Carnegie Mellon University

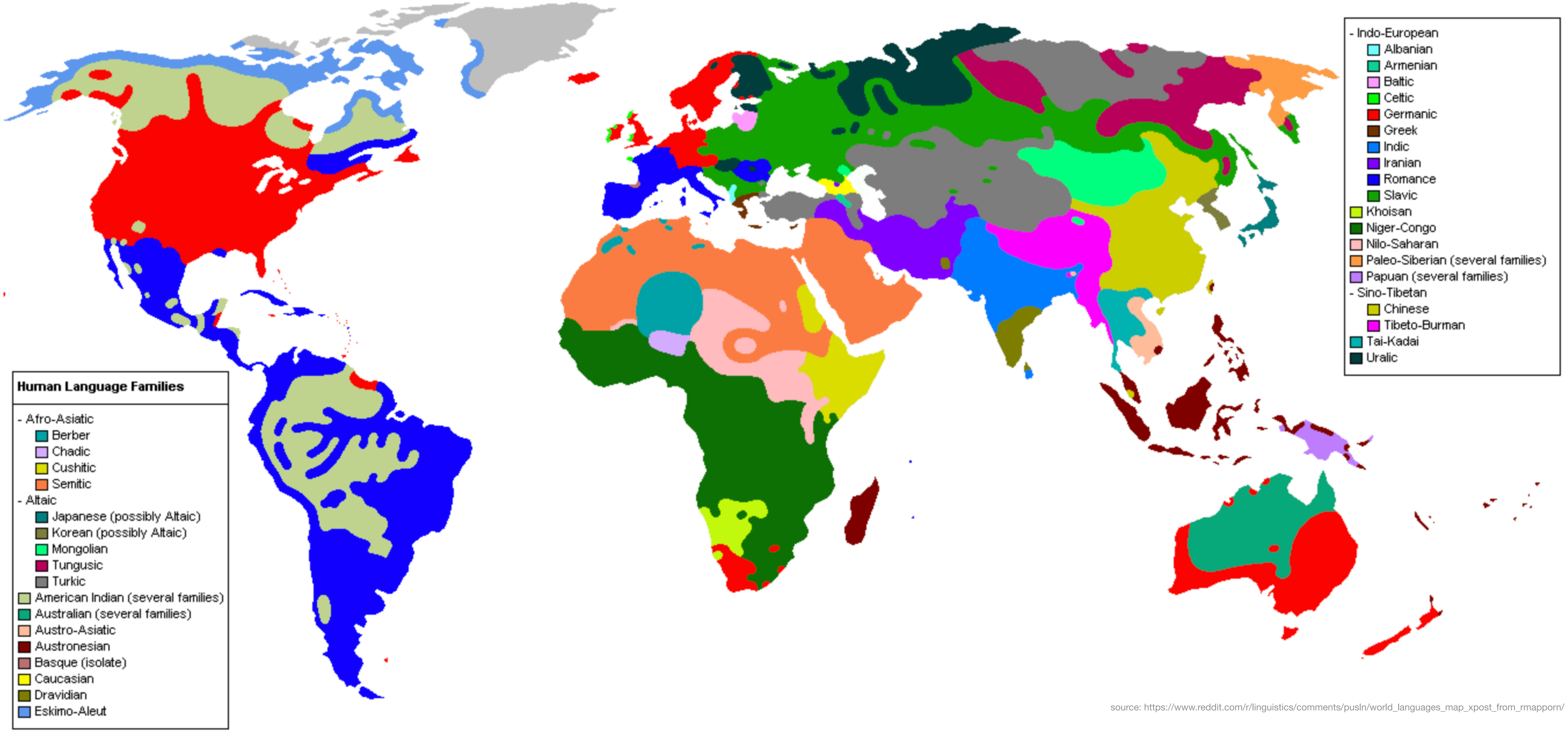
Can we Automatically Create Language-learning Textbooks?

https://www.autolex.co/

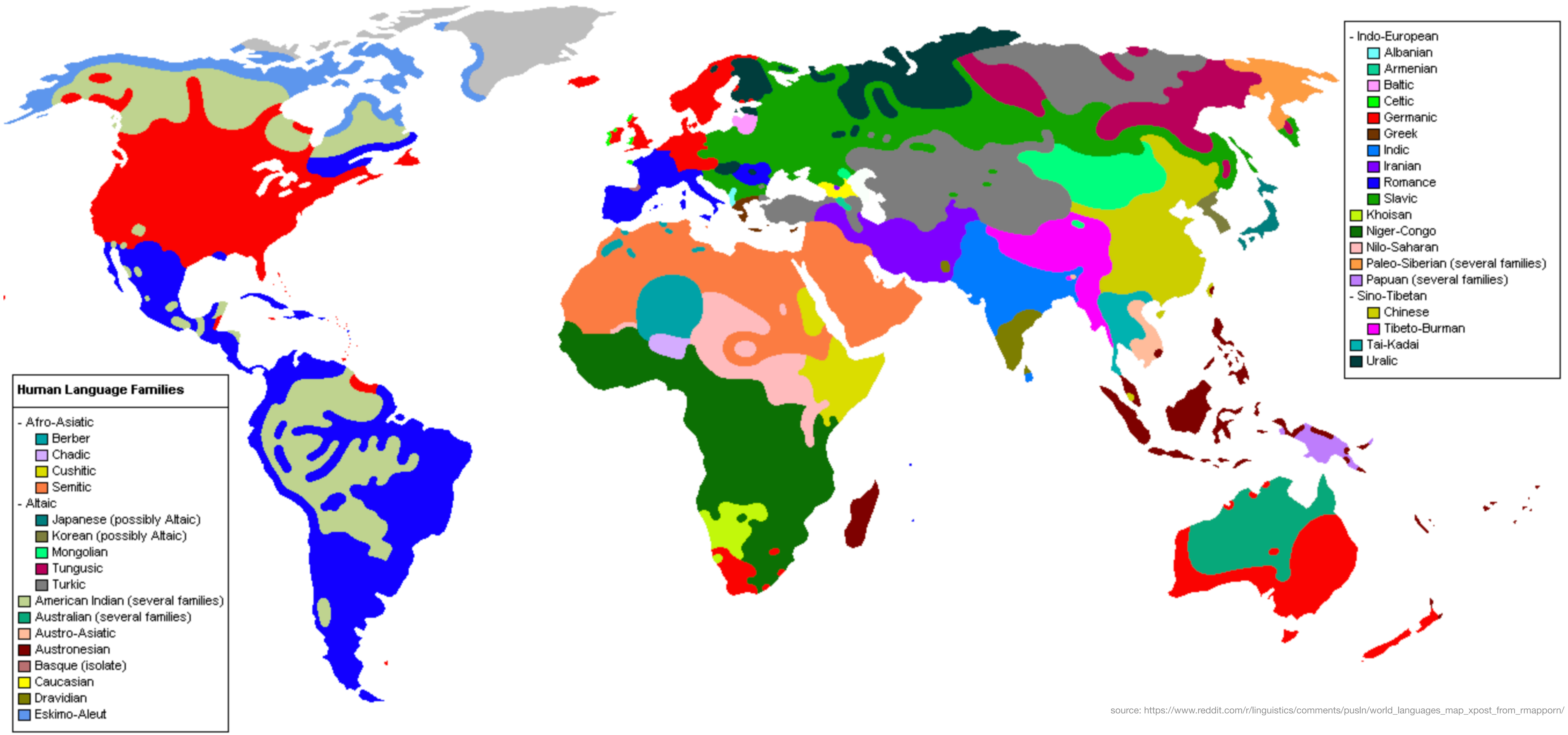
Graham Neubig

w/ Aditi Chaudhary

Antonios Anastasopoulos, David Mortensen, Zaid Sheikh, Arun Sampath, Ashwin Sheshadri, Adithya Pratapa, Yulia Tsvetkov, Kayo Yin

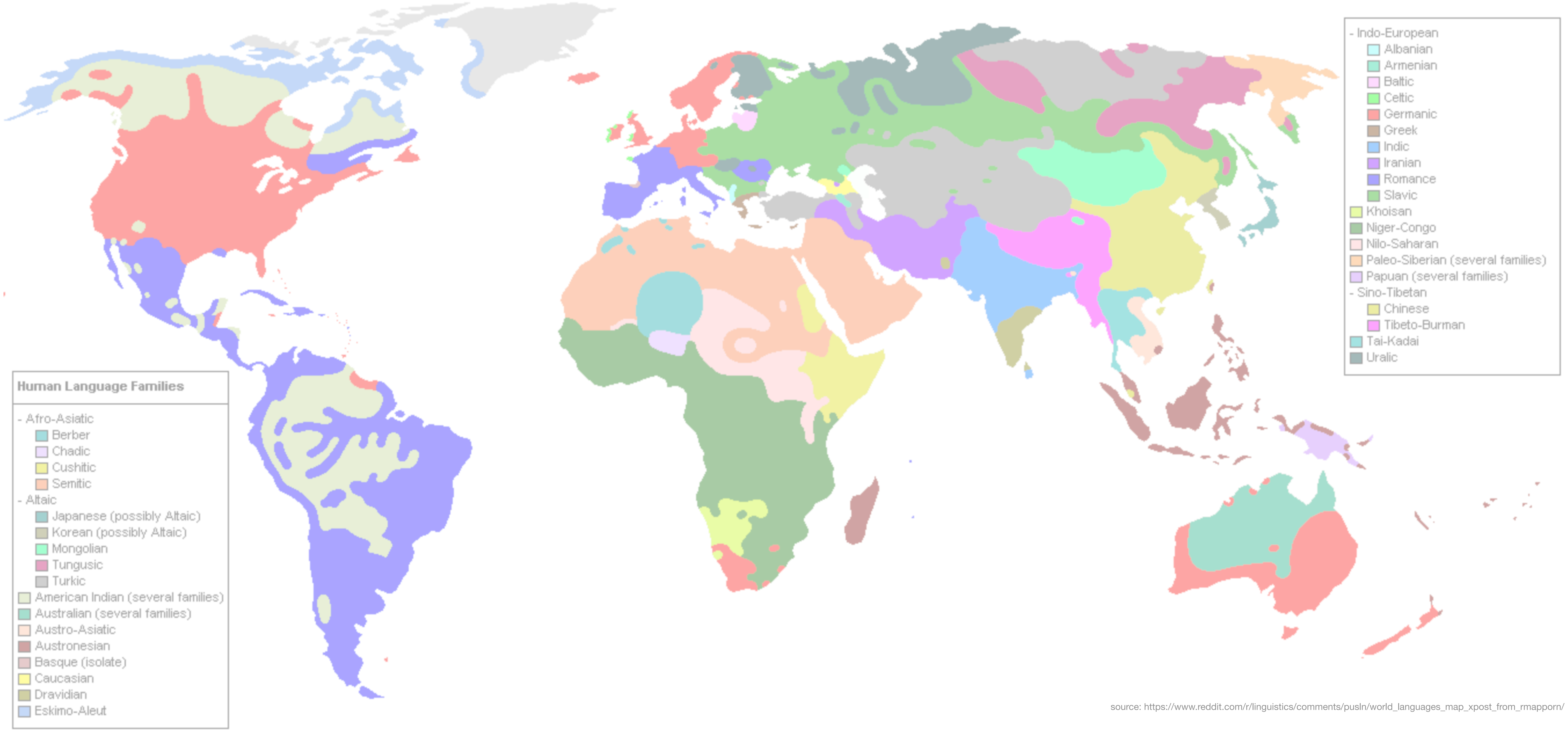


source: https://www.reddit.com/r/linguistics/comments/pusln/world_languages_map_xpost_from_rmapporn/



source: https://www.reddit.com/r/linguistics/comments/pusln/world_languages_map_xpost_from_rmapporn/

Remarkable language diversity!



source: https://www.reddit.com/r/linguistics/comments/pusln/world_languages_map_xpost_from_rmapporn/

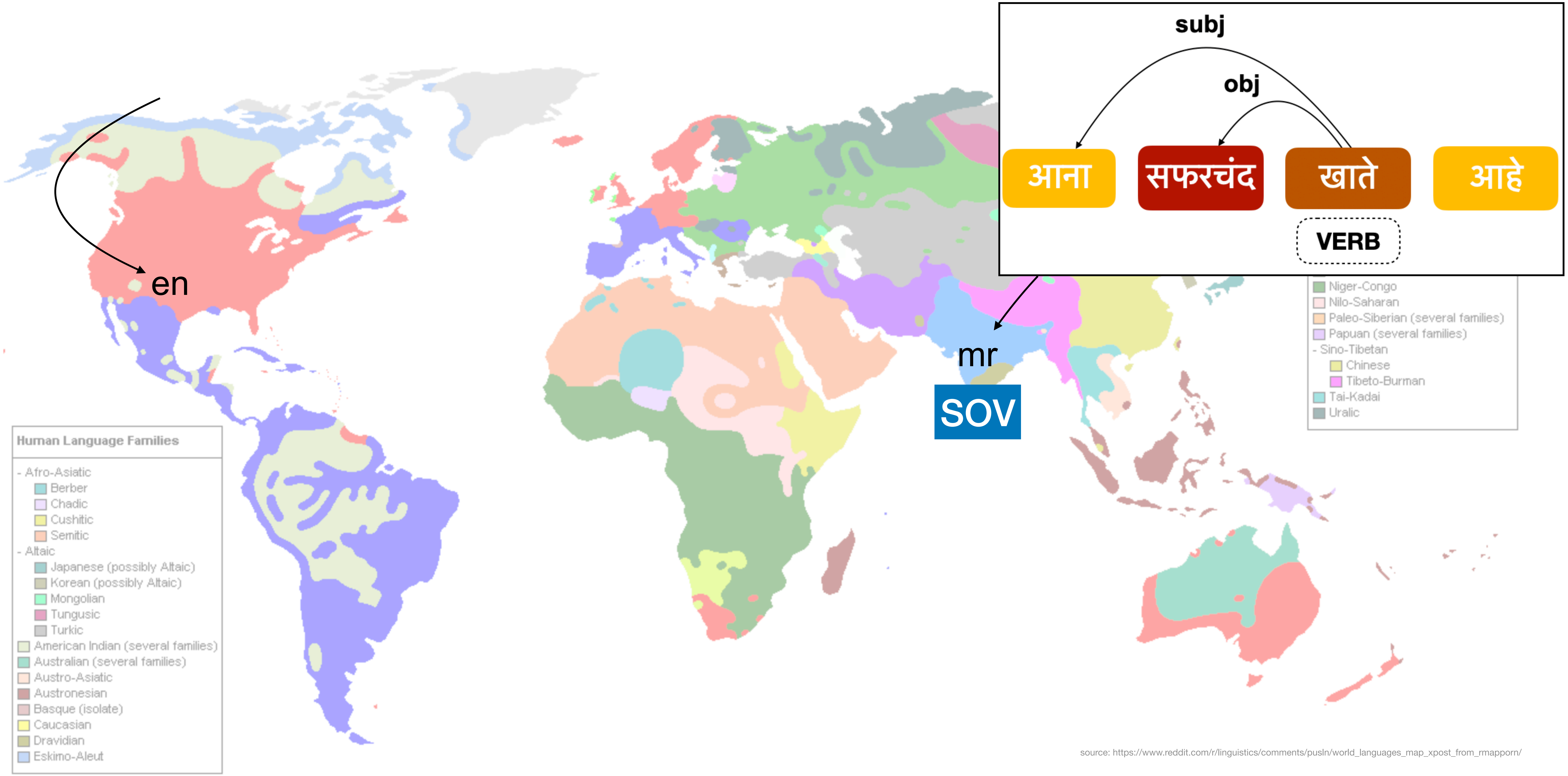
Remarkable language diversity!

Anna is eating an apple

आना सफरचंद खाते आहे

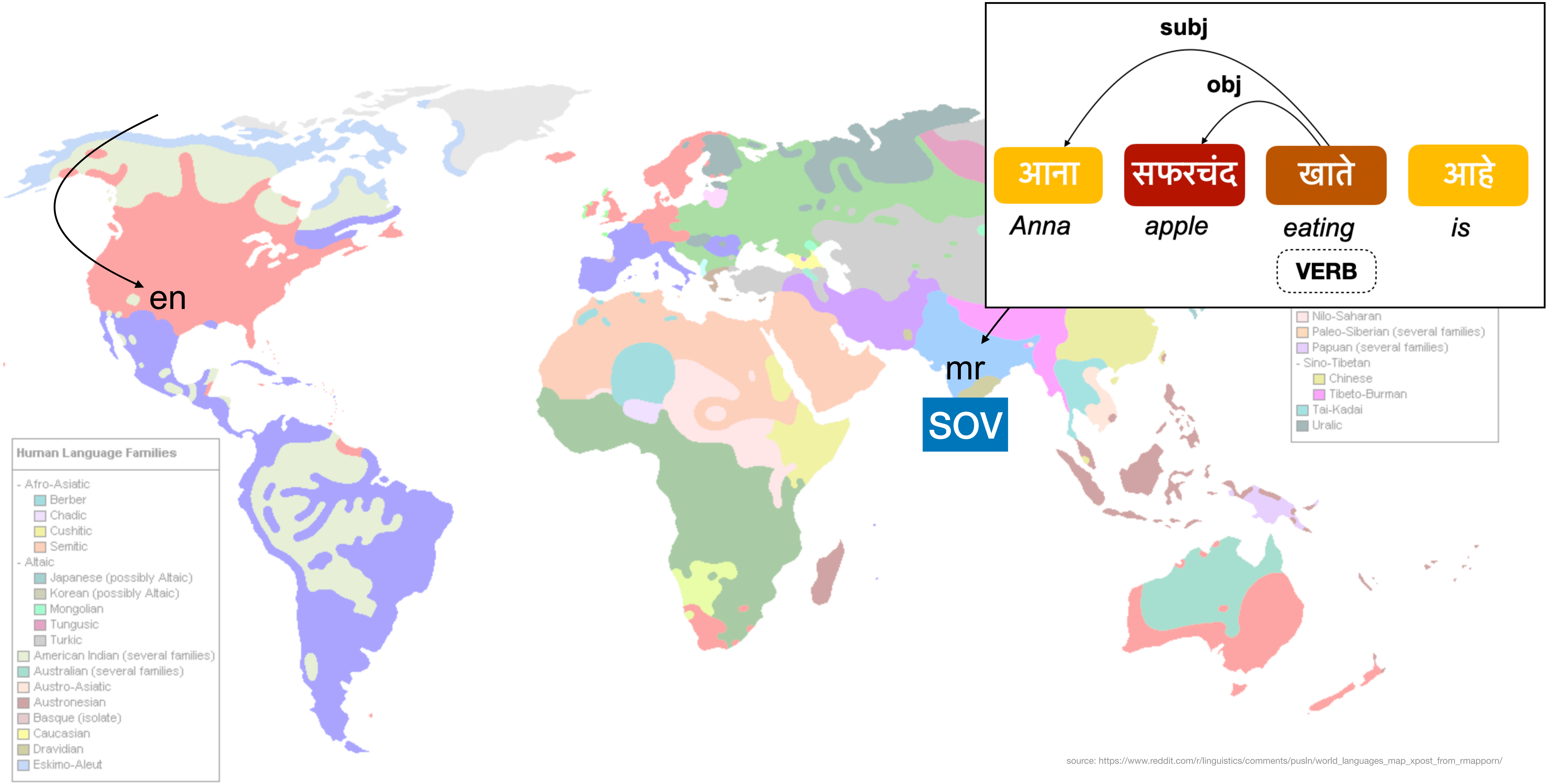
source: https://www.reddit.com/r/linguistics/comments/pusln/world_languages_map_xpost_from_rmapporn/

Remarkable language diversity!



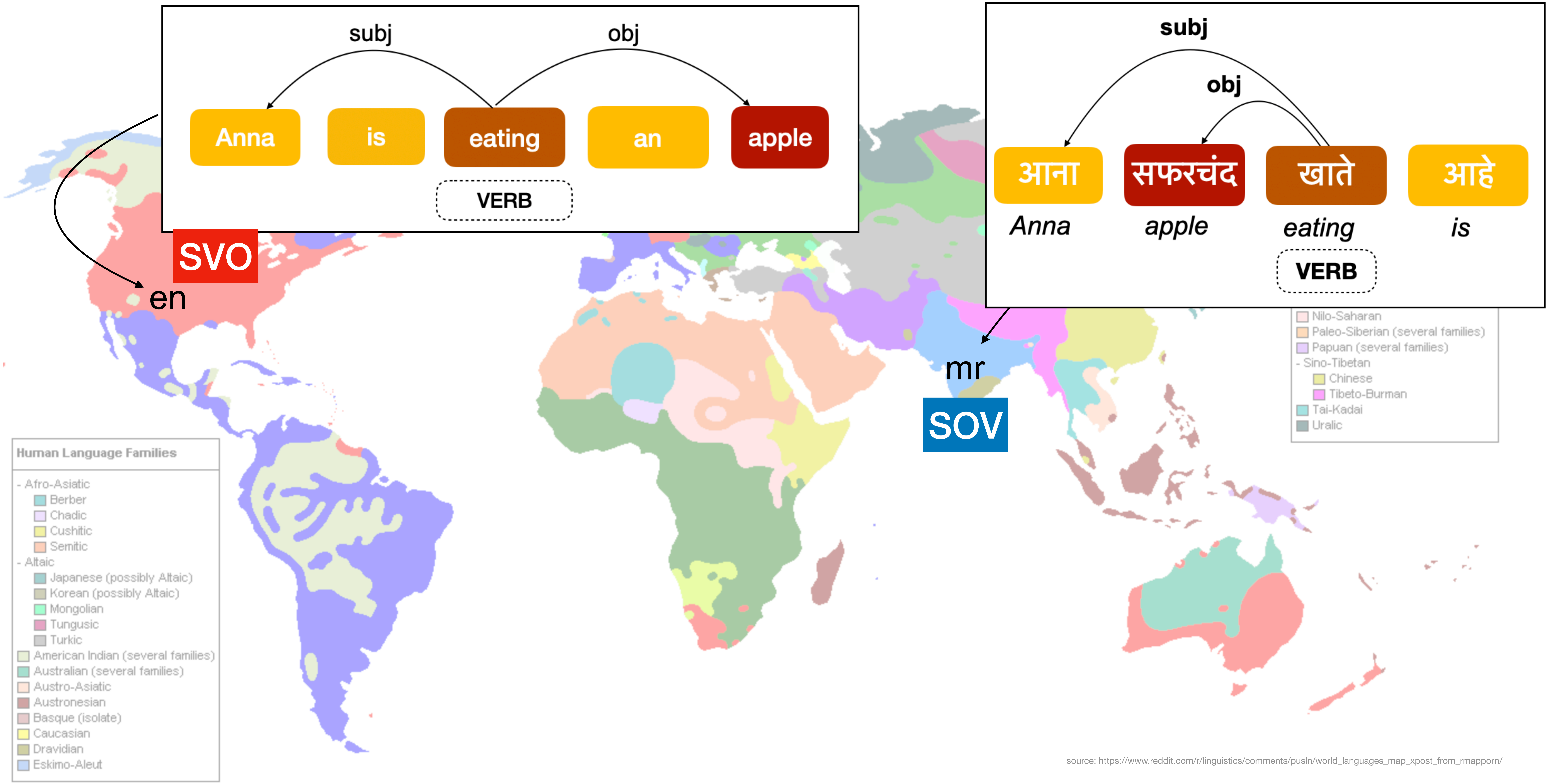
source: https://www.reddit.com/r/linguistics/comments/pusln/world_languages_map_xpost_from_rmapporn/

Remarkable language diversity!



source: https://www.reddit.com/r/linguistics/comments/pusln/world_languages_map_xpost_from_rmapporn/

Remarkable language diversity!

