

The background of the slide is a close-up photograph of a wood grain, showing wavy, concentric patterns in shades of brown and tan. A dark blue rectangular box is positioned on the right side of the image, containing the title and authors' names in white text. The title is written in a large, white, serif font, and the authors' names are in a smaller, white, sans-serif font.

Handling Idioms in Symbolic Multilingual Natural Language Generation

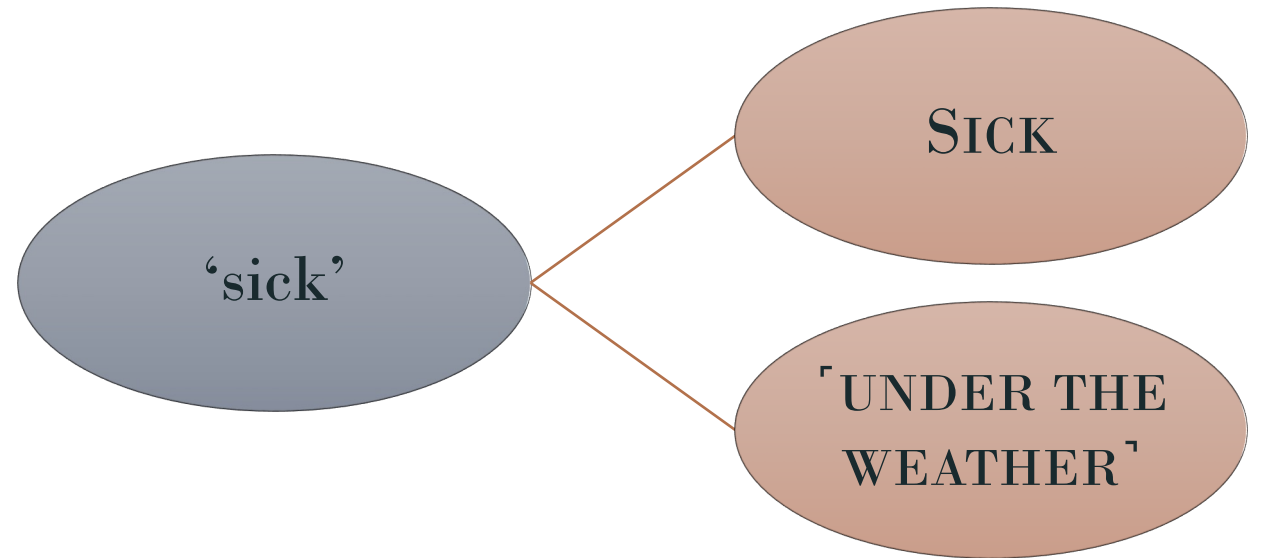
MICHAELLE DUBÉ, FRANÇOIS LAREAU

OLST, Université de Montréal

Introduction

IDIOMS IN RULED BASES NLG

- **Non-compositional MWE**
 - Single unit of meaning
 - Syntactic structure
- Syntactic cohesion
- Form flexibility



Form flexibility

1. Donner encore sa langue au chat
‘give up guessing again’, lit. ‘give again one’s tongue to the cat’
- modifier
2. Take it with a big grain of salt.
3. You must take whatever he says with a grain of salt.
- complement
4. He’s just pulling your leg.
5. ‘LIETUVOS APELIACINIS TEISMAS’
‘court of appeals of Lithuania’ lit. ‘appellate court of Lithuania’
- inflection
6. It was a pretty big bullet to bite

