

Inkrementelle Syntax

Synchroneous Tree Unification Grammar

Timm Lichte

HHU Düsseldorf, Germany

29.01.2014



Das ideale Grammatikmodell

- linguistisch adäquat:
 - **Phänomene:**
Linearisierung, Kongruenz, Diskontinuität, Ellipse, Koordination (RNR), ...
 - **Generalisierungen/Zusammenhänge:**
Valenzrahmen, Aktiv/Passiv-Diathese, Satztypen, Stellungsalternation, Syntax-Semantik-Schnittstelle, Kontext-/Diskursabhängigkeit
 - intuitive Implementierung
- computerlinguistisch adäquat:
 - explizit/formalisiert
 - entscheidbar, vielleicht sogar mit vertretbarem Aufwand
 - bidirektional
- psycholinguistisch adäquat
 - strikt inkrementelle Derivationen
 - richtige Vorhersagen bzgl. Verarbeitungskomplexität
 - Vorhersage & Verifikation (prediction & verification)
 - PiG

(strenge) Evaluation

- **ATN:** zu prozedural, zu mächtig (Jumps + nichtmonotone Registeroperationen)
- **LAG:** zu intransparent? zwingend linksassoziiert, keine Konstituentenstruktur
- **Dynamic Syntax:** formalisiert? zu intransparent/ zu wenig modularisiert?
- **IM:** nicht formalisiert; zwingend rechtsassoziiert
- **PLTAG:** zu modularisiert (LTAG + inkrementelle Erweiterung)

- STUG (Lichte, 2012):
Motivation: Verhältnis von Syntax und Valenz,
Diskontinuität, Ellipse, Grammatikfaktorisierung
Modelltyp: automatengestützte Baumunifikation

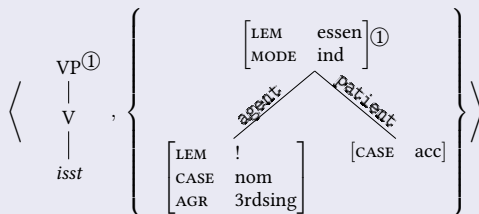
Ablauf:

- einführendes Beispiel
- Motivation:
 - Ellipsen
 - Grammatikfaktorisierung
 - Inkrementalität
 - Kohärenz
- Ausdrucksstärke
- Einordnung

Einführendes Beispiel

Synchrone Baumunifikationsgrammatik (STUG)

- Eine STUG besteht aus Syntax-Valenz-Paaren:



- Syntaxbaum und Valenzstruktur sind getrennt, aber verlinkt!
- Symmetrische Kombinationsoperation: Baumunifikation
- Linearisierungsbeschränkungen durch endlichen Automaten über syntaktischen Kategorien

Einführendes Beispiel

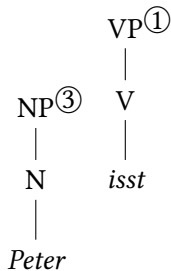
(1) *Peter isst einen Apfel.*

NP^③
|
N
|
Peter

{ [LEM Peter
CASE acc|nom
AGR 3rdsing]^③ }

Führendes Beispiel

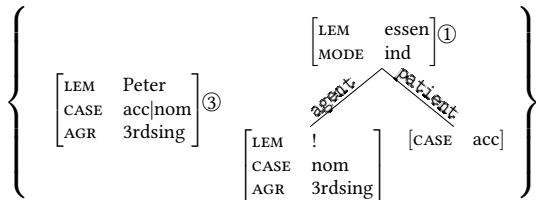
(1) *Peter isst einen Apfel.*



VP^①

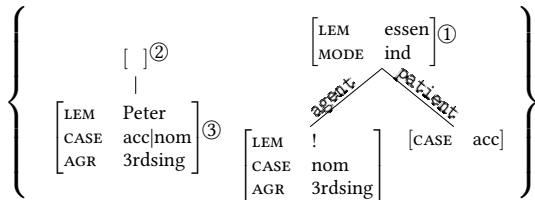
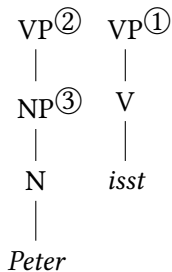
V