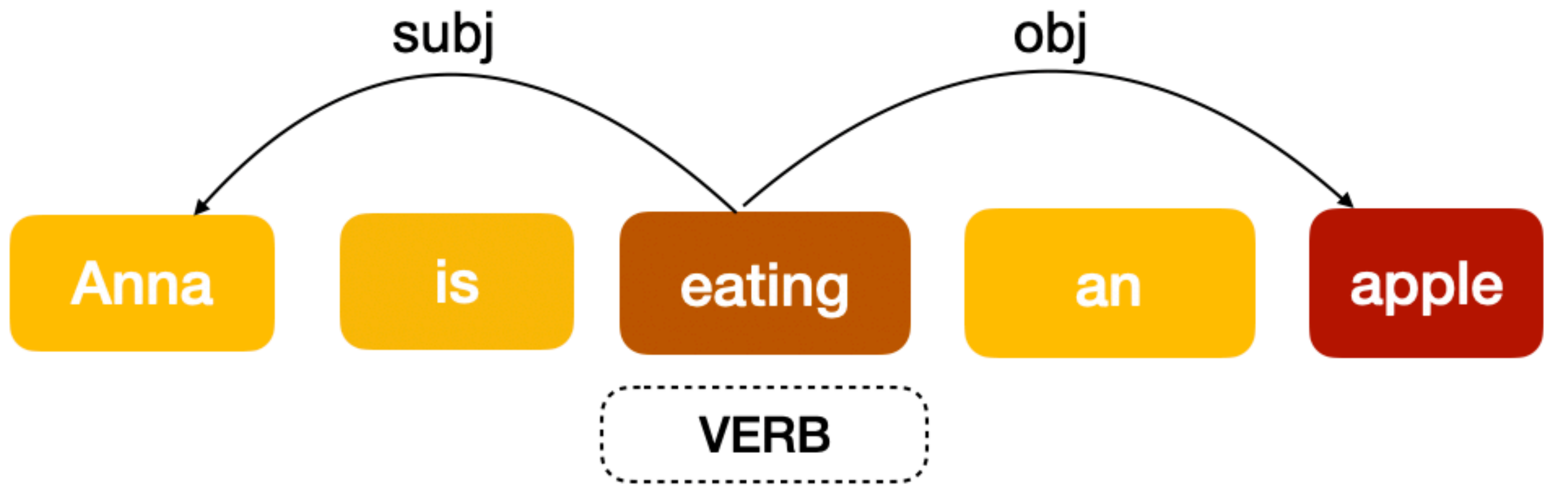


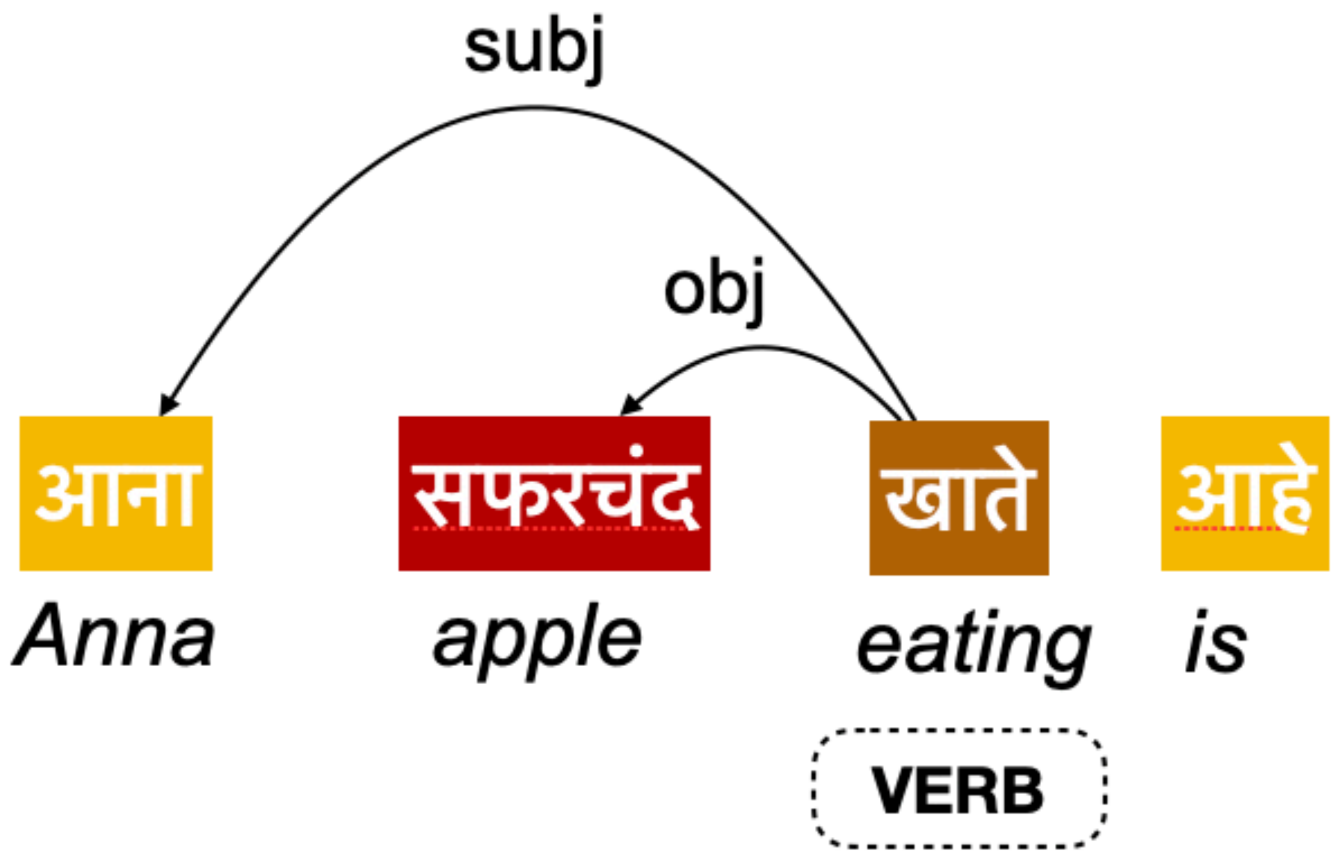
source: https://www.reddit.com/r/linguistics/comments/pusln/world_languages_map_xpost_from_rmapporn/

Remarkable language diversity!

