

Describing morphology with Metagrammars

Simon Petitjean

TreeGraSP meeting #3

Introduction

- XMG ([Crabbé et al., 2013], [Petitjean et al., 2016]): description tool for various types of linguistic resources
- Descriptions based on logic and constraints
- In this talk: morphological descriptions
- Interface with semantics (and syntax?)

Plan

- 1 Introduction
- 2 The morphological dimension(s)
- 3 Output and applications
- 4 Conclusion

Plan

- 1 Introduction
- 2 The morphological dimension(s)**
- 3 Output and applications
- 4 Conclusion

The morphological dimension(s)

- Describing the formation of words: several attempts with XMG
- Using the theory of topological fields → Ikota [Duchier et al., 2012]
- Describing the morpho-semantic interface → Arabic [Petitjean et al., 2015]
- Focusing on the semantics → polysemy in English [Andreou and Petitjean, 2017]
- Producing lexicons for syntactic parsing (previously lexConverter)

Ikota and topological fields

- Ikota: Bantu language of Gabon and the Democratic Republic of Congo
- Verbal morphology can be described using topological fields (following [Bech, 1955])

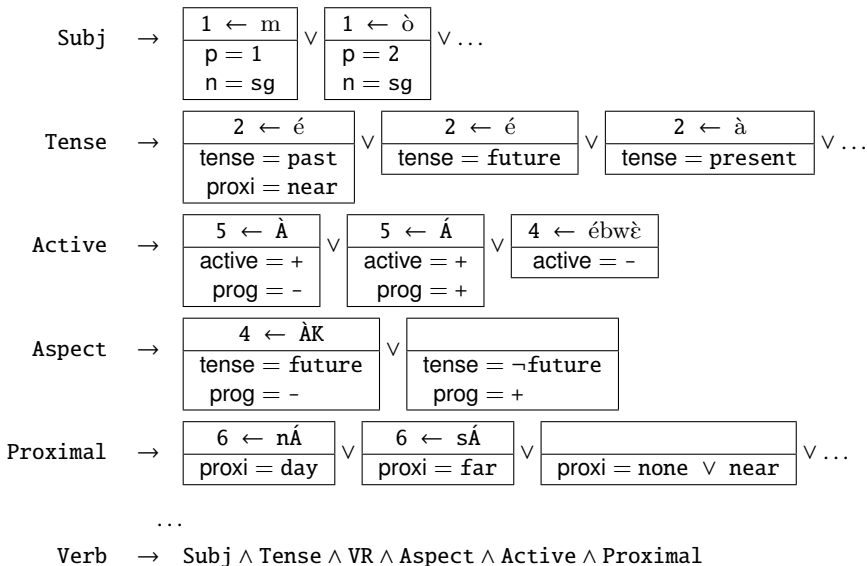
Table: Verb formation

- | | | | | | |
|-------|--------|----|-----------|---------|-------------|
| Subj- | Tense- | VR | -(Aspect) | -Active | -(Proximal) |
|-------|--------|----|-----------|---------|-------------|

Subj.	Tense	VR	Aspect	Active	Prox.	Value
m-	à-	ɔ̃		-á		present
m-	à-	ɔ̃		-á	-ná	past, yesterday
m-	à-	ɔ̃		-á	-sá	distant past
m-	é-	ɔ̃		-á		recent past
m-	é-	ɔ̃	-àk	-à		medium future
m-	é-	ɔ̃	-àk	-à	-ná	future, tomorrow
m-	é-	ɔ̃	-àk	-à	-sá	distant future
m-	ábí-	ɔ̃	-àk	-à		imminent future

Ikota: Metagrammatical description

- Fields are defined and ordered globally
- Classes can affect strings into fields
- Morphosyntactic features: similar to the interface dimension



Derivational morphology of Arabic verbs

- Words: combination of two abstract morphemes: a root and a pattern (Binyanim)

Pattern	Active	Passive
1	$C_1aC_2aC_3$	$C_1uC_2iC_3$
2	$C_1aC_2C_2aC_3$	$C_1uC_2C_2iC_3$
3	$C_1aaC_2aC_3$	$C_1uuC_2iC_3$
4	$\text{ʔ}aC_1C_2aC_3$	$\text{ʔ}uC_1C_2iC_3$
5	$taC_1aC_2C_2aC_3$	$tuC_1uC_2C_2ib$
6	$taC_1aaC_2aC_3$	$tuC_1uuC_2iC_3$
7	$nC_1aC_2aC_3$	$nC_1uC_2iC_3$
8	$C_1taC_2aC_3$	$C_1tuC_2iC_3$
10	$staC_1C_2aC_3$	$stuC_1C_2iC_3$