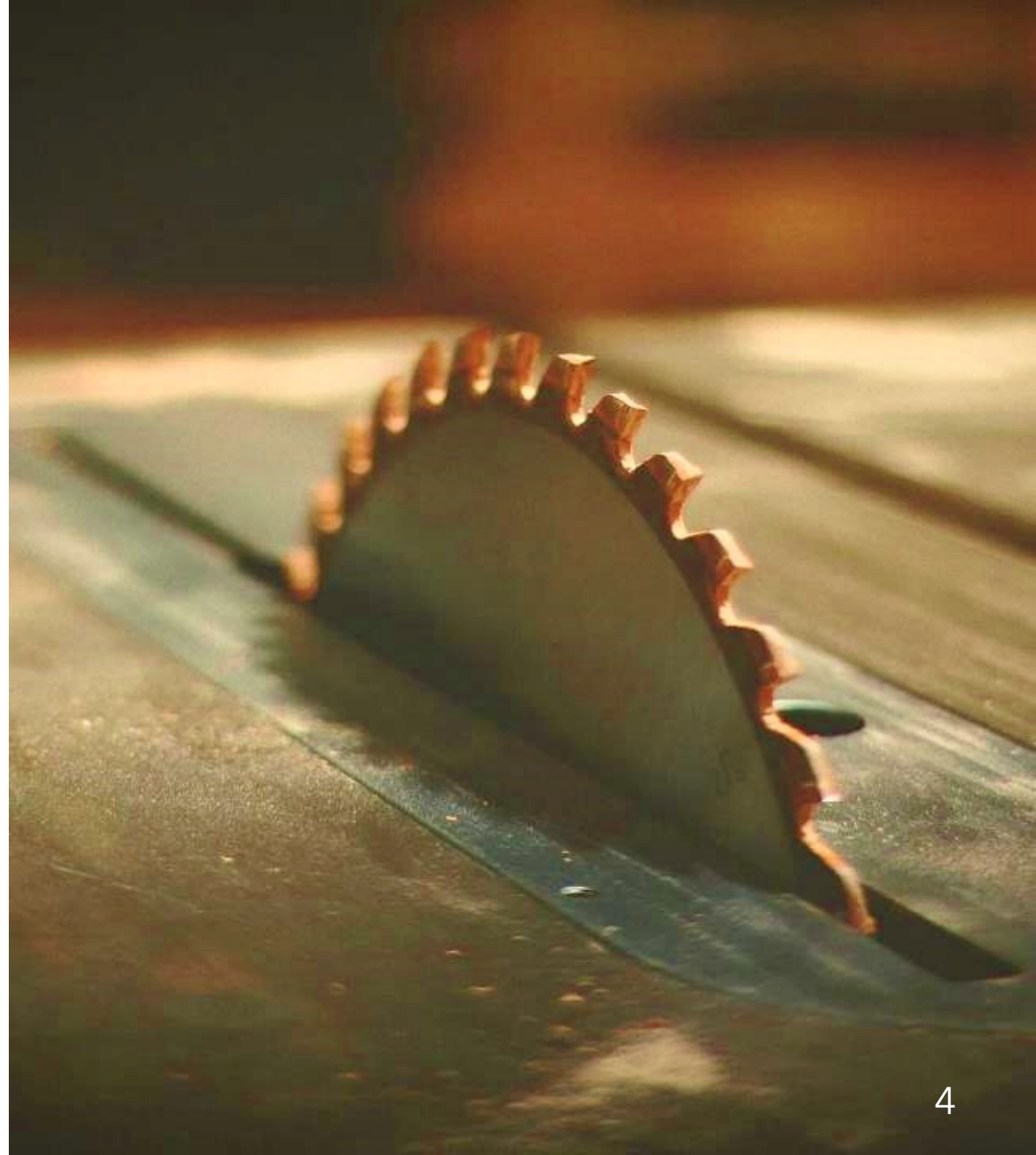
Overview

Linguistic Realizers

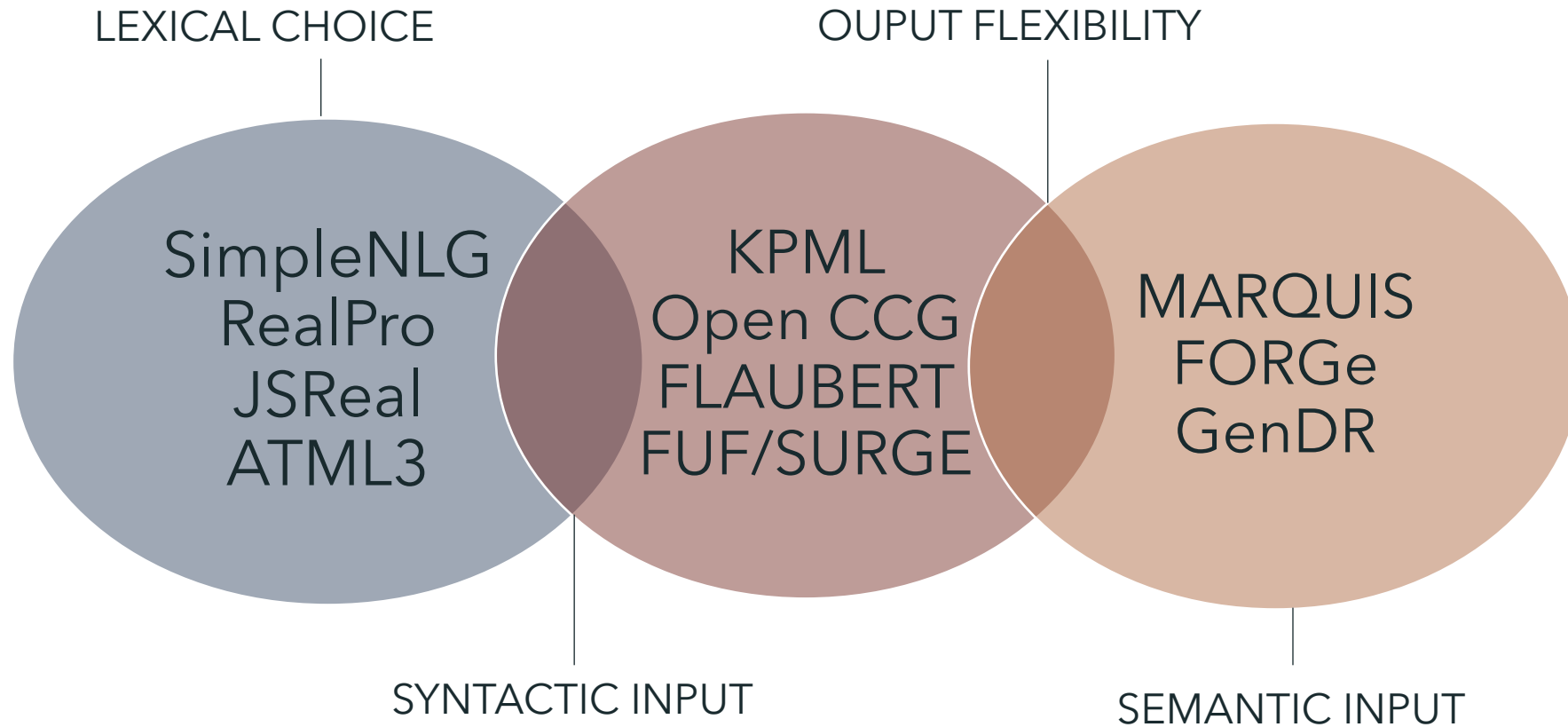
Lexical Data

Implementation

Evaluation



Linguistic Realizers



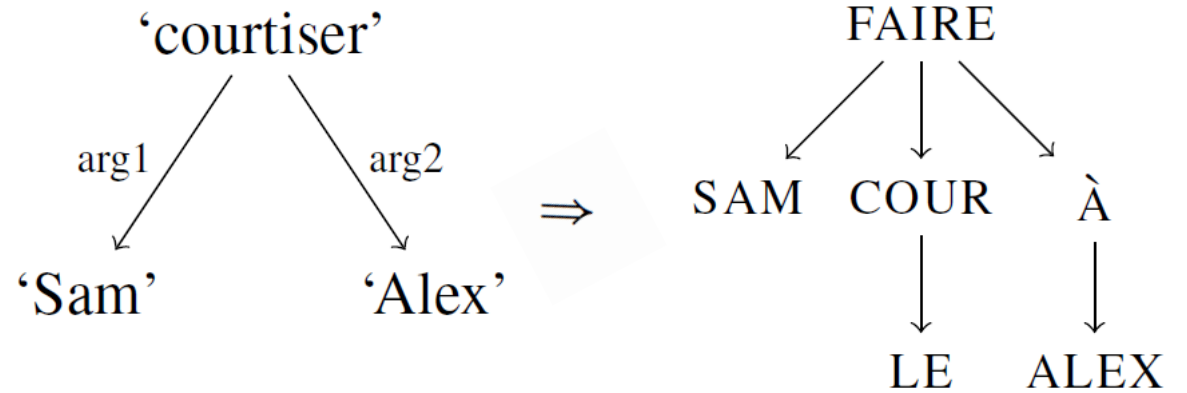
Lexical Data

- French Lexical Network (LN_{fr})
- Grammatical classification