Carnegie Mellon University

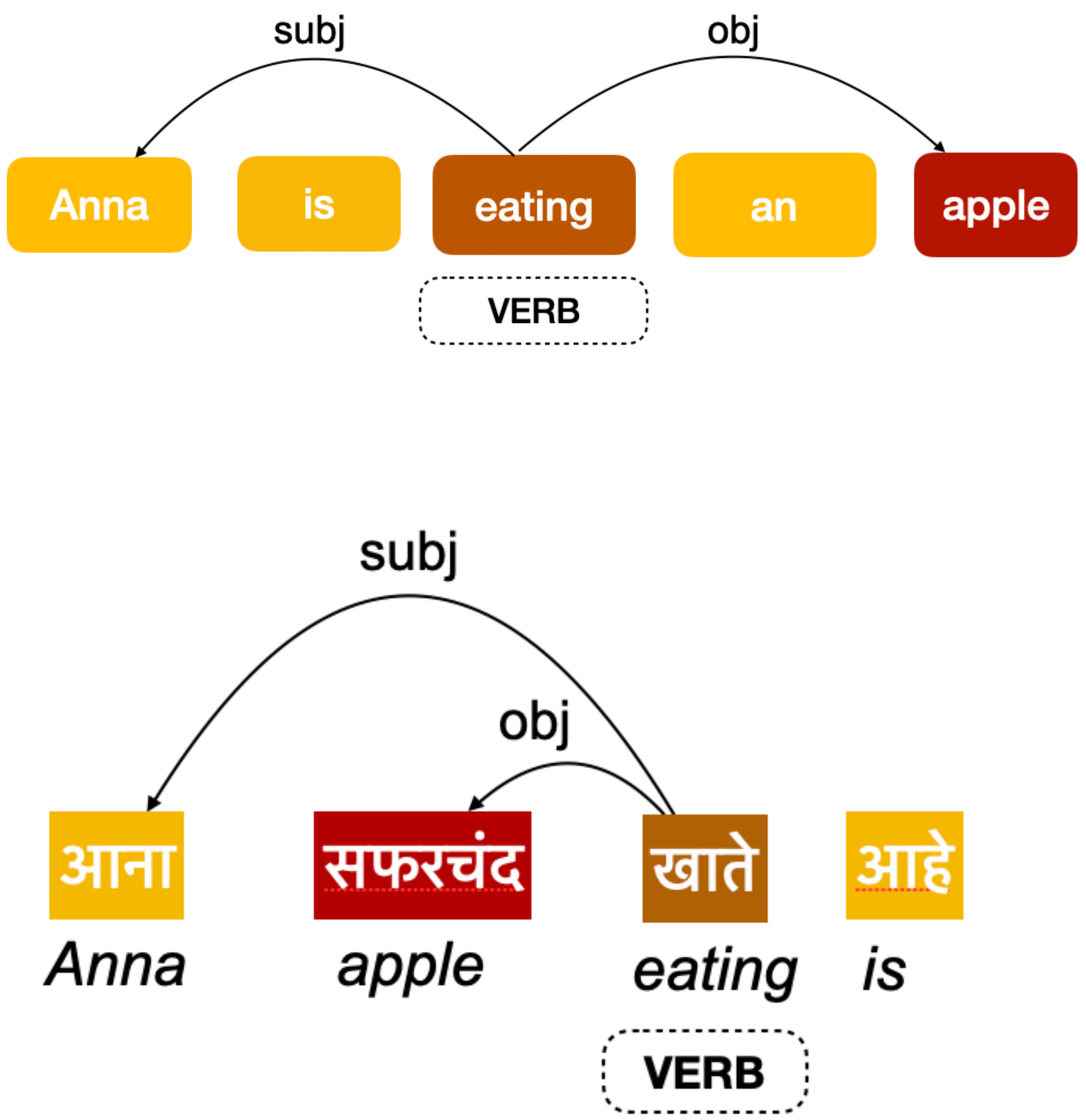
Different Systems Governing Language





Syntax

Different Systems Governing Language



subj obj

Anna is eating an apple

VERB

subj obj

आना सफरचंद खाते आहे

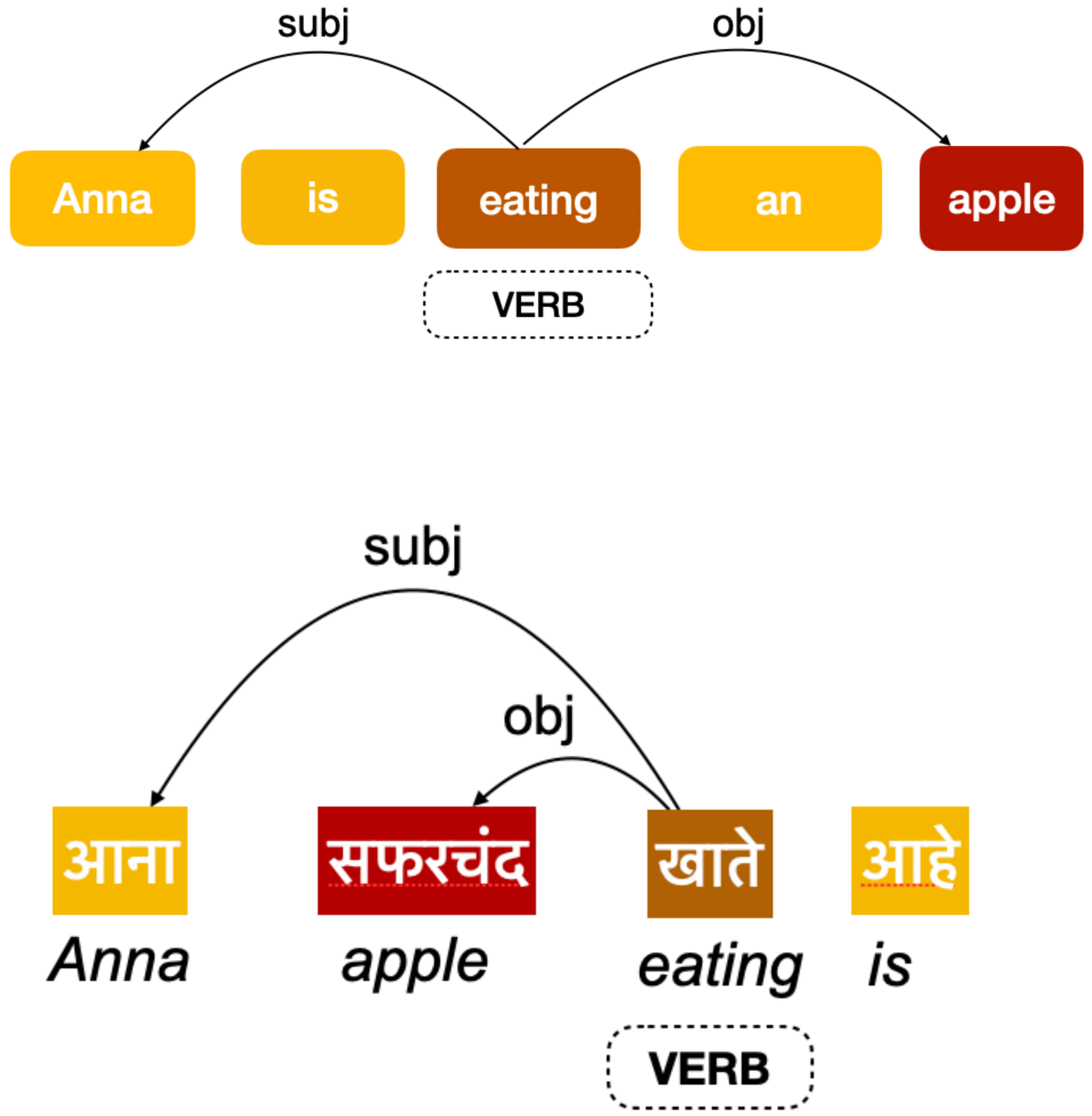
Anna apple eating is

VERB

Syntax

Morphology

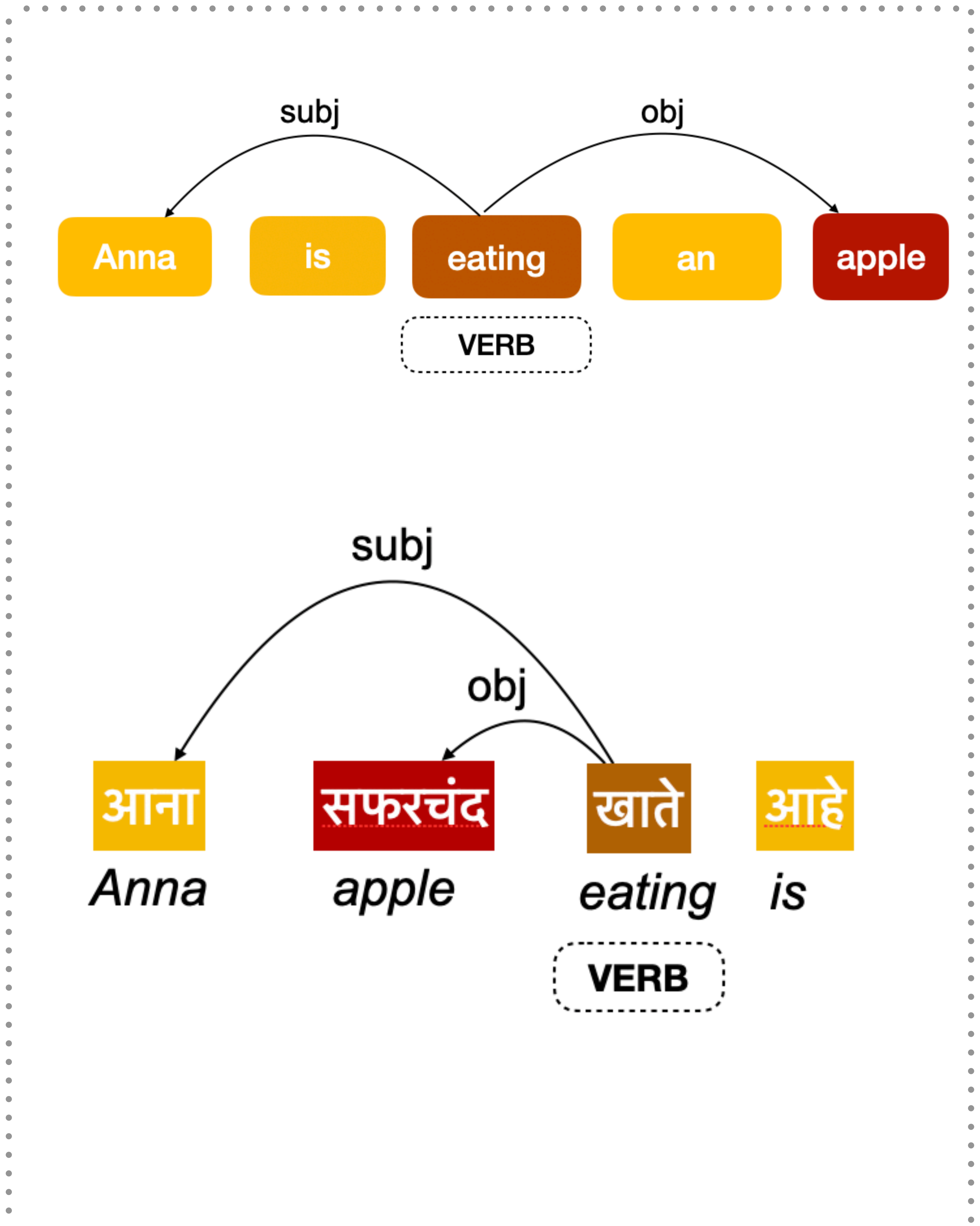
Different Systems Governing Language



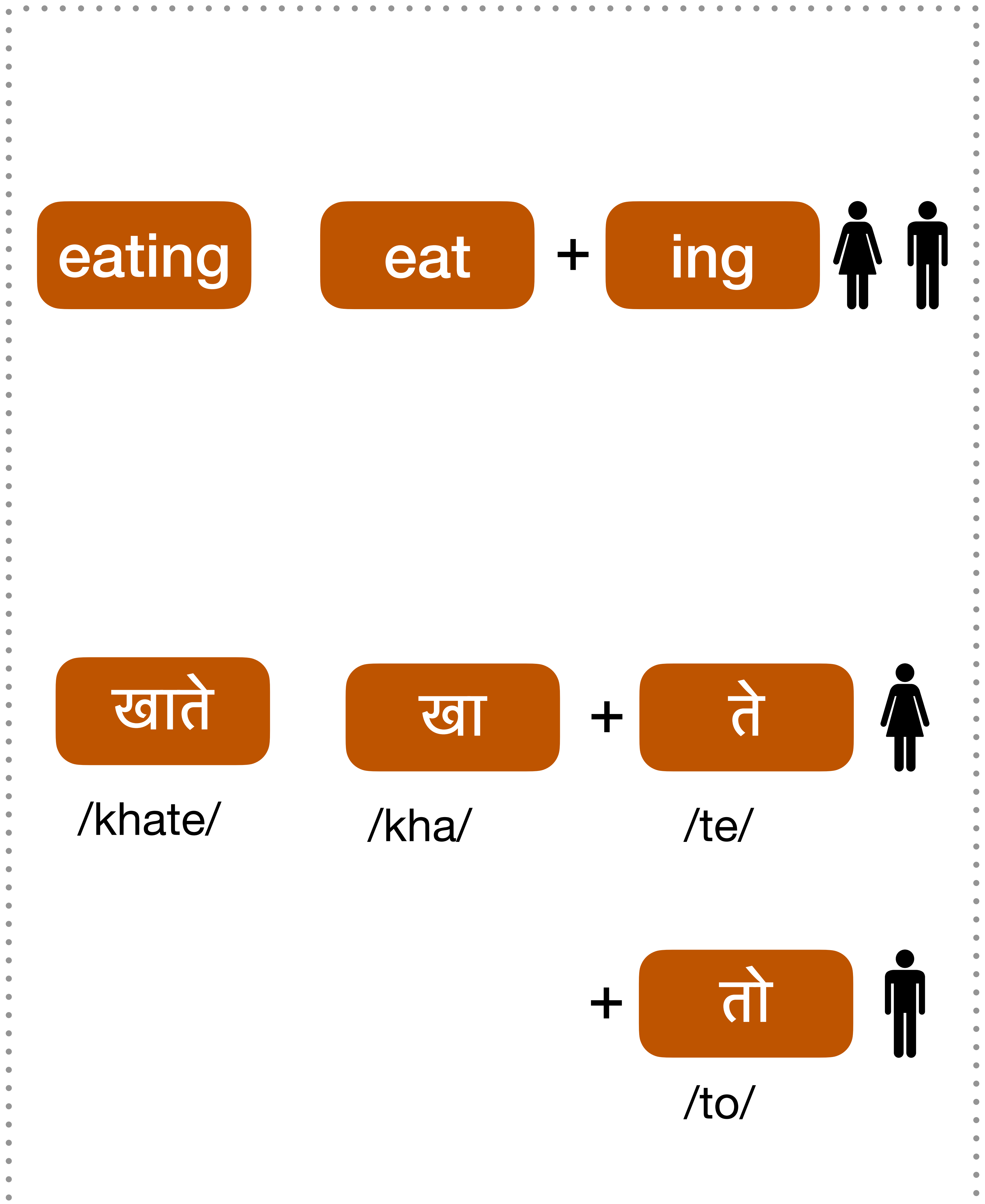
Syntax

Morphology

Different Systems Governing Language

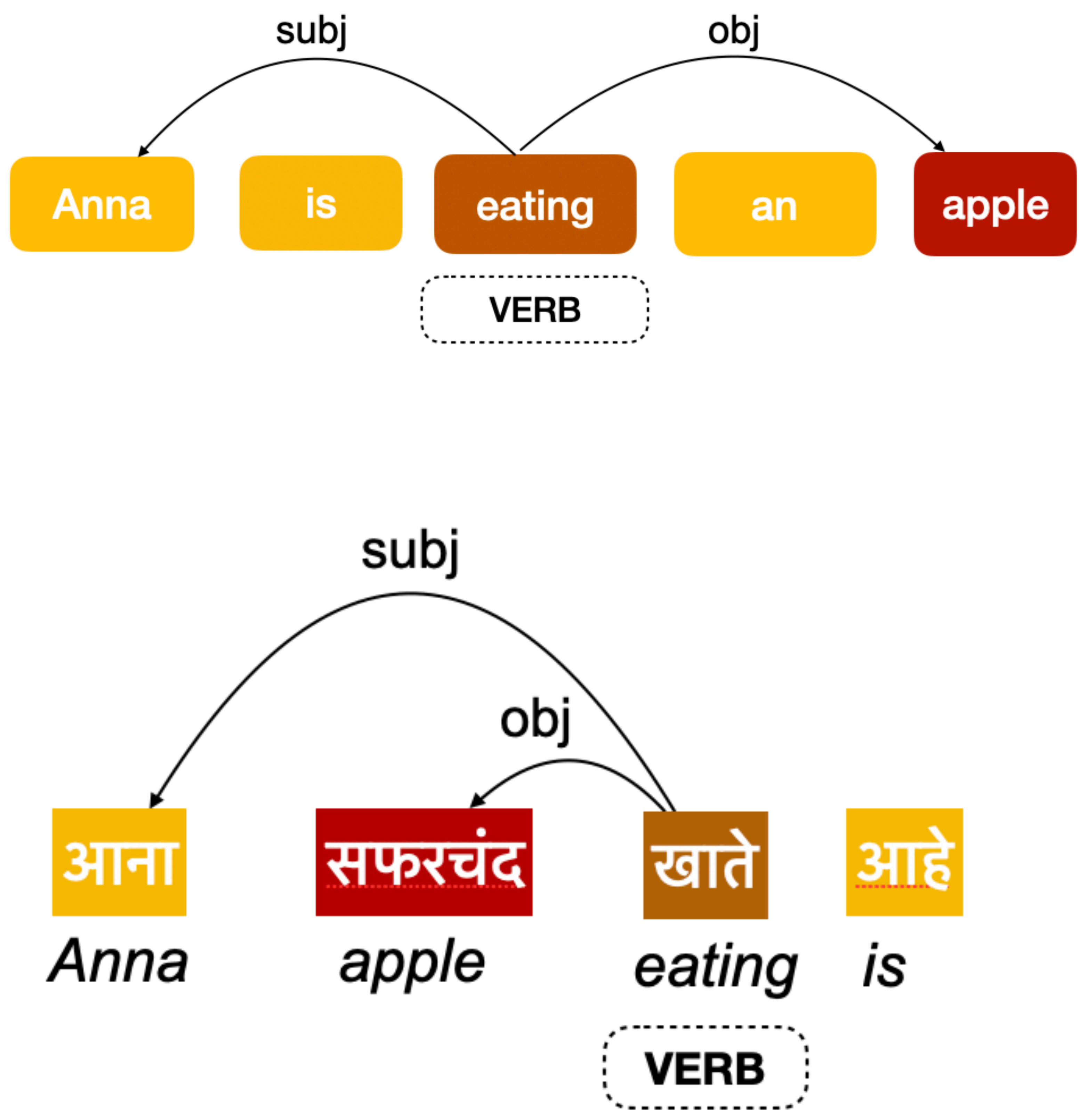


Syntax

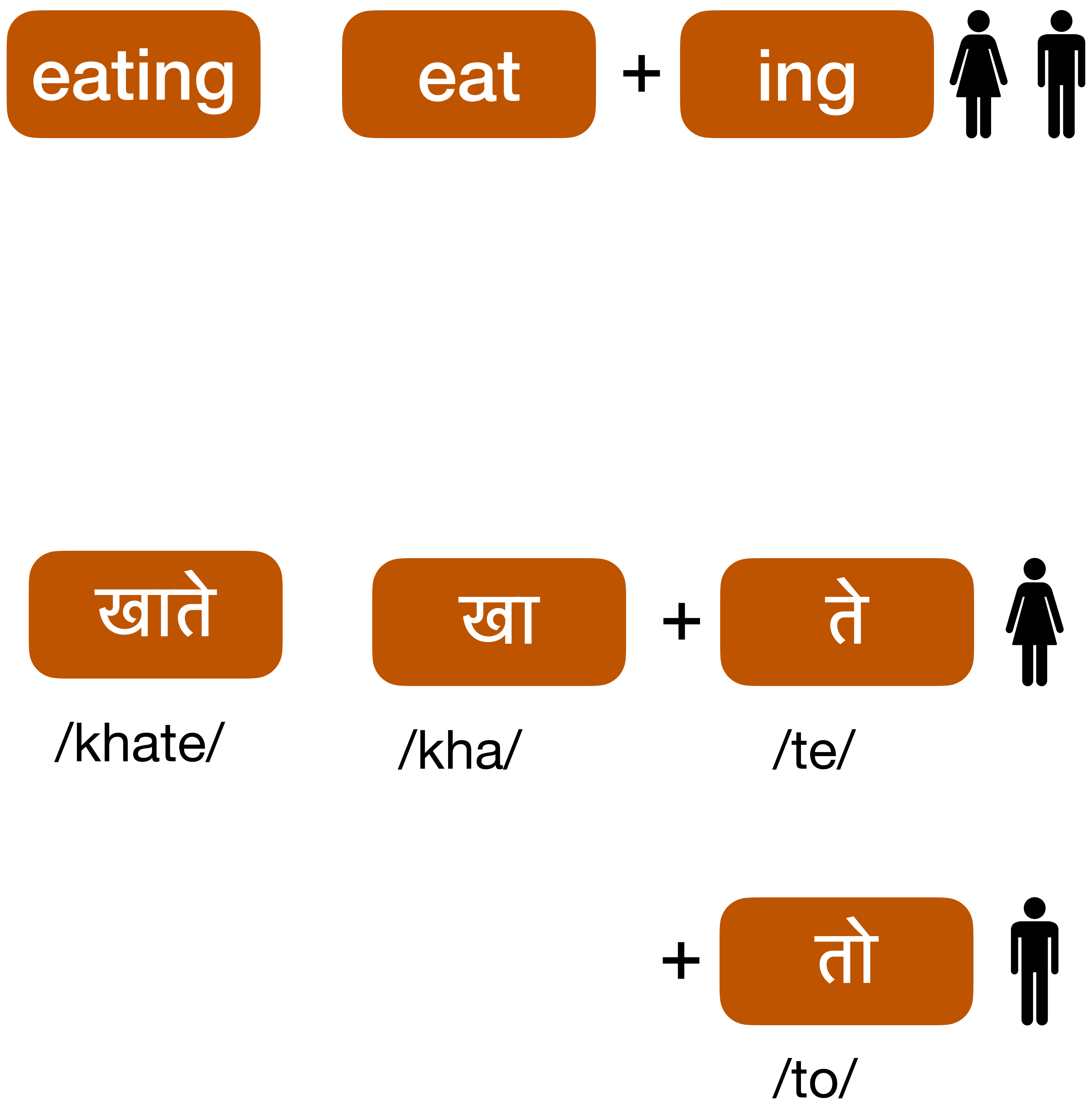


Morphology

Different Systems Governing Language



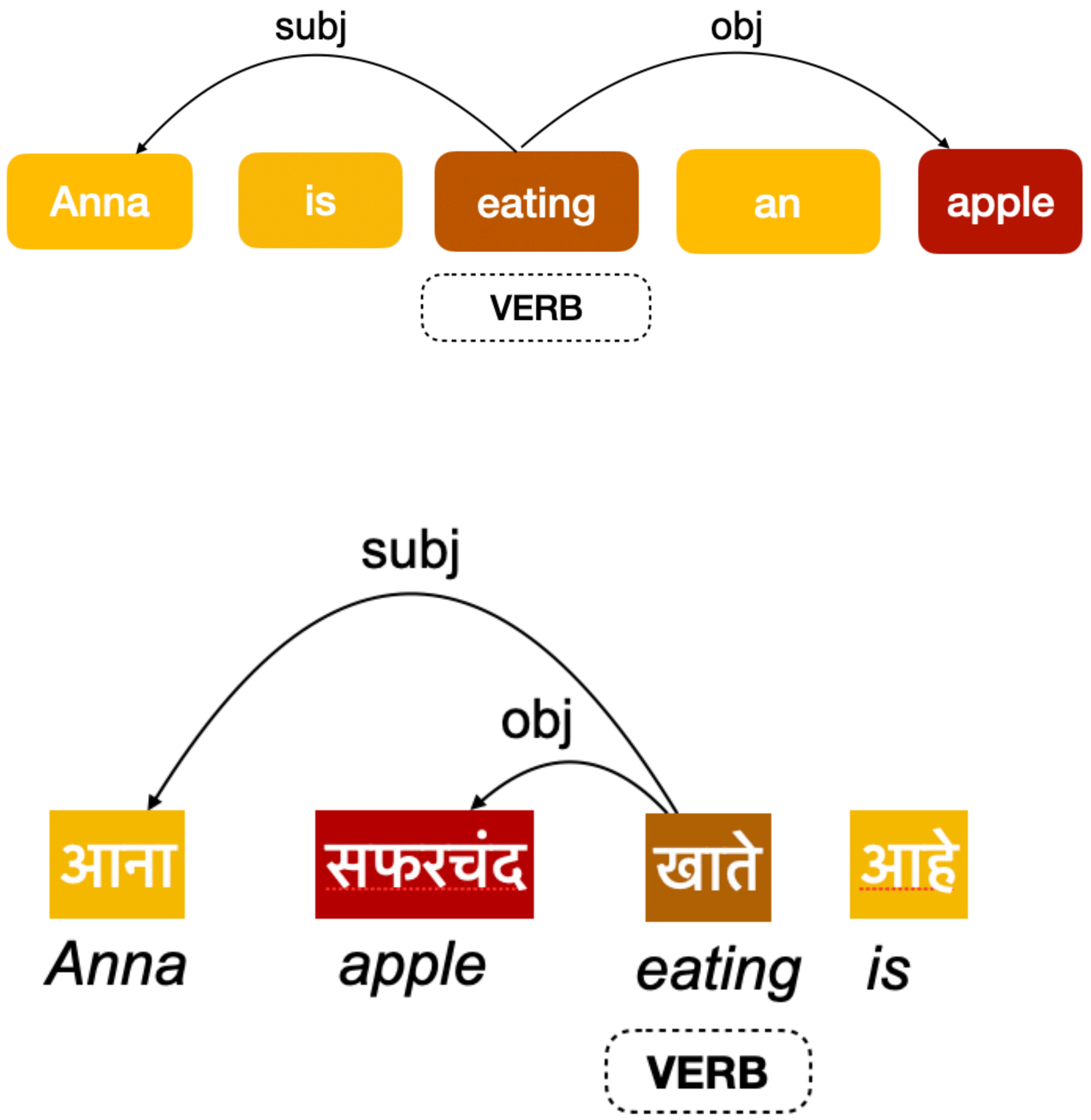
Syntax



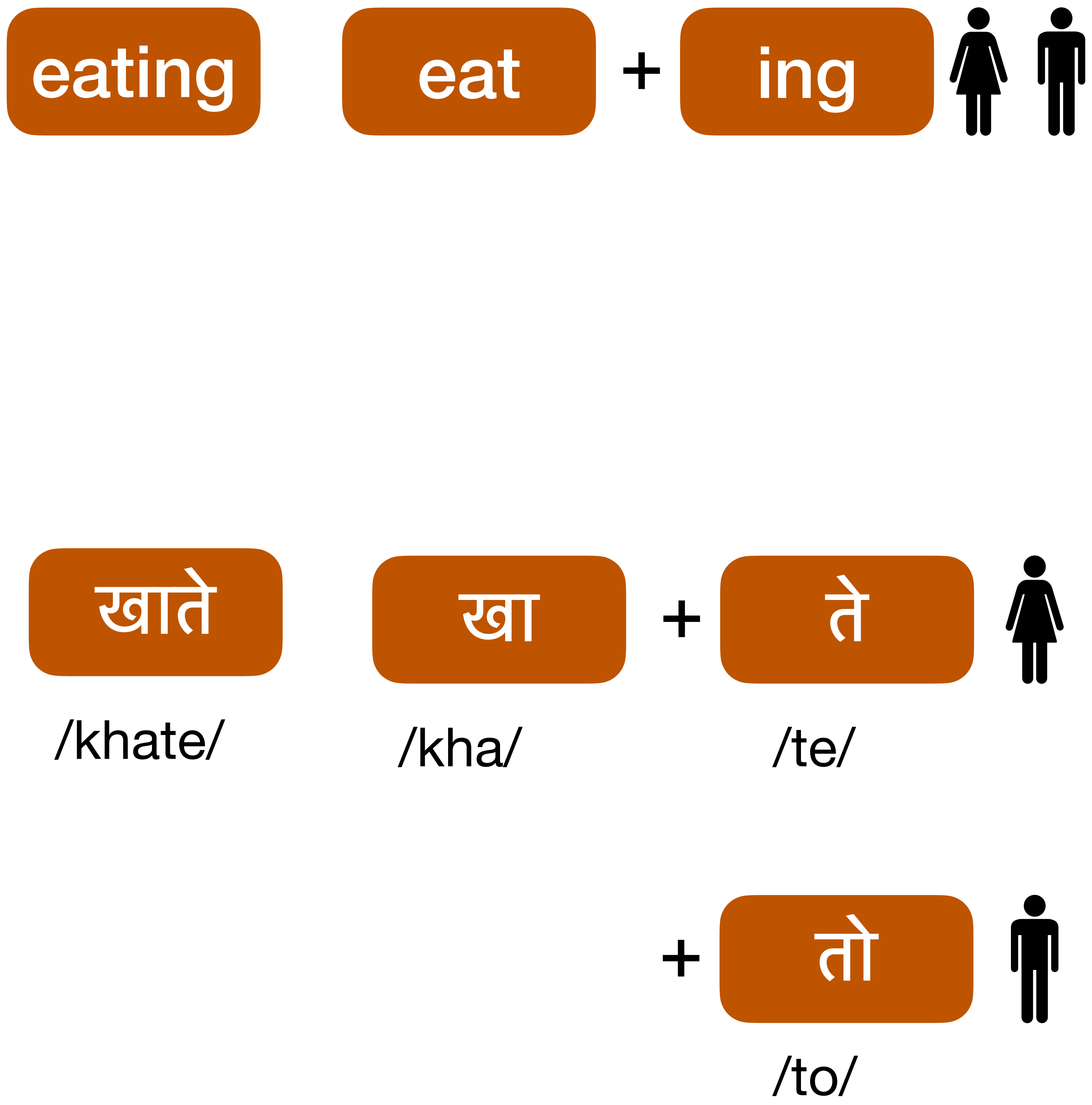
Morphology

Semantics

Different Systems Governing Language



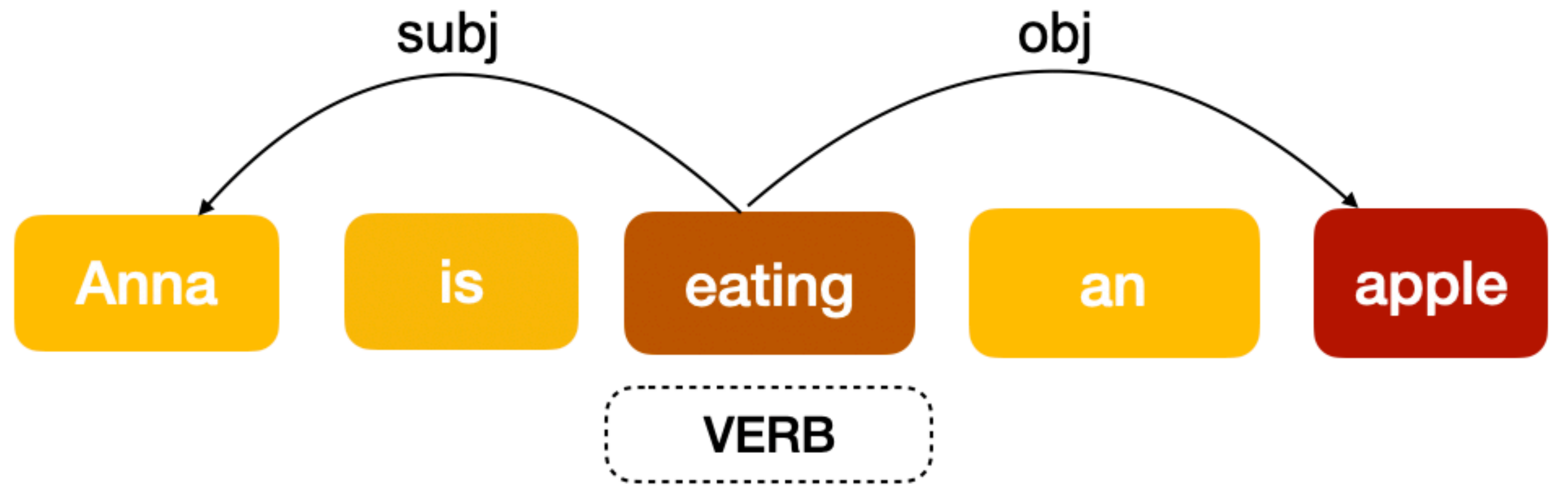
Syntax

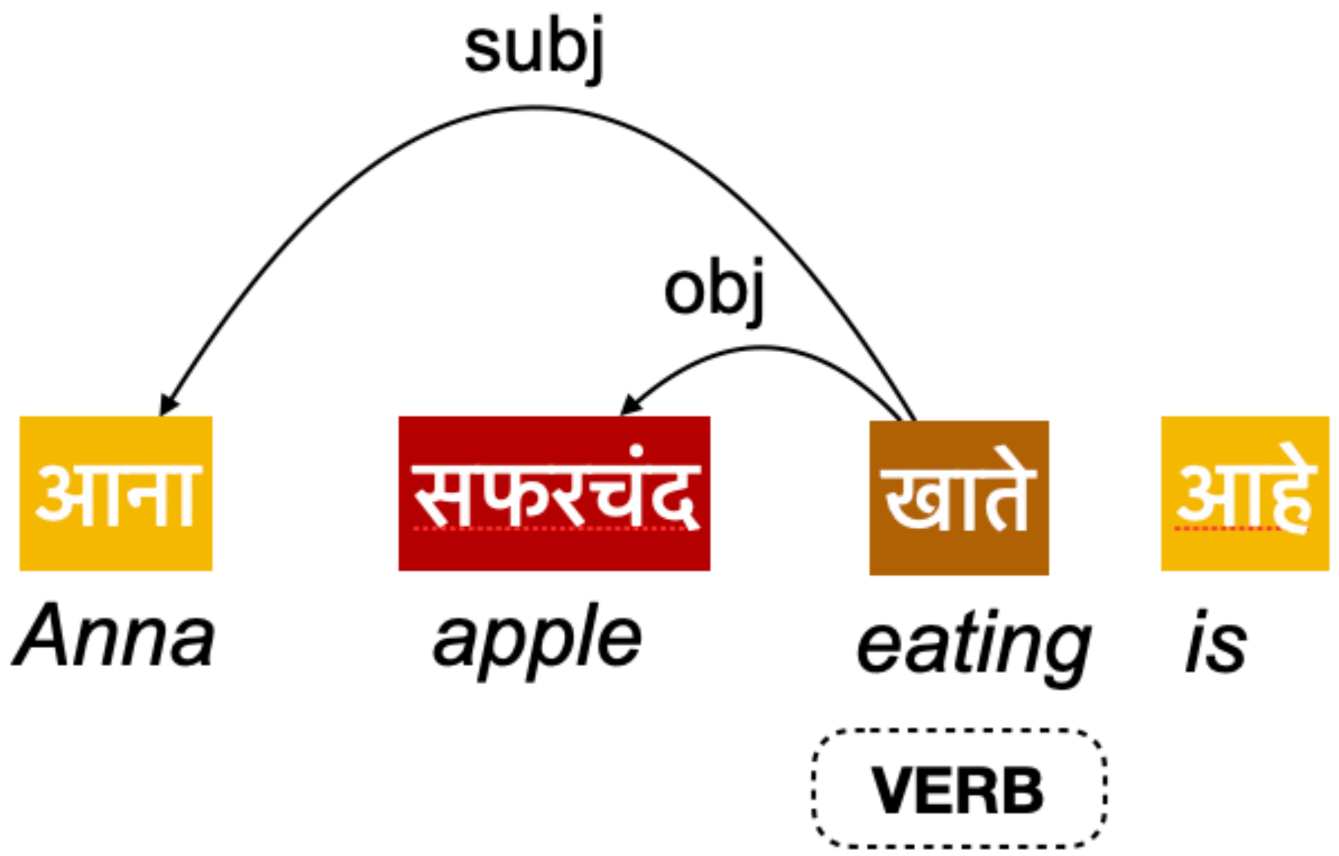


Morphology

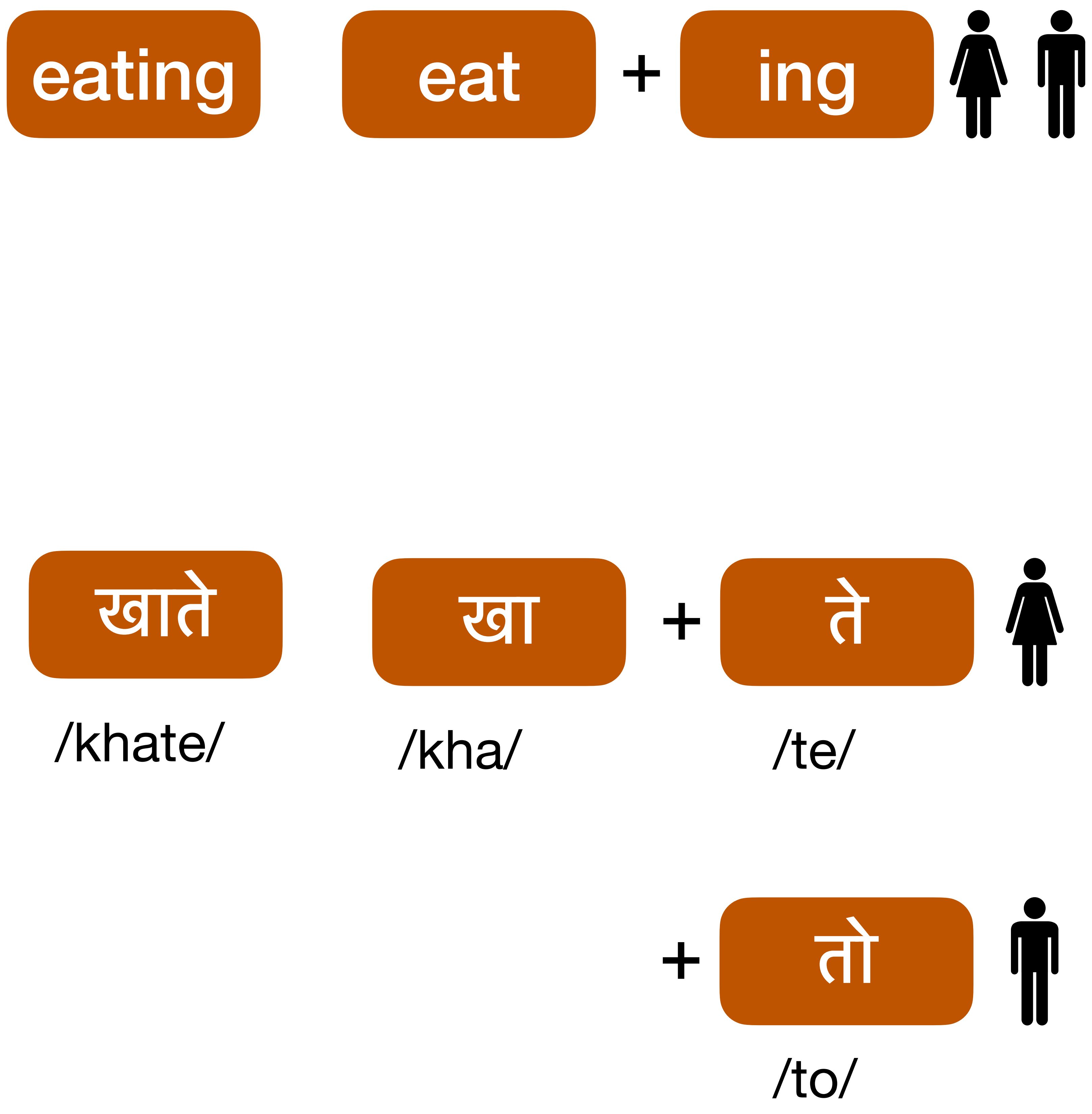
Semantics

Different Systems Governing Language





Syntax



Morphology

तांदूळ
tandul
'raw rice'

भात
bhaat
'cooked rice'

Semantics

Describing Systems of Language

Describing Systems of Language

Describing Systems of Language

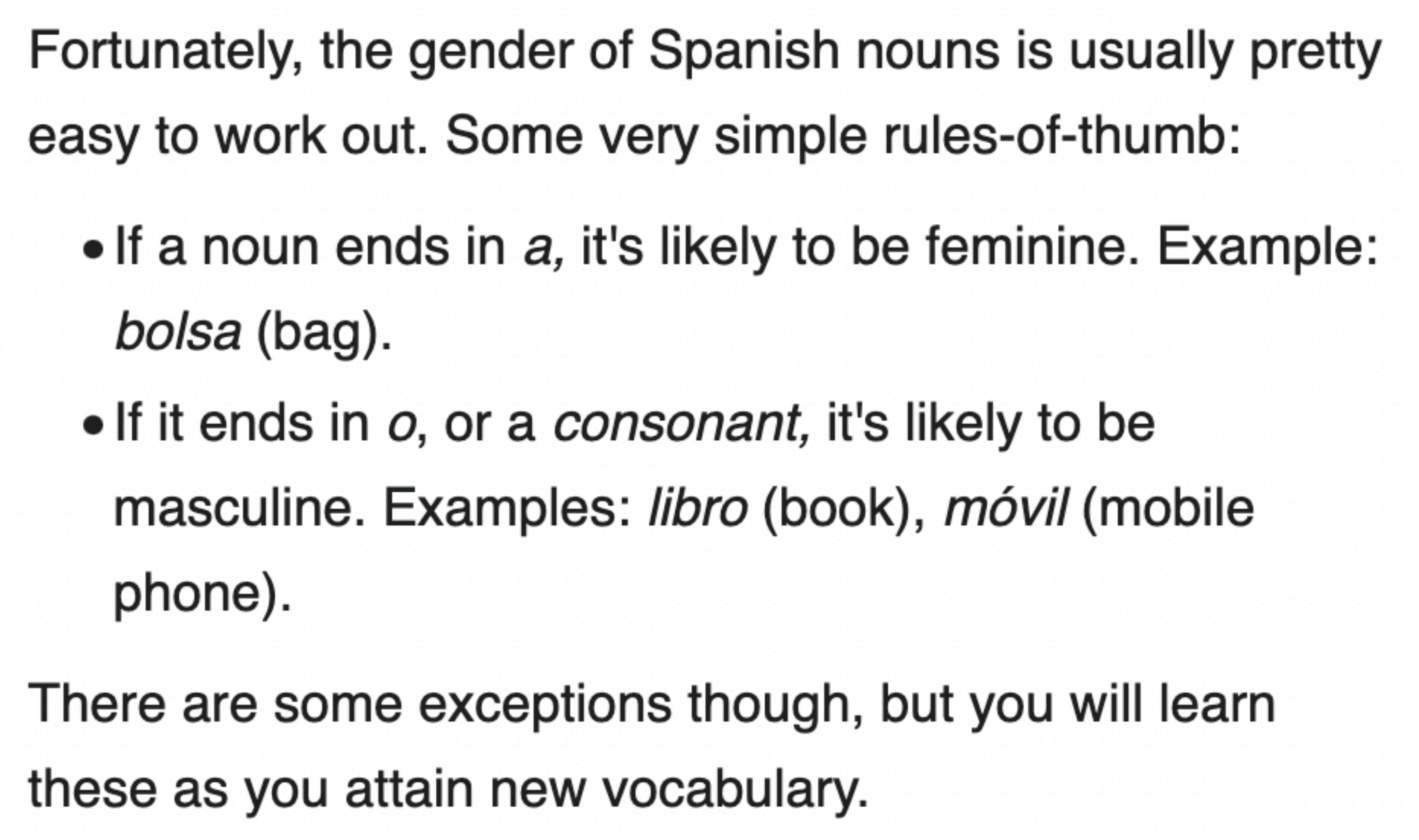
WIKIBOOKS
Open books for an open world

Main Page
Help

Book
Discussion

Spanish

Describing Systems of Language

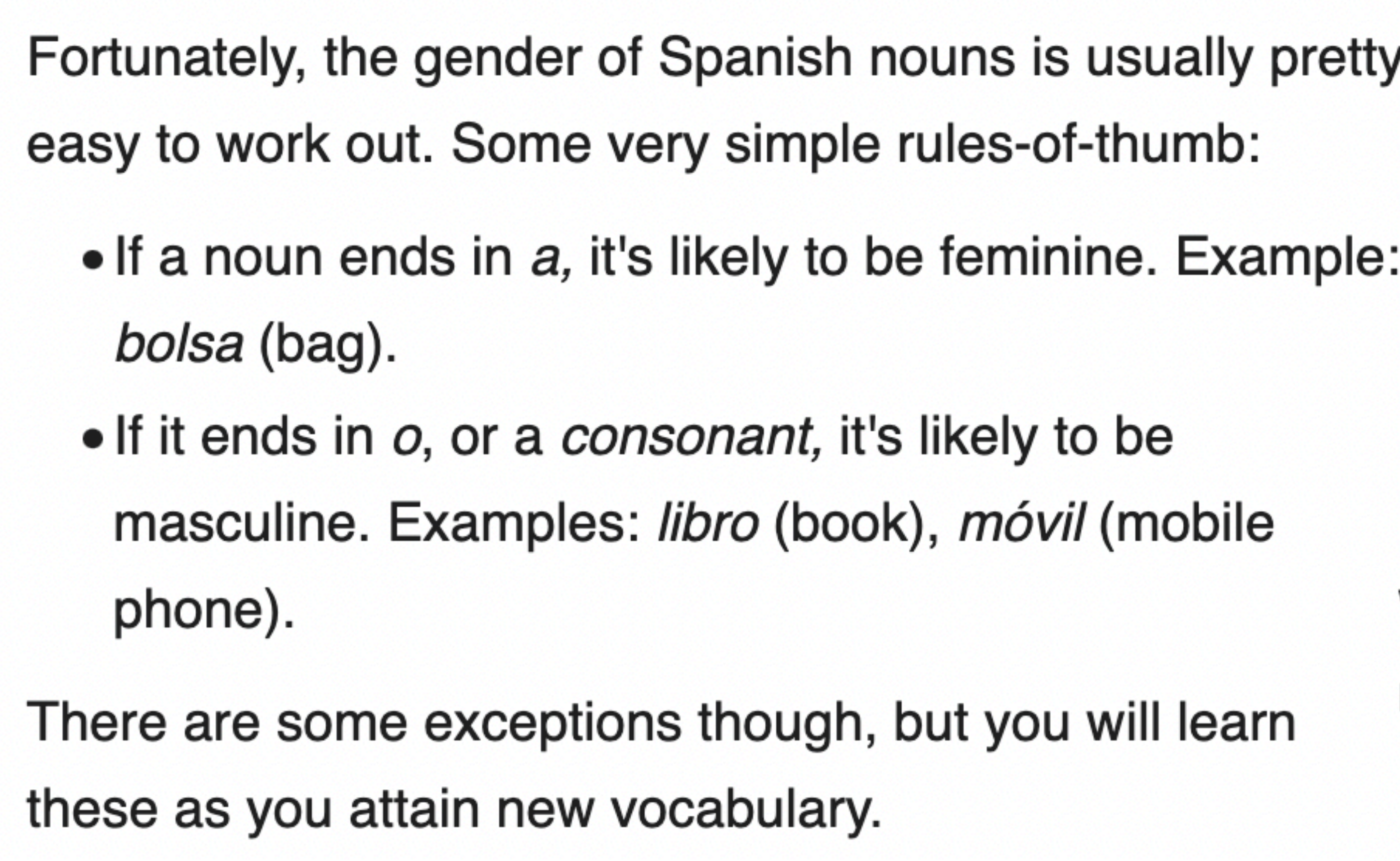


Fortunately, the gender of Spanish nouns is usually pretty easy to work out. Some very simple rules-of-thumb:

- If a noun ends in *a,* it's likely to be feminine. Example: *bolsa* (bag).
- If it ends in *o*, or a *consonant,* it's likely to be masculine. Examples: *libro* (book), *móvil* (mobile phone).