isst



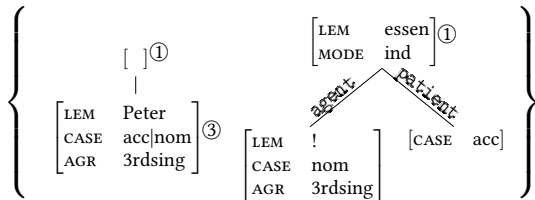
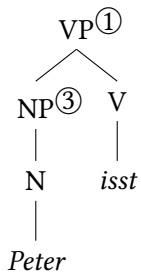
Führendes Beispiel

(1) *Peter isst einen Apfel.*



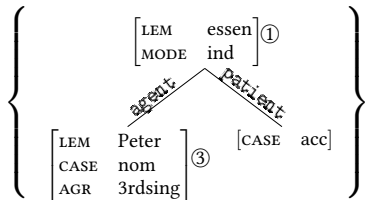
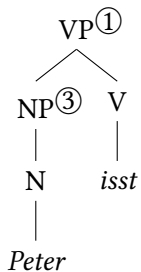
Einführendes Beispiel

(1) *Peter isst einen Apfel.*



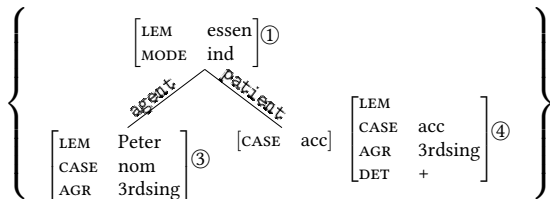
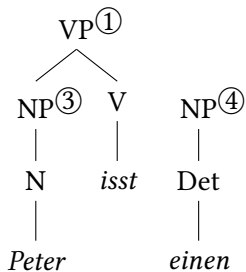
Einführendes Beispiel

(1) *Peter isst einen Apfel.*



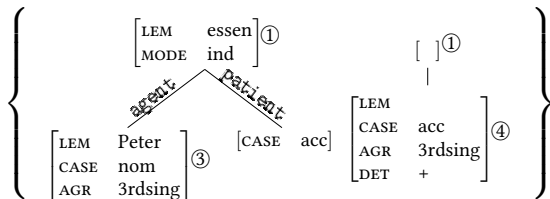
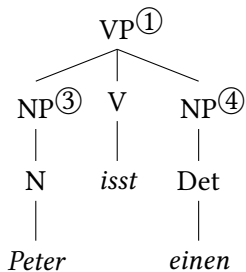
Einführendes Beispiel

(1) *Peter isst einen Apfel.*



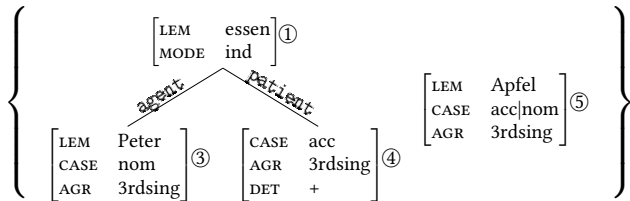
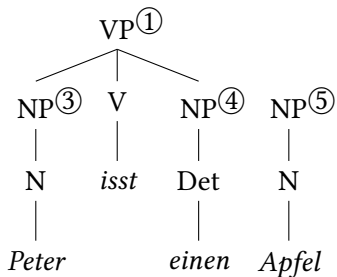
Einführendes Beispiel