- Linear syntactic patterns
- Sequences of part of speech (POS)

Linguistic patterns	Example	#	%
N Prep N	TÊTE DE MULE	409	14%
N Adj	TERRE FERME	377	13%
Prep N	DE JUSTESSE	222	8%
N Prep.circ N	CORPS À CORPS	138	5%
Adj N	JOLI COEUR	100	4%
Prep Det N	DANS LE VENT	98	3%
V Det N	LEVER LE PIED	97	3%
N Prep N	ART DE LA TABLE	97	3%
Others	...	1417	49%
Total		2919	100%

Implementation

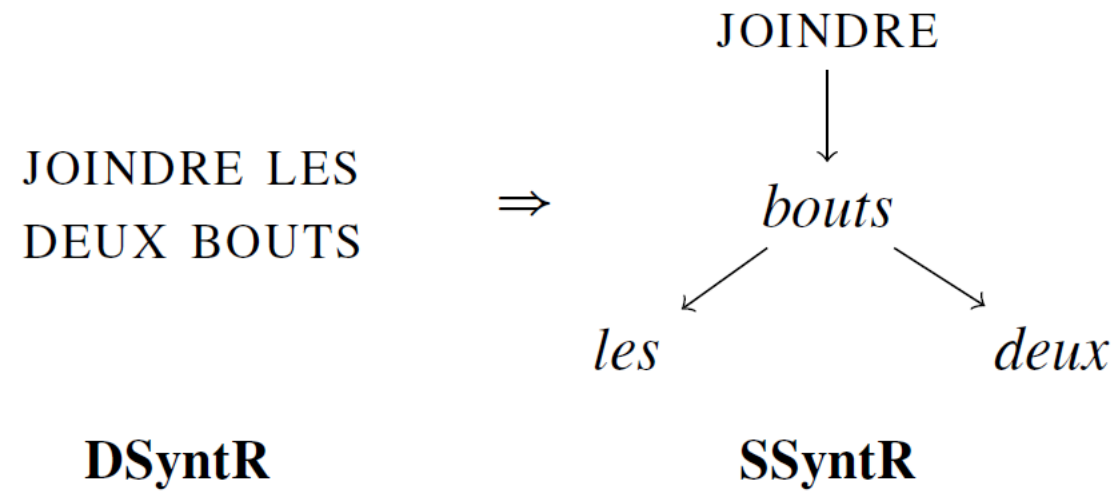
- Generic deep realizer (GenDR)
- Resource sharing
- MATE = graph transducer
- Meaning-Text theory (MTT)
- Semantics-syntax interface



