# An LTAG Fragment for German

—

# Hands-on Experience with XMG

Timm Lichte, Wolfgang Maier
Universität Tübingen

June 22, 2007

1. state of affairs: theory and implementation
2. extending XMG for MCTAG
3. XMG in action: tree fragments, colors, semantic dimension

# State of affairs: theory and implementation

- **Theory:**
  Syntax: free word order (scrambling), coherent constructions
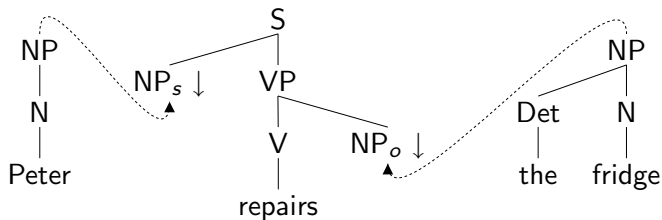  Semantics: underspecified flat semantics on the derivation tree

- **Implementation:**
  Preparatory work (extension of XMG)
  Implementation of some basic constructions

# TAG for English

Fixed complement order in English:

(1) Peter repairs the fridge. $\neq$ The fridge repairs Peter.
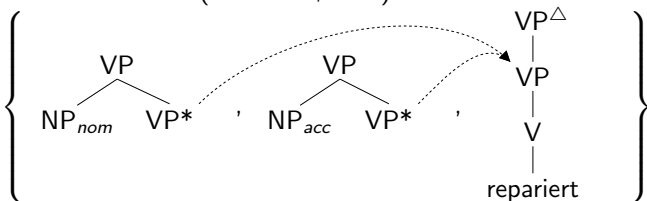
# German syntax is flexible

(1) dass Peter den Kühlschrank heute repariert
    dass den Kühlschrank Peter heute repariert
    ...
    ('that Peter repairs the fridge today')

⇒ linguistically sound analysis using TAG?
⇒ desire to abstract away from word order
⇒ Multi-Component TAG (MCTAG)

(2) dass den Kühlschrank Peter heute repariert
    ('that Peter repairs the fridge today')

- Vector MCTAG (Rambow,1994):



- MCTAG with shared nodes (Kallmeyer,2005)
- MCTAG with tuples (Lichte, 2007)

# An MCTAG for German

(2) dass den Kühlschrank Peter heute repariert
('that Peter repairs the fridge today')
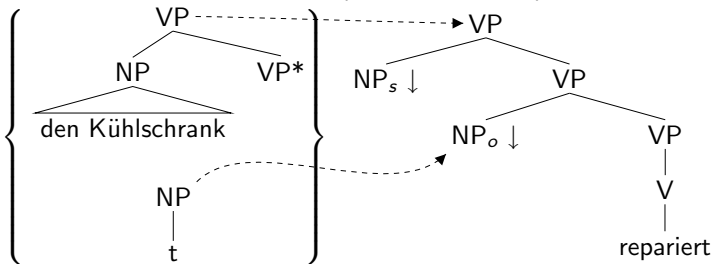
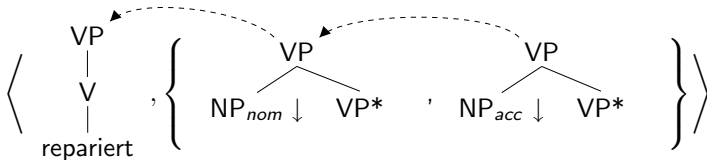- Vector MCTAG (Rambow,1994):
- MCTAG with shared nodes (Kallmeyer,2005)



- MCTAG with tuples (Lichte, 2007)

(2) dass den Kühlschrank Peter heute repariert
('that Peter repairs the fridge today')

- Vector MCTAG (Rambow,1994):
- MCTAG with shared nodes (Kallmeyer,2005)
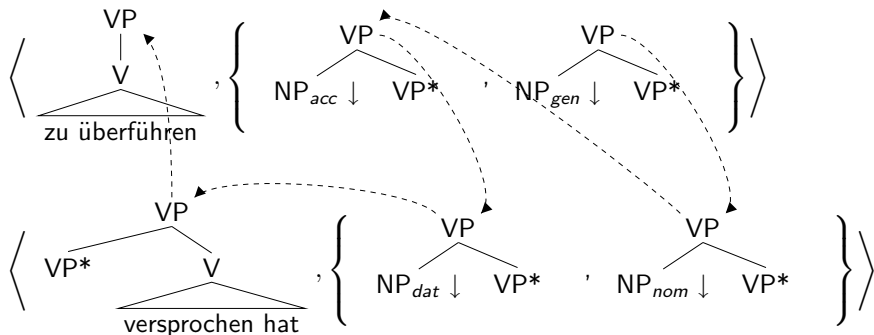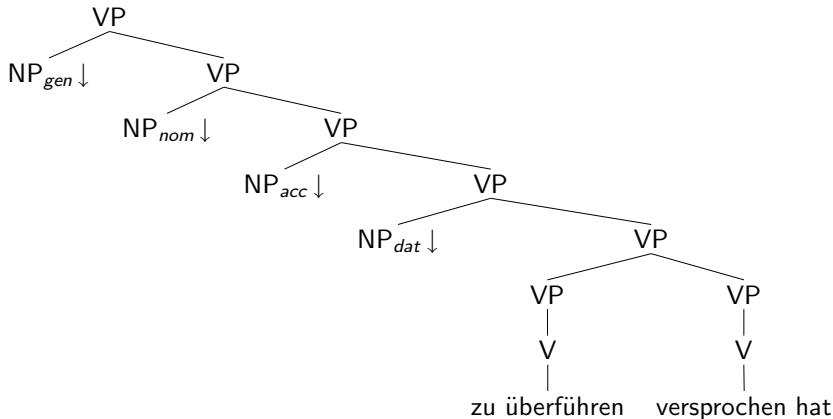- MCTAG with tuples (Lichte, 2007)

