

Topologische Dependenz- grammatik fürs Arabische

Abschlussbericht des Fortgeschrittenenpraktikums

Von Marwan Odeh
Betreut von Ralph Debusmann
Unter der Aufsicht von Prof. Gert Smolka

08. Dezember 2004

Inhaltsverzeichnis

Abschnitt 1.....	4
1. Einleitung	4
2. Phänomene.....	5
Phänomen 1: Verbstellung.....	5
Beispiel 2.1.1:	5
Beispiel 2.1.2:	6
Beispiel 2.1.3:	6
Beispiel 2.1.4:	7
Beispiel 2.1.5:	7
Phänomen 2: Nominalphrasen-Stellung.....	8
Beispiel 2.2.1:	8
Beispiel 2.2.2:	9
Phänomen 3: Adjektive und Adjektivartikel	9
Beispiel 2.3.1:	9
Beispiel 2.3.2:	10
Beispiel 2.3.3:	10
Phänomen 4: Mafool Motlak	11
Beispiel 2.4.1:	11
Beispiel 2.4.2:	11
Beispiel 2.4.3:	11
Phänomen 5: Verb-Subjekt-Genus-Agreement.....	12
Beispiel 2.5.1:	12
Beispiel 2.5.2:	12
Beispiel 2.5.3:	12
Beispiel 2.5.4:	12
Abschnitt 2.....	13
Arabisch mit Hilfe von Topologischer Dependenzgrammatik.....	13
1. ID-Teil.....	13
1.1 Verben.....	13
Intransitive Verben	14
Transitive Verben	15
Ditransitive Verben.....	15
1.2 Artikel	16
1.3 Adjektive.....	17
1.4 Appositionen	17
1.5 Korrelationen.....	17
1.6 Adverbien	17
1.7 Mafool Motlak.....	18
1.8 Präpositionen	18
1.9 Nomina.....	18
1.10 Zusammengesetzte Nomina	19
Beispiel 1.10.1:	19

Beispiel 1.10.2:	20
Beispiel 1.10.3:	20
Beispiel 1.10.4:	21
1.11 Namen.....	21
2. LP-Teil.....	22
2.1 Verben.....	23
2.2 Artikel	23
2.3 Das Adjektive	24
2.4 Appositionen	24
2.5 Korrelationen.....	24
2.6 Adverbien	25
2.7 Mafool Motlak.....	25
2.8 Präpositionen	26
2.9 Nomina.....	26
2.10 Zusammengesetzte Nomina	26
2.11 Namen.....	27
Abschnitt 3.....	27
Abschluss.....	27
Literatur	28

Abschnitt 1

1. Einleitung

In meinem Fortgeschrittenenpraktikum habe ich mir vorgenommen, mit dem Grammatikformalismus Topologische Dependenzgrammatik (TDG) (Duchier und Debusmann 2001, ACL-Konferenz) und (Debusmann 2001, Diplomarbeit) einige interessante Phänomene des Arabischen zu modellieren.

Arabisch ist in vielerlei Hinsicht eine interessante Sprache:

- a) Arabisch wird statt von links nach rechts von rechts nach links geschrieben,
- b) es besitzt einzigartige grammatische Kategorien (Adjektivartikel und „Mafool Motlak“),
- c) hat eine reiche Morphologie,
- d) und auf Satzebene eine sehr freie Wortstellung, freier sogar als im Deutschen.

Während Punkt 1 bei der grammatischen Modellierung keine Schwierigkeiten bereitet, stellen die Punkte 2-4 herkömmliche Grammatikformalismen, die auf kontextfreier Grammatik basieren, vor einige Probleme. Besonders die reiche Morphologie und die daraus resultierende freie Wortstellung sind hier problematisch.

In diesem Fopra zeige ich, wie diese Schwierigkeiten auf elegante Weise mit Topological Dependency Grammar (TDG) gelöst werden können. Dazu habe ich eine „toy grammar“ entwickelt, die u.a. die einzigartigen Kategorien des Adjektivartikels, des „Mafool Motlak“, die reiche Morphologie, und die auf Satzebene freie Wortstellung adäquat behandelt, und die ich in dieser Ausarbeitung ausführlich erkläre.

Ich beginne mit einer Beschreibung einiger interessanter Phänomene (Teil 1), die ich in meiner Grammatik behandle, bevor ich die Grammatik selbst beschreibe, wobei ich besonderen Wert auf den Aufbau des Lexikons und der lexikalischen Klassen legen werde (Teil 2). In Teil 3 schließe ich die Ausarbeitung ab.

2. Phänomene

Phänomen 1: Verbstellung

In einem arabischen Satz kann das Verb an jeder Stelle stehen, solange es nicht dafür sorgt, dass eine Nominalphrase unterbrochen wird:

Beispiel 2.1.1:

- Das Verb „tuhibu“ ist im folgenden Satz an erster Stelle (am Anfang des Satzes¹). Das Suffix „-u“ von „fatat-u“ zeichnet dieses als Subjekt aus:

mohammed-a	fatat-u	al	tuhibu
محمد	فتاة	ال	تحب
Mohammed	Mädchen	das	liebt

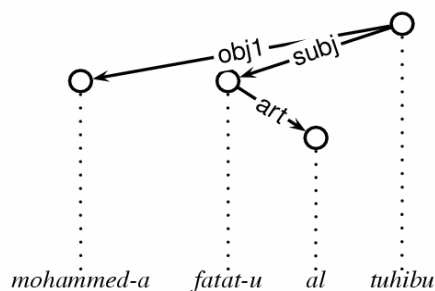


Abbildung 1 : ID-Baum

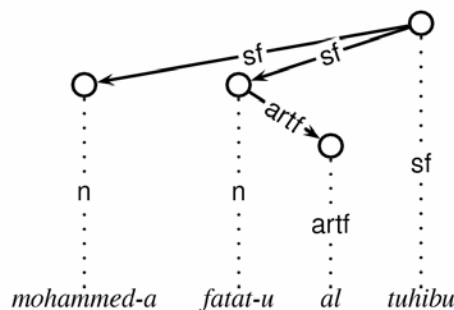


Abbildung 2 :LP-Baum

¹Ich werde die arabischen Beispielsätze konsequent von rechts nach links schreiben. Daher ist in Beispiel II.1.1 das Verb „tuhibu“ ganz rechts an erster Stelle.