(1) *Peter isst einen Apfel.*



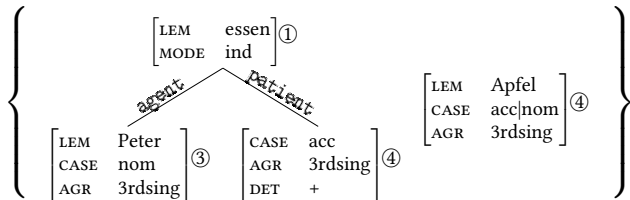
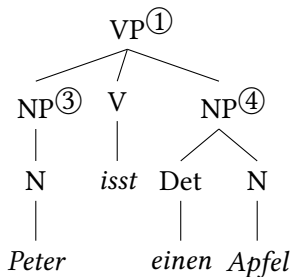
Einführendes Beispiel

(1) *Peter isst einen Apfel.*



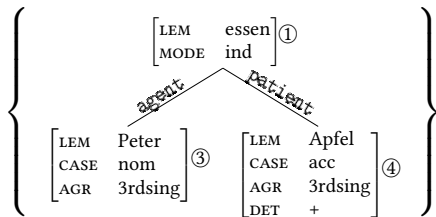
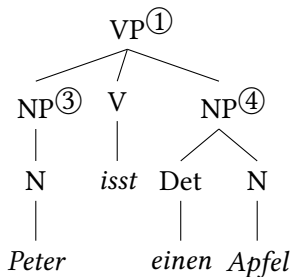
Einführendes Beispiel

(1) *Peter isst einen Apfel.*



Einführendes Beispiel

(1) *Peter isst einen Apfel.*



Warum noch ein weiterer Grammatikformalismus?

⇒ Er ist anders (vgl. LTAG):

- Trennung von Syntax und Valenz (ATN?)
- flache syntaktische Strukturen
- Ableitungsbäume uninteressant

⇒ attraktive Eigenschaften:

- Ellipse ohne zusätzliche Regeln
- substantielle Verringerung der Grammatikgröße (vgl. LTAG)
- inkrementelles Parsen ohne Hilfsstrukturen (vgl. PLTAG)

Motivation: Ellipse

- (2) a. *Mary always laughs, and John sometimes.*
b. Q: *Who doesn't laugh?* A: *John sometimes.*
c. [Sue is pointing to knitting things.] Mary: *John sometimes.*

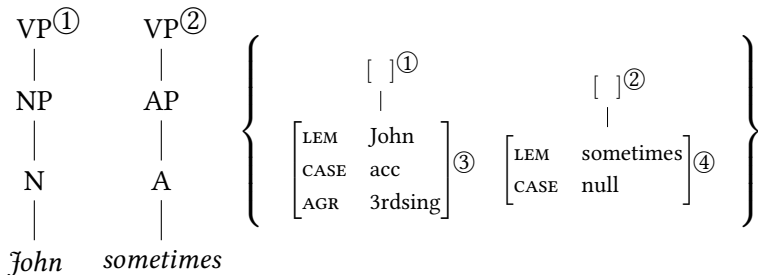
übliche Strategien:

- Verschmelzung (vollständig, aber Überschneidung)
- Tilgung (vollständig, aber phonetisch reduziert)
- Füllung (vollständig, aber mit leeren Worten)
- spezielle Phrase/Lexikoneinträge (unvollständig, weil defekt)

STUG-Strategie: **keine** speziellen Regeln, **direkte** Modellierung

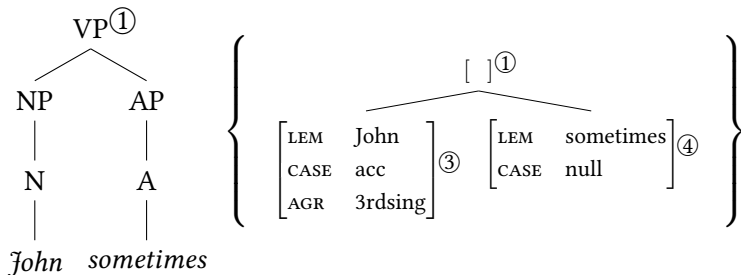
Motivation: Ellipse

(3) [Sue is pointing to knitting things.] Mary: *John sometimes*.



