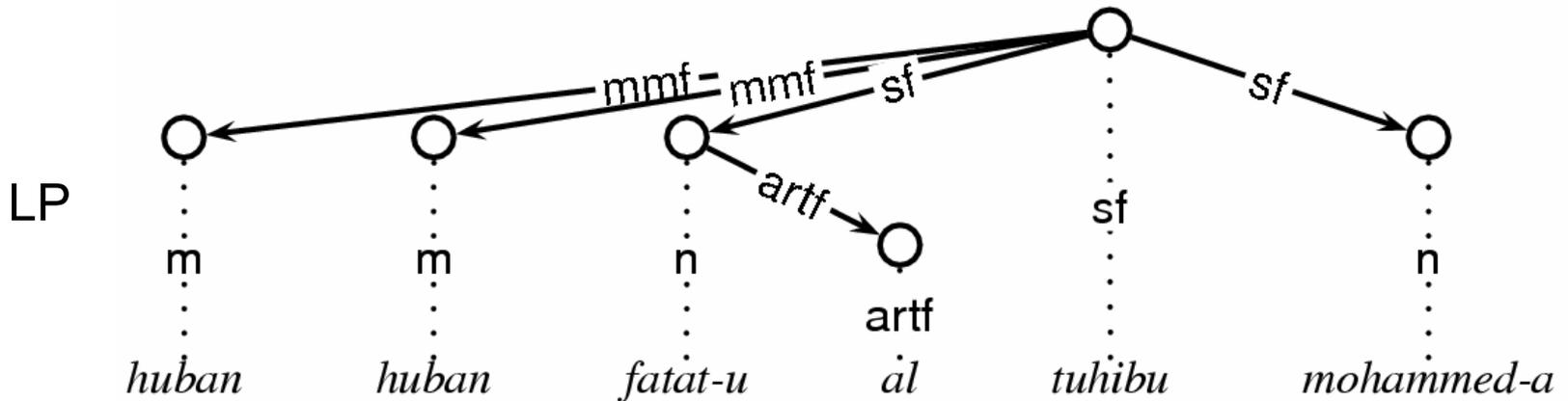
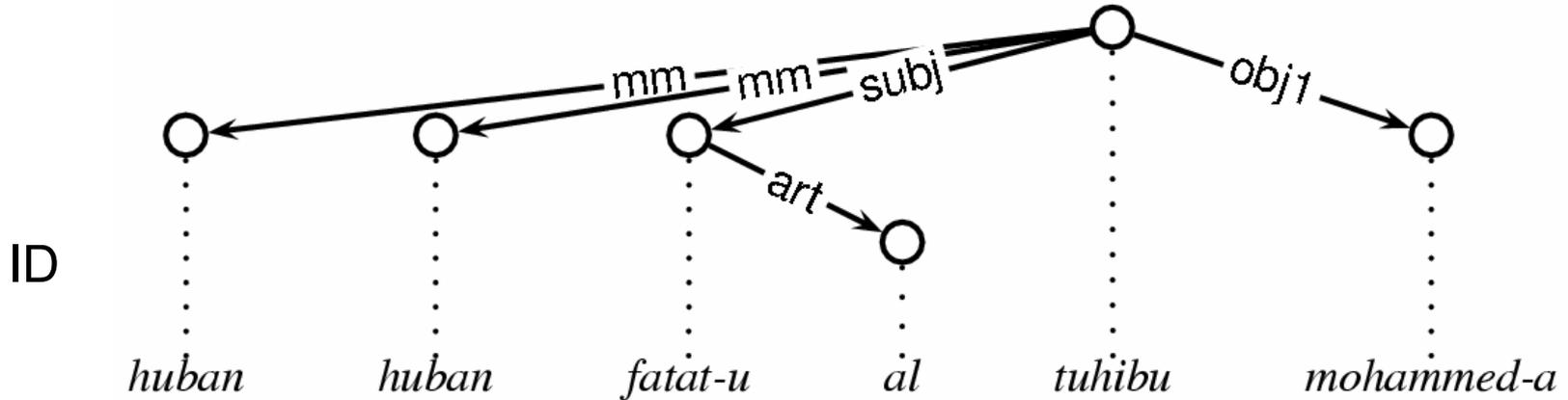
- Nutzung der sauberen ID/LP-Trennung in XDG
- Beschreibung der Wortstellung losgelöst von grammatischer Funktion
- topologische Satzstruktur fürs Arabische:

Mafool Motlaks (mmf)	Verb und andere Verb-Dependenten (sf)
----------------------	---------------------------------------

# Beispielanalyse



# Weitere Beispielanalyse



# Beispiel-Lexikoneintrag

$$tuhibu = \left[ \begin{array}{l} id : \left[ \begin{array}{l} in : \{\} \\ out : \{subj!, obj!!, mm^*, adv^*\} \end{array} \right] \\ lp : \left[ \begin{array}{l} in : \{\} \\ out : \{mmf^*, sf^*\} \\ on : \{sf\} \end{array} \right] \end{array} \right]$$