Beispiel 2.1.2:

- In diesem Beispiel steht das Verb an zweiter Stelle:

mohammed-a	Tuhibu	fatat-u	al
محمد	تحب	فتاة	ال
Mohammed	Liebt	Mädchen	das

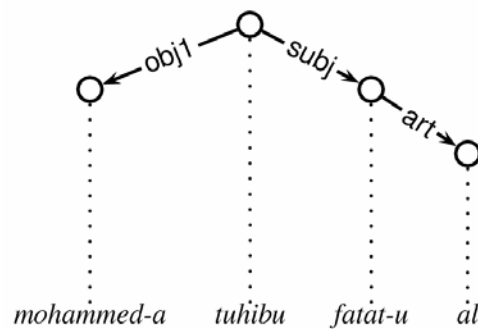


Abbildung 3:ID-Baum

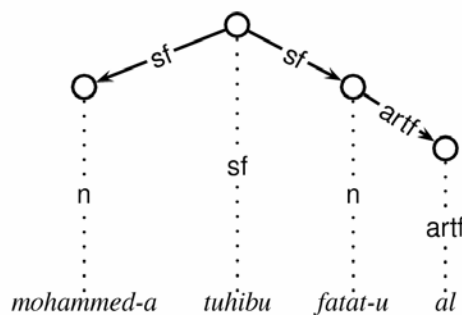


Abbildung 4:LP-Baum

Beispiel 2.1.3:

- Und in diesem Beispiel an dritter Stelle:

tuhibu	mohammed-a	fatat-u	al
تحب	محمد	فتاة	ال
liebt	Mohammed	Mädchen	das

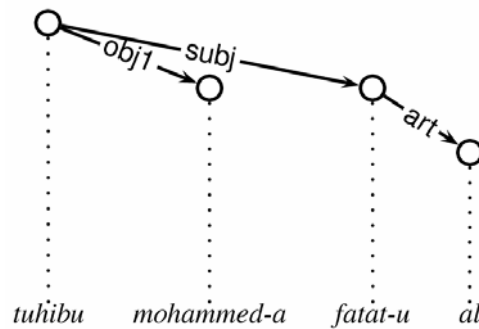


Abbildung 5: ID-Baum

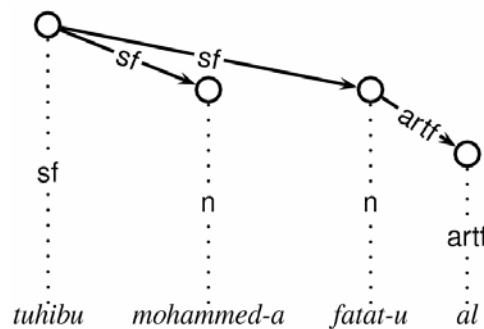


Abbildung 6: LP-Baum

Wie bereits oben angedeutet können Nomina im Arabischen mit Kasus-Markierern suffixiert werden. So zeichnet die Endung „-u“ ein Nomen als Subjekt, und die Endung „-a“ als Objekt aus. Da das transitive Verb „tuhibu“ genau ein Subjekt und genau ein Objekt verlangt, sind die beiden folgenden Sätze dann ungrammatisch:

Beispiel 2.1.4:

*Mohammed-u fatat-u al tuhibu

Beispiel 2.1.5:

*Mohammed-a fatat-a al tuhibu

In Beispiel 2.1.4 sind beide Nomina als Subjekt markiert, und in 2.1.5 als Objekt. Also gibt es in 2.1.4 ein Subjekt zuviel und ein Objekt zuwenig. 2.1.5 verhält sich symmetrisch.

Phänomen 2: Nominalphrasen-Stellung

Nicht nur das Verb, sondern auch die Nominalphrasen können innerhalb des Satzes frei bewegt werden. Die Voraussetzung dafür ist, dass aus ihrer morphologischen Markierung eindeutig hervorgeht, welche syntaktische Funktion sie ausüben. Im untenstehenden Beispiel steht so „Mohammed-a“ vor „al fatat-u“:

Beispiel 2.2.1:

fatat-u	al	mohammed-a	tuhibu
فتاة	ال	محمد	تحب
Mädchen	das	Mohammed	liebt

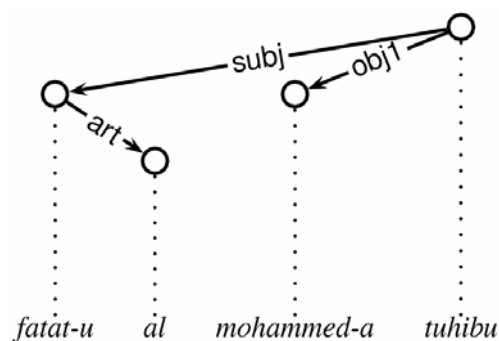


Abbildung 7: ID-Baum

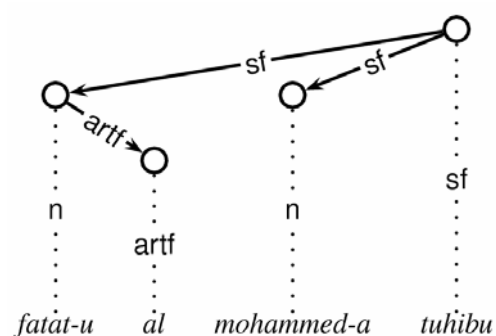


Abbildung 8: LP-Baum

Genauso grammatikalisch ist aber auch der folgende Satz, in dem die Ordnung der beiden Nominalphrasen vertauscht ist. Dieses Phänomen ist übrigens vergleichbar mit dem sogenannten „Scrambling“ im Deutschen, das die vielen möglichen Permutationen von Verbkomplementen im deutschen „Mittelfeld“ beschreibt:

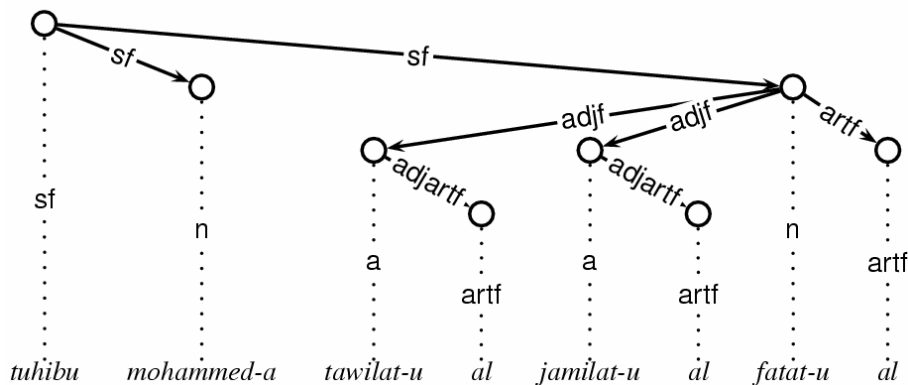


Abbildung 10 : LP-Baum

Hier wird das Nomen „fatat-u“ von zwei Adjektiven („jamilat-u“ und „tawilat-u“) modifiziert, die jeweils einen eigenen Adjektivartikel bekommen. Außerdem haben die Adjektive die gleiche Endung „-u“, wie das entsprechende Nomen „fatat-u“, d.h. sie kongruieren.

Ich werde die Kongruenz zwischen Adjektiven und Nomina ebenso wie andere morphologische Phänomene mit der Idee des „Agreement“ behandeln. Hierbei ordne ich jedem Wort eine Menge von möglichen Agreement-Tupeln zu, die aus den drei Komponenten Kasus (Nominativ, Akkusativ, jar für Objekte von Präpositionen), Genus (Maskulinum und Femininum) und Numerus (Singular und Plural) bestehen. Jeder Knoten in der letztendlichen Analyse wählt aus dieser Menge genau ein Agreement-Tupel aus.

Nun kann ich für alle Dependenzkanten von Nomina zu Adjektiven die Agreement-Tupel beider beteiligten Knoten gleichsetzen, um den Effekt der Kongruenz zu erzielen. Damit schließe ich korrekterweise die folgenden, ungrammatischen Sätze aus:

Beispiel 2.3.2:

*tuhibu mohammed-a tawilat-u al jamilat-a al fatat-u al

Beispiel 2.3.3:

*tuhibu mohammed-a tawilat-a al jamilat-u al fatat-u al

In Beispiel 2.3.2 stimmt die Markierung von „jamilat-a“ (Akkusativ-Suffix) nicht mit der des Nomens „fatat-u“ (Nominativ-Suffix) überein. In 2.3.3 stimmt die Markierung von „tawilat-a“ nicht.

Phänomen 4: Mafool Motlak

Das Arabische bietet das einzigartige Phänomen des „Mafool Motlak“. Hierbei wird das finite Verb am Ende eines Satz wiederholt, um seine Bedeutung zu verstärken. Hier ein Beispiel:

Beispiel 2.4.1:

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

Die Bedeutung dieses Satzes ist in etwa „Das Kind spielte sehr intensiv“. Das wiederholte Verb am Ende des Satzes hat genau die Kategorie oder syntaktische Funktion „Mafool Motlak“. Diese können nur von manchen Verben gebildet werden. Morphologisch werden Verben dieser Kategorie mit dem Suffix „-an“ ausgestattet. Hier ist eine ID-Analyse des obigen Beispiels:

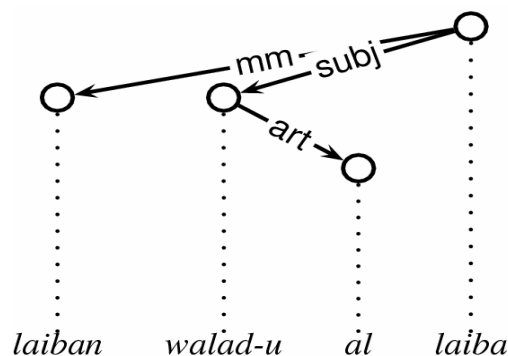


Abbildung 11: ID-Baum

Interessanterweise können auch mehrere „Mafool Motlaks“ benutzt werden, um das modifizierte Verb noch mehr zu verstärken:

Beispiel 2.4.2:

laiban laiban walad-u al laiba

Beispiel 2.4.3:

laiban laiban laiban walad-u al laiba

Phänomen 5: Verb-Subjekt-Genus-Agreement

Interessanterweise können die finiten Verben im Arabischen ihren Subjekten deren Genus vorschreiben. So bedeuten die beiden Verbformen „tuhibu“ und „yuhibu“ beide „liebt“, erstere kann jedoch nur für feminine Subjekte, und letztere für maskuline Subjekte verwendet werden.

Beispiel 2.5.1:

fatat-a al mohammed-u yuhibu

Beispiel 2.5.2:

*fatat-a al mohammed-u tuhibu

Beispiel 2.5.3:

mohammed-a fatat-u al tuhibu

Beispiel 2.5.4:

*mohammed-a fatat-u al yuhibu

In Beispiel 2.5.1 ist das Subjekt „mohammed-u“ korrekterweise maskulin und kongruiert mit dem geforderten Genus des Verbs „yuhibu“. Beispiel 2.5.2 ist ungrammatikalisch, weil das Verb „tuhibu“ ein feminines Subjekt fordert, jedoch ein maskulines („mohammed-u“) erhält. Die Beispiele 2.5.3 und 2.5.4 verhalten sich symmetrisch.