Motivation: Ellipse

(3) [Sue is pointing to knitting things.] Mary: *John sometimes.*



Motivation: Grammatikfaktorisierung

LTAG-Motto

- “complicate locally, simplify globally” (Joshi, 2004)
- eine Baumschablone pro Konstruktion
(= Valenzrahmen \times Linearisierung \times Aktiv/Passiv-Diathese)

Peter aß den Apfel.

Den Apfel aß Peter.

Gestern aß Peter den Apfel.

Gestern aß den Apfel Peter.

Peter aß.

Gestern aß Peter.

dass Peter aß

dass Peter den Apfel aß

dass den Apfel Peter aß

Der Apfel wurde gegessen.

Der Apfel wurde von Peter gegessen

...

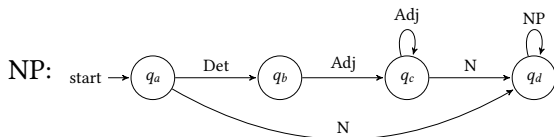
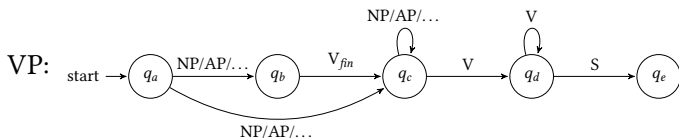
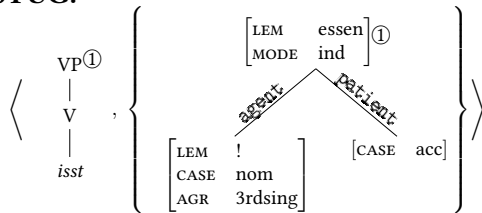
optionale Argumente
Linearisierungsalternativen } \rightsquigarrow viele Baumschablonen

LTAG:

- XTAG enthält 1008 verbale Baumschablonen (Prolo, 2002).
- XTAG benutzt 50 syntaktische Merkmale, um Linearisierung, Kongruenz und Valenz zu beschränken.
 - ausgetüftelte Metagrammatik-Systeme (metarules, XMG)
 - bestimmte Generalisierungen in LTAG nicht ausdrückbar
- lexikalische Ambiguität: *laughs* mit optionalem PP-Argument kann wenigstens 10 Baumschablonen lexikalisieren.
 - ausgeklügelte Disambiguierungsmethoden zur Beschleunigung des Parsers (Gardent et al., 2011)

Motivation: Grammatikfaktorisierung

STUG:



Motivation: Direkte Inkrementalität

PLTAG = LTAG +

- prediction trees
- Unifikation/Verifikation

STUG:

- unlexikalisierte Verbindungsbäume (~ prediction trees)
 - Vorhersage durch Verbindungsbäume und Valenzbaum
 - nur Unifikation
- ⇒ keine speziellen Komponenten für Inkrementalität

Weiteres Beispiel: Kohärente Konstruktionen

(4) (dass) *ihn Peter zu reparieren versuchte*

NP^③

|

N

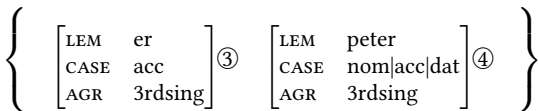
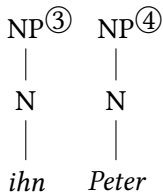
|

ihn

{ [LEM er
CASE acc
AGR 3rdsing]^③ }

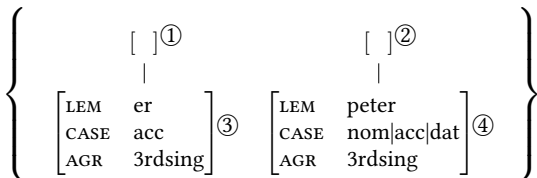
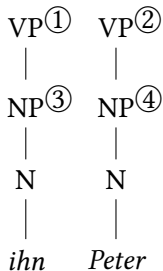
Weiteres Beispiel: Kohärente Konstruktionen

