

# **Topologische Abhängigkeitsgrammatik, eine Einführung**

Ralph Debusmann

C-Tag, 1. und 2.7. 2001

# Ausgangspunkt

Constraint-basierter Abhängigkeitsparser (Denys Duchier).

- Vorteile:
  - keine Probleme mit diskontinuierlichen Konstruktionen, damit geeignet für Sprachen mit freierer Wortstellung (z.B. Deutsch)
  - effizient
- Nachteile:
  - Wortstellungs-Constraints ad-hoc, Untergenerierung
  - kein deklarativer Grammatikformalismus

# Zielsetzung der Diplomarbeit

Beide Nachteile beheben, dazu Entwicklung von:

- einer Theorie für Wortstellungs-Constraints. Grundidee: Reapes Wortstellungsdomänen (Word Order Domains)
- einem deklarativen Grammatikformalismus

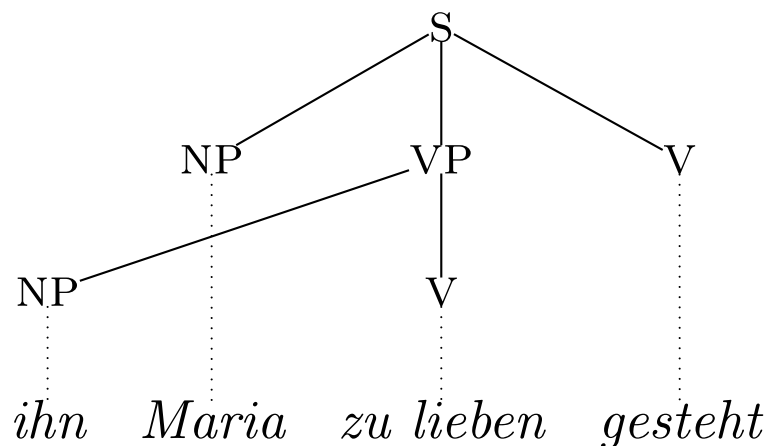
Herausgekommen ist *Topologische Dependenzgrammatik* (TDG).

# Wortstellungsdomänen: Motivation

Problem: diskontinuierliche Konstituenten, z.B. Scrambling (Ross 67) im Deutschen:

*(dass) ihn Maria zu lieben gesteht.*

PS-Analyse:



Kreuzende Kanten sorgen u.a. für schlechte Parse-Eigenschaften. Lösung in vielen Ansätzen: kreuzende Kanten nicht erlauben und tree hacking betreiben.

# Wortstellungsdomänen: Idee

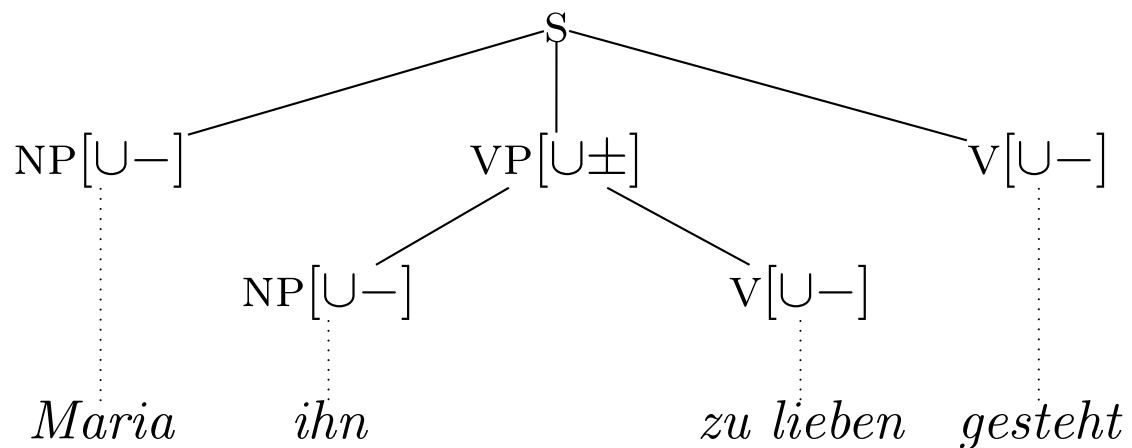
Reape (93, 94) postuliert zwei Analyse-Ebenen:

- Syntaxbaum
- Wortstellungsdomänen-Baum

Eigenschaften:

- Syntaxbaum: ungeordnet, kreuzende Kanten erlaubt
- Wortstellungsdomänen-Baum: flachere Version des Syntaxbaums, geordnet, keine kreuzenden Kanten

# Wortstellungsdomänen: Syntaxbaum

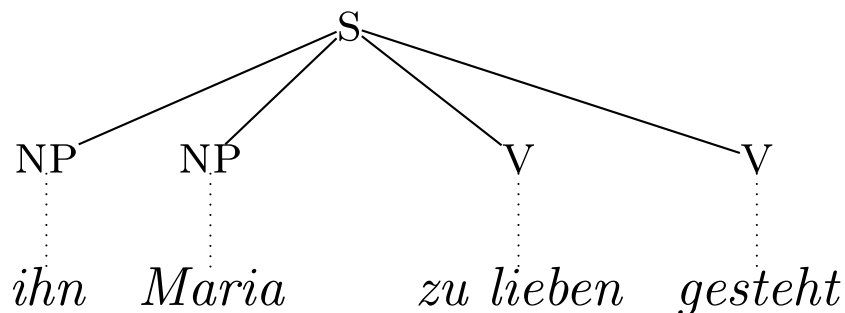
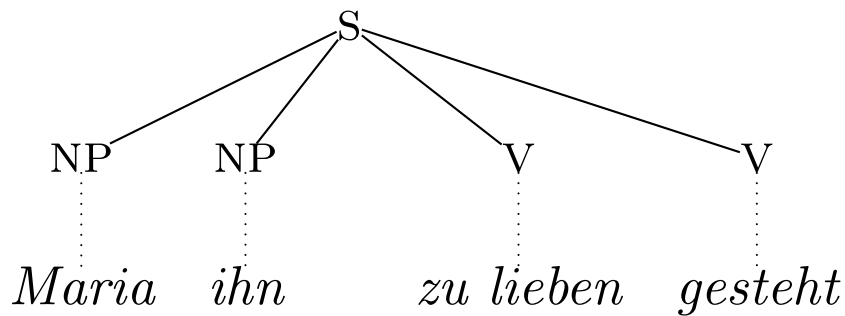


Jeder Knoten erhält  $[U\pm]$ -Wert zugeordnet:

- $[U-]$ : Knoten ist auch in entsprechendem Wortstellungsdomänen-Baum
- $[U+]$ : Knoten ist nicht in entsprechendem Wortstellungsdomänen-Baum

# Wortstellungsdomänen: VP in kanonischer Position

VPs in kanonischer Position (nicht extraponiert) sind [U+]. Ermöglicht Scrambling. Lizenzierte Wortstellungsdomänen-Bäume:

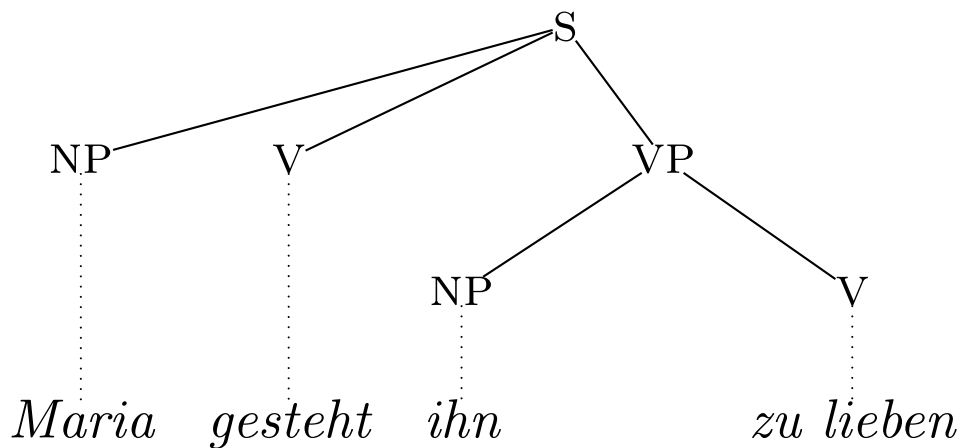


Zutreffende LP-Regeln:

$$\begin{array}{lcl}
 & \text{NP} & \prec \text{V} \\
 \text{v(Komplement)} & \prec & \text{v(Kopf)}
 \end{array}$$