Ich modelliere die Forderung des Verbs nach Genus-Kongruenz wie folgt. Jedes Verb bekommt im Lexikon eine Abbildung von syntaktischen Funktionen auf Mengen von Agreement-Tupeln zugeordnet. Abhängig von dieser Abbildung und jeweiligen Kantenlabel bestimme ich nun das Agreement-Merkmal der Töchter. Z.B. lege ich fest, dass Verben wie „yuhibu“ zum einen Nominativ-Kasus vom Subjekt fordern, und zudem, dass es maskulin ist. „tuhibu“ fordert ebenfalls Nominativ-Kasus, und zudem, dass das Subjekt feminin ist.

Abschnitt 2

Arabisch mit Hilfe von Topologischer Dependenzgrammatik

In diesem Abschnitt beschreibe ich die Grammatik, die ich im Laufe meines Forschungspraktikums geschrieben habe. Ich beginne mit dem ID-Teil (Immediate Dominance) in Abschnitt 1, bevor ich weiter gehe mit dem LP-Teil (Linear Precedence) in Abschnitt 2.

1. ID-Teil

Zunächst definiere ich die Menge der syntaktischen Funktionen (grammatische Rollen):

Rolle	Steht für
<i>adj</i>	Adjektiv
<i>adjart</i>	Adjektivartikel
<i>adv</i>	Adverb
<i>app</i>	Apposition
<i>art</i>	Artikel
<i>cor</i>	Korrelation
<i>gen</i>	Genitiv
<i>mm</i>	Mafool Motlak
<i>obj1</i>	Erste Objekt
<i>obj2</i>	Zweite Objekt
<i>pp</i>	Präposition Phase
<i>prep</i>	Präposition
<i>prepobj</i>	Präposition Objekt
<i>subj</i>	Subjekt

1.1 Verben

Ich definiere zunächst die Klasse *t-fin-id* für finite Verben:

$$t - fin - id = \left[\begin{array}{l} labels_{id} : \{ \} \\ valency_{id} : \{ subj, adv^*, pp^*, mm^* \} \end{array} \right]$$

- a) *labels_{id}*: Da ich keine Nebensätze behandle, ist jedes finite Verb die Wurzel des Baums (kann keine eingehenden Kanten haben).
- b) *valency_{id}*: Finite Verben verlangen ein Subjekt, und können von jeweils beliebig vielen Adverbien, präpositionalen Modifikatoren, und Mafool Motlaks modifiziert werden.

Finite Verben bestimmen zusätzlich die Menge der möglichen Agreement-Tupel ihrer Subjekte (wie in Abschnitt 1 erläutert).

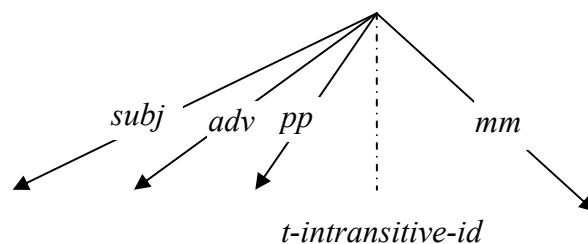
Intransitive Verben

Der ID-Teil des lexikalischen Typs für intransitiven Verben sieht so aus:

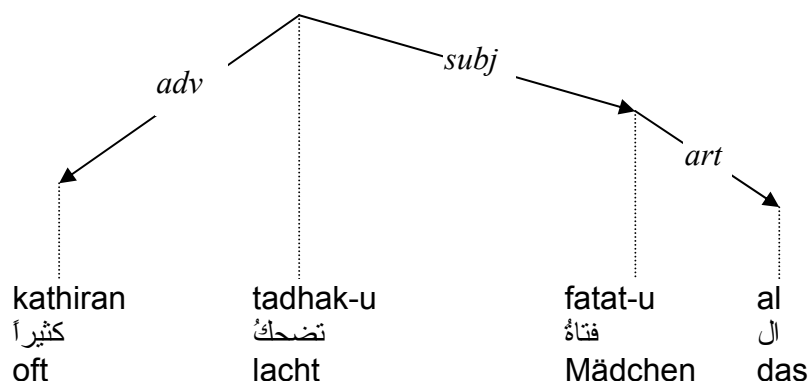
$$t-intransitiv-id = t-fin-id \& \left[\begin{array}{l} labels_{id} : \{ \} \\ valency_{id} : \{ \} \end{array} \right]$$

D.h., intransitive Verben erben von der Klasse für finite Verben, und fügen keine weiteren Argumente hinzu. Ich schreibe das kaufmännische und & für die Kombination von lexikalischen Klassen.

Die folgende Grafik zeigt die möglichen ausgehenden Kanten eines Knotens, der einem intransitiven Verb entspricht:



Da ID-Bäume ungeordnet sind, ist die Ordnung der Kanten hier nicht relevant. Untenstehend zeige ich nun eine Beispielanalyse eines Satzes mit intransitivem Verb:



Transitive Verben

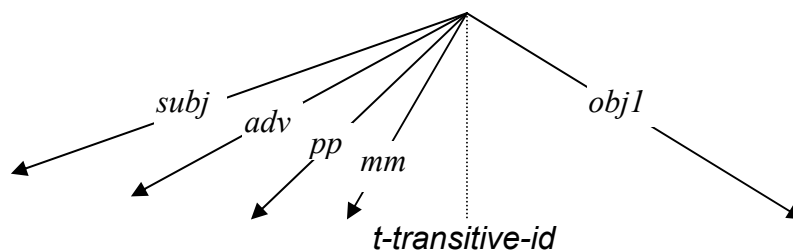
Transitive Verben erben von intransitiven Verben und verlangen zusätzlich entweder ein direktes Objekt (*obj1*) oder ein präpositionales Komplement (*prep*):

$$t\text{-transitive-id} = t\text{-intransitive-id} \& [valency_{id} : \{(obj1 | prep)\}]$$

Das Ergebnis sieht so aus:

$$t\text{-transitive-id} = \left[\begin{array}{l} labels_{id} : \{\} \\ valency_{id} : \{subj, adv^*, pp^*, mm^*, (obj1 | prep)\} \end{array} \right]$$