(3) dass des Verbrechens der Detektiv den Verdächtigen dem Klienten |
   zu überführen versprochen hat
   ('that the detective has promised the client to indict the suspect of
   the crime')

(3) dass des Verbrechens der Detektiv den Verdächtigen dem Klienten |
zu überführen versprochen hat
('that the detective has promised the client to indict the suspect of
the crime')

# The formalism of choice: TT-MCTAG

(4) dass des Verbrechens der Detektiv den Verdächtigen dem Klienten |
zu überführen versprochen hat
('that the detective has promised the client to indict the suspect of
the crime')

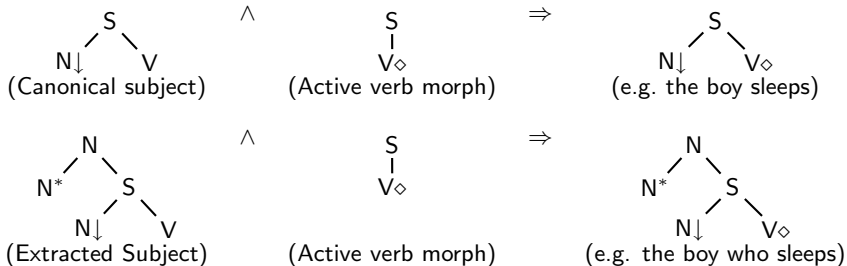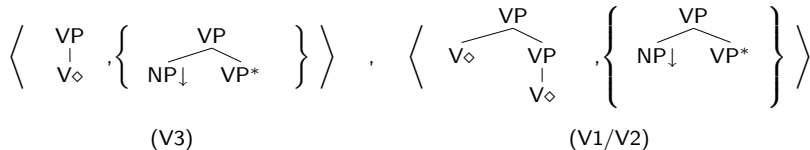- Until January 2007: XMG compiler only for standard TAG, no support for tree sets as present in MCTAG and its variants
- A Metagrammar looked like this:



(Canonical subject) ∧ (Active verb morph) ⇒ (e.g. the boy sleeps)



(Extracted Subject) ∧ (Active verb morph) ⇒ (e.g. the boy who sleeps)

- We use (TT-)MCTAG for our German grammar fragment
- XMG should allow for descriptions of sets/tuples, like

$$
\left\langle \quad \begin{array}{c} VP \\ | \\ V_\diamond \end{array} \quad , \left\{ \begin{array}{c} VP \\ \overline{NP_\downarrow \quad VP^*} \end{array} \right\} \right\rangle \quad , \quad \left\langle \begin{array}{c} VP \\ \overline{V_\diamond \quad \begin{array}{c} VP \\ | \\ V_\diamond \end{array}} \end{array} \quad , \left\{ \begin{array}{c} VP \\ \overline{NP_\downarrow \quad VP^*} \end{array} \right\} \right\rangle
$$

$$\qquad \qquad \text{(V3)} \qquad \qquad \qquad \qquad \qquad \text{(V1/V2)}$$

$\rightsquigarrow$ **Extension of the compiler!**
(Joint work with Yannick Parmentier)

# Explicit specification of sets

- To deal with sets, we would need to extend the XMG description language
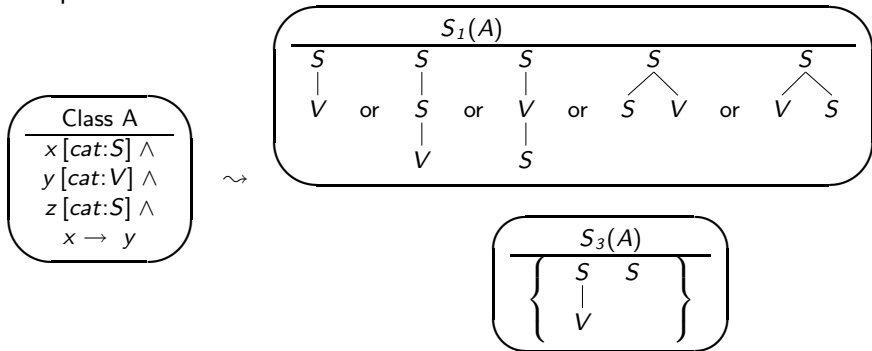- Core parts of the compiler would have to be touched