There are some exceptions though, but you will learn these as you attain new vocabulary.

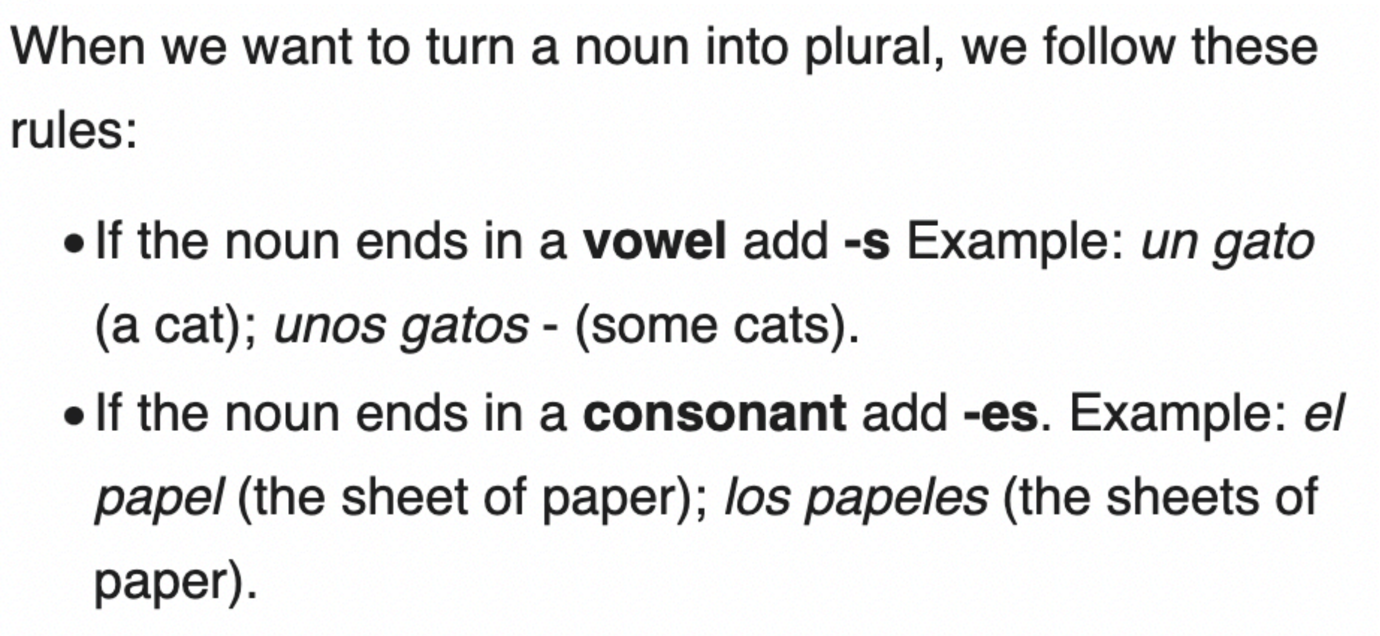
Describing Systems of Language



Fortunately, the gender of Spanish nouns is usually pretty easy to work out. Some very simple rules-of-thumb:

- If a noun ends in *a,* it's likely to be feminine. Example: *bolsa* (bag).
- If it ends in *o*, or a *consonant,* it's likely to be masculine. Examples: *libro* (book), *móvil* (mobile phone).

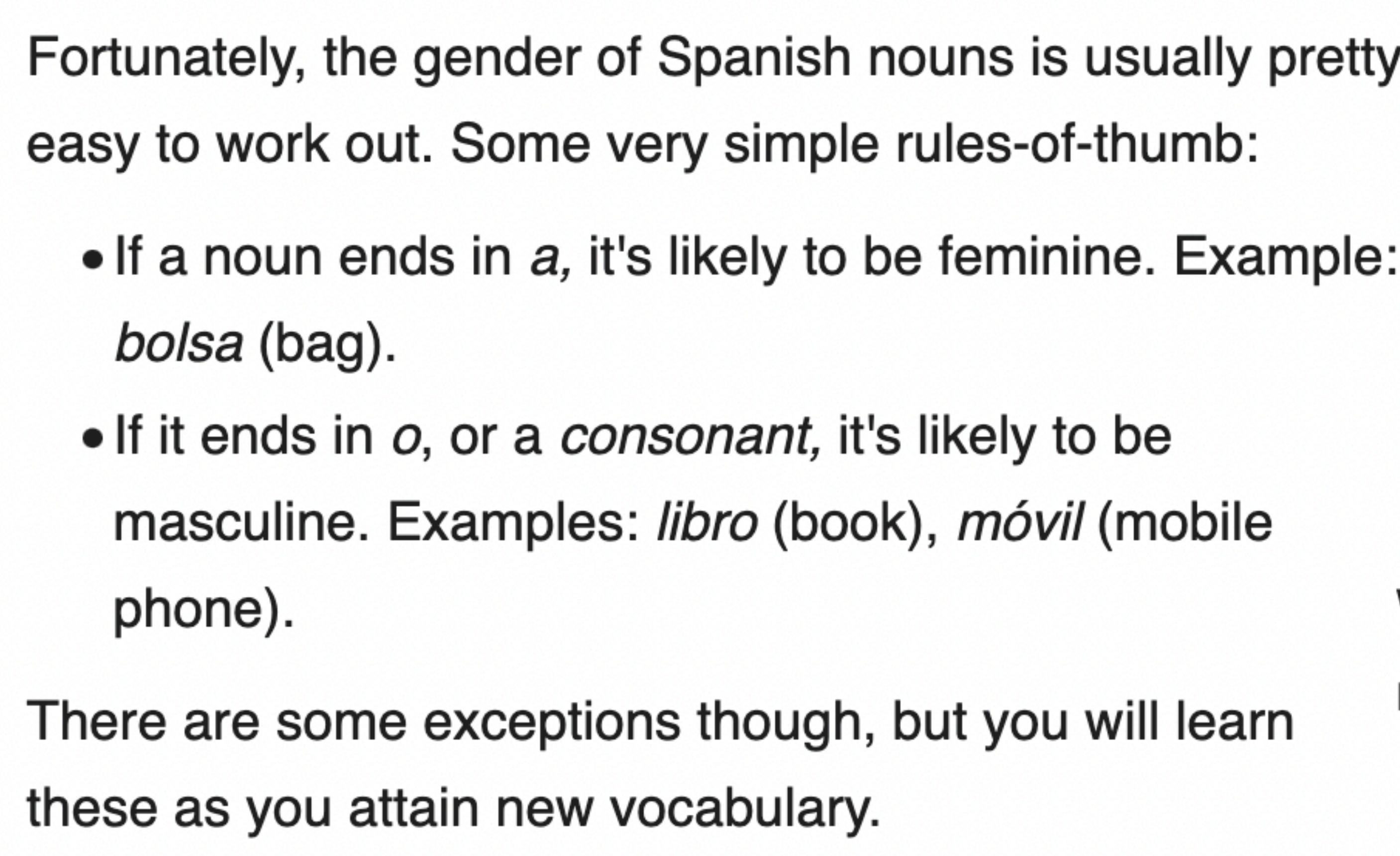
There are some exceptions though, but you will learn these as you attain new vocabulary.



When we want to turn a noun into plural, we follow these rules:

- If the noun ends in a **vowel** add **-s** Example: *un gato* (a cat); *unos gatos* - (some cats).
- If the noun ends in a **consonant** add **-es**. Example: *el papel* (the sheet of paper); *los papeles* (the sheets of paper).

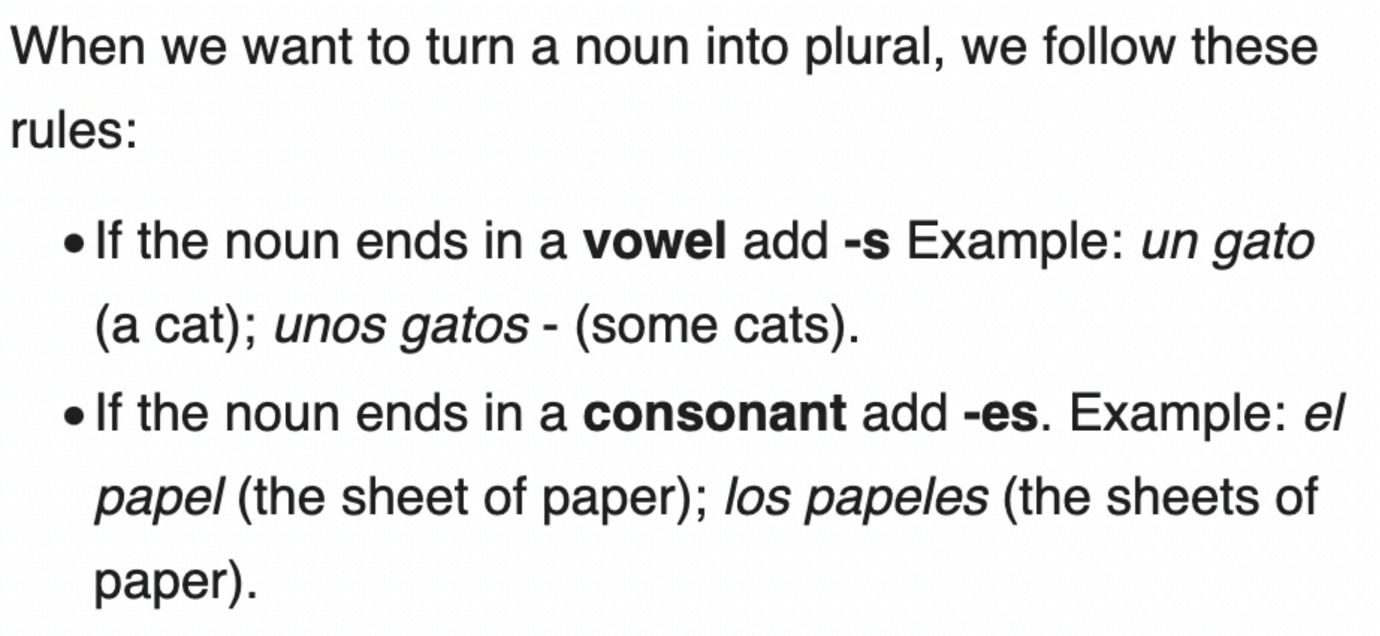
Describing Systems of Language



Fortunately, the gender of Spanish nouns is usually pretty easy to work out. Some very simple rules-of-thumb:

- If a noun ends in *a,* it's likely to be feminine. Example: *bolsa* (bag).
- If it ends in *o*, or a *consonant,* it's likely to be masculine. Examples: *libro* (book), *móvil* (mobile phone).

There are some exceptions though, but you will learn these as you attain new vocabulary.



When we want to turn a noun into plural, we follow these rules:

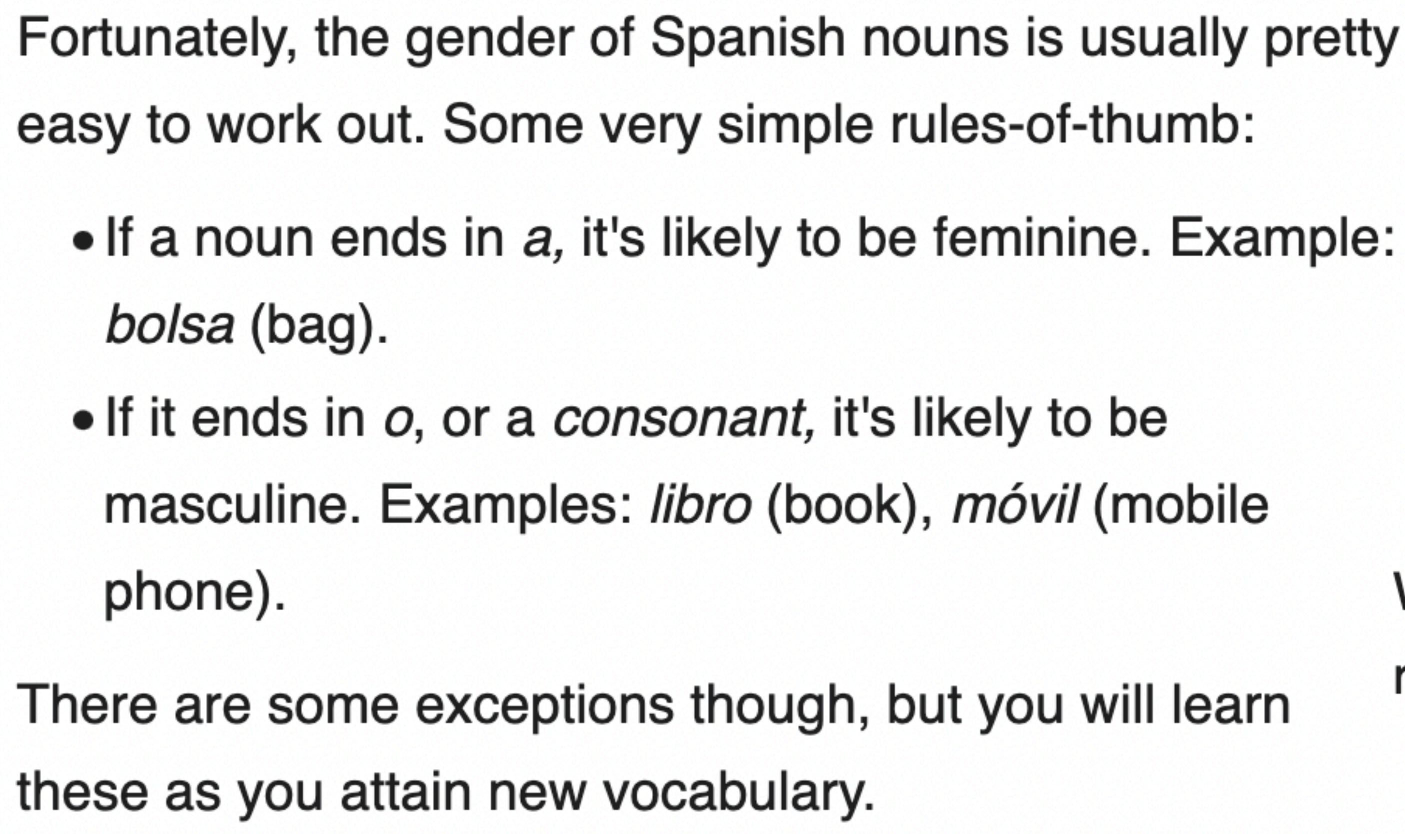
- If the noun ends in a **vowel** add **-s** Example: *un gato* (a cat); *unos gatos* - (some cats).
- If the noun ends in a **consonant** add **-es**. Example: *el papel* (the sheet of paper); *los papeles* (the sheets of paper).

Language Learning

Language Documentation

Describing Systems of Language

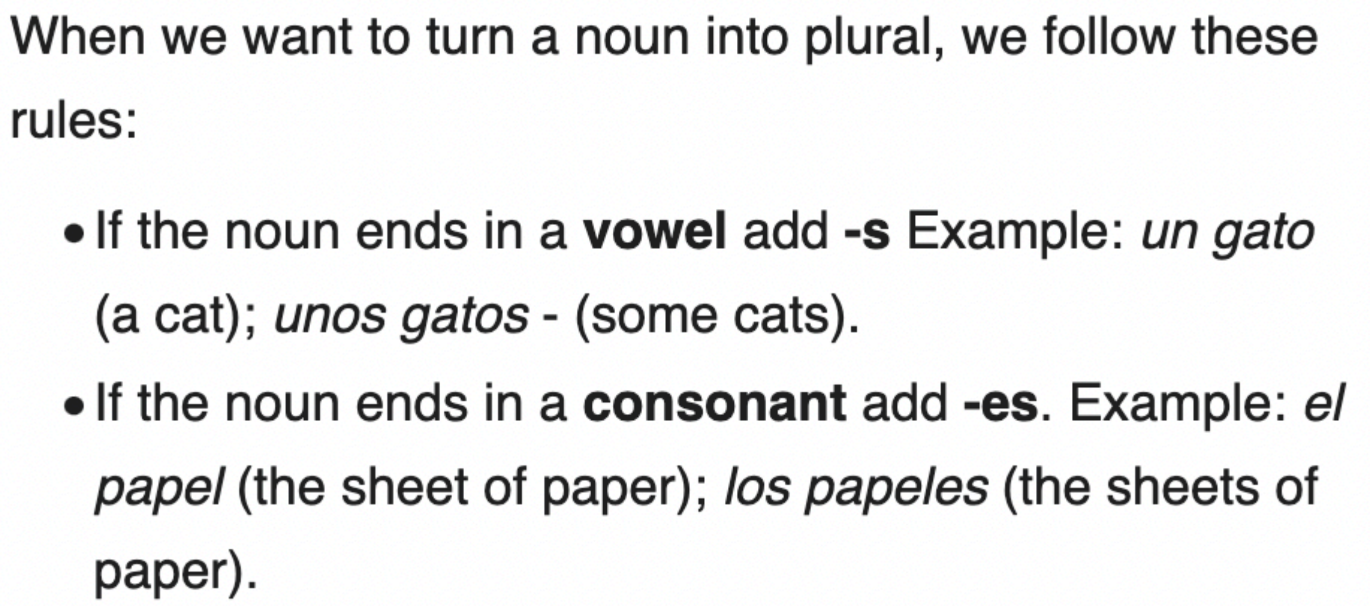
Spanish



Fortunately, the gender of Spanish nouns is usually pretty easy to work out. Some very simple rules-of-thumb:

- If a noun ends in *a,* it's likely to be feminine. Example: *bolsa* (bag).
- If it ends in *o*, or a *consonant,* it's likely to be masculine. Examples: *libro* (book), *móvil* (mobile phone).

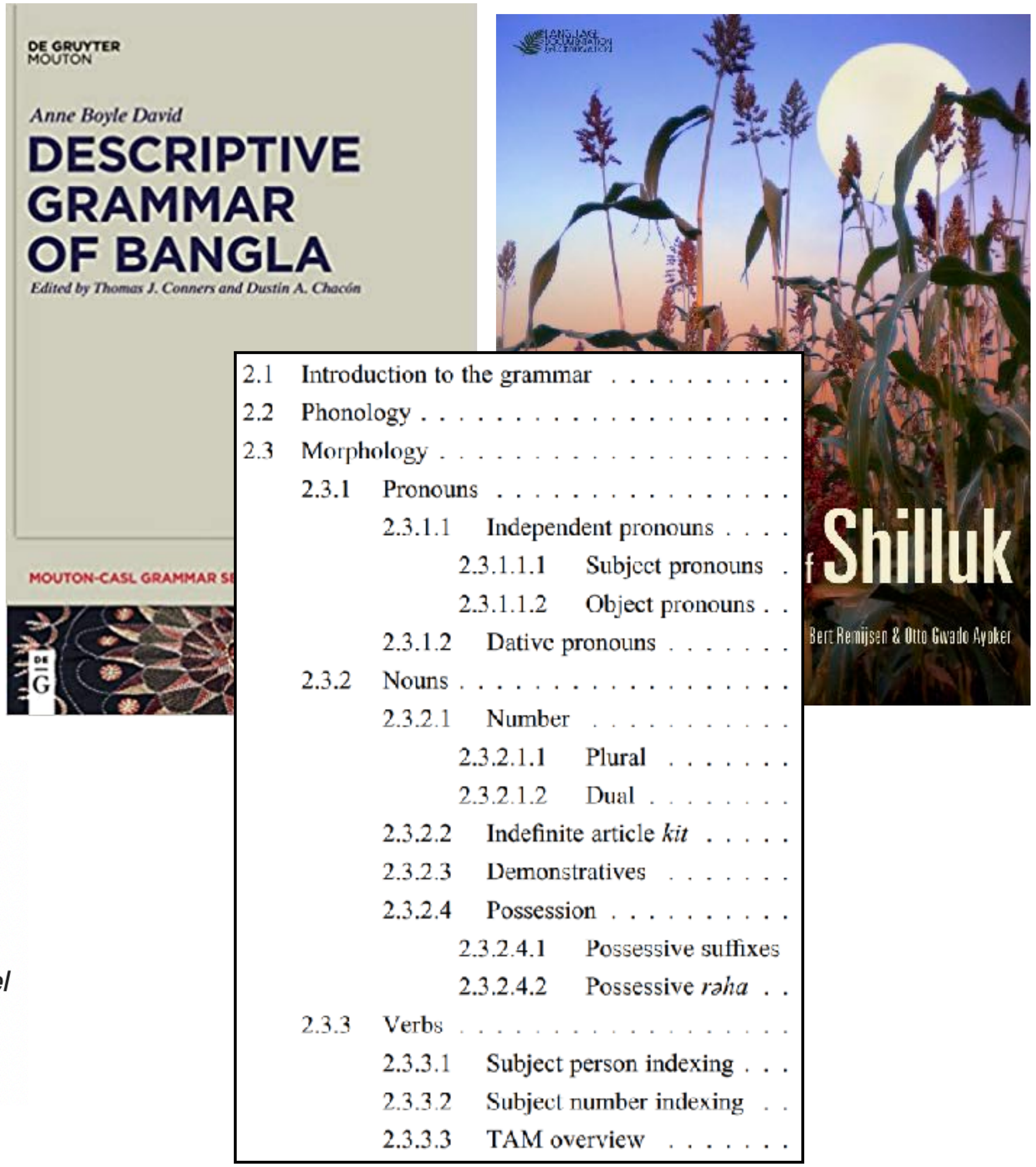
There are some exceptions though, but you will learn these as you attain new vocabulary.