Die folgende Grafik beinhaltet die möglichen ausgehenden Kanten eines Knotens, der einem transitiven Verb entspricht (im Beispiel mit *obj1*-Komplement):



Ditransitive Verben

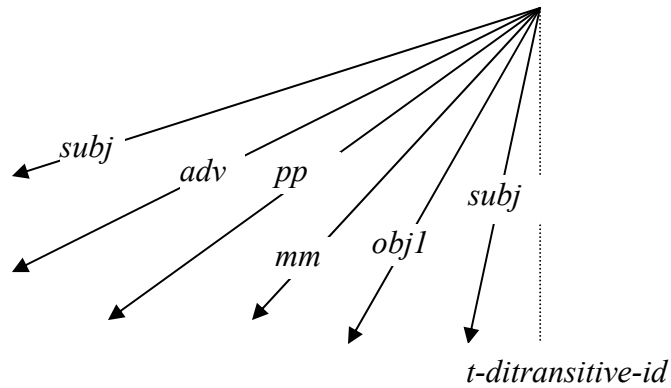
Ditransitive Verben erben von transitiven Verben, und ergänzen ein indirektes Objekt-Komplement (*obj2*):

$$t\text{-ditransitive-id} = t\text{-transitive-id} \& [valency_{id} : \{obj2\}]$$

Das Ergebnis sieht man wie folgt:

$$t\text{-ditransitive-id} = \left[\begin{array}{l} labels_{id} : \{\} \\ valency_{id} : \{subj, adv^*, pp^*, mm^*, (obj1 | prep), obj2\} \end{array} \right]$$

Hier ist eine Grafik, die die möglichen ausgehenden Kanten eines ditransitiven Verbs zeigt:



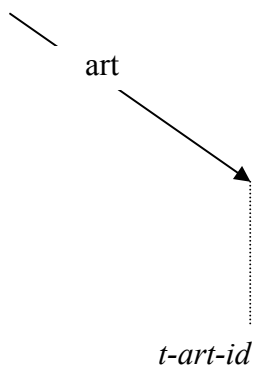
1.2 Artikel

Artikel erben vom lexikalischen Typ *t-art-id*:

$$t-art-id = \left[\begin{array}{l} labels_{id} : \{art, adjart\} \\ valency_{id} : \{\} \end{array} \right]$$

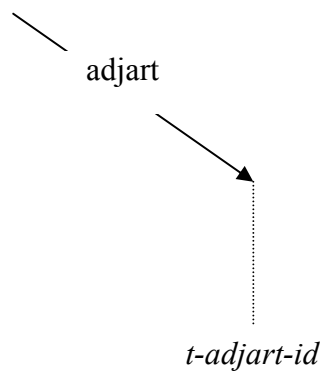
Abhängig von seiner eingehenden Kante kann ein Artikel entweder von einem Nomen oder von einem Adjektiv abhängig sein:

normaler Artikel
(von Nomen abhängig)



Adjektivartikel
(von Adjektiv abhängig)

oder

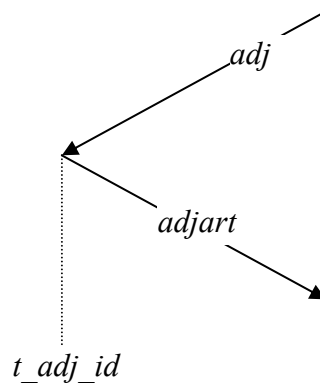


1.3 Adjektive

Adjektive erben vom folgenden lexikalischen Typ:

$$t - adj - id = \left[\begin{array}{l} labels_{id} : \{adj\} \\ valency_{id} : \{adjart\} \end{array} \right]$$

D.h., Adjektive haben eine eingehende Kante, die mit *adj* beschriftet ist, und verlangen einen Adjektivartikel:



1.4 Appositionen

Die eingehende Kante von Appositionen ist mit *app* gelabelt:

$$t - app - id = \left[\begin{array}{l} labels_{id} : \{app\} \\ valency_{id} : \{\} \end{array} \right]$$

1.5 Korrelationen

Korrelationen haben eingehende Kante *cor*:

$$t - cor - id = \left[\begin{array}{l} labels_{id} : \{cor\} \\ valency_{id} : \{\} \end{array} \right]$$

1.6 Adverbien

Und Adverbien haben eine eingehende Kante, die mit *adv* gelabelt ist:

$$t - adv - id = \left[\begin{array}{l} labels_{id} : \{adv\} \\ valency_{id} : \{\} \end{array} \right]$$

1.7 Mafool Motlak

Mafool-Motlak erbt vom folgenden lexikalischen Typ:

$$t - mm - id = \left[\begin{array}{l} labels_{id} : \{mm\} \\ valency_{id} : \{adv^*, pp^*\} \end{array} \right]$$

1.8 Präpositionen

Ich unterscheide zwei Klassen von Präpositionen:

a. Präpositionale Komplemente (eingehendes Kantenlabel: *prep*):

$$t - prep - id = \left[\begin{array}{l} labels_{id} : \{prep\} \\ valency_{id} : \{prepobj\} \\ outagrs_{id} : \{prepobj : \$jar\} \end{array} \right]$$

Hier ist *prepobj* das von der Präposition abhängige Nomen. Die Menge *outagrs* dient zur Festlegung der Abbildung von ausgehenden Kantenlabels auf mögliche Agreement-Tupel. Hier lege ich fest, dass das Objekt der Präposition den Kasus *jar* erhält.

b. Präpositionale Adjunkte (eingehendes Kantenlabel *pp*):

$$t - pp - id = \left[\begin{array}{l} labels_{id} : \{pp\} \\ valency_{id} : \{prepobj\} \\ outagrs_{id} : \{prepobj : \$jar\} \end{array} \right]$$

Diese Klasse verhält sich bis auf das eingehende Kantenlabel analog.