(4) (dass) *ihn Peter zu reparieren versuchte*



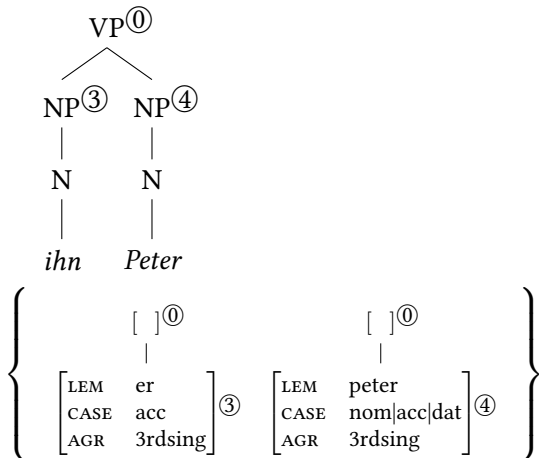
Weiteres Beispiel: Kohärente Konstruktionen

(4) (dass) *ihn Peter zu reparieren versuchte*



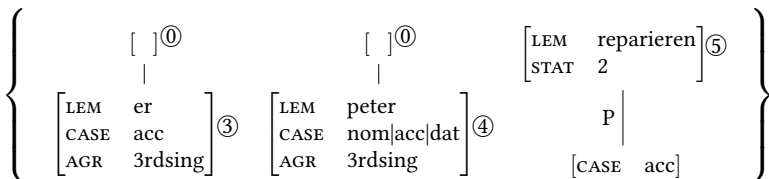
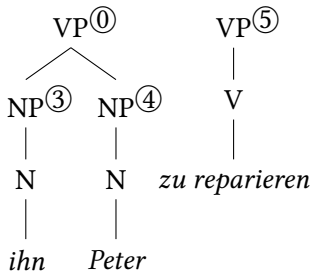
Weiteres Beispiel: Kohärente Konstruktionen

(4) (dass) *ihn Peter zu reparieren versuchte*



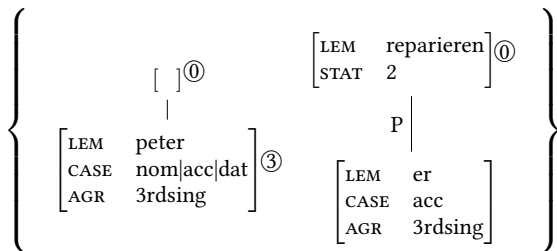
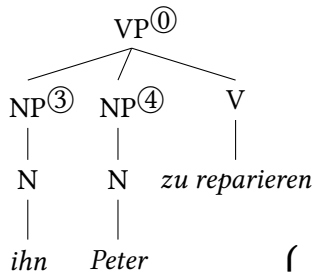
Weiteres Beispiel: Kohärente Konstruktionen

(4) (dass) *ihn Peter zu reparieren versuchte*



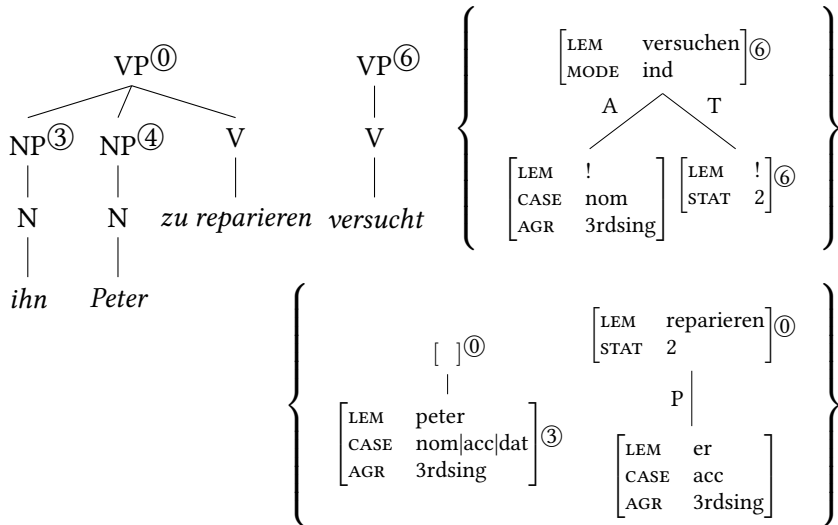
Weiteres Beispiel: Kohärente Konstruktionen