When we want to turn a noun into plural, we follow these rules:

- If the noun ends in a **vowel** add **-s** Example: *un gato* (a cat); *unos gatos* - (some cats).
- If the noun ends in a **consonant** add **-es**. Example: *el papel* (the sheet of paper); *los papeles* (the sheets of paper).

Language Learning



2.1 Introduction to the grammar
2.2 Phonology
2.3 Morphology
  2.3.1 Pronouns
    2.3.1.1 Independent pronouns
      2.3.1.1.1 Subject pronouns
      2.3.1.1.2 Object pronouns
    2.3.1.2 Dative pronouns
  2.3.2 Nouns
    2.3.2.1 Number
      2.3.2.1.1 Plural
      2.3.2.1.2 Dual
    2.3.2.2 Indefinite article *kìt*
    2.3.2.3 Demonstratives
    2.3.2.4 Possession
      2.3.2.4.1 Possessive suffixes
      2.3.2.4.2 Possessive *rəha*
  2.3.3 Verbs
    2.3.3.1 Subject person indexing
    2.3.3.2 Subject number indexing
    2.3.3.3 TAM overview

Language Documentation

What are Language Descriptions?

आना सफरचंद खाते आहे

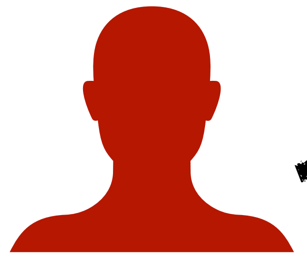
What are Language Descriptions?

आना	सफरचंद	खाते	आहे
Anna	*apple*	*eating*	*is*

How are subjects ordered wrt verbs in Marathi?

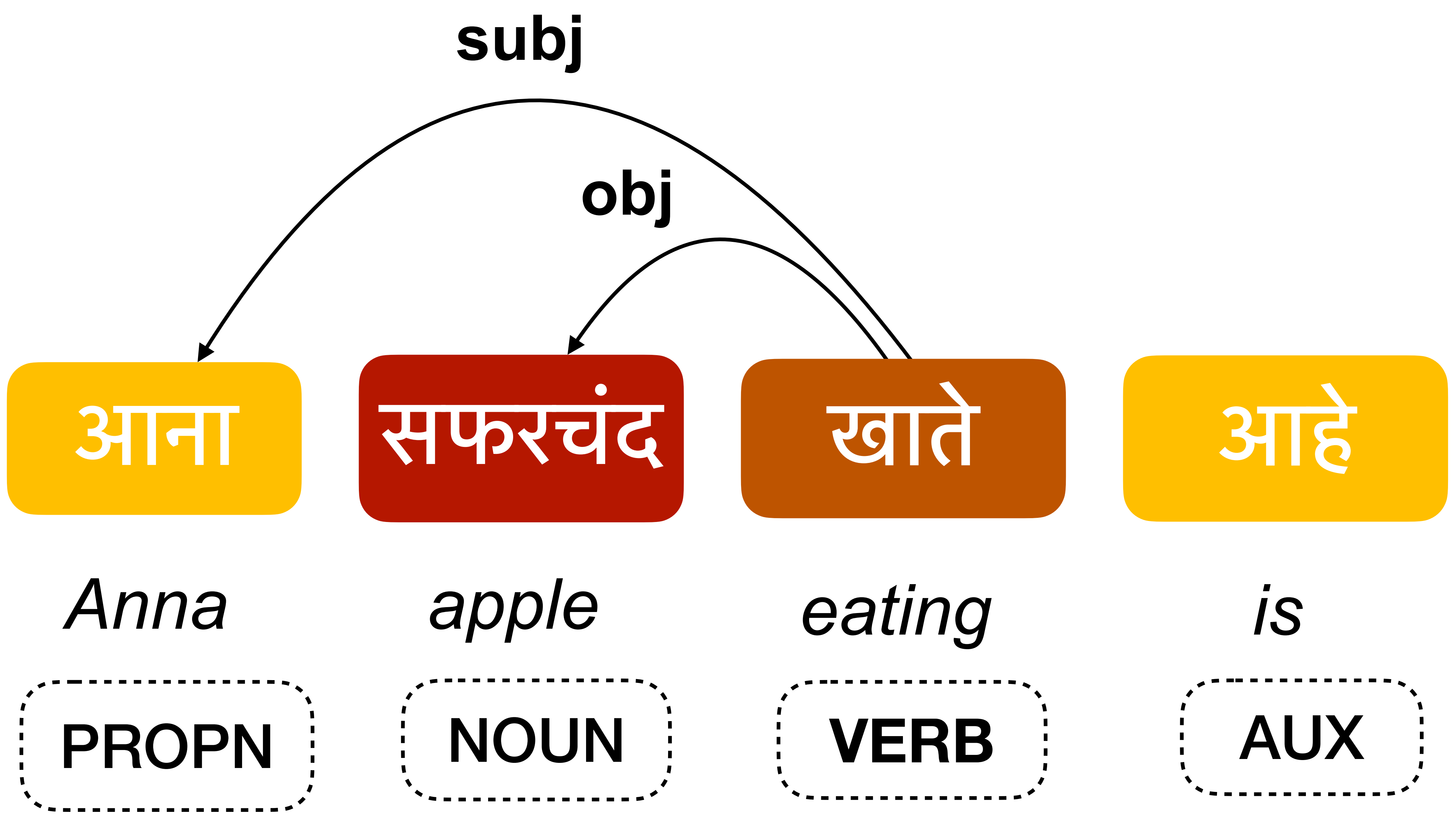
What are Language Descriptions?

आना	सफरचंद	खाते	आहे
Anna	*apple*	*eating*	*is*
PROPN	NOUN	**VERB**	AUX



How are subjects ordered wrt verbs in Marathi?

What are Language Descriptions?



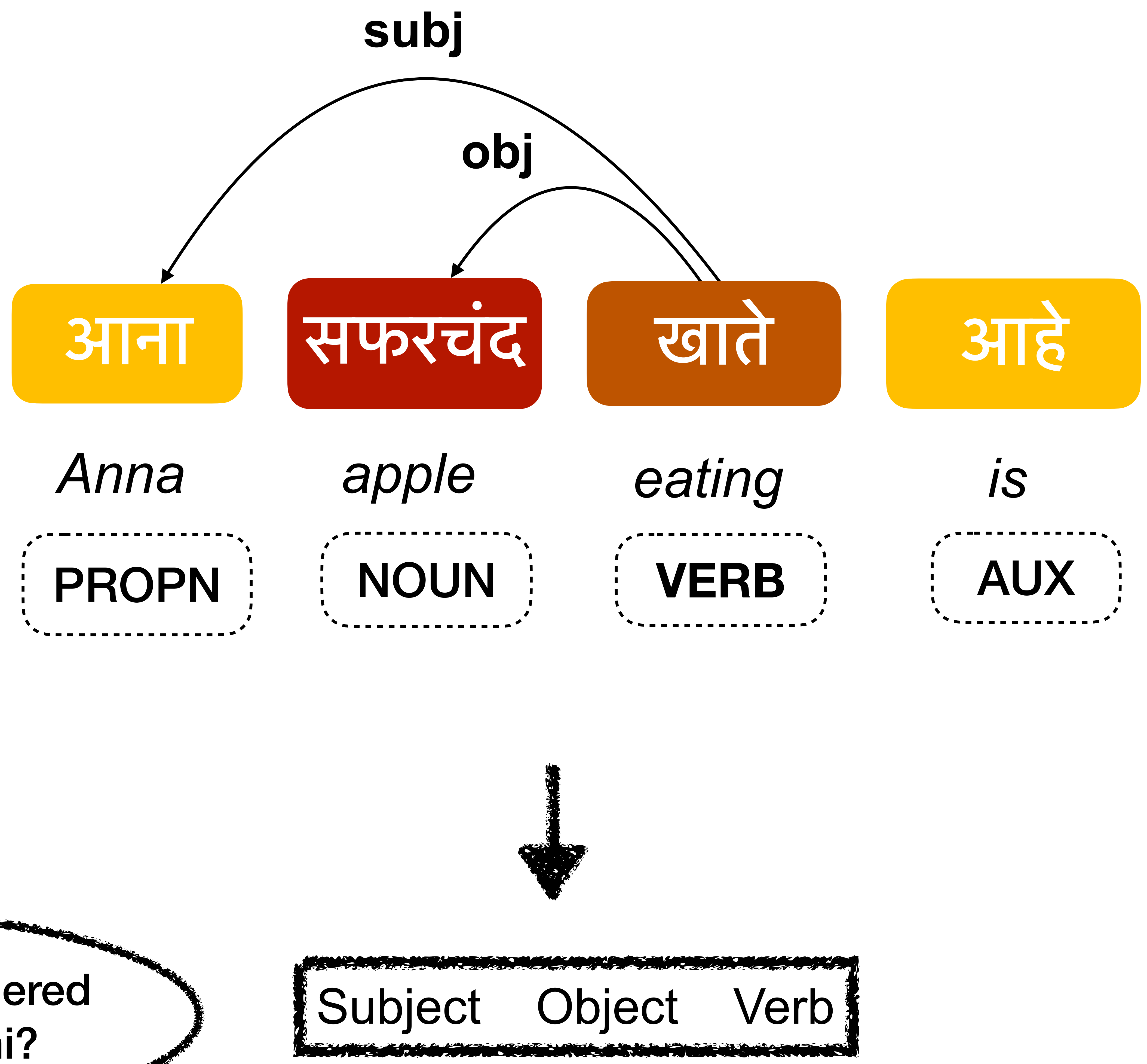
subj

obj

आना	सफरचंद	खाते	आहे
Anna	*apple*	*eating*	*is*
PROPN	NOUN	**VERB**	AUX

How are subjects ordered
wrt verbs in Marathi?

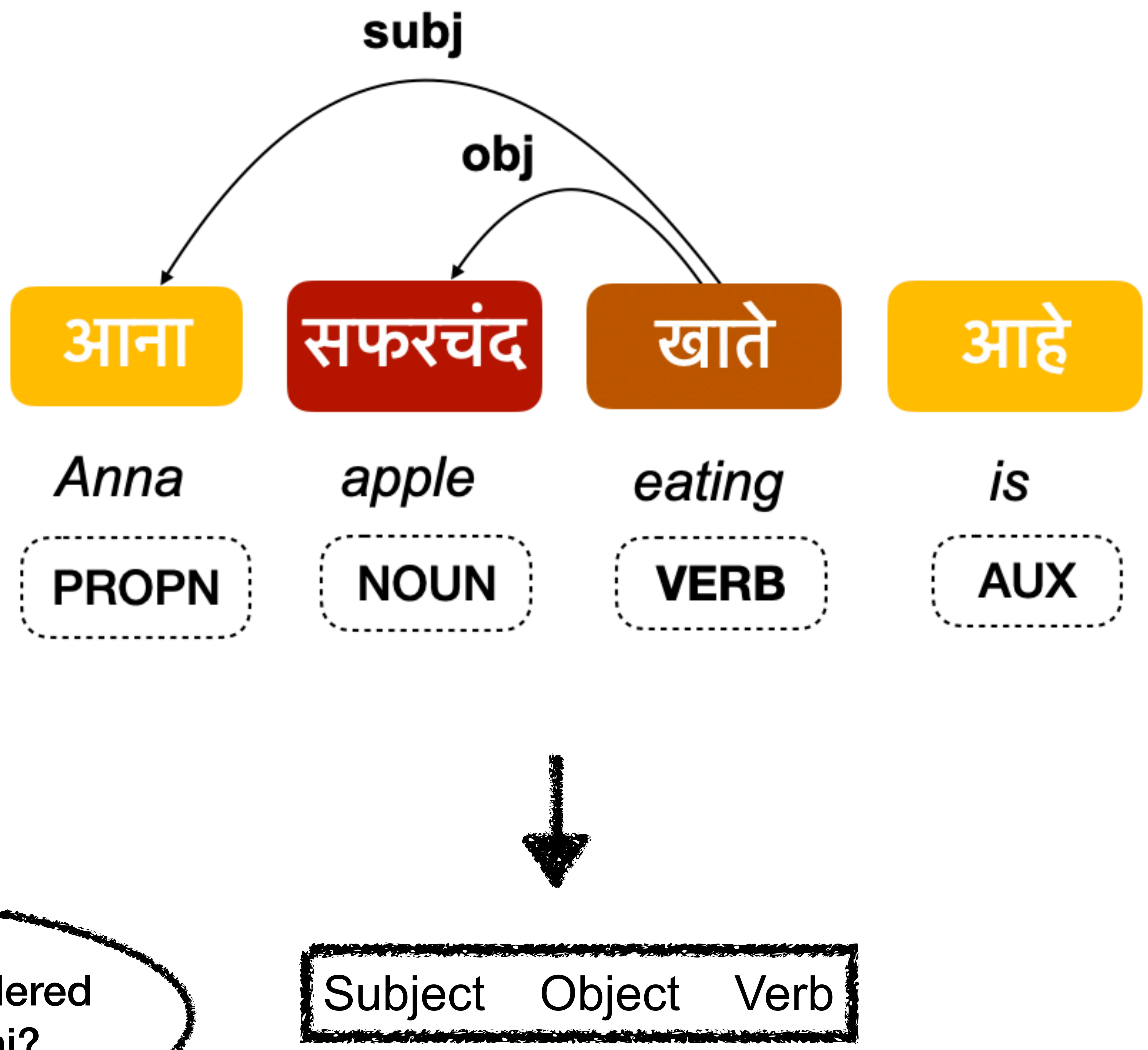
What are Language Descriptions?



Subject Object Verb

How are subjects ordered wrt verbs in Marathi?

Why is this challenging?



subj

obj

आना सफरचंद खाते आहे

Anna *apple* *eating* *is*

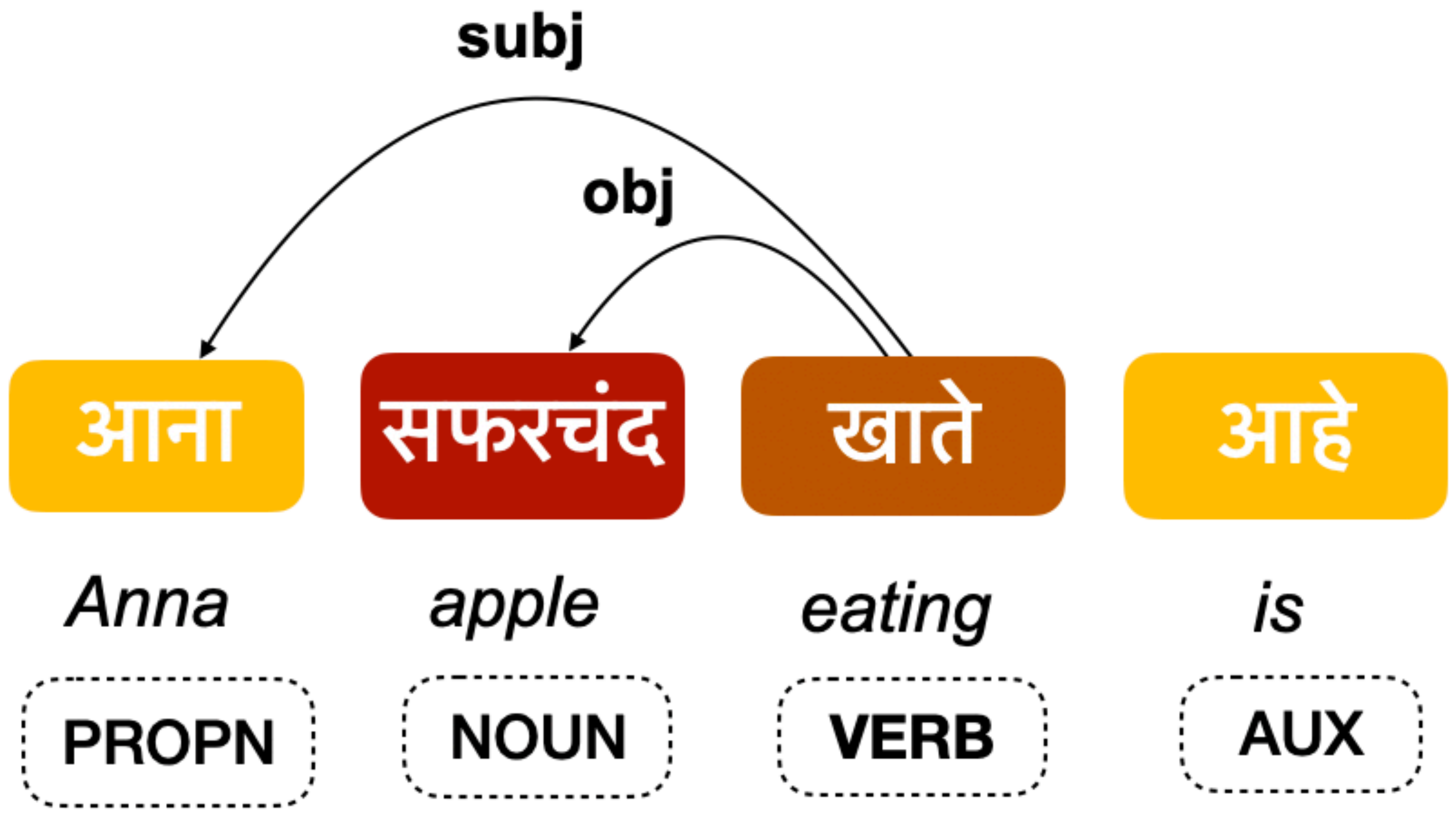
PROPN NOUN VERB AUX

Subject Object Verb

How are subjects ordered wrt verbs in Marathi?

Why is this challenging?

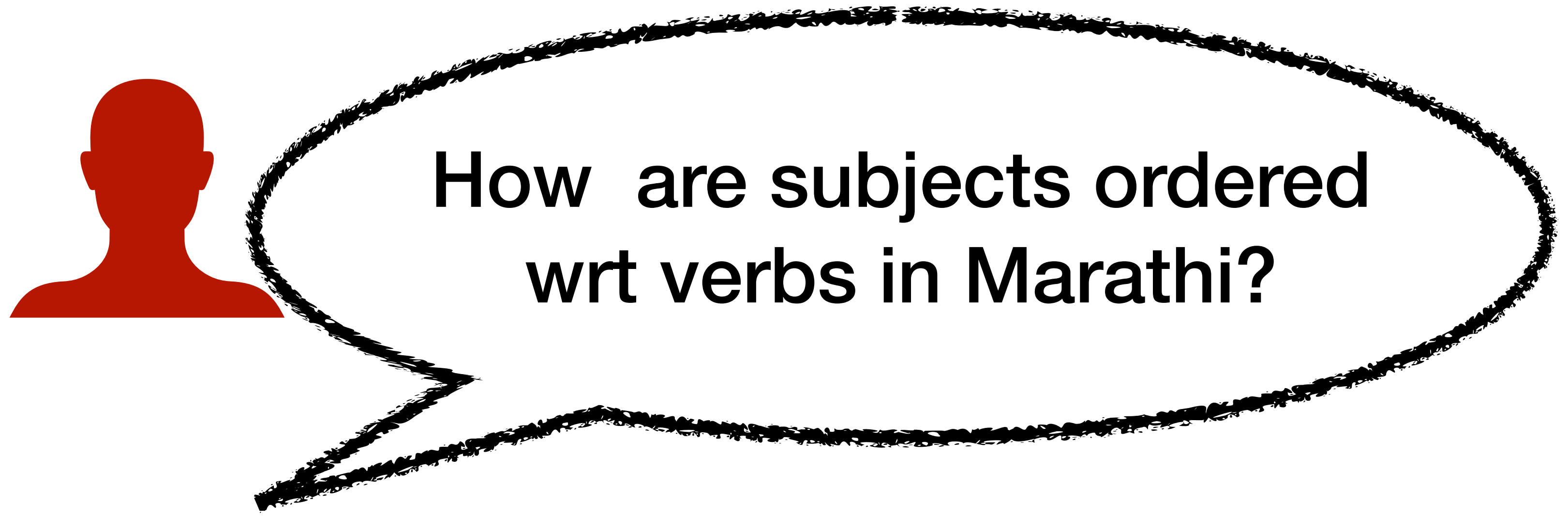
How are subjects ordered wrt verbs in Marathi?

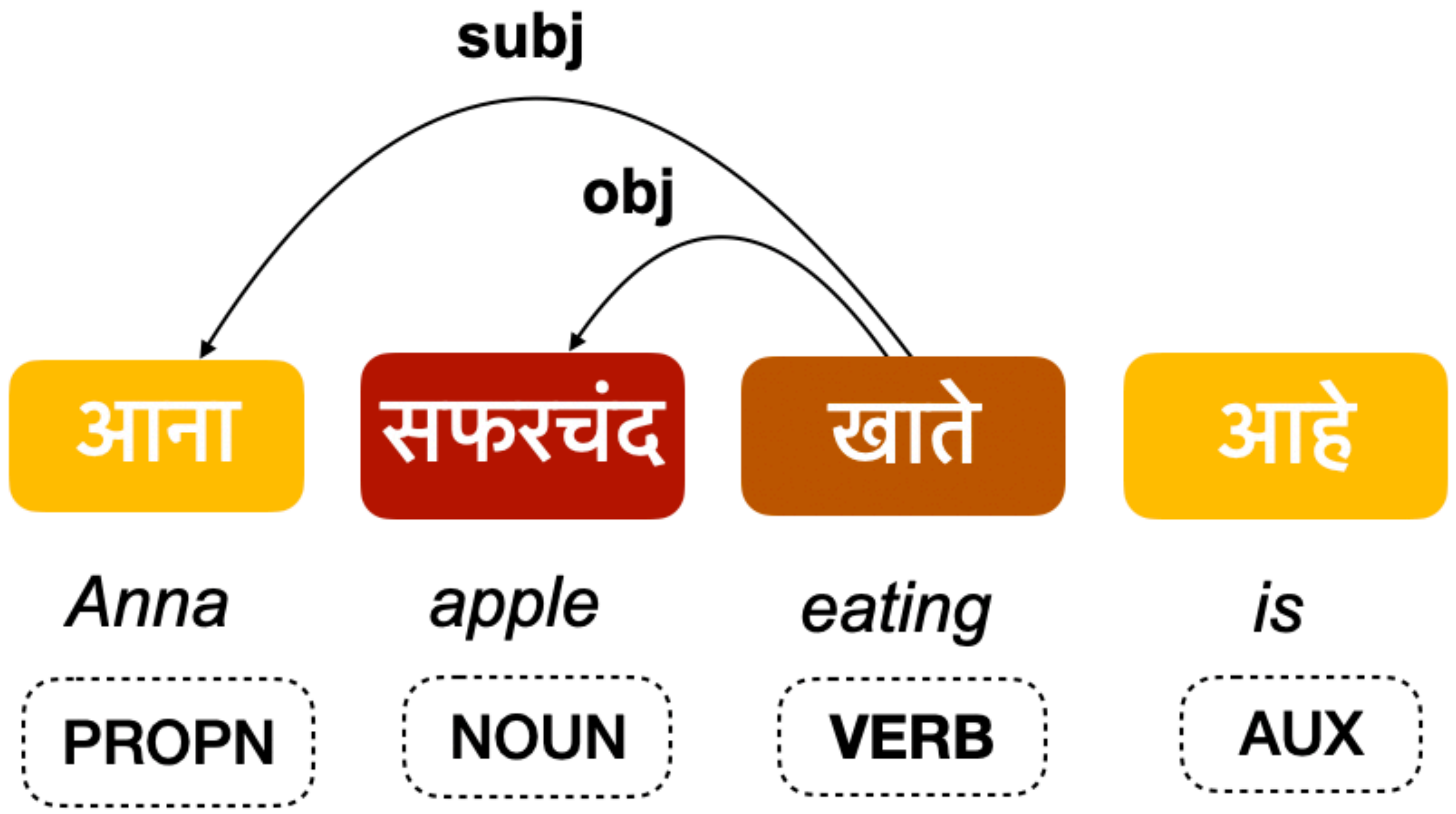


आना	सफरचंद	खाते	आहे
Anna	*apple*	*eating*	*is*
PROPN	NOUN	VERB	AUX

Subject Object Verb

Why is this challenging?