1.9 Nomina

$$t - nomen - id = [labels_{id} : \{subj, obj1, obj2, gen, prepobj\}]$$

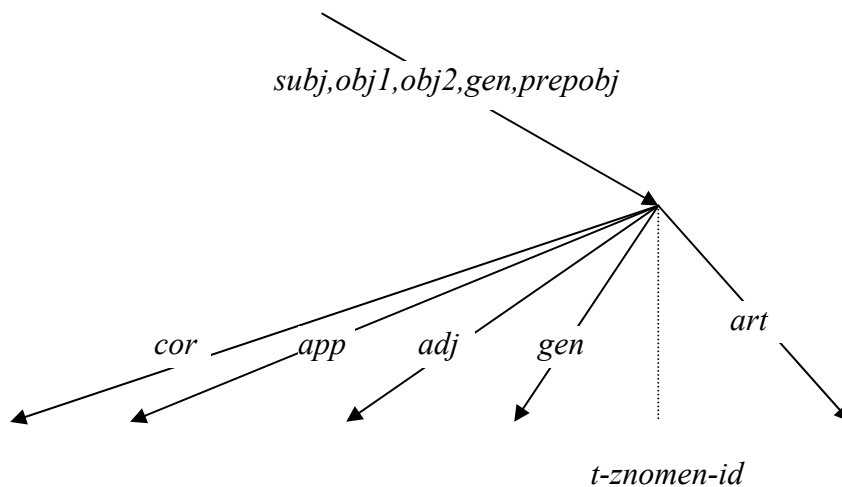
Ein Nomen kann *subj*, *obj1*, *obj2*, *gen*, oder *prepobj* werden. Ich unterscheide zwischen zusammengesetzten Nomina und Namen. Beide unterscheiden sich in ihren ausgehenden Kanten.

1.10 Zusammengesetzte Nomina

Zusammengesetzte Nomina erben vom Typ *t-nomen-id*, und ergänzen diesen um eine Reihe von optionalen Dependents: Korrelationen, Appositionen, Adjektive, Genitive und Artikel:

$$t\text{-znamen-id} = t\text{-nomen-id} \& [\text{valency}_{id} : \{cor^*, app^*, adj^*, gen^?, art^?\}]$$

Die unten stehende Grafik zeigt die möglichen eingehenden bzw. ausgehenden Kanten:



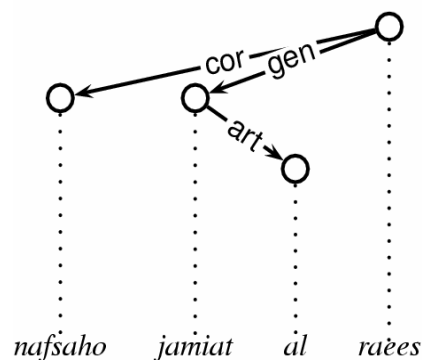
Beispiel 1.10.1:

In diesem Beispiel wird das Nomen „raees“ (Präsident) vom Genitiv „al jamiat“ (der Universität) und von der Korrelation „nafsaho“ (selbst) modifiziert:

nafsaho
نفسه
selbst

jamiat al
الجامعة
Universität der

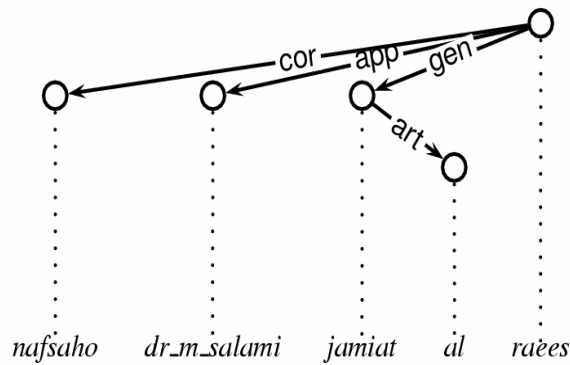
raees
رئيس
Präsident



Beispiel 1.10.2:

Im folgenden Beispiel wird „raees“ zusätzlich von der Apposition „dr-m-salami“ modifiziert:

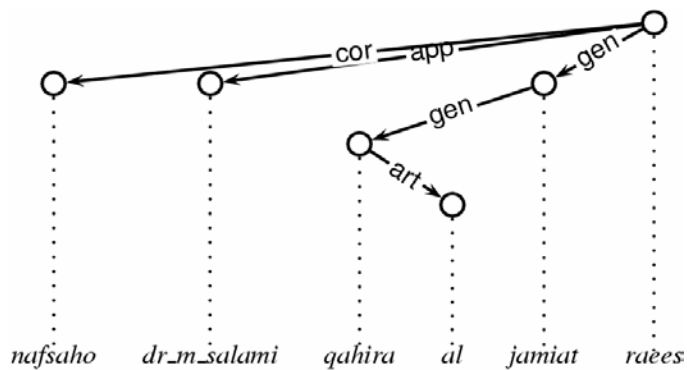
nafsaho	dr-m-salami	jamiat al	raees
نفسه	د.م. سلامه	الجامعة	رئيس
selbst	Dr. M. Salami	Universität der	Präsident



Beispiel 1.10.3:

In diesem Beispiel kommt der eingebettete Genitiv „al qahira“ (von Kairo) hinzu:

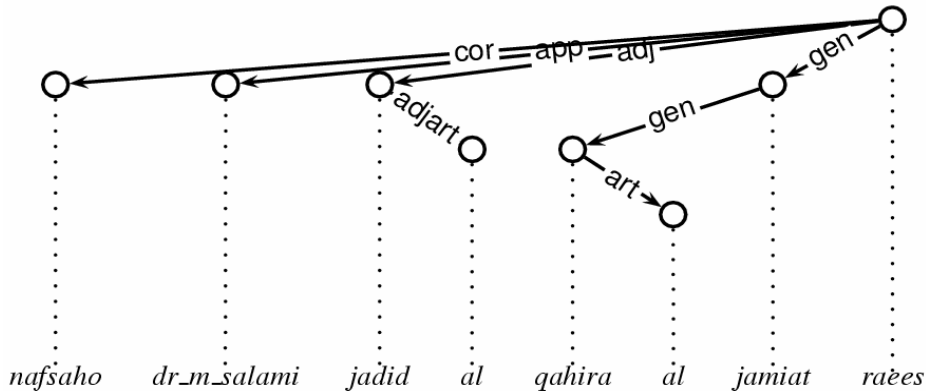
nafsaho	dr-m-salami	qahira al	jamiat	raees
نفسه	د.م. سلامه	القاهرة	جامعة	رئيس
selbst	Dr. M. Salami	Kairo von	Universität	Präsident



