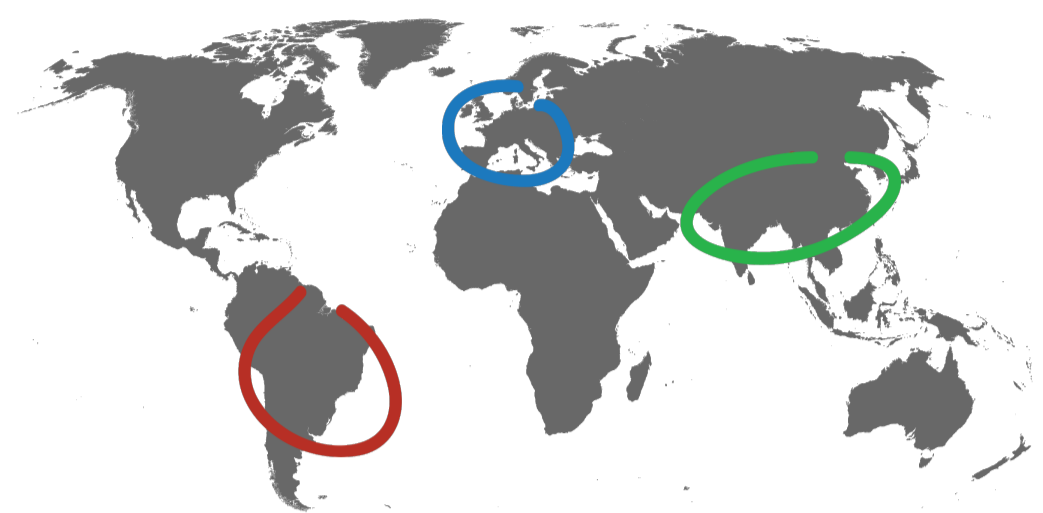
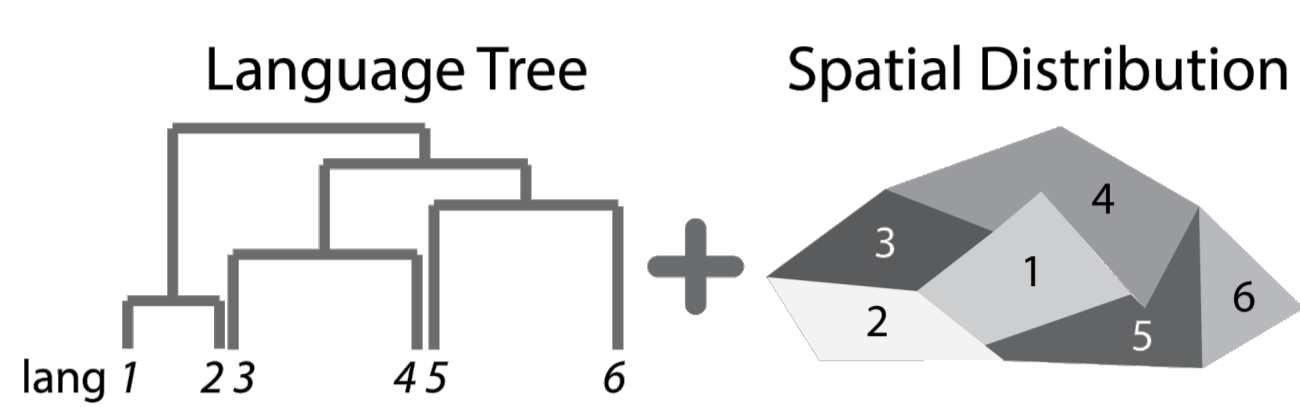


Next-Generation Databases for Linguistics: Capturing Complex Patterns in Grammars

Taras Zakharko, Rik van Gijn
University of Zurich



LIMITS – Linguistic Morphology In Time and Space

How does morphology develop in different genealogical and geographical contexts and to what extent is this development affected by language contact?

Multivariate, probabilistic, bottom-up approach

Problem: how do we capture complex morphological patterns?

- Heterogenous structures and types
- Complex, transitive, nested relations
- Descriptions constantly change and evolve

Traditional database designs do not scale!

LiMiTslang – a novel knowledge representation system

- A domain-specific language for encoding morphosyntactic patterns
- Powerful abstractions model paradigmatic and syntagmatic relations between arbitrary linguistic entities
- Flexible, adaptable and compact
- Designed for linguists by linguists
- Implemented as DSL on top of R programming language

Phrase	%is_a%	Category()
NP	%->%	Phrase
Word	%is_a%	Category()
N	%->%	Word
N	%->%	NP[.]
NP	%->%	[N]

A knowledge base for grammars

Is equivalent to

or

```
verbal.prefix %<->% 'm-' | 'g-' | 'v-' | 'gv-'
verbal.suffix %<->% '-m' | '-t' | '-en'
```

contains

```
(verbal.prefix | verbal.suffix) %->% V[.]
```

implies

- A verbal prefix is exactly one of 'm-', 'g-', 'v-' or 'gv-'
- A verbal suffix is exactly one of '-m', '-t' or '-en'
- Something that is either a verbal suffix or a verbal prefix only occurs as part of a verb (V)

uniqueness quantifier

Immediate sequence

```
NumP[.1:!!Numeral(n>1) - .2:N(animate)] %->% .[
  .1[!!`denaa`],
  .2[!!`pl.marker`]
]
```

anchor label

- If a numeral phrase (NumP) with a numeral(>1) contains an animate noun, then this noun carries obligatory plural marker and the numeral contains the classifier marker 'denaa'