Subject Object Verb

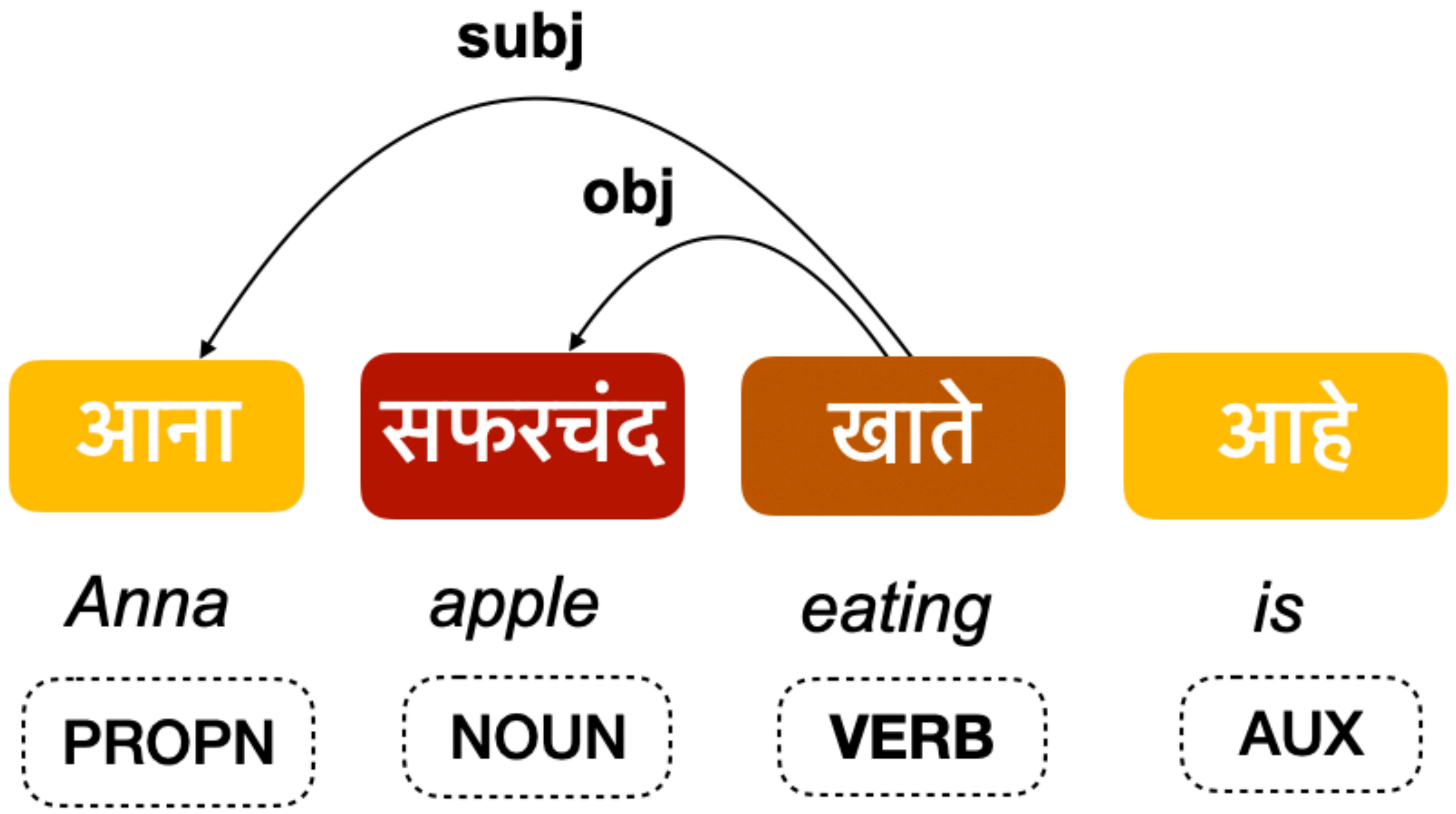
Where Languages Are Dying

Languages classified as threatened/ endangered in 2022, by region

* Including the Caribbean
** Including the Caucasus
Source: Endangered Languages Project

Why is this challenging?

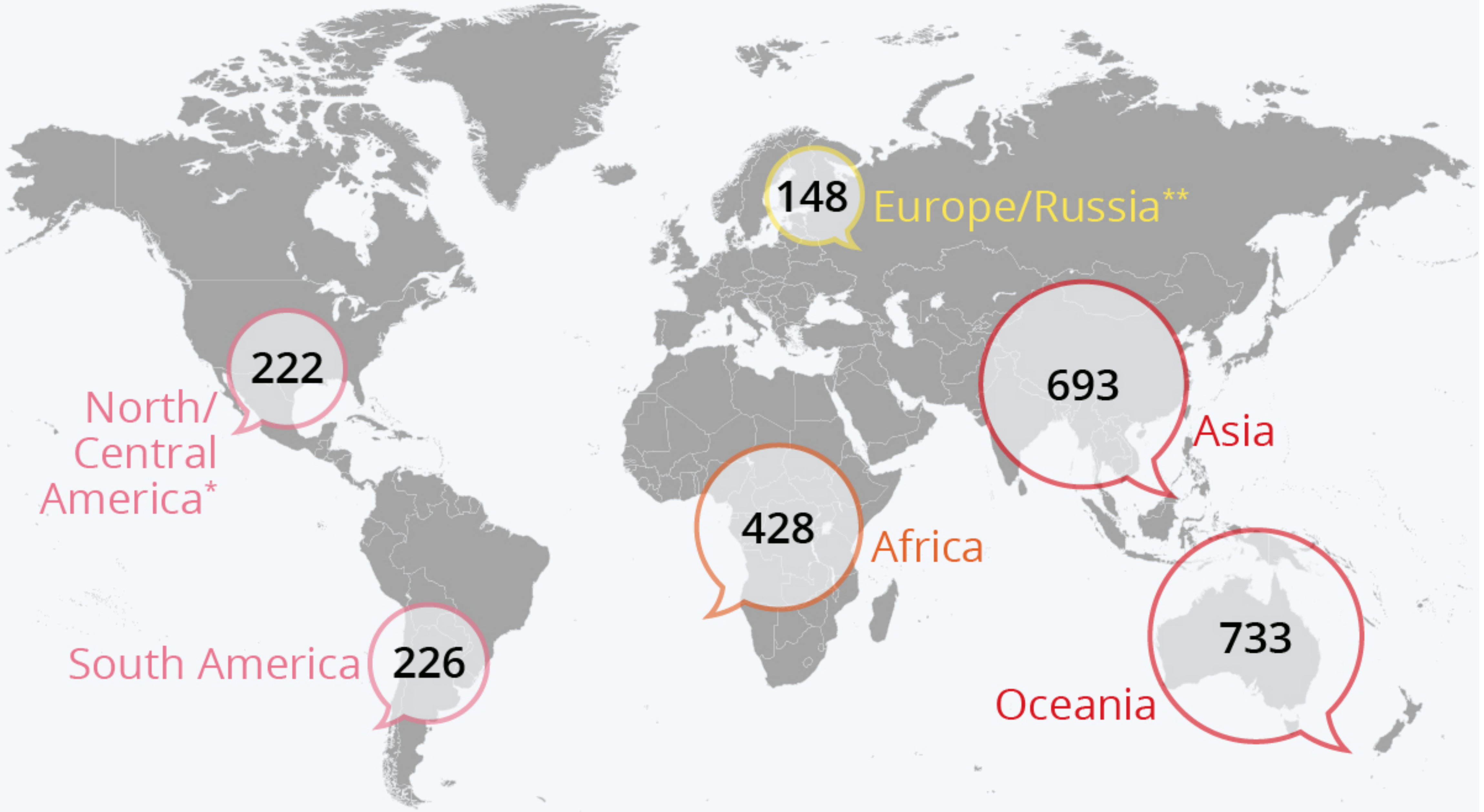
How are subjects ordered wrt verbs in Marathi?



आना	सफरचंद	खाते	आहे
Anna	*apple*	*eating*	*is*
PROPN	NOUN	VERB	AUX

Where Languages Are Dying

Languages classified as threatened/ endangered in 2022, by region

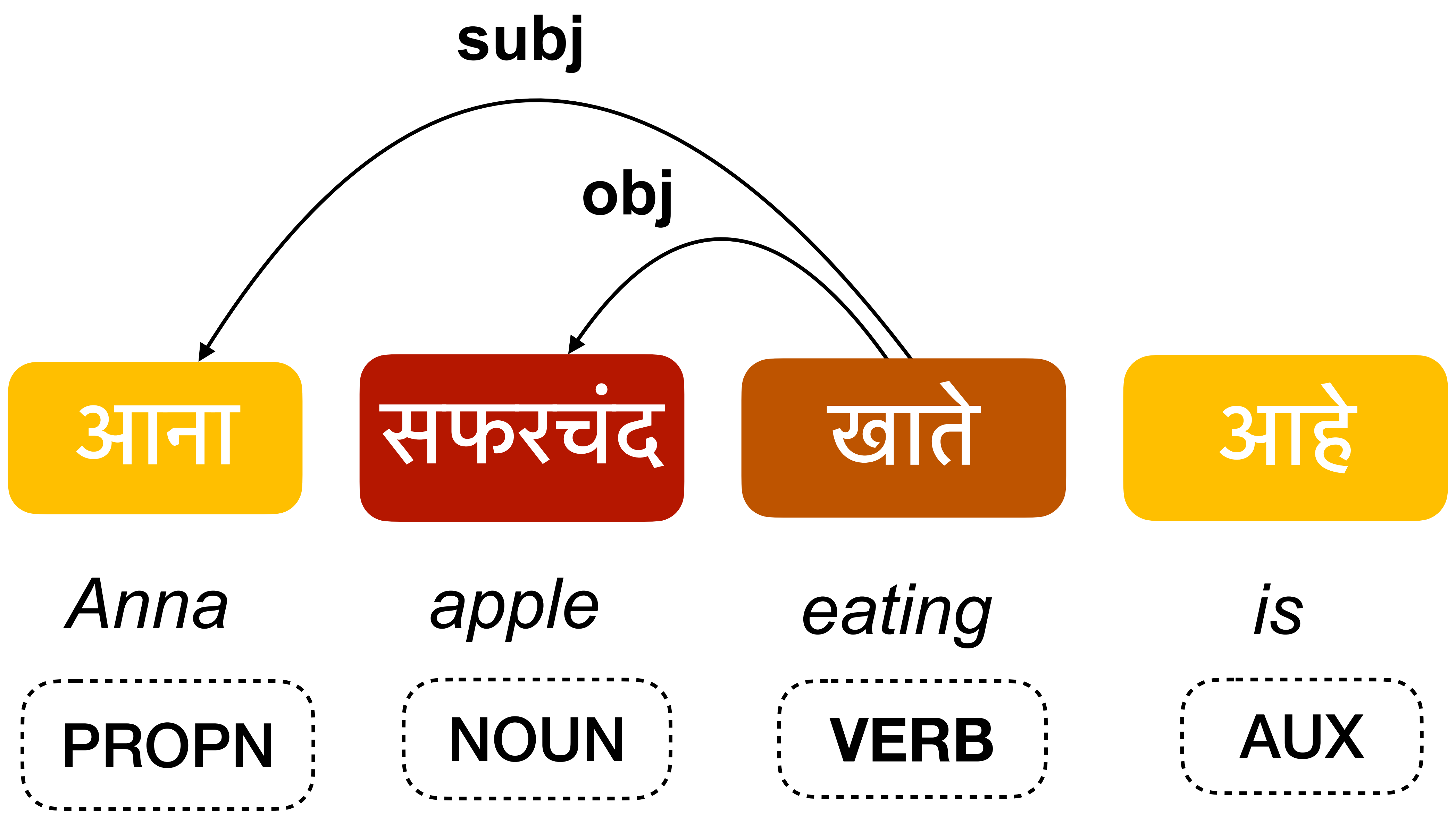


* Including the Caribbean
** Including the Caucasus
Source: Endangered Languages Project

Subject Object Verb

Could we help in some way?

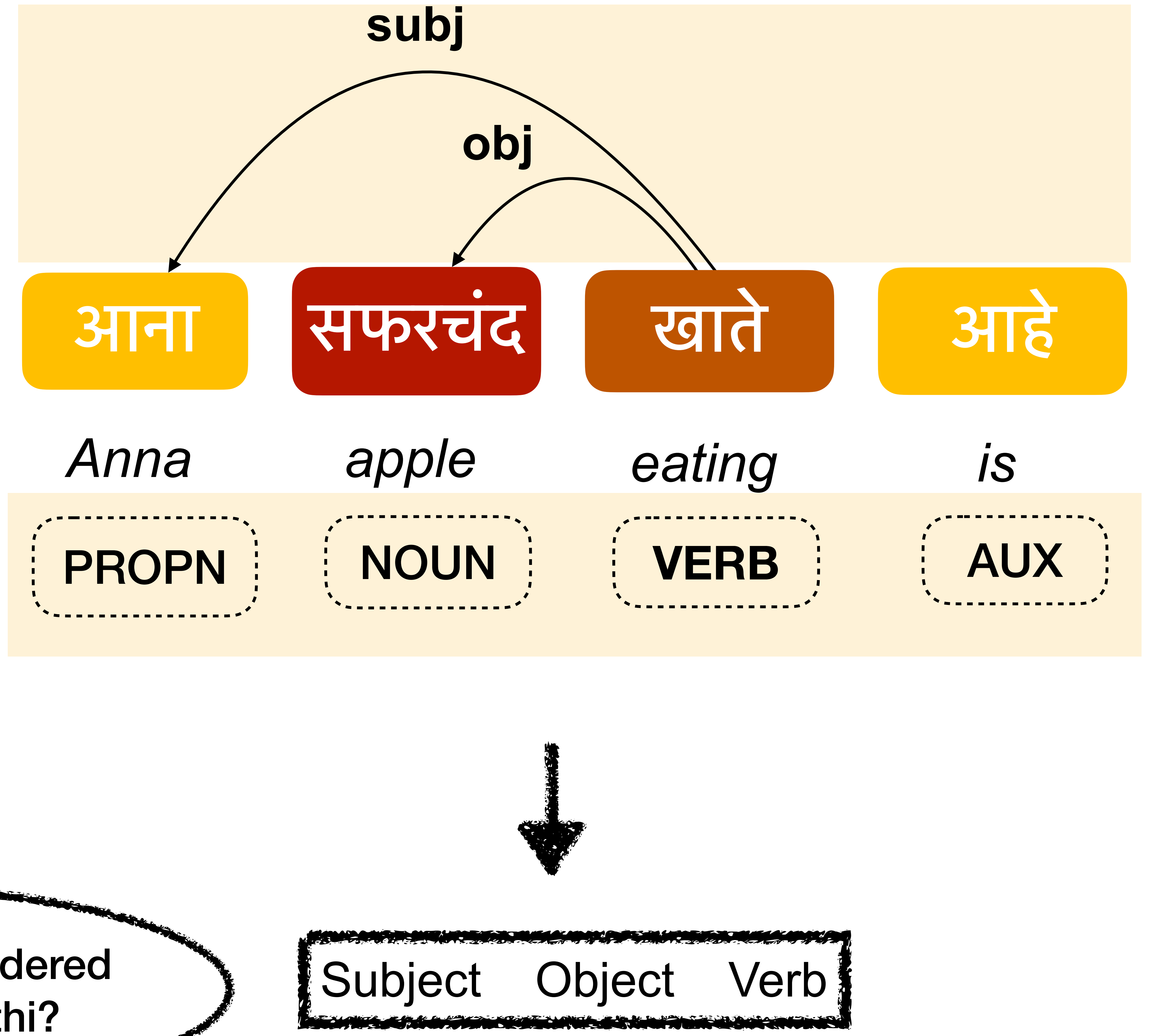
NLP already does some of this!



Subject Object Verb

How are subjects ordered wrt verbs in Marathi?

NLP already does some of this!



subj

obj

आना सफरचंद खाते आहे

Anna apple eating is

PROPN NOUN VERB AUX

Dependency Parsing

POS Tagging

Subject Object Verb

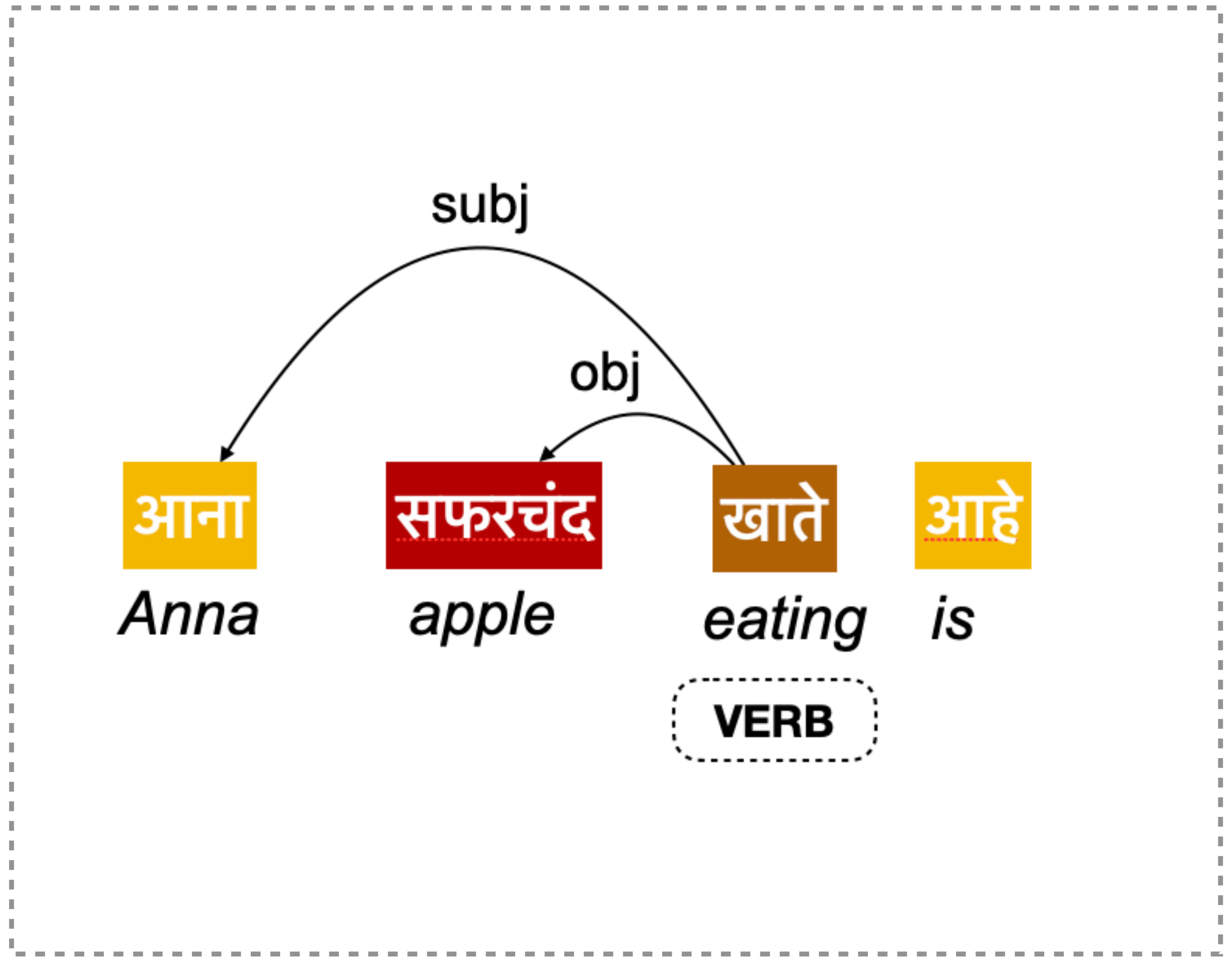
How are subjects ordered wrt verbs in Marathi?

Our Proposed Pipeline

Chaudhary, et al. "Automatic extraction of rules governing morphological agreement." EMNLP (2020).
Chaudhary, et al. "When is Wall a Pared and when a Muro?--Extracting Rules Governing Lexical Selection." EMNLP (2021).
Chaudhary, et al. "AUTOLEX: An Automatic Framework for Linguistic Exploration." arXiv (2022).
Chaudhary, et al. "Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning." arXiv (2022).

Our Proposed Pipeline

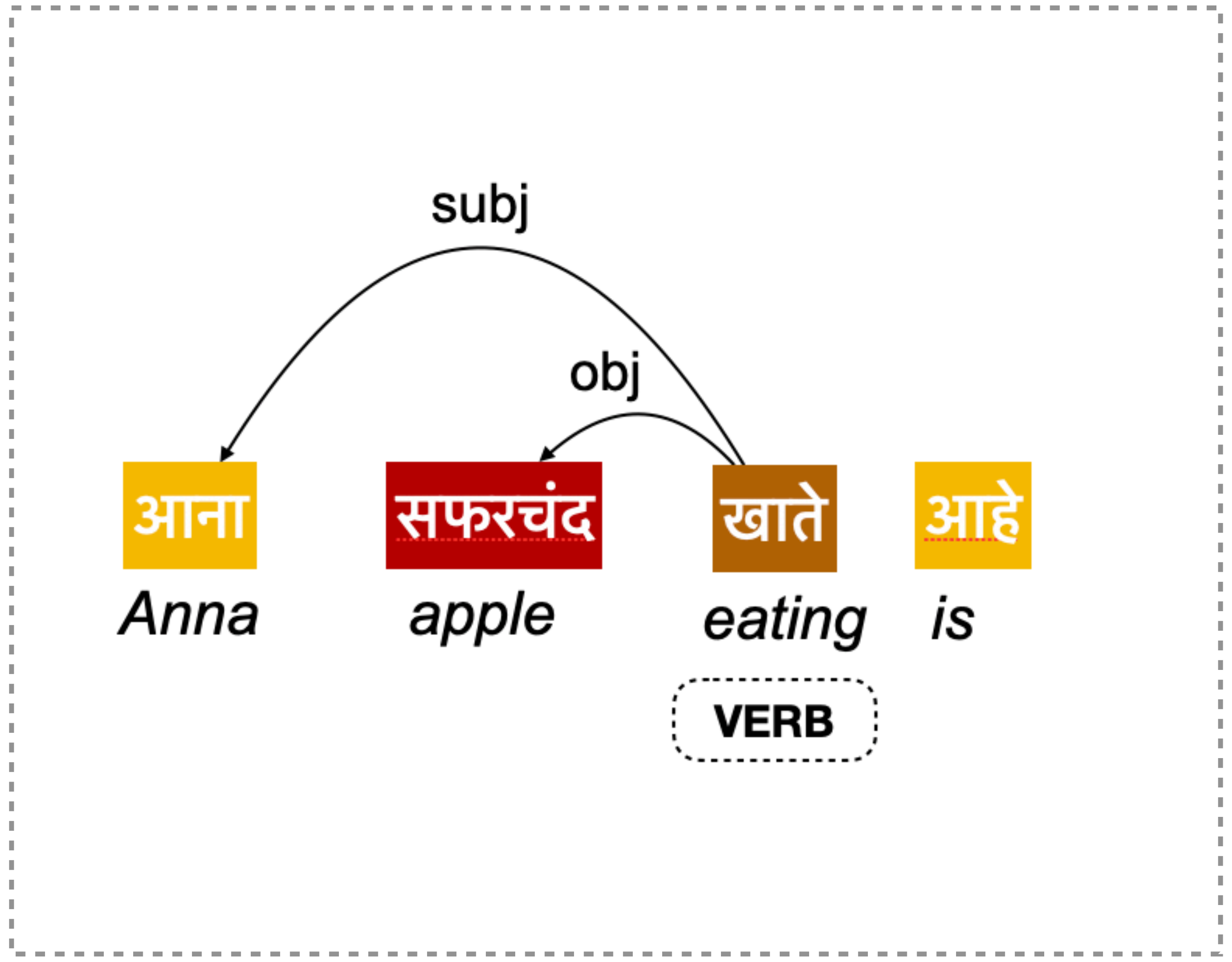
(Low-resource) Language Analysis



Chaudhary, et al. "Automatic extraction of rules governing morphological agreement." EMNLP (2020).
Chaudhary, et al. "When is Wall a Pared and when a Muro?--Extracting Rules Governing Lexical Selection." EMNLP (2021).
Chaudhary, et al. "AUTOLEX: An Automatic Framework for Linguistic Exploration." arXiv (2022).
Chaudhary, et al. "Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning." arXiv (2022).

