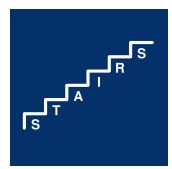
Formalizing How Concepts are Expressed in Natural Language

Jan Frederik Schaefer, Marcel Schütz

FAU Erlangen-Nürnberg







Verbalizations

Concept: divisibility

Symbolic notation: "$x \mid y$"

Verbalizations: "x *divides* y"

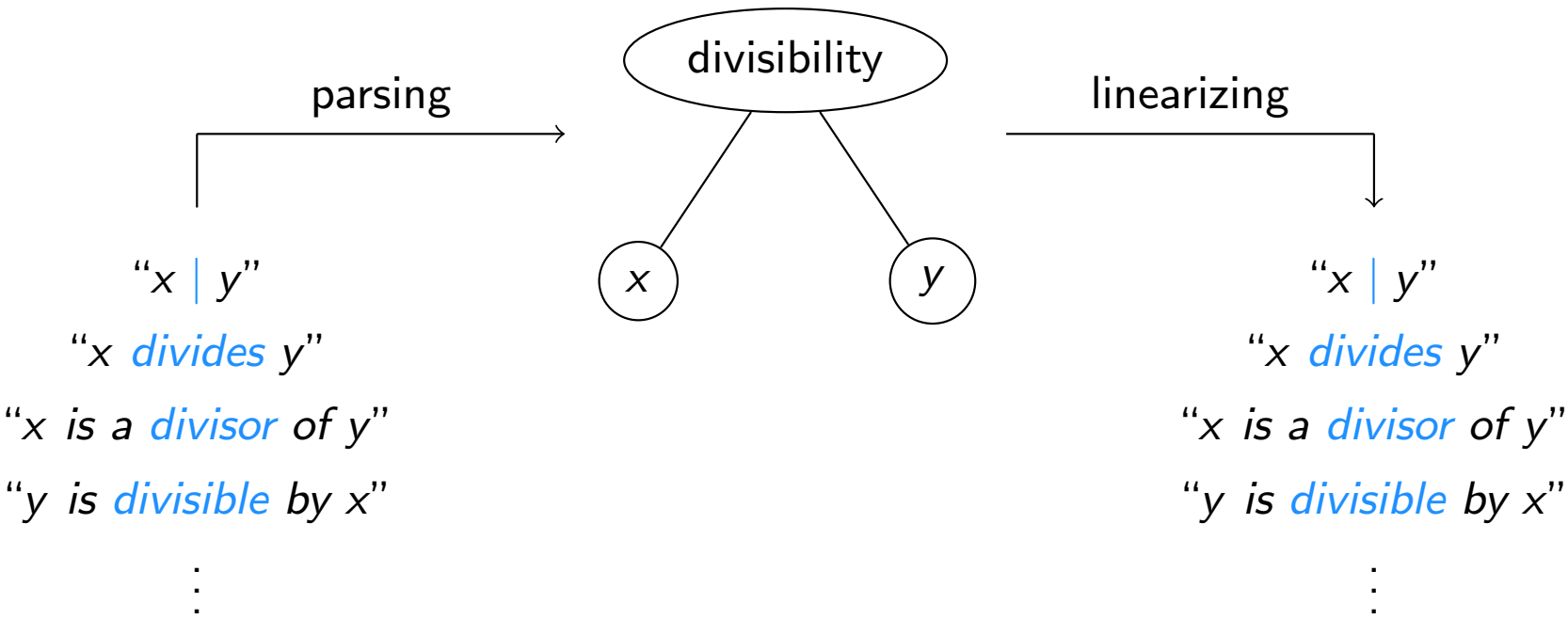
"x *is a divisor of* y"

"y *is divisible by* x"

⋮

Applications

- Auto-(in)formalization
- Natural language generation/translation/...
- Adapting learning material (e.g. in ALEA)
- ...



Representing Verbalizations in sTeX

sTeX = a LaTeX package for semantic annotations

Concept Declaration:

```
\symdecl{div}[args=2]
```

Symbolic Notation:

```
\notation{div}{#1 \mid #2}
```
→ "$x \mid y$"

Verbalizations (new):

```
\verbalization{div}{divide:V ?2}
```
→ "x *divides* y"

```
\verbalization{div}{divisor:N of ?2}
```
→ "x *is a divisor of* y"

```
\verbalization{div}{divisible:A by ?1}
```
→ "y *is divisible by* x"

Generating a GF Lexicon *

GF (Grammatical Framework) = a grammar-based NLP tool

```
\verbalization{div}{divisor:N of ?2}
```