Types of linguistic objects (morphemes, words, constructions) represented as abstract *categories*

All categories are freely defined and endowed with a specific interpretation by the linguist

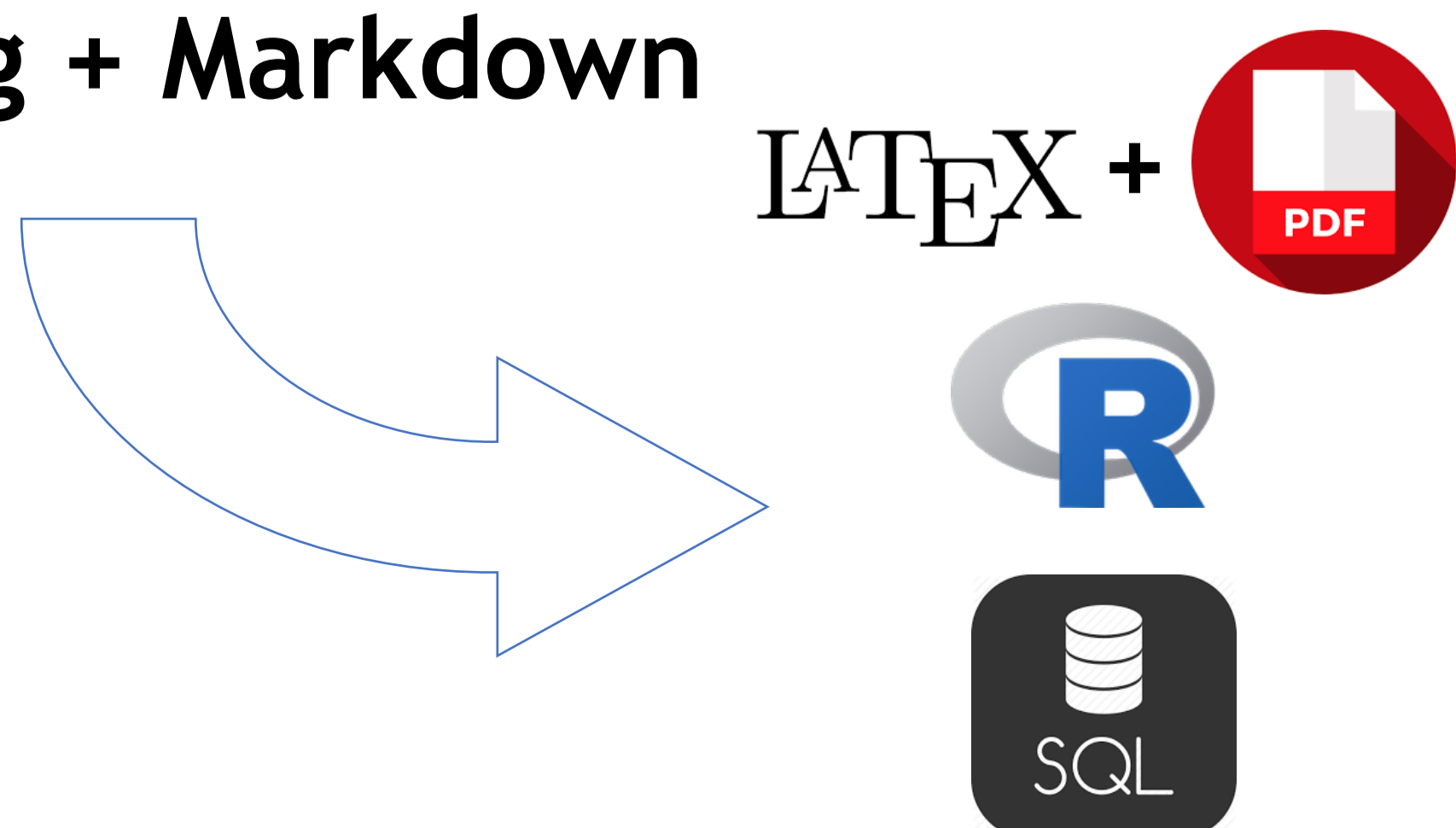
N	Category of all nouns
N(animate)	Category of all animate nouns
N(pl)	Category of noun plural forms
N(masc)	Category of all masculine nouns

Relationships between categories described in form of *pattern entailments*

N %->% Word	Every noun is a word
N[pl.m] %->% N(pl)	Every noun that contains a plural marker is a plural noun form

Rich syntax to describe category combinations

LiMiTslang + Markdown



- LiMiTslang descriptions are text and can be processed with industry-standard text-editing and management tools (e.g. version control systems)
- LiMiTslang descriptions can be freely mixed with markdown for integrated language reports and papers that combine informal and formal descriptions
- Generated knowledge bases can be exploited with a plethora of tools