Our Proposed Pipeline

(Low-resource) Language Analysis

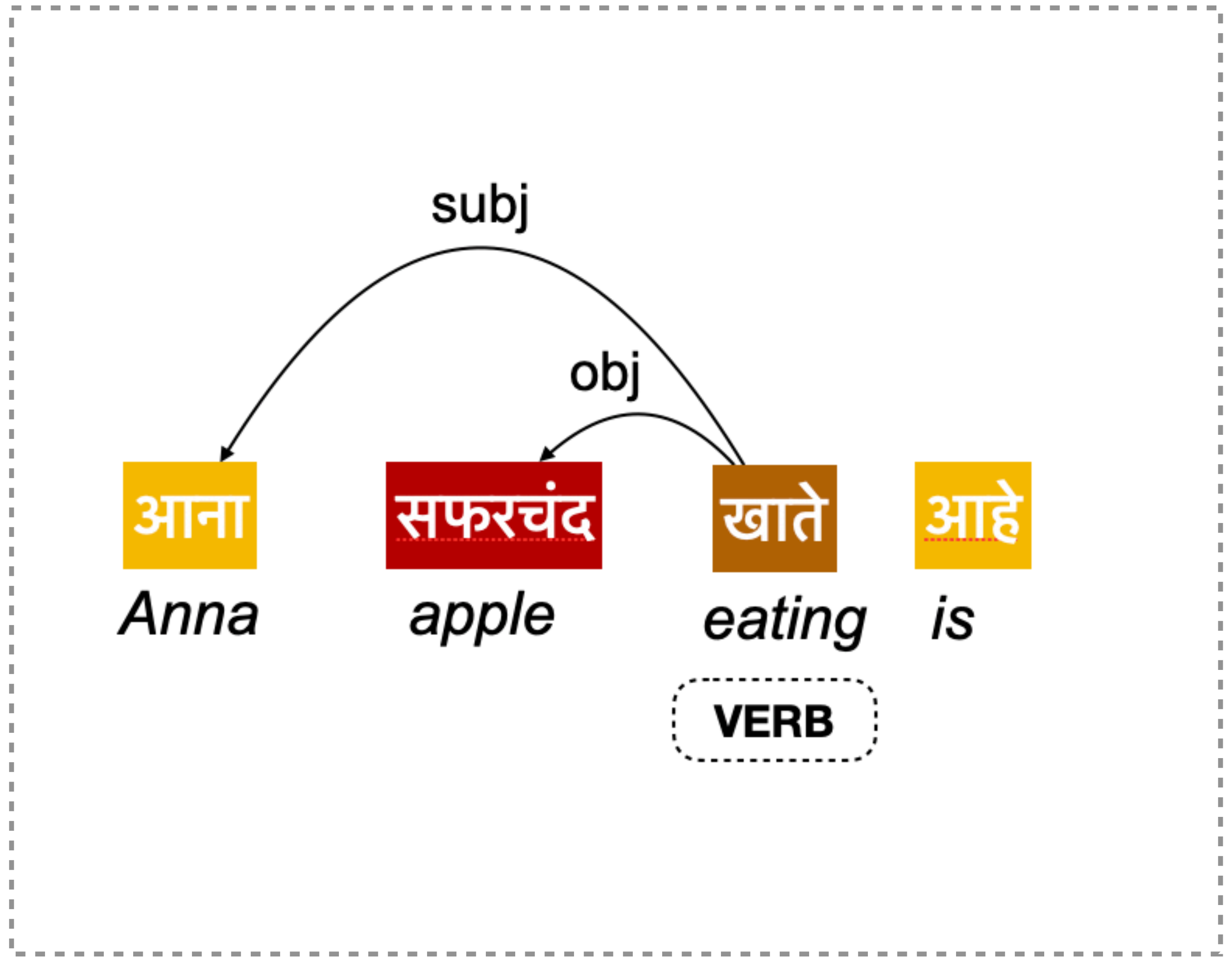


AutoLEX: Automatic Language Explorer

Chaudhary, et al. "Automatic extraction of rules governing morphological agreement." EMNLP (2020).
Chaudhary, et al. "When is Wall a Pared and when a Muro?--Extracting Rules Governing Lexical Selection." EMNLP (2021).
Chaudhary, et al. "AUTOLEX: An Automatic Framework for Linguistic Exploration." arXiv (2022).
Chaudhary, et al. "Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning." arXiv (2022).

Our Proposed Pipeline

(Low-resource) Language Analysis

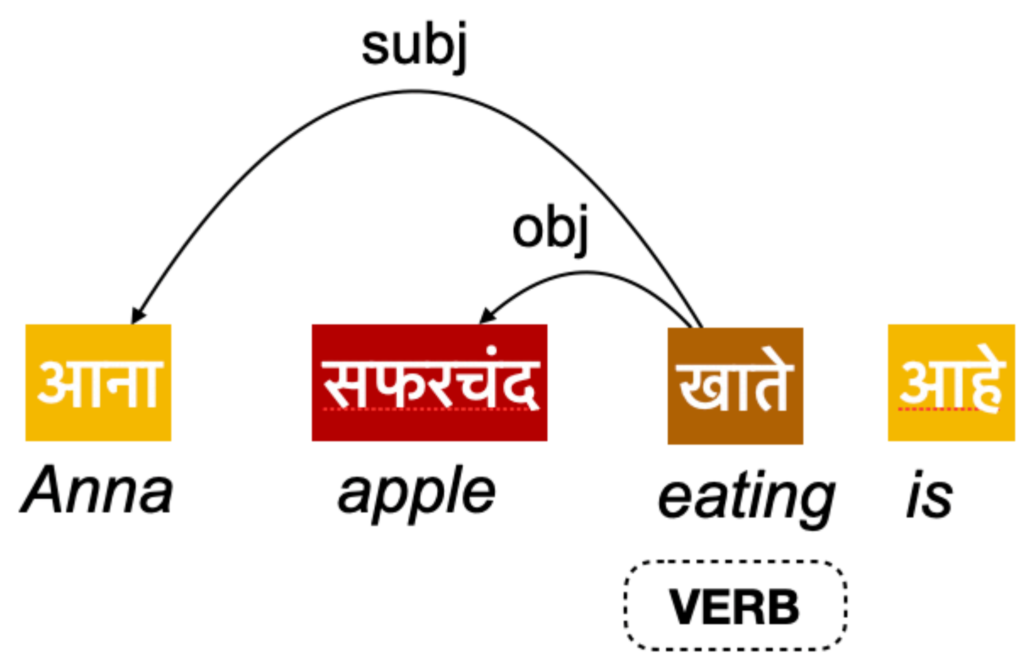


AutoLEX: Automatic Language Explorer

Chaudhary, et al. "Automatic extraction of rules governing morphological agreement." EMNLP (2020).
Chaudhary, et al. "When is Wall a Pared and when a Muro?--Extracting Rules Governing Lexical Selection." EMNLP (2021).
Chaudhary, et al. "AUTOLEX: An Automatic Framework for Linguistic Exploration." arXiv (2022).
Chaudhary, et al. "Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning." arXiv (2022).

Our Proposed Pipeline

(Low-resource) Language Analysis



→

AutoLEX: Automatic Language Explorer

Subject Object Verb

तांदूळ
/tandul/
‘raw rice’

भात
/bhaat/
‘cooked rice’

→

Applications

Language Education and Documentation

Chaudhary, et al. "Automatic extraction of rules governing morphological agreement." EMNLP (2020).
Chaudhary, et al. "When is Wall a Pared and when a Muro?--Extracting Rules Governing Lexical Selection." EMNLP (2021).
Chaudhary, et al. "AUTOLEX: An Automatic Framework for Linguistic Exploration." arXiv (2022).
Chaudhary, et al. "Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning." arXiv (2022).

AutoLEX: Automatic Language Explorer

AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats

AutoLEX: Automatic Language Explorer

How are subjects
ordered in Marathi?

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats

AutoLEX: Automatic Language Explorer

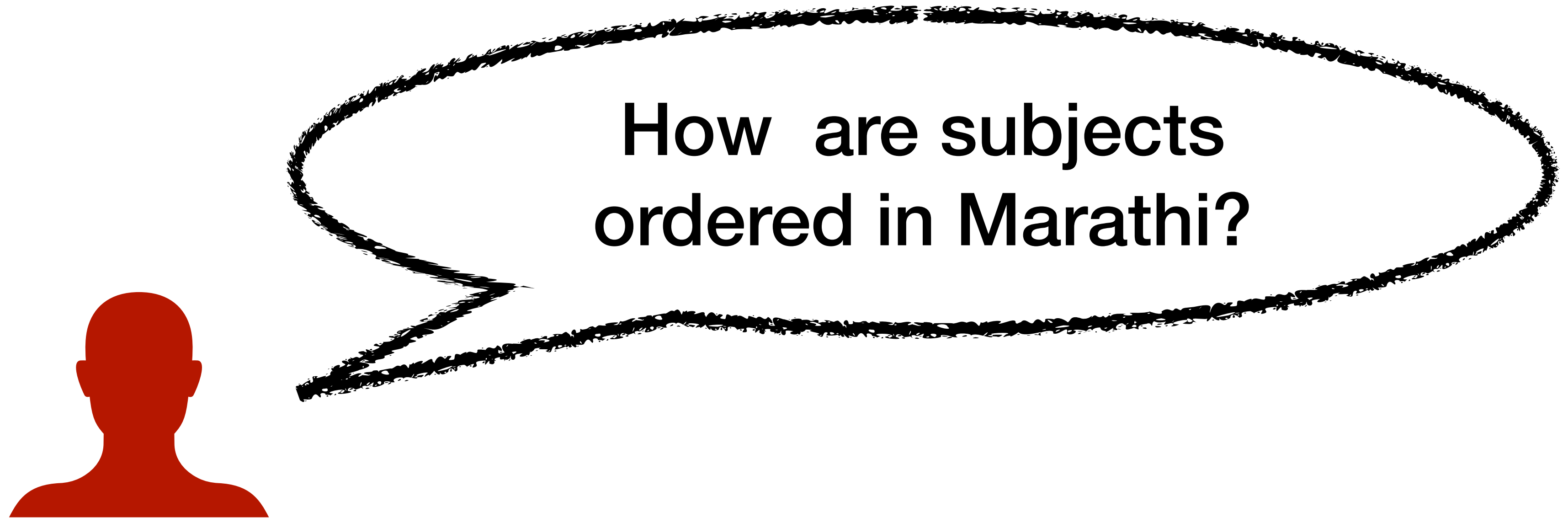
Extract and visualize answers to different linguistic questions in both human- and machine-readable formats

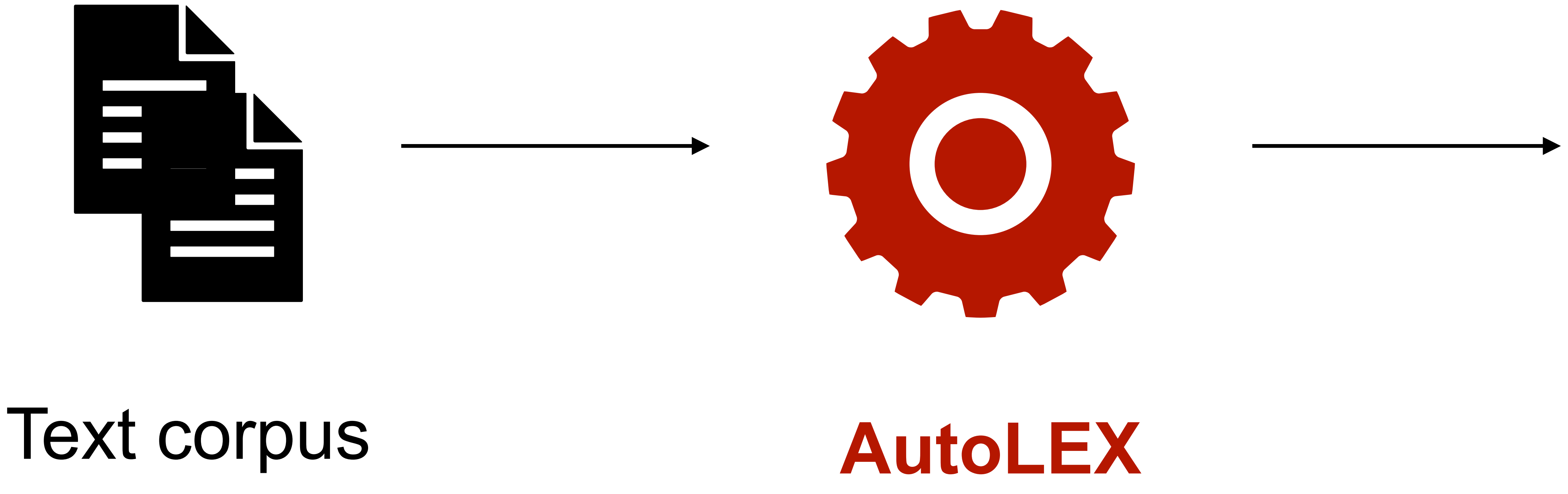
How are subjects ordered in Marathi?

Text corpus

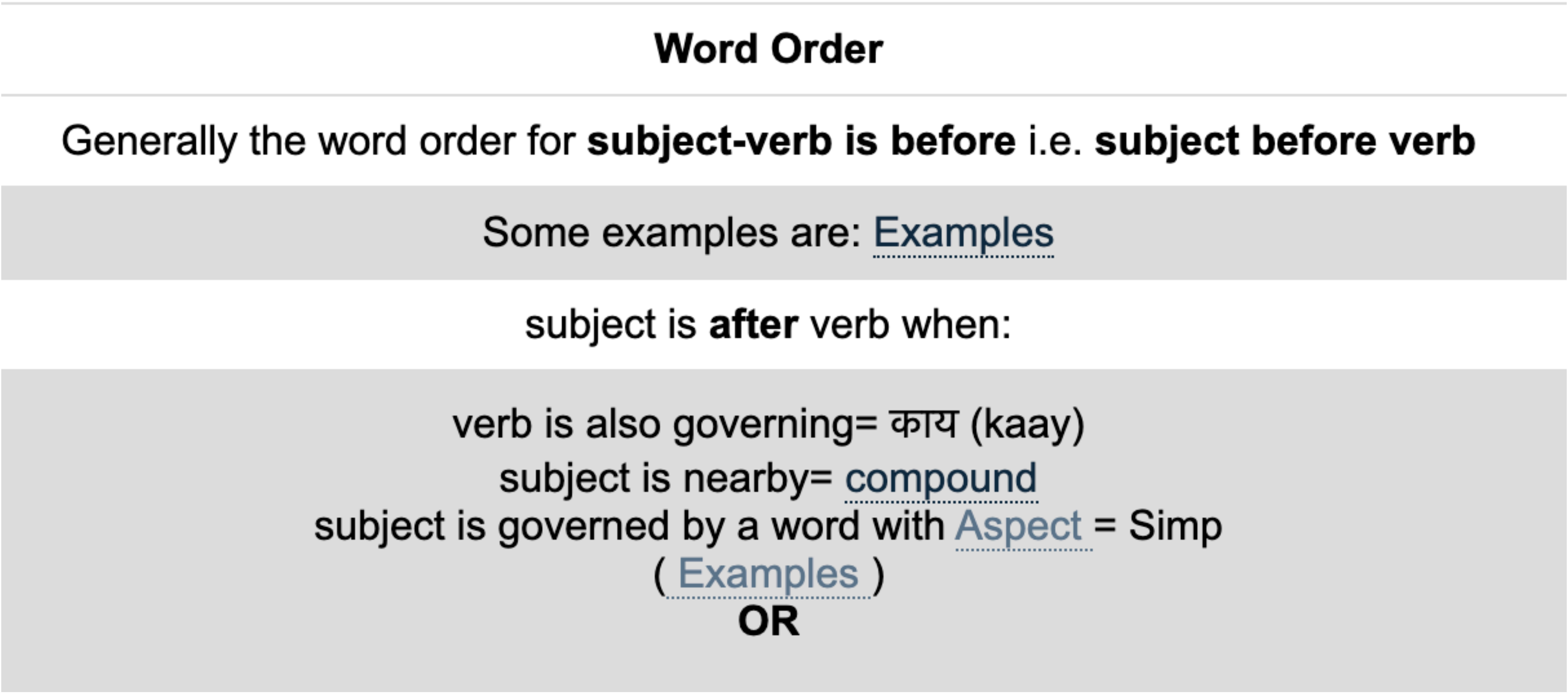
AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats




Text corpus

AutoLEX


Word Order

Generally the word order for **subject-verb is before** i.e. **subject before verb**

Some examples are: Examples

subject is **after** verb when:

verb is also governing= काय (kaay)
subject is nearby= compound
subject is governed by a word with Aspect = Simp
(Examples)
OR

AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats

How are subjects ordered in Marathi?

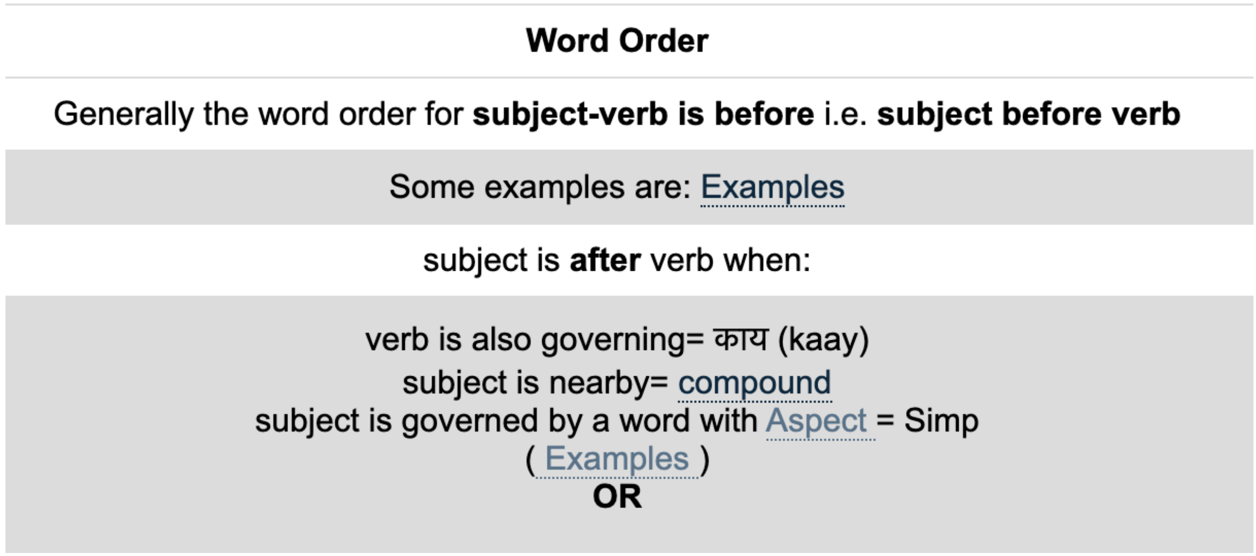
Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Text corpus



Word Order

Generally the word order for **subject-verb is before** i.e. **subject before verb**

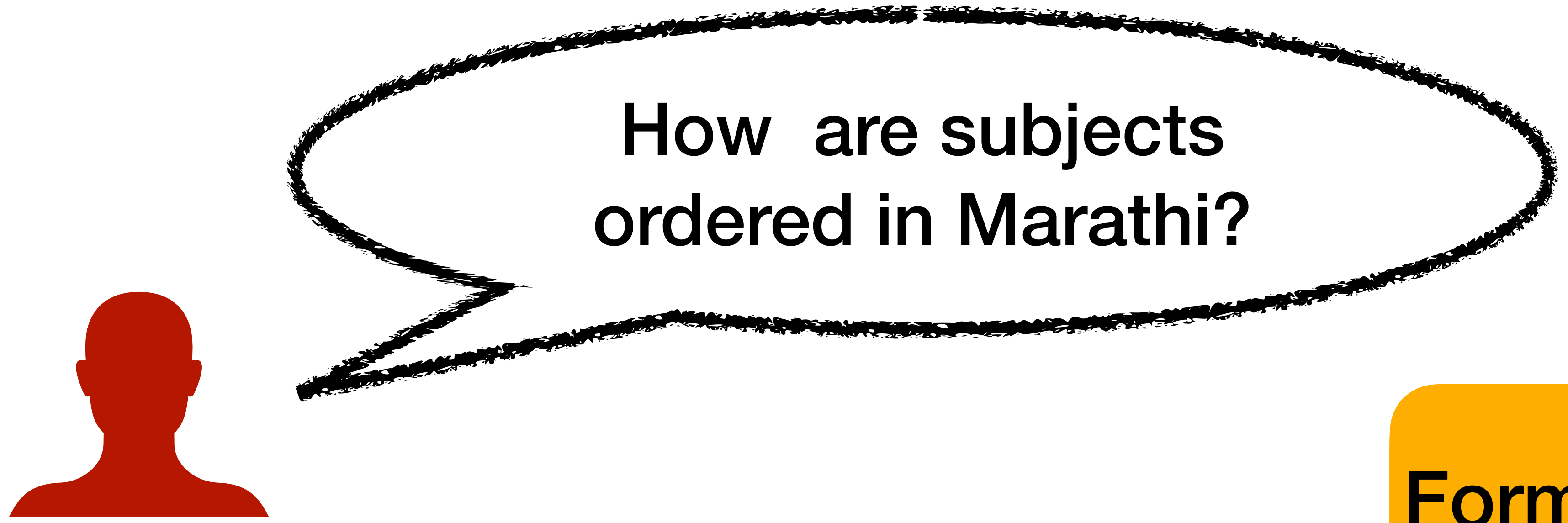
Some examples are: Examples

subject is **after** verb when:

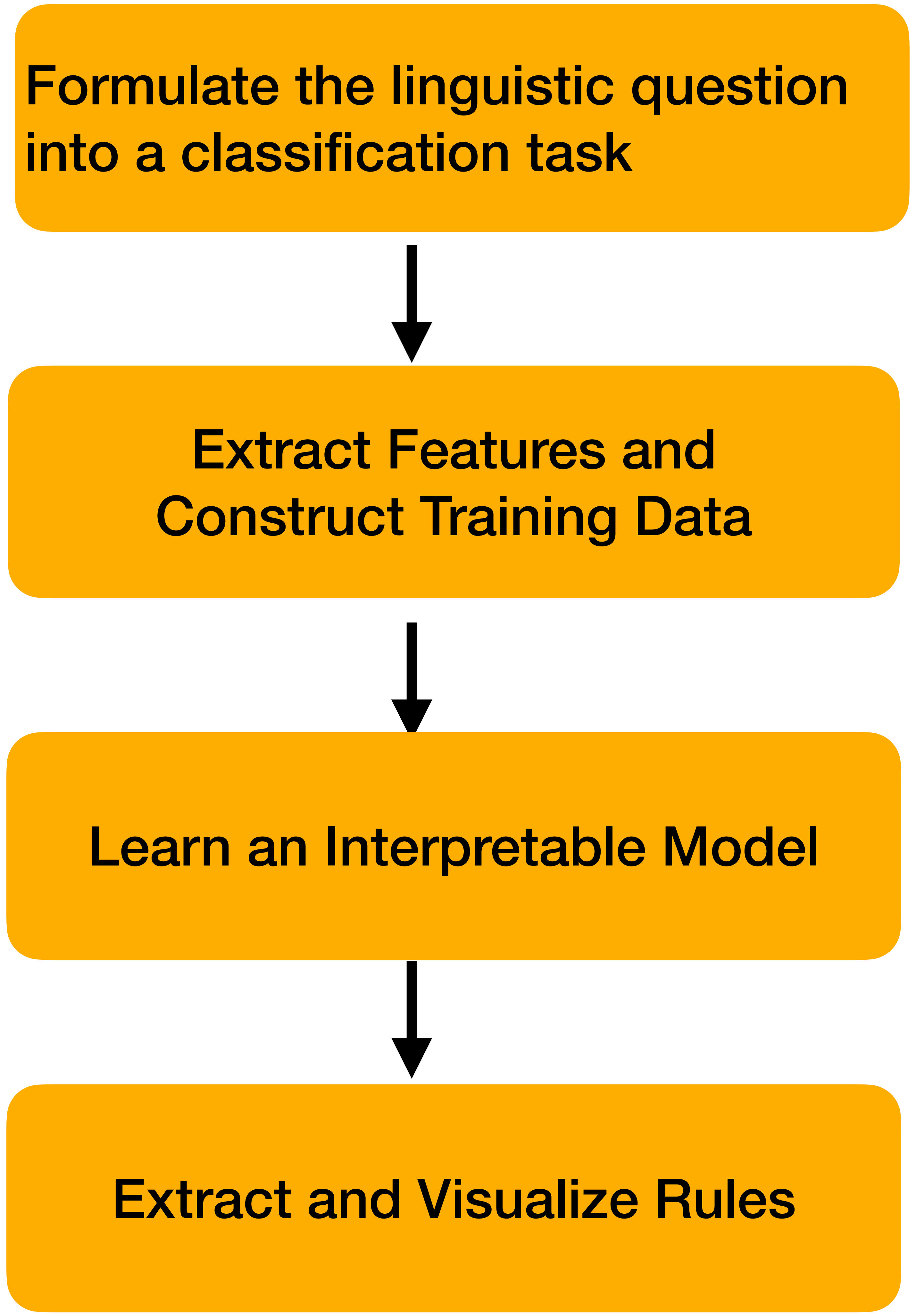
verb is also governing= काय (kaay)
subject is nearby= compound
subject is governed by a word with Aspect = Simp
(Examples)
OR

AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats



ext corpus

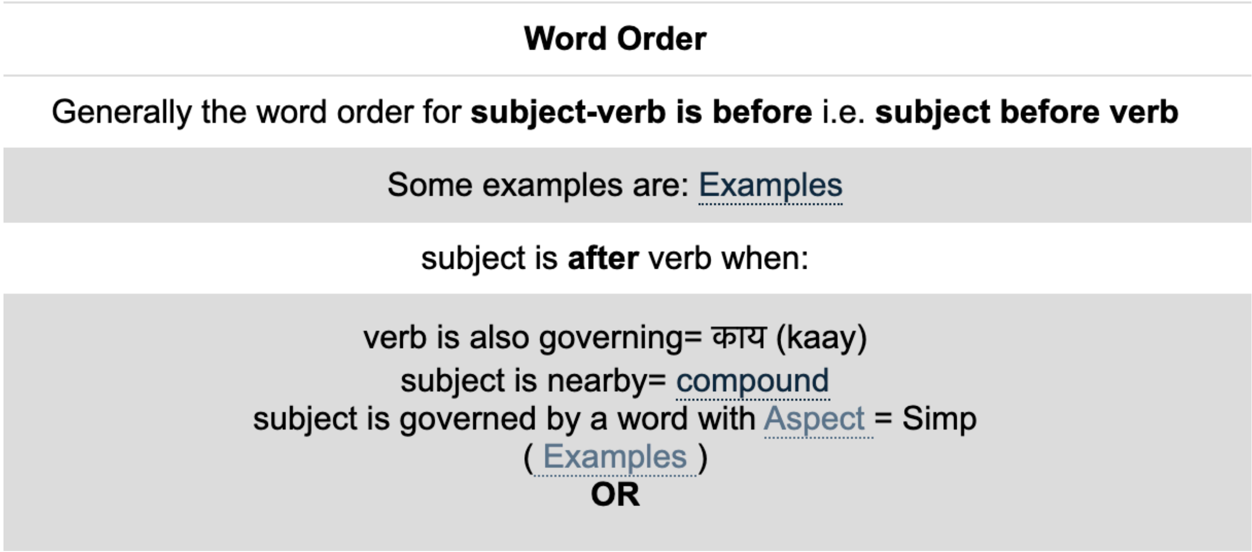


Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules



Word Order

Generally the word order for **subject-verb is before** i.e. **subject before verb**

Some examples are: Examples

subject is **after** verb when:

verb is also governing= काय (kaay)
subject is nearby= compound
subject is governed by a word with Aspect = Simp
(Examples)
OR

AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats

How are subjects ordered in Marathi?