Table: Patterns for verbal roots with three consonants

Derivational morphology of Arabic verbs

- Compute the derived semantics: underspecified semantic descriptions associated to (at least) some patterns

Pattern	Meaning
2	Causative of transitive 1
3	Associative
4	Causative of 1
5	Reflexive of 2 (mediopassive)
6	Reciprocal of 3
7	Reflexive / resultative / passive / mediopassive
8	Reflexive / mediopassive
10	Requestative

Table: Semantics for verbal patterns according to [Ryding, 2005]

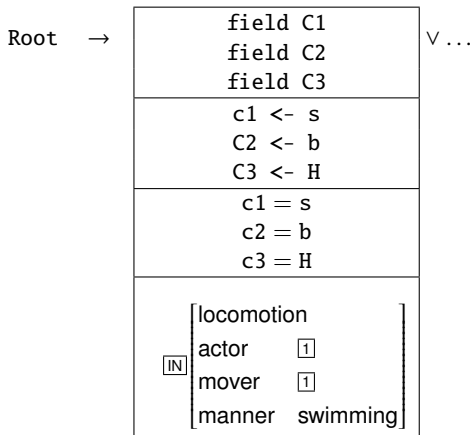
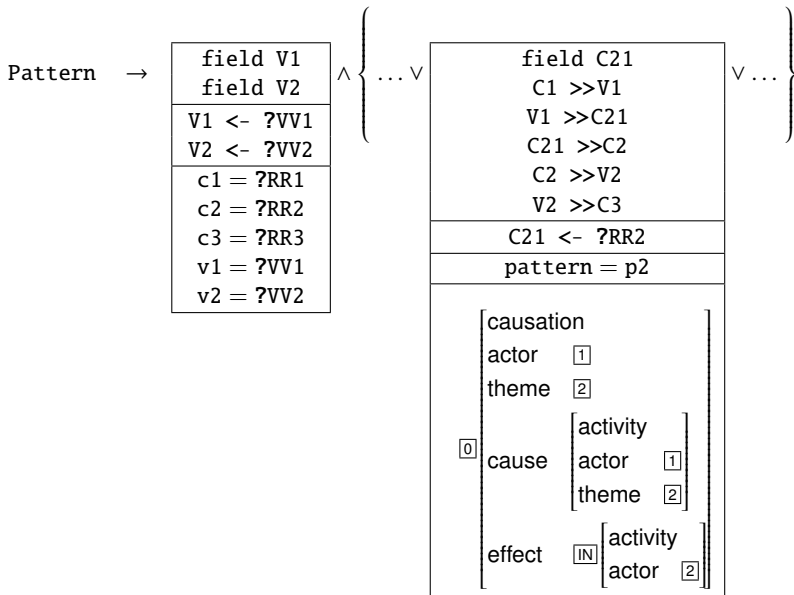


Figure: Metagrammar of the morphology-semantics interface



Arabic: Metagrammatical description

- Fields are defined and ordered in classes (vs globally for Ikota), as their number/order is not fixed
- Classes can affect strings into fields and typed feature structures into the frame dimension

```

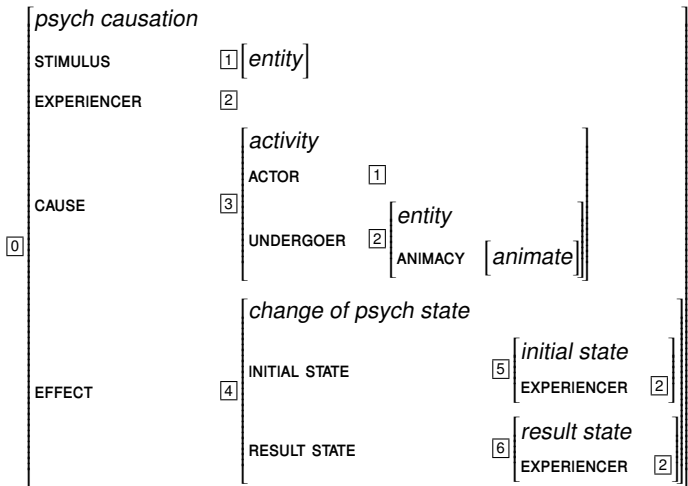
class sbh
export CN
declare ?CN ?X1
{
  <frame>{
    ?CN[locomotion,
      actor:?X1,
      mover:?X1 ,
      manner: swimming]}
  ; <morph>{field R1; field R2; field R3}
  ; <iface>{[r1="s", r2="b", r3="H"]}
}