⤳ **lack of generality**

# Specification of tree sets

**Introduce a different solver module $S_3$**

- No root uniqueness constraint
- Two nodes of a model can be in different trees
- New keyword `setvalue`
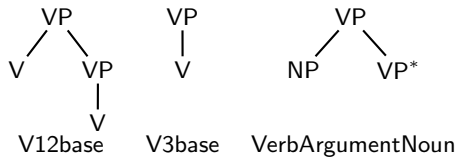
Example:

**Drawbacks (?)**

- Unmarked TT-MCTAG head tree
- (So far) no colors available

**Advantages**

- Different solvers ($S_1$, $S_3$) applyable on the same grammar via different axiom definitions
- Maintained flexibility: Other solvers can be integrated
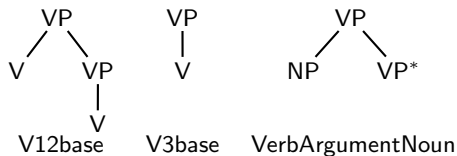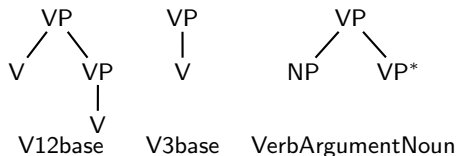- XMG for MCTAG has wider range of potential users

# An example

```
      VP            VP              VP
     /  \           |             /    \
    V    VP         V           NP      VP*
         |
         V
   V12base        V3base    VerbArgumentNoun
```

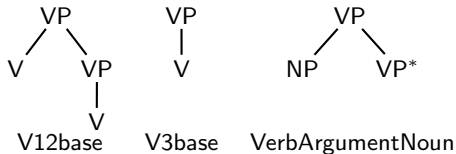- Class Vn1 (intransitive verbs) uses all of the above trees:

# An example



- Class `Vn1` (intransitive verbs) uses all of the above trees:
    - V12base: Vorfeld + Mittelfeld
    - V3base: Mittelfeld
    - VerbArgumentNoun

# An example



- Class `Vn1` (intransitive verbs) uses all of the above trees:
    - V12base: Vorfeld + Mittelfeld
    - V3base: Mittelfeld
    - VerbArgumentNoun
- Case feature of NP is exported

## An example



VP

V VP
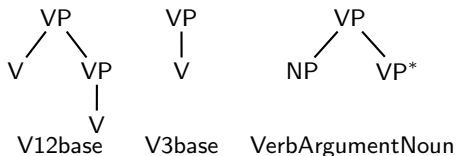
V
V12base

VP

V
V3base

VP

NP VP*
VerbArgumentNoun

- Class `Vn1` (intransitive verbs) uses all of the above trees:
  - V12base: Vorfeld + Mittelfeld
  - V3base: Mittelfeld
  - VerbArgumentNoun
- Case feature of NP is exported
- Call `setvalue Vn1`. Result:

# An example



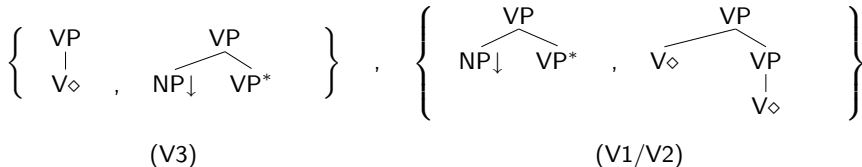V12base    V3base    VerbArgumentNoun

- Class `Vn1` (intransitive verbs) uses all of the above trees:
    - V12base: Vorfeld + Mittelfeld
    - V3base: Mittelfeld
    - VerbArgumentNoun
- Case feature of NP is exported
- Call `setvalue Vn1`. Result:



(V3)                          (V1/V2)

- The modified compiler is available in the XMG subversion repository (branch XMG-TUEBINGEN)

# The agenda

**Include more linguistic phenomena**

- Partial fronting phenomena
- Extraposed relative clauses
- Treatment of verbal complex
- Coordination and ellipsis
- Semantics

# The agenda

**Lexicon**

- More subcat frames
  - (Hand-written) list of all subcat frames present in the TüBa-D/Z treebank is available
- Larger Lexicon

**Parsing**

- restricting the formalism to keep/make it mildly context-sensitive
- transforming the MCTAG into a simple RCG for parsing