SemR

SSyntR

Template Lexicalization Rules

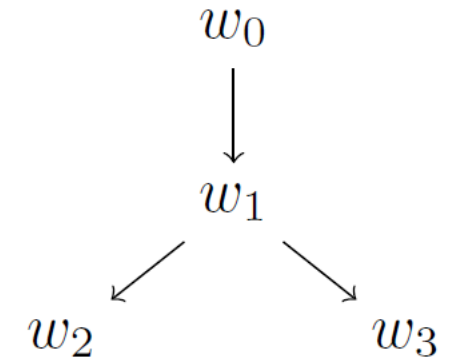


Generic Tree Patterns

「JOINDRE LES DEUX BOUTS」
「ENFONCER UNE PORTE OUVERTE」
「DANS DE BEAUX DRAPS」

V Det Num N
V Det N Adj
Prep Det Adj N

\Rightarrow



Linguistic

Generic

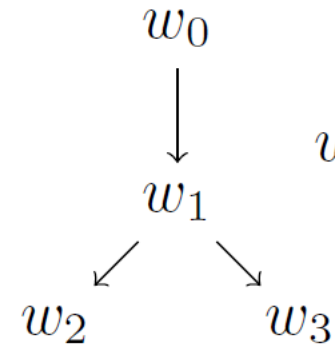
Generic Tree Patterns

	Partition	# tree
4 nodes	3	2
	1 + 2	1
	1 + 1 + 1	1
		=4