Beispiel 1.10.4:

Und im folgenden Beispiel kommt noch das Adjektiv (samt Adjektivartikel) „al jadid“ (der neue) dazu:

nafsaho	dr-m-salami	jadid al	qahira al	jamiat	raees
نفسه	د.م.سلامه	الجدید	القاهرة	جامعة	رئيس
selbst	Dr. M. Salami	neue	Kairo von	Universität	Präsident



1.11 Namen

Namen erben von folgendem lexikalischen Typ:

$$t\text{-namen-id} = t\text{-nomen-id} \& [\text{valency}_{id} : \{cor^*, app^*, adj^*, gen^*\}]$$

Das Ergebnis ist derselbe Typ wie *t-znomen-id*, ohne Artikel (Namen können im Arabischen keinen Artikel haben).

2. LP-Teil

Wie im ID-Teil definiere ich zunächst die Menge der Kantenlabels. In diesem Fall sind dies nicht syntaktische Funktionen, sondern topologische Felder:

Top. Feld	Steht für
<i>a</i>	Adjektiv-Kopf
<i>adjartf</i>	Adjektivartikel
<i>adjf</i>	Adjektiv-Dependent
<i>advf</i>	Adverb
<i>appf</i>	Apposition
<i>artf</i>	Artikel
<i>corf</i>	Korrelation
<i>genf</i>	Genitiv
<i>m</i>	Mafool Motlak-Kopf
<i>mmf</i>	Mafool Motlak-Dependent
<i>n</i>	Nomen-Kopf
<i>p</i>	Präposition-Kopf
<i>prepobjf</i>	Präpositionsobjekt
<i>sf</i>	Satzfeld

In dieser Menge gibt es drei Gruppen von Labels:

- Labels, die genau den ID-Kantenlabels entsprechen (*adjartf*, *adjf*, *advf*, *appf*, *artf*, *corf*, *genf*, *mmf*, *prepobjf*)
- Labels, die nur Knotenlabels sein können und den jeweiligen topologischen Kopf respektive seiner topologischen Dependents ordnen (*a*, *m*, *n*, *p*)
- das Label *sf*, das „Satzfeld“, in dem alle verbalen Dependents landen können, und die dort permutiert werden können. *Sf* kann auch Knotenlabel sein, damit Verben frei angeordnet werden können.

Die Menge der topologischen Felder ist folgendermaßen total geordnet:

$$\begin{aligned} \text{corf} < \text{appf} < \text{adjf} < \text{genf} < n < \text{artf} < \\ & a < \text{adjartf} < \\ & \text{prepobjf} < p < \\ & \text{mmf} < \text{sf} < m \end{aligned}$$

Die vier Zeilen partitionieren die Ordnung grob in vier Teile: der oberste für Nominalphrasen, der zweite für Adjektivartikel, der dritte für Präpositionen, und der vierte für die Ordnung auf Satzebene:

- a) Innerhalb von Nominalphrasen müssen also Artikel vor dem nominalen Kopf selbst, dieser vor den Genitiven, diese vor den Adjektiven, diese vor den Appositionen und diese vor den Korrelationen stehen.
- b) Der Adjektiv-Kopf folgt seinem Adjektivartikel.
- c) Der präpositionale Kopf muss vor seinem präpositionalen Objekt stehen.
- d) Auf Satzebene folgen die Mafool Motlak-Formen allem anderen, und stehen ihrerseits wieder vor eingebetteten Mafool Motlaks.

Im folgenden gehe ich auf die lexikalischen Typen auf der LP-Dimension ein.

2.1 Verben

Dies ist der lexikalische Typ für finite Verben:

$$t - fin - lp = \left[\begin{array}{l} labels_N : \{sf\} \\ valency_{lp} : \{mmf^*, sf^*\} \end{array} \right]$$

Finite Verben erben also vom Typ $t-fin-lp$. Sie haben das Knotenlabel sf , sie bieten die Felder mmf für Mafool Motlaks und sf für alle anderen Dependenden an.

2.2 Artikel

Artikel erben vom folgenden lexikalischen Typ, der erneut, wie auf der ID-Dimension, zwischen normalen Artikeln (eingehende Kante $artf$) und Adjektivartikeln ($adjartf$) unterscheidet:

$$t - art - lp = \left[\begin{array}{l} labels_{lp} : \{artf, adjartf\} \\ labels_N : \{artf\} \\ valency_{lp} : \{\} \end{array} \right]$$

Da Artikel keine weiteren Dependenden haben, die sie anordnen müssten, ist das Knotenlabel nicht relevant (hier habe ich $artf$ verwendet).

2.3 Das Adjektive

Adjektive erben vom folgenden lexikalischen Typ:

$$t - adj - lp = \begin{bmatrix} labels_{lp} : \{adjf\} \\ labels_N : \{a\} \\ valency_{lp} : \{adjartf\} \\ blocks : \{adjart\} \end{bmatrix}$$

D.h., sie haben eingehendes Kantenlabel *adjf*, Knotenlabel *a*, und verlangen einen Adjektivartikel *adjartf*. Das *blocks*-Merkmal sichert, dass die vom Adjektiv abhängigen Adjektivartikel nicht über das Adjektiv hinaus nach oben klettern können.

Durch die obige totale Ordnung steht der Artikel hinter dem Adjektivartikel.

2.4 Appositionen

Appositionen erben vom folgenden lexikalischen Typen:

$$t - app - lp = \begin{bmatrix} labels_{lp} : \{appf\} \\ labels_N : \{appf\} \\ valency_{lp} : \{\} \end{bmatrix}$$

D.h., sie haben eine eingehende Kante *appf*. Ihr Knotenlabel ist erneut uninteressant, weil Appositionen nie Dependents haben. Weil ich aber irgendein Knotenlabel wählen muss, habe ich mich für *appf* entschieden.