# Wortstellungsdomänen: VP-Extrapolation

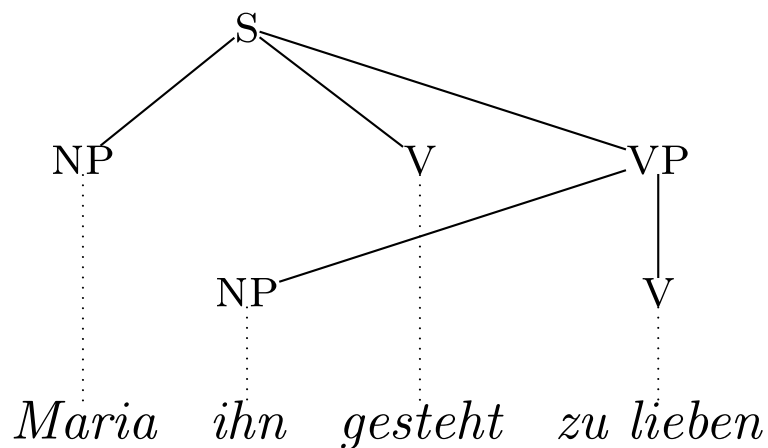
VPs in extraponierter Position sind [U-]. Entsprechender Wortstellungsdomänen-Baum:





# Wortstellungsdomänen: Partielle VP-Extrapolation

Partielle VP-Extrapolation: VP extrapoliert und damit [U-]. VP *ihn zu lieben* darf aber nicht von *gesteht* unterbrochen werden:



Relativsatz-Extrapolation kann Reapes Theorie ebenfalls nicht behandeln.

# Übergang zu TDG

Hauptunterschiede von TDG zu Reape:

- DG statt PSG als Basistheorie
- verwendet Ideen aus Theorie topologischen Felder (Herling 1821, Erdmann 1886, Höhle 1986): Baumflachung feinkörniger steuerbar, beherrscht partielle VP-Extrapolation und Relativsatz-Extrapolation.
- constraint-basierte Spezifikation, dadurch auch direkt implementierbar (siehe Demo)

# TDG: ID Baum und LP Baum

- Syntaktischer Dependenzbaum (ID Baum): ungeordnet, nicht-projektiv
- Topologischer Dependenzbaum (LP Baum): partiell geordnet, projektiv

Wohlgeformtheitsbedingungen:

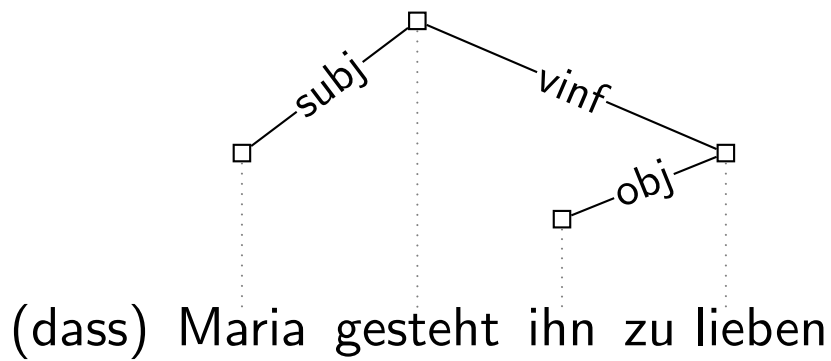
- Baumheit
- lexikalisierte Constraints: Valenz und akzeptierte Rollen bzw. Felder

Prinzipien zur Inbeziehungssetzung von ID und LP Baum:

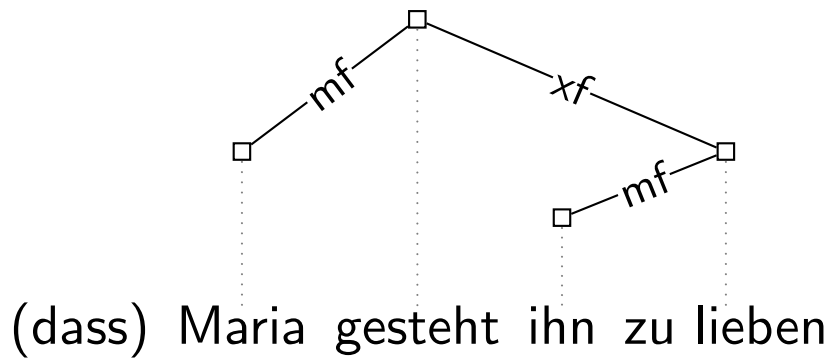
- Emanzipation
- Barrieren

# TDG: Beispiel

ID Baum:



LP Baum:



ID Baum mit syntaktischen Rollen, LP Baum mit topologischen Feldern beschriftet.

# TDG: Topologische Felder

In diesem Vortrag benutzen wir folgende topologische Felder:

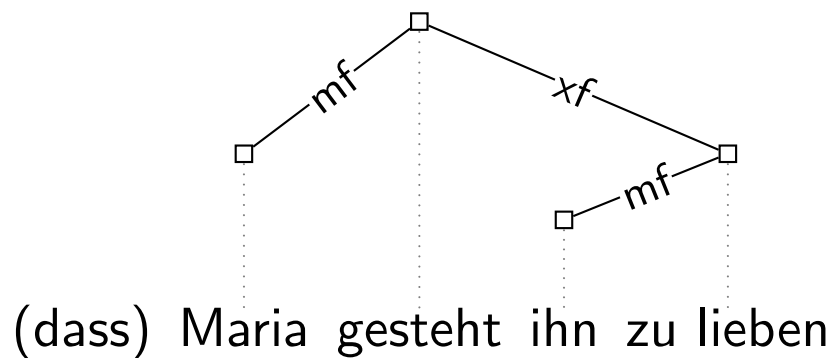
- mf: Mittelfeld. Platz für alle nicht-verbale Verbkomplemente.
- vc: verb canonical position. nicht-finite Verben in kanonischer Position (links von Kopf).
- xf: extraposition field. nicht-finite Verben in extraponierter Position (rechts von Kopf).

Menge der Felder total geordnet:

$$mf \prec vc \prec xf$$

# TDG: Ordnen der Knoten, externe Felder

Beispiel LP Baum:



Aufgrund von:

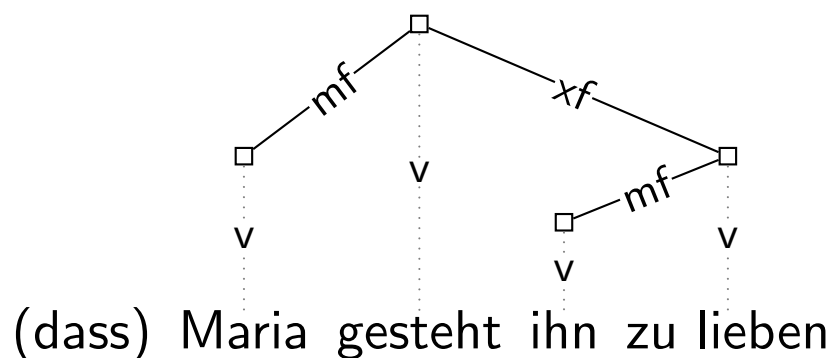
$$mf \prec xf$$

muss *Maria* im *mf* vor *ihn zu lieben* im *xf* stehen:

$$Maria \prec \text{ihn zu lieben}$$

# TDG: Ordnen der Knoten, interne Felder

Beispiel LP Baum:



*gesteht* soll zwischen *Maria* und *zu lieben* angeordnet werden. Einführung von internen Feldern. *gesteht* bekommt internes Feld (Knotenlabel) *v* zugeordnet. Dann folgt aus:

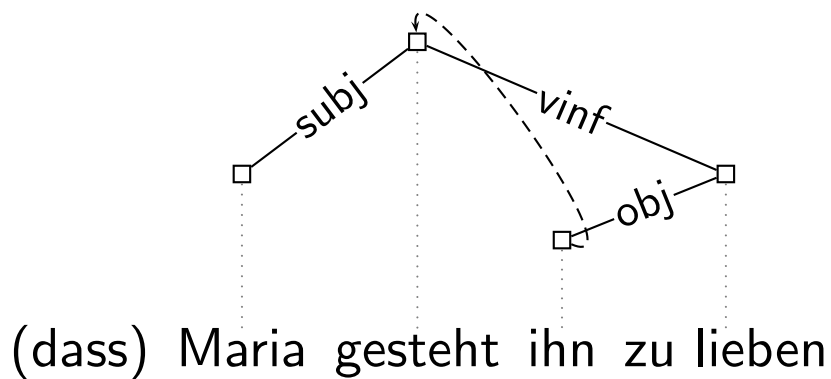
$$mf \prec v \prec xf$$

die folgende Abfolge:

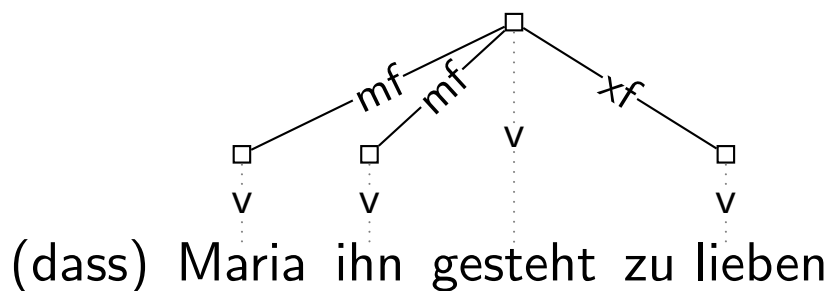
$$Maria \prec gesteht \prec ihm \text{ zu lieben}$$

# TDG: Emanzipation

ID und LP Bäume durch Emanzipation in Beziehung gesetzt. Knoten können sich emanzipieren bzw. hochklettern. Beispiel ID Baum:



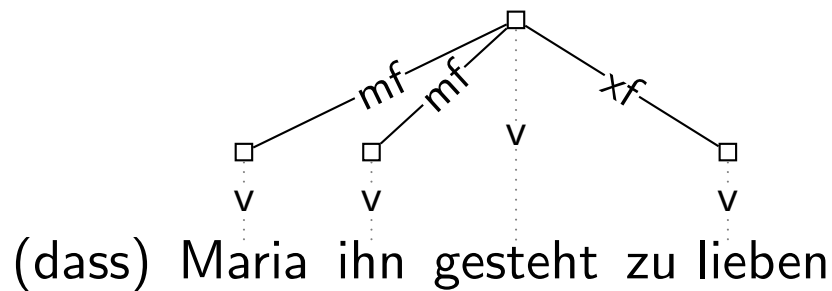
Durch Emanzipation flacherer LP Baum:





# TDG: Ordnen der Knoten, partielle Ordnung

Menge der externen und internen Felder total geordnet, daraus resultierende Ordnung aber partiell. Falls mehrere Knoten in demselben Feld landen, bleiben sie untereinander ungeordnet:



*Maria* und *ihn* beide im *mf* von *gesteht*. Damit untereinander ungeordnet, lizenzierte Abfolgen:

*(dass) Maria ihn gesteht zu lieben.*  
*(dass) ihn Maria gesteht zu lieben.*

# TDG: Valenz und akzeptierte Label

Lexikalisierte Constraints:

- Valenz: für alle Labels  $\ell$ , wie viele Kanten mit Label  $\ell$  angeboten?
- Akzeptierte Labels: welche Labels  $\ell$  darf eingehende Kante beschriften?

Anbieten und akzeptieren: Kante  $w-\ell\rightarrow w'$  von Knoten  $w$  zu Knoten  $w'$  mit Label  $\ell$  nur dann zugelassen, wenn sowohl:

- $\ell$  von  $w$  *angeboten*
- $\ell$  von  $w'$  *akzeptiert*

## TDG: Lexikoneinträge

*gesteht*

|                       |   |                 |
|-----------------------|---|-----------------|
| valency <sub>ID</sub> | : | {subj, vinf}    |
| accepts <sub>ID</sub> | : | {}              |
| valency <sub>LP</sub> | : | {mf*, vc?, xf?} |
| accepts <sub>LP</sub> | : | {}              |