```

Polysemy: the suffix *-ment*

- Assumption: the semantics of a noun derived from a verb (here using the suffix *-ment*) can be inferred from the one of the verb
- Semantic frames: the derived reading will target a part of the verb's frame
- [Kawaletz and Plag, 2015] : *-ment* on psych verbs derives 5 readings
 - EVENT 'transpositional'
 - STIMULUS
 - ACTIVITY
 - CHANGE OF PSYCH STATE
 - RESULT STATE readings

Attribute-value matrix for psych verbs



Polysemy: Metagrammatical description

- Earliest version based on fields (identified by variables) and feature structures :

```
|           field ?ID [f1=v1, ..., fn=vn]
```

- Linear precedence constraints between fields:

```
|           ?ID1 >> ?ID2
```

- Affect strings to fields:

```
|           ?ID1 <- 'bla'
```

```
class ment
import amuse[]
declare ?Ref ?Suffix
{
  <morph>{field ?Suffix [e=?Ref]; ?RootField >> ?Suffix;
          ?Suffix <- "ment"};
  <frame>{ [ment-lexeme,
           m-base:[event,
                    sem:?Root]
           ref:?Ref           ];
          ?Root >* ?Ref
          ...}
```


Plan

- 1 Introduction
- 2 The morphological dimension(s)
- 3 Output and applications**
- 4 Conclusion

Output and applications

- What is generated? A lexicon of inflected forms, with associated semantics
- Visualization in the WebGUI
(http://xmg.phil.hhu.de/index.php/upload/upload_viewer)
- In a better world: usage as a lexicon for syntactic parsing

Conclusion

- 2 ways to describe morphology with XMG
- For visualization: the `morph1p` dimension based on fields
- For syntactic parsing: the `mph` dimension
- Current work: towards a more unified solution
- Future work: compute fully inflected (surface) forms

Thank you



Andreou, M. and Petitjean, S. (2017).

Describing derivational polysemy with xmg.

In Eshkol, I. and Antoine, J.-Y., editors, *Actes de TALN 2017, 24e Conférence sur le Traitement Automatique des Langues Naturelles*, volume 2, pages 94–101.



Bech, G. (1955).

Studien über das deutsche Verbum infinitum.

Det Kongelige Danske videnskabernes selskab. Historisk-Filosofiske Meddelelser, bd. 35, nr.2 (1955) and bd. 36, nr.6 (1957). Munksgaard, Copenhagen.

2nd unrevised edition published 1983 by Max Niemeyer Verlag, Tübingen (Linguistische Arbeiten 139).



Crabbé, B., Duchier, D., Gardent, C., Le Roux, J., and Parmentier, Y. (2013).

XMG : eXtensible MetaGrammar.

Computational Linguistics, 39(3):1–66.



Duchier, D., Magnana Ekoukou, B., Parmentier, Y., Petitjean, S., and Schang, E. (2012).

Describing Morphologically-rich Languages using Metagrammars: a Look at Verbs in Ikota.

In *Workshop on "Language technology for normalisation of less-resourced languages", 8th SALT MIL Workshop on Minority Languages and the 4th workshop on African Language Technology*, page –, Istanbul, Turkey.



Kawaletz, L. and Plag, I. (2015).

Predicting the Semantics of English Nominalizations: A Frame-Based Analysis of -ment Suffixation, pages 289–319.

Springer International Publishing, Cham.



Petitjean, S., Duchier, D., and Parmentier, Y. (2016).

XMG 2: Describing Description Languages.

In Logical Aspects of Computational Linguistics. Celebrating 20 Years of LACL (1996–2016) 9th International Conference, LACL 2016, Nancy, France, December 5-7, 2016, Proceedings 9, pages 255–272. Springer Berlin Heidelberg.



Petitjean, S., Samih, Y., and Lichte, T. (2015).

Une métagrammaire de l'interface morpho-sémantique dans les verbes en arabe.

In Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles, pages 473–479, Caen, France.



Ryding, K. (2005).

A Reference Grammar of Modern Standard Arabic.

A Reference Grammar of Modern Standard Arabic. Cambridge University Press.