(4) (dass) *ihn Peter zu reparieren versuchte*



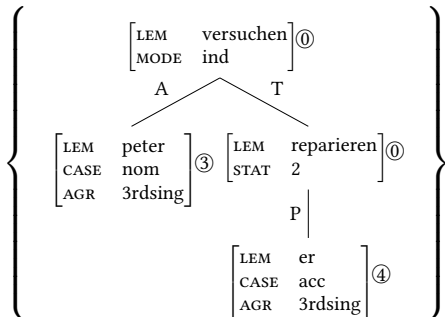
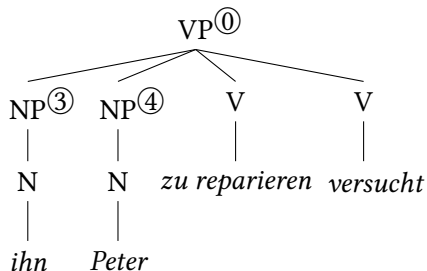
Weiteres Beispiel: Kohärente Konstruktionen

(4) (dass) *ihn Peter zu reparieren versuchte*



Weiteres Beispiel: Kohärente Konstruktionen

(4) (dass) *ihn Peter zu reparieren versuchte*



mit STUG generierbar:

- nicht-wohleingebettete Abhängigkeiten
- Abzählsprache $a^n b^n$ mit eingebetteten Abhängigkeiten
- MIX-Sprache
- Abzählsprachen $a_1^n \dots a_k^n$

mit STUG nicht generierbar (Vermutung):

- Abzählsprache $a^n b^n$ mit kreuzenden Abhängigkeiten (bzw. Kopiersprache)
- exponentielle Abzählsprache a^{2^n}
- $SCR^{ind} = \{ \sigma(NP_1, \dots, NP_m) V_m \dots V_1 \mid m \geq 1 \text{ and } \sigma \text{ is a permutation} \}$

⇒ STUG scheint mehr **und gleichzeitig** weniger Ausdrucksstärke als LTAG zu besitzen! (vgl. LAG)

STUG \approx ATN mit Bäumen und (monotoner) Unifikation ...

- linguistisch adäquat?
 - ⇒ Linearisierung, Kongruenz, Ellipse, Koordination?, MWE?, weite Abhängigkeiten?
 - ⇒ Linearisierungsalternation, Valenzrahmen, Aktiv/Passiv, Diskursabhängigkeit?
 - ⇒ intuitiv?
- computerlinguistisch adäquat
 - ⇒ formalisiert, aber Mächtigkeit unbekannt
 - ⇒ möglicherweise bidirektional
- psycholinguistisch adäquat
 - ⇒ strikt inkrementell, Vorhersagen & Verifikation
 - ⇒ richtige Vorhersagen?

Es sind noch viele Fragen offen!

- Gardent, Claire, Yannick Parmentier, Guy Perrier & Sylvain Schmitz. 2011. Lexical Disambiguation in LTAG using Left Context. In *5th language and technology conference - ltc'11*, 395–399. Poznan, Poland.
<http://hal.archives-ouvertes.fr/hal-00629902/en/>.
- Joshi, Aravind K. 2004. Starting with complex primitives pays off: complicate locally, simplify globally. *Cognitive Science* 28(5). 637–668.
- Lichte, Timm. 2012. Synchronous tree unification grammar. In *Proceedings of the 11th international workshop on tree adjoining grammars and related formalisms (tag+11)*, 46–54. Paris, France.
<http://www.aclweb.org/anthology-new/W/W12/W12-4606>.
- Prolo, Carlos A. 2002. Generating the XTAG English grammar using metarules. In *Proceedings of the 19th international conference on computational linguistics (coling 2002)*, 814–820. Taipei. Taiwan.