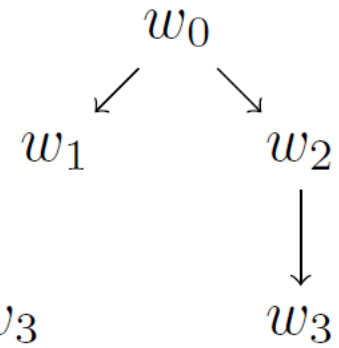
4_01



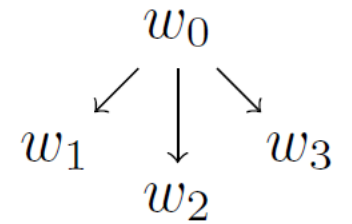
4_02



4_03

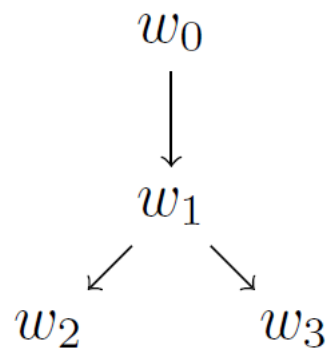


4_04



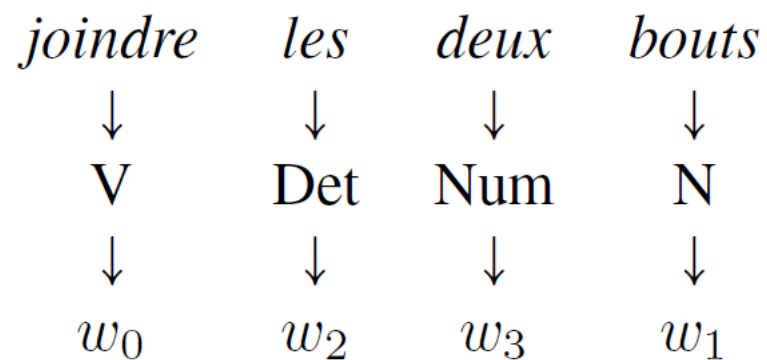
Mapping Patterns

Input

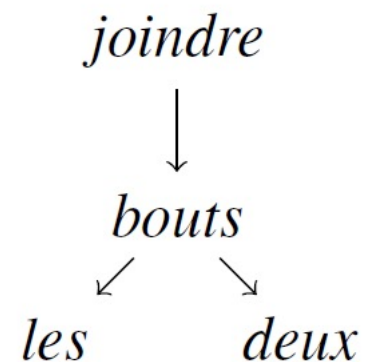


V Det Num N

Mapping



Output



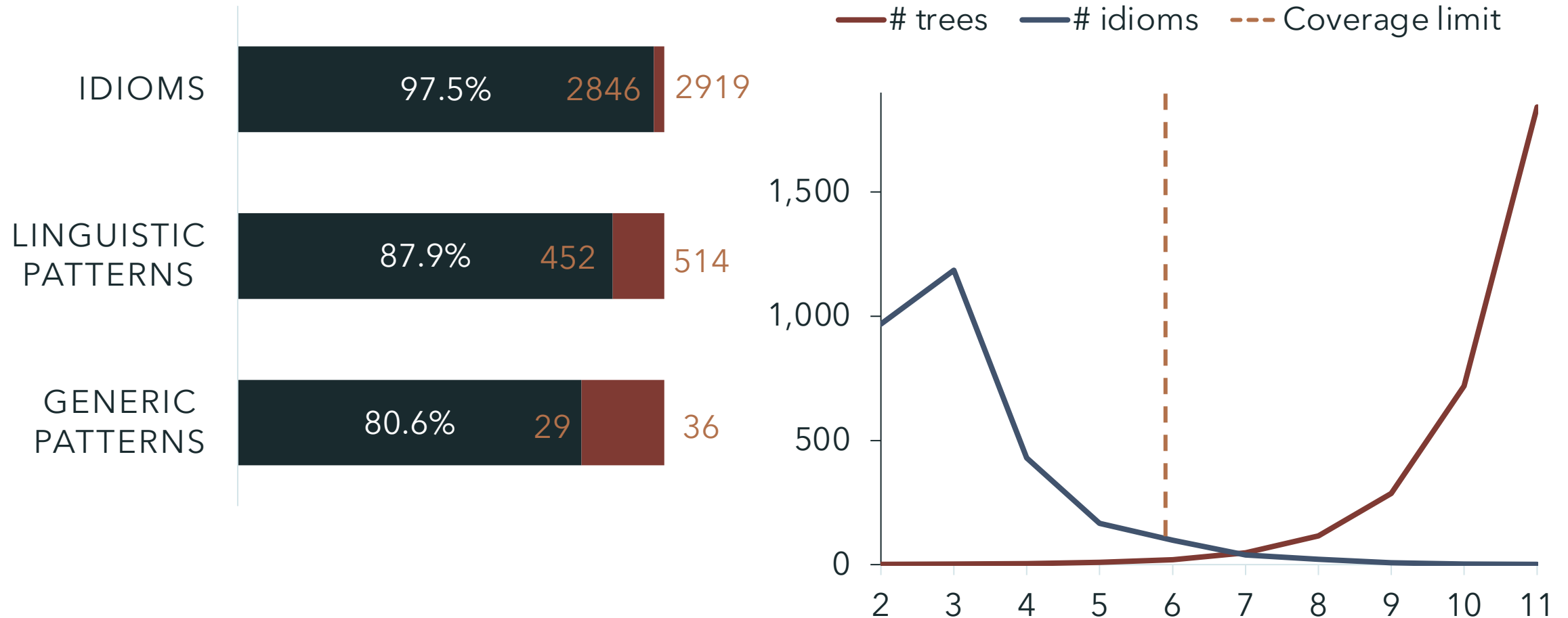


5. Evaluation

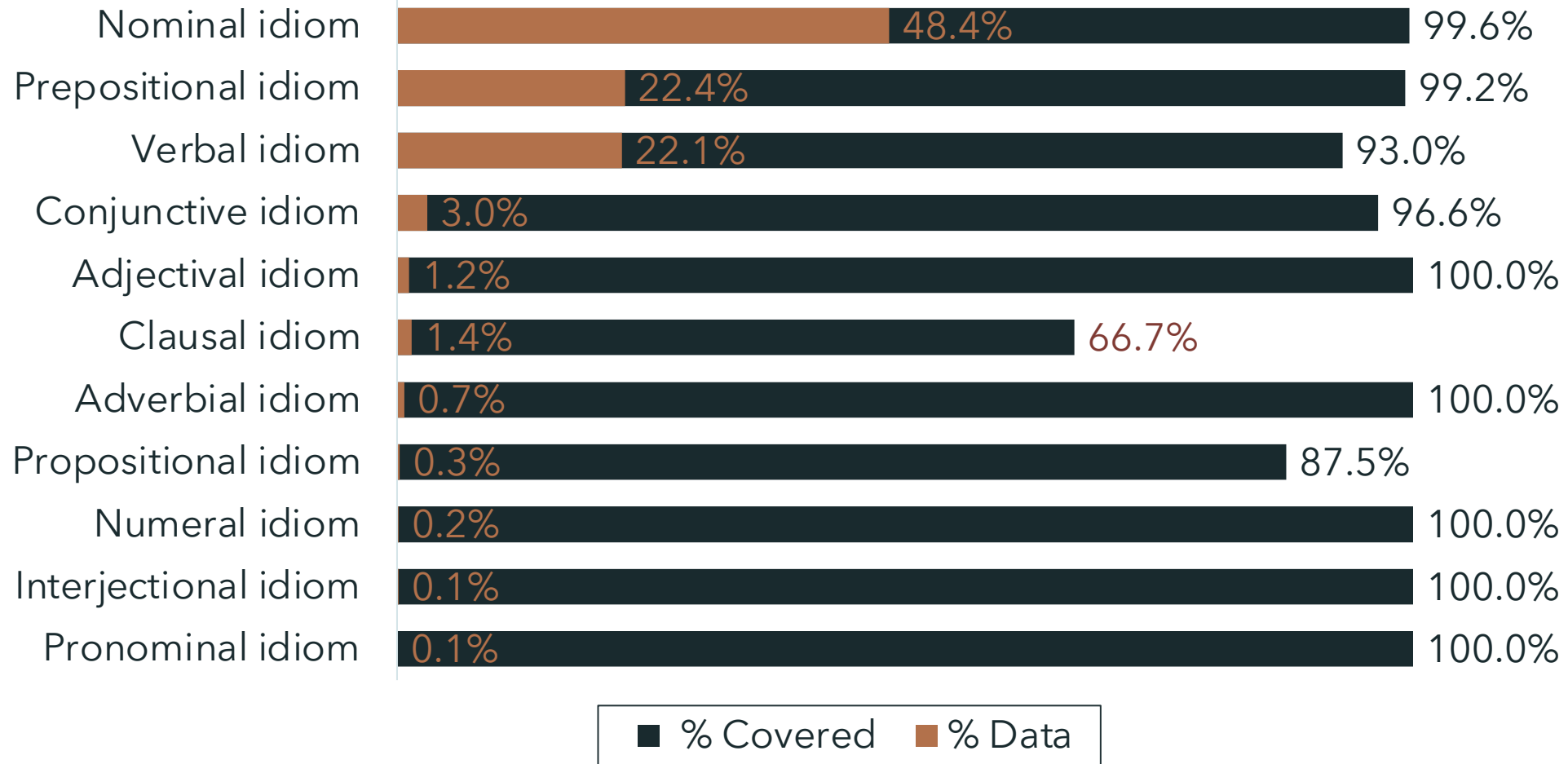
5.1 Coverage

5.2 Precision

Results : Coverage



COVERAGE BY POS



Precision

97.8 % OVERALL PRECISION

Sample	n	Judge A	Judge B	k
Sample 1	100	98		
Sample 2	100		97	
Sample 3	300	295	294	0.91

PROBLEMS / SOLUTIONS

- Pattern mapping
 - Vague linguistic pattern
 - Node inflection
- Data collection (inflection)
- LNfr precision
 - Government patterns & coreferences



