Parsing morphological causatives with XMG

Valeria Generalova
Simon Petitjean

Heinrich Heine University of Dusseldorf
TreeGraSP

June 25, 2021

Overview

1 Motivation

2 Data

3 Solution

- Overview
- Morph
- Lemma
- Frames
- Syntax

4 Conclusion

Motivation

- Many languages have rich morphology
- Usually, grammatical morphemes specify the value of one (several) features precisely
- The values of features for the whole word can be deduced from morphemes
- → There is a necessity account for morphology in order to facilitate parsing in such languages

Data

(1)	*Satu*	*tapa-tt-i*	*etana-n*	*Diane-lla.*
	Satu.NOM	kill-CAUS-PST.3SG	slug-ACC	Diane-ADE

‘Satu had Diane kill the slug.’ Manninen and Nelson 2004, p. 222, (20)

- Semantically and syntactically transitive base verb
- A causative morpheme attaches to the verb → derivation
- Three arguments are marked with different morphological cases → inflection
- Some morphemes can be realized as ∅

Solution overview

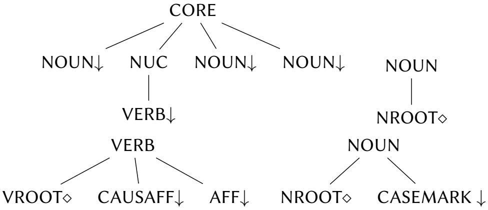
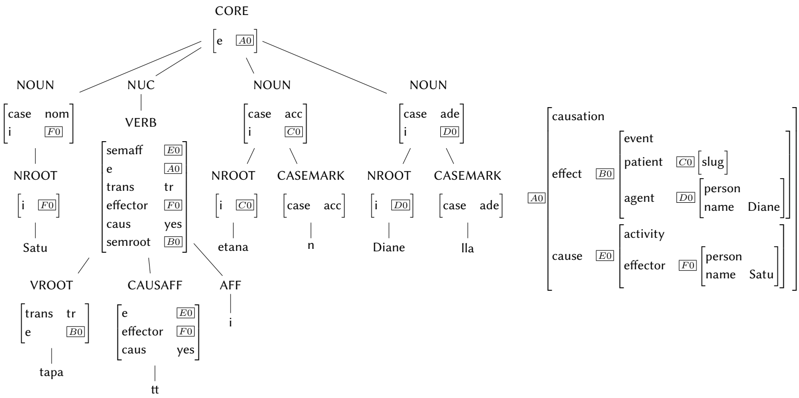


Figure 1: Main trees of the grammar

- Three classes producing lexicalized trees
- The root morpheme is marked as anchor
- Other morphemes are marked as substitution nodes
- The tree for the whole core is non-lexicalized

Solution output



The *morph* dimension

```
class VRoot_kill{
<morpho>{
  morph <- "tapa";
  lemma <- "tapa";
  trans <- tr;
  cat <- vroot
}}
```

Figure 2: Entry for the morpheme *tapa*

```
class CaseMarker_Ade{
<morpho>{
  {morph <- "lla"
  |
  morph <- "llä"};
  lemma <- "ADE";
  case <- ade;
  cat <- casemark
}}
```

Figure 3: Entry for the ADE marker

- Assigns category necessary for anchoring and substitution
- Relates allomorphs to a lemma
- Specifies the values of relevant features

The *lemma* dimension

```
class Lemma_VRoot_kill{
<lemma>{
  entry <- "tapa";
  cat <- vroot;
  fam <- Verb_Tapa-tt-i
}}
```

Figure 4: Lemma entry for the verb *tapa*

- Relates the lemma to the family, i.e., the range of trees it can be used in

Semantics with *frame* dimension

```
class AnchoringCausaff
declare ?eaff ?F0
{
 <syn>{
  node (mark = anchor)[cat = causaff,
           e = ?eaff, effector = ?F0]
 };
 <frame>{?eaff[activity,
           effector: ?F0]}
}
```

Figure 5: Syntactic and semantic descriptions for the causative affix

- Can share features with the syntactic dimension

Anchoring nouns in *syn* dimension

```
class AnchoringNoun
declare ?Noun ?NRoot ?CaseMarker ?VarCase ?I
{
 <syn>{
  node ?NRoot (mark=anchor) [cat = nroot, i = ?I];
  {
   node ?Noun [cat = noun, case = ?VarCase, i = ?I];
   node ?CaseMarker (mark=subst) [cat = casemark,
                                  case = ?VarCase];
   ?Noun -> ?NRoot; ?Noun -> ?CaseMarker;
   ?NRoot >> ?CaseMarker
  }
  |
  {
   node ?Noun [cat = noun, case = nom, i = ?I];
   ?Noun -> ?NRoot
  }
 };
 <iface>{ [i = ?I] }
}
```

Anchoring verbs in *syn* dimension

```
class Verb_Tapa-tt-i
declare ?Trans ?Caus ?NodeVerb ?NodeVRoot ?
    NodeCausAff ?NodeOtherAff
{
  <syn>{
    node ?NodeVerb [cat=verb, trans=?Trans, caus=?
        Caus];
    node ?NodeVRoot (mark=anchor) [cat=vroot, trans=?
        Trans];
    node ?NodeCausAff (mark=subst) [cat=causaff, caus
        =?Caus];
    node ?NodeOtherAff (mark=subst) [cat=aff];
    ?NodeVerb -> ?NodeVRoot; ?NodeVerb -> ?NodeCausAff;
         ?NodeVerb -> ?NodeOtherAff;
    ?NodeVRoot >> ?NodeCausAff; ?NodeCausAff >> ?
        NodeOtherAff
  }
}
```

Building the 3-argument core

```
class core_spine
declare ?Core ?Nuc ?UnderNuc ?NArg1 ?NArg2 ?NArg3
{
  <syn>{
    node ?Core [cat=core];
    node ?Nuc [cat=nuc];
    node ?UnderNuc (mark = subst) [cat = verb, trans
        = tr, caus = yes];
    node ?NArg1 (mark = subst) [cat = noun, case =
        nom];
    node ?NArg2 (mark = subst) [cat = noun, case =
        acc];
    node ?Narg3 (mark = subst) [cat = noun, case =
        ade];
    ?Core -> ?Nuc; ?Nuc -> ?UnderNuc;
    ?Core -> ?NArg1; ?Core -> ?NArg2; ?Core -> ?NArg3;
    ?NArg1 >> ?Nuc ; ?Nuc >> ?NArg2; ?NArg2 >> ?NArg3
  }
}
```

Conclusions

- Our solution handles both inflection and derivation
- Captures the meaning and function of each particular morpheme
- Allows precise parsing of morphologically rich languages
- The lexicon can be created once for different parts of the metagrammar, but from scratch for each new language
- Further step: add automatic segmentation

Thank you!

Your feedback is very welcome!

References

Manninen, Satu and Diane Nelson (2004). "What is a passive? The case of Finnish." In: *Studia linguistica* 58.3, pp. 212–251.