# Überblick

- Einführung
  - Phänomene
  - Modellierung in XDG
  - Abschluss und Demonstration
-

# Abschluss

- Proof-of-concept: Arabisch kann mit XDG modelliert werden
  - Für Kombination reiche Morphologie + freie Wortstellung sehr gut geeignet
  - Vermutung: Topologische Felder funktionieren auch fürs Arabische
  - Grundlage für größere Grammatik fürs Arabische (mit großem Lexikon)
-

# Demonstration

The image displays the XDK (XDG Development Kit) graphical user interface. The main window shows project settings for 'Arabic ul' and a list of examples. An 'About' dialog box is open, providing version information and developer details. Below the main window, the 'Oz Explorer' shows a file tree with a single file icon. On the right side, two 'XDK: 1' windows display linguistic analysis diagrams. Each diagram consists of two rows of nodes representing words and their grammatical functions. Blue arrows indicate dependencies between nodes in the top row and the bottom row. The top row nodes are labeled with grammatical functions like 'nmf', 'nmf', 'subj', and 'art'. The bottom row nodes are labeled with morphological forms like 'm', 'n', 'o', 'a', 'sf', and 'o'. The 'Oz Inspector' window at the bottom right shows a JSON-like structure representing the selected node's data, including attributes, entry information, model, and index.

**XDK: Main window**

Project Search Dimensions Principles Outputs Extras  
Grammar: Arabic ul  
Examples: Arabic.txt  
Inspect lexical entries

-----  
\* Teil 1

**XDK: About**

XDG Development Kit: Graphical User Interface  
Copyright 2001-2004  
Version 1.0.2  
by Ralph Debusmann <rade@ps.uni-sb.de> (Saarland University) and  
Dennis Duchier <duchier@ps.uni-sb.de> (LORIA, Nancy)

-----  
\* Beispiel 2.1.4: Ungrammatikalisch (zwei Subjekte, kein Objekt)  
mohammed-u fatat-u al tuhibu

-----  
\* Beispiel 2.1.5: Ungrammatikalisch (zwei Objekte, kein Subjekt)  
mohammed-a fatat-a al tuhibu

-----  
\* Beispiel 2.2.1: mohammed-a vor fatat-u  
fatat-u al mohammed-a tuhibu

-----  
\* Beispiel 2.2.2: fatat-u vor mohammed-a  
mohammed-a fatat-u al tuhibu

-----  
\* Beispiel 2.3.1: Adjektive und Adjektivartikel  
tuhibu mohammed-a tawilat-u al janilat-u al fatat-u al

Solve | huban huban fatat-u al tuhibu mohammed-a

**Oz Explorer**

Explorer Move Search Nodes Hide Options

Time: 47ms 0 0 1 0 Depth: 1

**XDK: 1**

id  
huban 1 huban 2 mohammed-a 3 fatat-u 4 al 5 tuhibu 6

lp  
m huban 1 m huban 2 n mohammed-a 3 o fatat-u 4 a fatat-u 5 o tuhibu 6

**XDK: 1**

id  
huban 1 huban 2 fatat-u 3 al 4 tuhibu 5 mohammed-a 6

lp  
m huban 1 m huban 2 o fatat-u 3 a fatat-u 4 sf tuhibu 5 o mohammed-a 6

**Oz Inspector**

Inspector Selection Options

```
o (entryIndex: 1
  id: o (attrs: o (agr: ' ' (' $' ))
    entry: o (' in': [nm])
      out: o (adv: ' *' pp: ' '))
    model: o (eq: [2]
      eqdown: [2]
      equip: [2 6]
      index: 2
      labels: [nm]
      mothers: [6]
      up: [6] ))
  index: 2
  word: huban)
```

# Literatur

- Ralph Debusmann 2001: A Declarative Grammar Formalism for Dependency Grammar, Diplomarbeit, Computerlinguistik, Universität des Saarlandes, Saarbrücken/GER
  - Ralph Debusmann, Denys Duchier, Alexander Koller, Marco Kuhlmann, Gert Smolka und Stefan Thater 2004: A Relational Syntax-Semantics Interface Based on Dependency Grammar, Proceedings of COLING 2004, Genf/CH
  - Denys Duchier und Ralph Debusmann 2001: Topological Dependency Trees: A Constraint-Based Account of Linear Precedence, Proceedings of ACL 2001, Toulouse/FRA
  - Igor Melcuk 1988: Dependency Syntax: Theory and Practice, State University Press of New York, Albany/USA
  - Lucien Tesniere 1959: Elements de Syntaxe Structurale, Klincksiek, Paris/FRA
  - Mahier s. Ahmed 1988: *An Expert System for Understanding Arabic Sentences*, Cairo University
  - W. Wright 1967: *A Grammar of the Arabic Language*, Third Edition, Cambridge, England
-