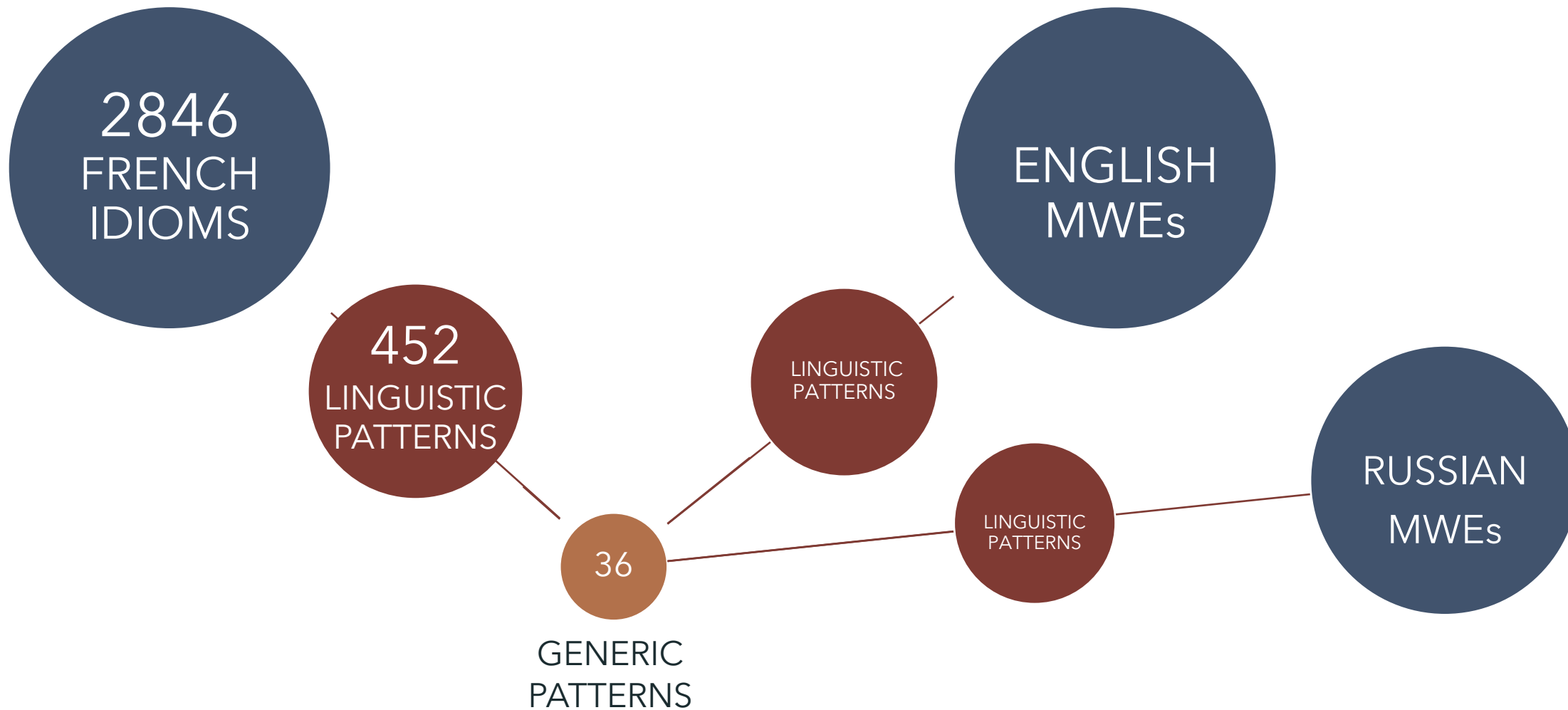
6. Conclusion

Handling MWEs in MNLG

- Language independent
- Linguistic patterns
- Generic patterns

Future work

- Graph transducer
 - GREW
- Extend to English and Russian
 - LN_{en} LN_{ru}



A close-up photograph of a tree trunk showing concentric growth rings in shades of brown, tan, and blue.

Thank you

Michaëlle Dubé

michaëlle.dube@umontreal.ca

www.olst.ling.umontreal.ca