*ihn, Maria*

|                       |   |             |
|-----------------------|---|-------------|
| valency <sub>ID</sub> | : | {}          |
| accepts <sub>ID</sub> | : | {subj, obj} |
| valency <sub>LP</sub> | : | {}          |
| accepts <sub>LP</sub> | : | {mf}        |

*zu lieben*

|                       |   |        |
|-----------------------|---|--------|
| valency <sub>ID</sub> | : | {obj}  |
| accepts <sub>ID</sub> | : | {vinf} |
| valency <sub>LP</sub> | : | {mf*}  |
| accepts <sub>LP</sub> | : | {xf}   |

## TDG: Kanonische und extraponierte Position

Lexikoneintrag für *zu lieben* auf voriger Folie zutreffend in extraponierter Position:

*zu lieben*

$$\left[ \begin{array}{l} \text{valency}_{\text{ID}} : \{\text{obj}\} \\ \text{accepts}_{\text{ID}} : \{\text{vinf}\} \\ \text{valency}_{\text{LP}} : \{\text{mf*}\} \\ \text{accepts}_{\text{LP}} : \{\text{xf}\} \end{array} \right]$$

In kanonischer Position bieten nicht-finite Verben kein Mittelfeld mf an:

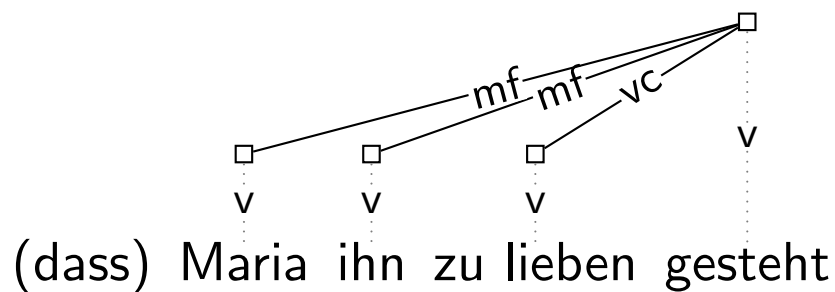
*zu lieben*

$$\left[ \begin{array}{l} \text{valency}_{\text{ID}} : \{\text{obj}\} \\ \text{accepts}_{\text{ID}} : \{\text{vinf}\} \\ \text{valency}_{\text{LP}} : \{\} \\ \text{accepts}_{\text{LP}} : \{\text{vc}\} \end{array} \right]$$

Objekt von *zu lieben* in kanonischer Position muss damit ins Mittelfeld des finiten Verbs klettern.

# TDG: Beispiel für Verb in kanonischer Position

LP Baum mit *zu lieben* in kanonischer Position:



Lizenzierte Abfolgen:

*(dass) Maria ihn zu lieben gesteht.*

*(dass) ihn Maria zu lieben gesteht.*

# TDG: Behandelte Phänomene

U.a. (bisher nur fürs Deutsche):

- Satztypen: Verb-erst, -zweit und -letzt
- Scrambling
- Verb-cluster: u.a. auxiliary flip, V-projection raising, Zwischenstellung (ACL-Papier)
- Relativsätze, inklusive Rattenfänger-Konstruktion (pied piping):

*ein Mann, den zu lieben Maria gesteht.*

# TDG: Parser

Im Rahmen der Diplomarbeit entstanden: constraint-basierter Parser für TDG-Grammatiken mit folgenden Eigenschaften:

- erlaubt Schreiben von TDG-Grammatiken ohne Oz-Kenntnisse (unter Benutzung eines Grammatikformalismus für TDG)
- bietet Funktion zum Generieren aller lizenzierten Abfolgen einer Menge von Wörtern
- Plugin-System für zusätzlich benötigte Constraints
- bereits eingesetzt in Softwareprojekt an der Computerlinguistik (englische Grammatik, Vorstufe zu einfacher Semantikkonstruktion)

# TDG: Ausblick

Es gibt noch viel zu tun, u.a.:

- Verbesserung der Abdeckung fürs Deutsche
- Anwendung auf andere Sprachen
- Aufbau von größeren Grammatiken auf Korpora-Basis (Christian Korthals)
- Anbindung von statistischen Suchheuristiken
- Spezifikation einer Syntax-Semantik-Schnittstelle
- Parser: Optimierung und umfangreichere Dokumentation