2.5 Korrelationen

Korrelationen erben vom folgenden lexikalischen Typ:

$$t - cor - lp = \begin{bmatrix} labels_{lp} : \{corf\} \\ labels_N : \{corf\} \\ valency_{lp} : \{\} \end{bmatrix}$$

Hier ist das eingehende Kantenlabel *corf*, und ich entscheide mich beim bedeutungslosen Knotenlabel auch für *corf*.

2.6 Adverbien

Adverbien erben vom folgenden lexikalischen Typ:

$$t - adv - lp = \begin{bmatrix} labels_{lp} : \{advf\} \\ labels_N : \{advf\} \\ valency_{lp} : \{\} \end{bmatrix}$$

Dieser hat eingehendes Kantenlabel und Knotenlabel *advf*.

2.7 Mafool Motlak

Mafool Motlaks erben vom folgenden lexikalischen Typ:

$$t - mm - lp = \begin{bmatrix} labels_{lp} : \{mmf\} \\ labels_N : \{m\} \\ valency_{lp} : \{sf^*\} \\ blocks : \{adv, pp\} \end{bmatrix}$$

Mafool Motlaks können als degenerierte eingebettete Verben betrachtet werden: Sie können ihrerseits Adverbien und präpositionale Modifikatoren haben, für die sie das Feld *sf* anbieten. Sie selbst haben das eingehende Kantenlabel *mmf*, und das Knotenlabel *m*. Durch die blocks-Spezifikation verhindere ich, dass eventuelle Modifikatoren der Mafool Motlaks nicht in das Satzfeld des übergeordneten Satzes klettern können.

Durch die obige totale Ordnung stehen Mafool Motlaks selbst vor ihren Dependents.

2.8 Präpositionen

Die Präpositionen erben vom folgenden lexikalischen Typ:

$$t - prep - lp = t - pp - lp = \left[\begin{array}{l} labels_{lp} : \{sf\} \\ labels_N : \{p\} \\ valency_{lp} : \{prepobjf\} \end{array} \right]$$

Auf der LP-Ebene unterscheide ich zunächst nicht mehr zwischen präpositionalen Komplementen und Modifikatoren. Beide haben eingehendes Kantenlabel *sf*, Knotenlabel *p*, und verlangen ein präpositionales Objekt im *prepobjf*.

2.9 Nomina

Nomina haben Knotenlabel *n*. Ihre eingehende Kante ist entweder:

1. *sf*, falls ihre eingehende Kante auf der ID-Dimension *subj*, *obj1* oder *obj2* ist,
2. *genf*, falls die eingehende ID-Kante *gen* ist, und
3. *prepobjf*, falls die eingehende ID-Kante *prepobj* ist.

Ich verwende Disjunktion (hier geschrieben als Choice |) aus der TDG-Lexikonsprache, um die obigen Bedingungen ausdrücken zu können (hier kombiniere ich ID- und LP-Beschreibungen):

$$t - nomen = \left[\begin{array}{l} labels_{id} : \{subj, obj1, obj2\} \\ labels_{lp} : \{sf\} \end{array} \right] \parallel \left[\begin{array}{l} labels_{id} : \{gen\} \\ labels_{lp} : \{genf\} \end{array} \right] \parallel \left[\begin{array}{l} labels_{id} : \{prepobj\} \\ labels_{lp} : \{prepobjf\} \end{array} \right]$$

2.10 Zusammengesetzte Nomina

Zusammengesetzte Nomina erben von der folgenden lexikalischen Klasse:

$$t - znomen - lp = t - nomen \ \& \ \left[\begin{array}{l} labels_N : \{n\} \\ valency_{lp} : \{corf^*, appf^*, adjf^*, genf?, artf?\} \\ blocks : \{cor, app, adj, gen, art\} \end{array} \right]$$

D.h., sie eröffnen Felder für Korrelationen (*corf*), Appositionen (*appf*), Adjektive (*adjf*), Genitive (*genf*) und Artikel (*artf*). Mithilfe des *blocks*-Features verhindere ich, dass die entsprechenden Dependents von der ID-Dimension über das Nomen hinausklettern.

2.11 Namen

Namen bieten die gleichen ausgehenden Kanten an wie zusammengesetzte Nomina in 2.10, außer dem Artikelfeld (da Namen keine Artikel haben dürfen). Analog blockieren sie auch alles außer Artikeln.

Abschnitt 3

Abschluss

In meinem Forschungspraktikum habe ich erstmals mithilfe von Topologischer Dependenzgrammatik das Arabische modelliert. Es hat sich gezeigt, dass TDG geeignet ist, um Arabisch elegant fassen zu können. Besonders der hohe Grad an morphologischer Markierung, der indirekt für die relativ freie Wortstellung im Arabischen verantwortlich ist, ist mittels TDG einfach modellierbar (durch die Agreement-Idee), außerdem ist die freie Wortstellung auf Satzebene sehr leicht modellierbar.

Ich hoffe damit einen kleinen Beitrag zur Weiterentwicklung dieses Dependenzgrammatik-Formalismus getan zu haben. Übrigens befindet sich die neueste Version meiner Grammatik fürs Arabische bereit zum Ausprobieren in der aktuellen Distribution des XDG Development Kit (XDK).

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

Mohammed Eid 1972: *Al-Naho al- Mosaffa*, Cairo University

Mahier s. Ahmed 1988: *An Export System for Understanding Arabic Sentences*, Cairo University

W. Wright 1967: *A Grammar of the Arabic Language*, Third Edition, Cambridge, England

XDG Development Kit (XDK):

<http://www.ps.uni-sb.de/~rade/mogul/publish/doc/debusmann-xdk/>

Papers:

<http://www.ps.uni-sb.de/~rade/papers.html>

Arabic:

<http://www.schoolarabia.net/arabic/>