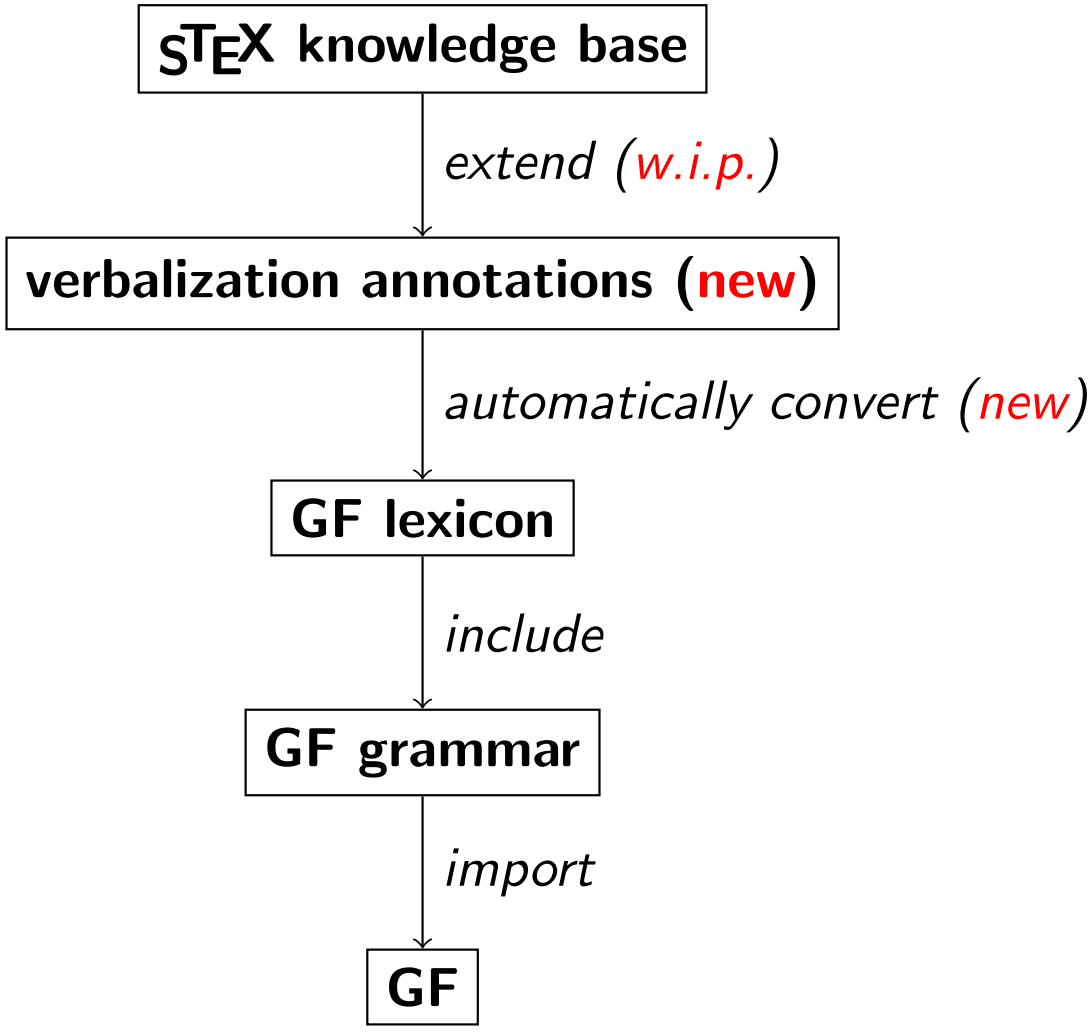
```
divisor_Noun : Noun;
divisor_Noun = mkNoun "divisor" "divisors";

divisor_of : Property;
divisor_of = mkProp (mkNP divisor_Noun) of_Prep;
```

- Generate GF rules via heuristics (default)
- Write GF rules manually (fallback)

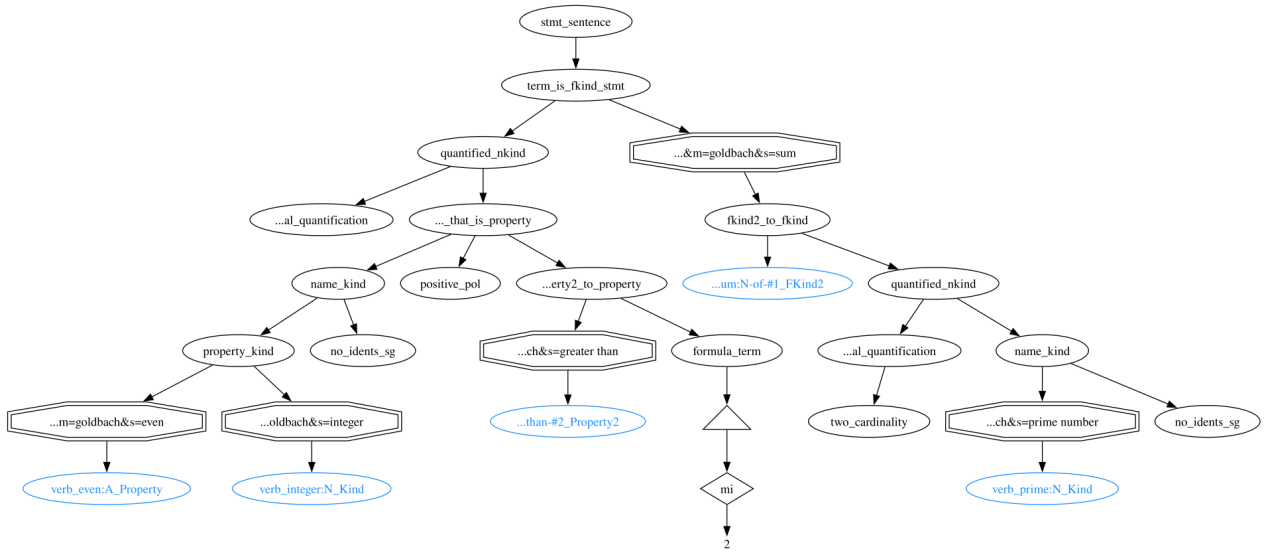
*Inspired by Aarne Ranta and his Informath project

Operationalizing Verbalizations



Verbalizations in NLP via sTEX and GF

- Formalize verbalizations of concepts in sTEX
- Generate GF grammar rules from these verbalizations
- Use these grammar rules for NLP tasks



"*Every even integer that is greater than 2 is the sum of two primes.*"