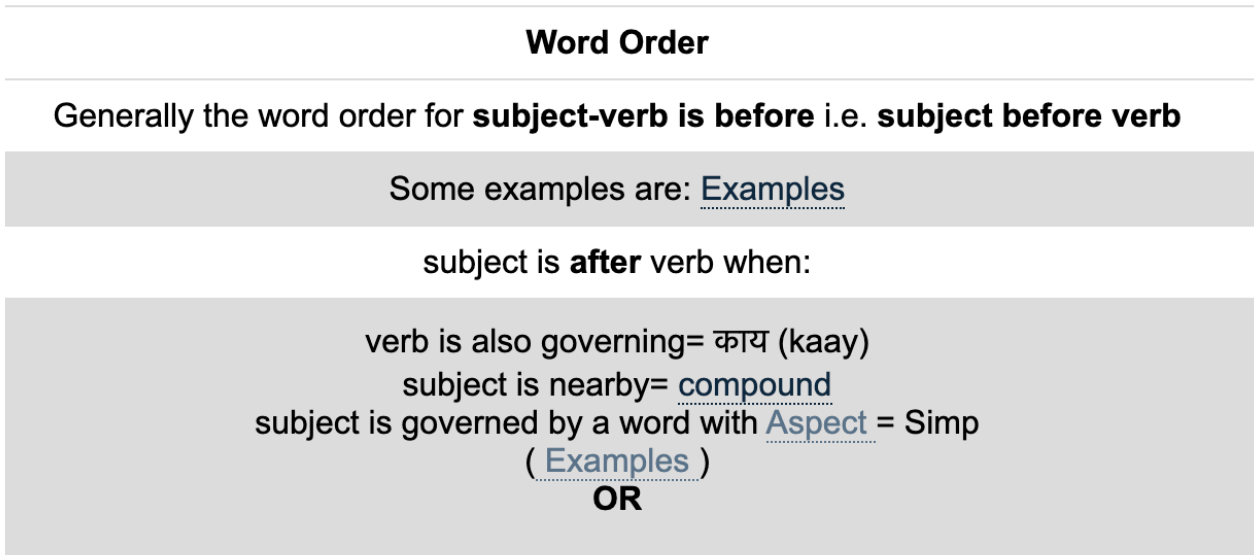
Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Text corpus



Word Order

Generally the word order for **subject-verb is before** i.e. **subject before verb**

Some examples are: Examples

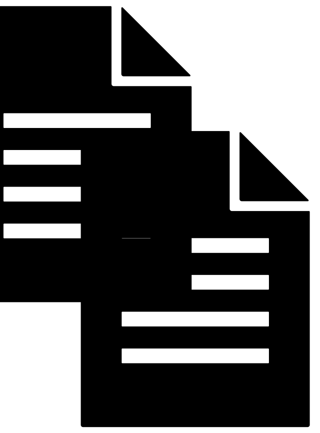
subject is **after** verb when:

verb is also governing= काय (kaay)
subject is nearby= compound
subject is governed by a word with Aspect = Simp
(Examples)
OR

AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats

How are subjects ordered in Marathi?



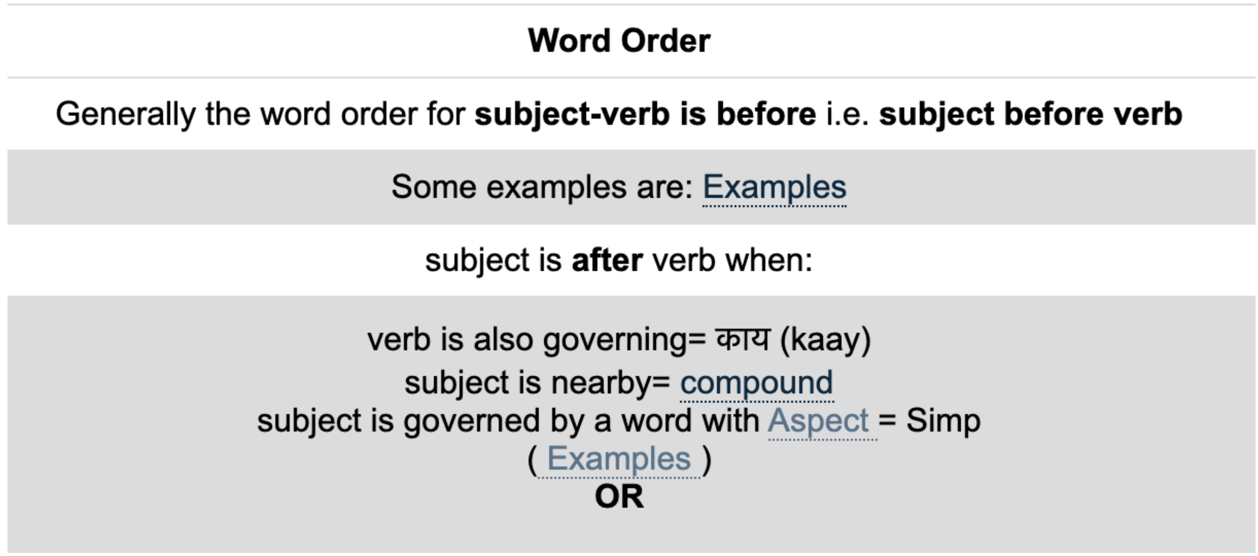
Text corpus

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules



Word Order

Generally the word order for **subject-verb is before** i.e. **subject before verb**

Some examples are: Examples

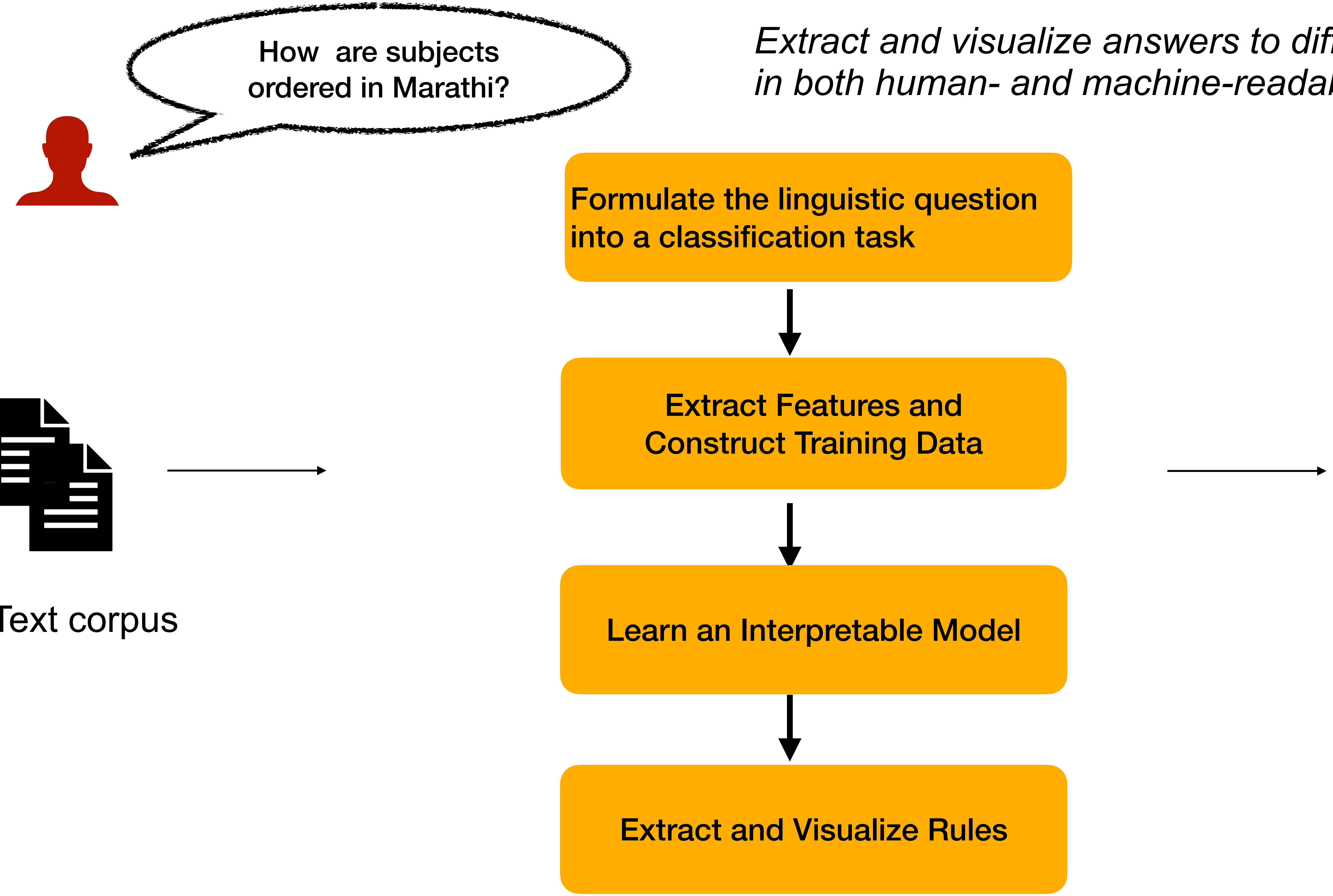
subject is **after** verb when:

verb is also governing= काय (kaay)
subject is nearby= compound
subject is governed by a word with Aspect = Simp
(Examples)
OR

Carnegie Mellon University

AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats



How are subjects ordered in Marathi?

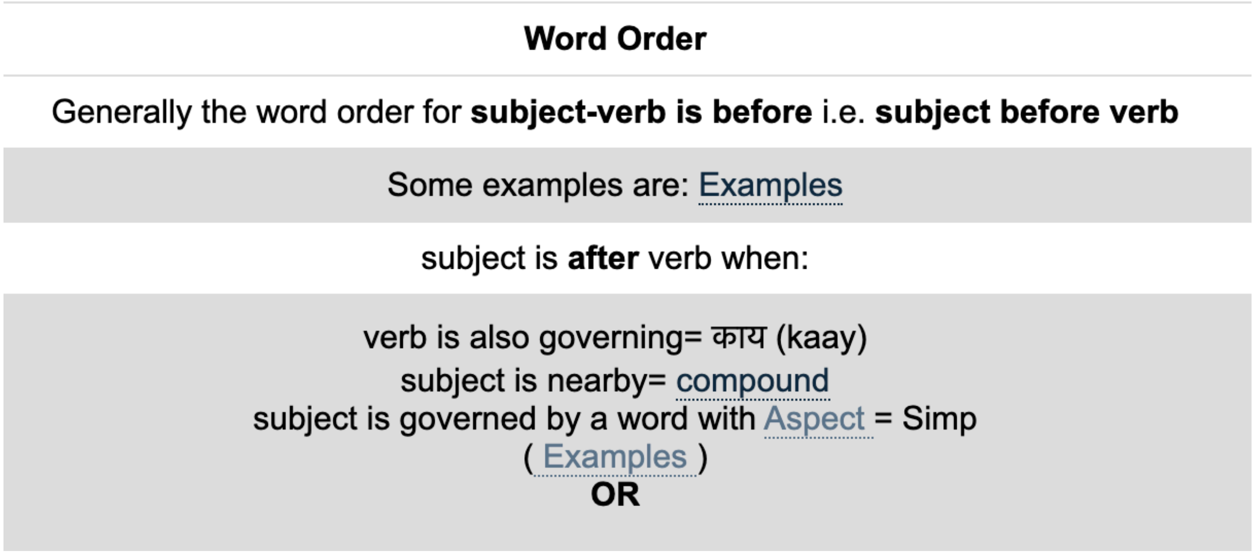
Text corpus

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules



Word Order

Generally the word order for **subject-verb is before** i.e. **subject before verb**

Some examples are: Examples

subject is **after** verb when:

verb is also governing= काय (kaay)
subject is nearby= compound
subject is governed by a word with Aspect = Simp
(Examples)
OR

AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats

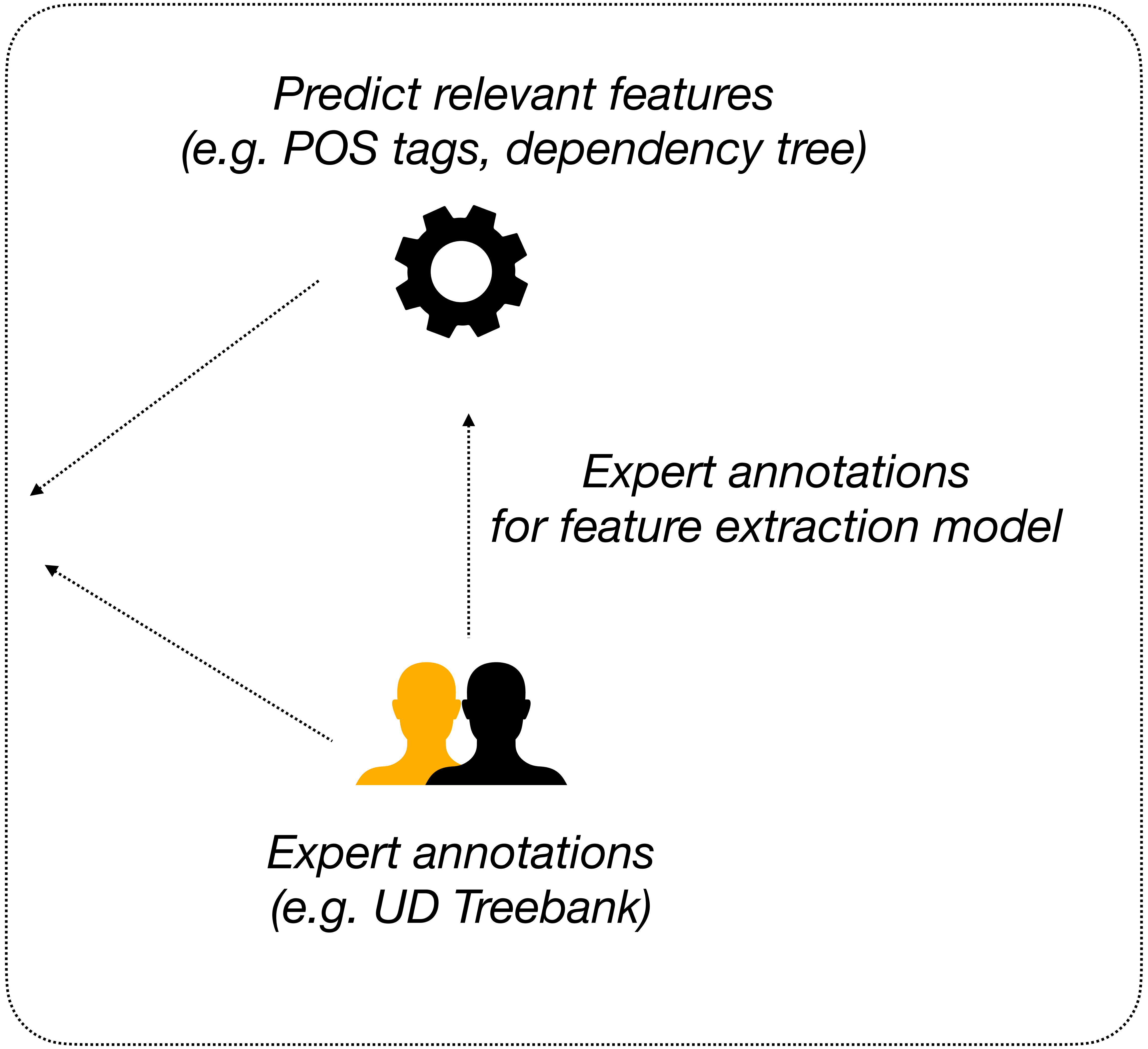
How are subjects ordered in Marathi?

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules



AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats

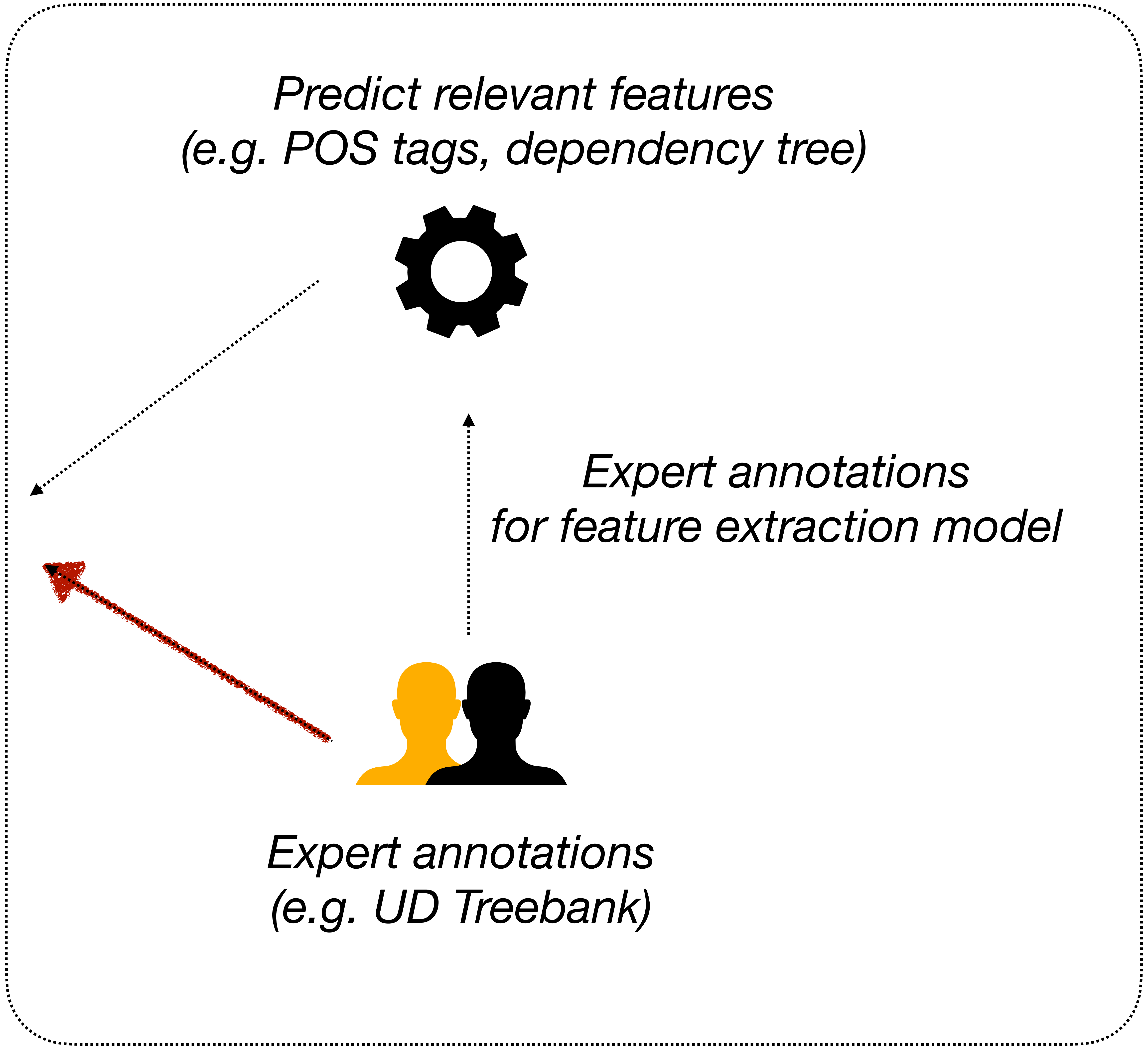
How are subjects ordered in Marathi?

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules



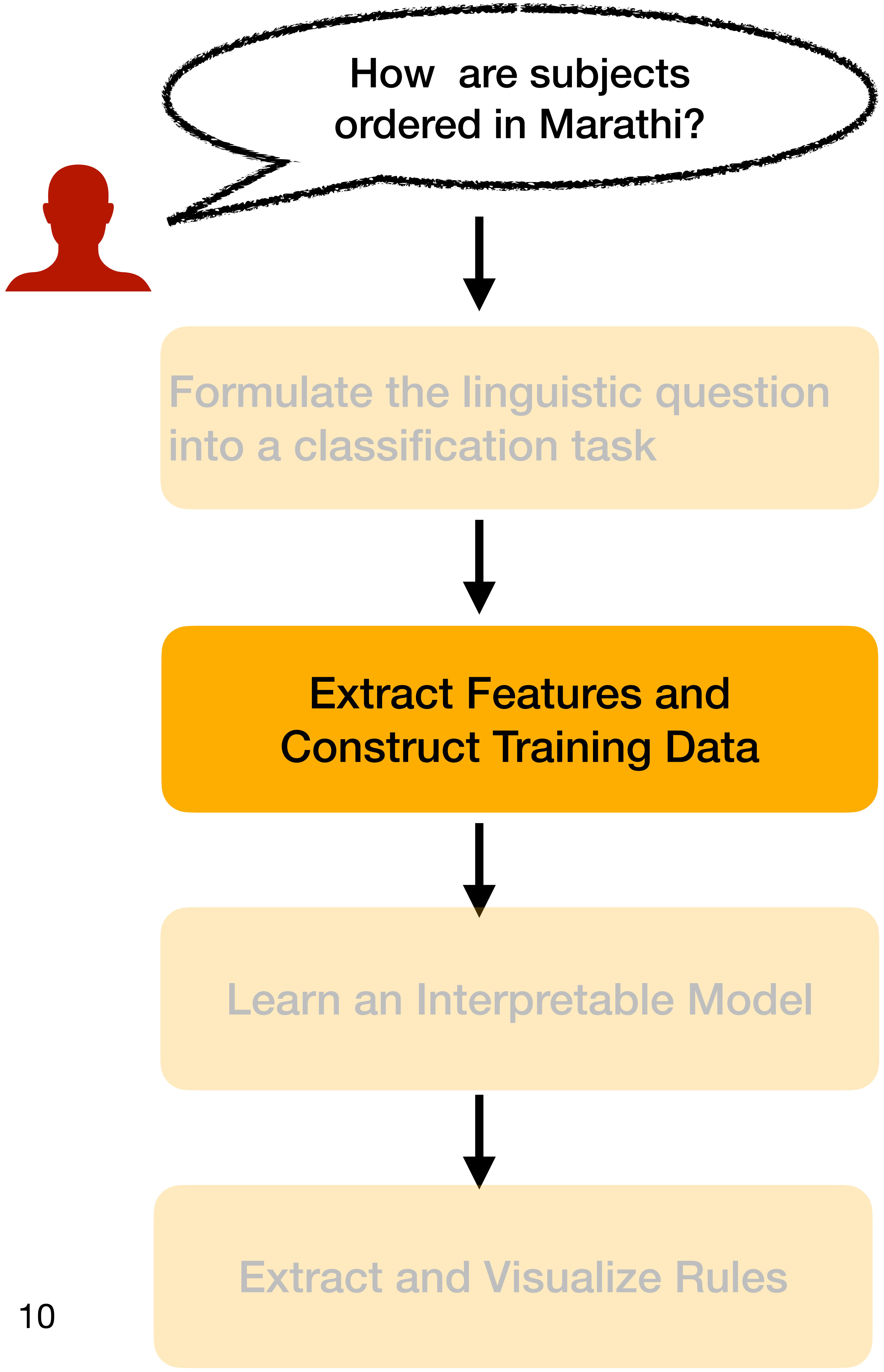
Predict relevant features (e.g. POS tags, dependency tree)

Expert annotations for feature extraction model

Expert annotations (e.g. UD Treebank)

AutoLEX: Automatic Language Explorer

Extract and visualize answers to different linguistic questions in both human- and machine-readable formats



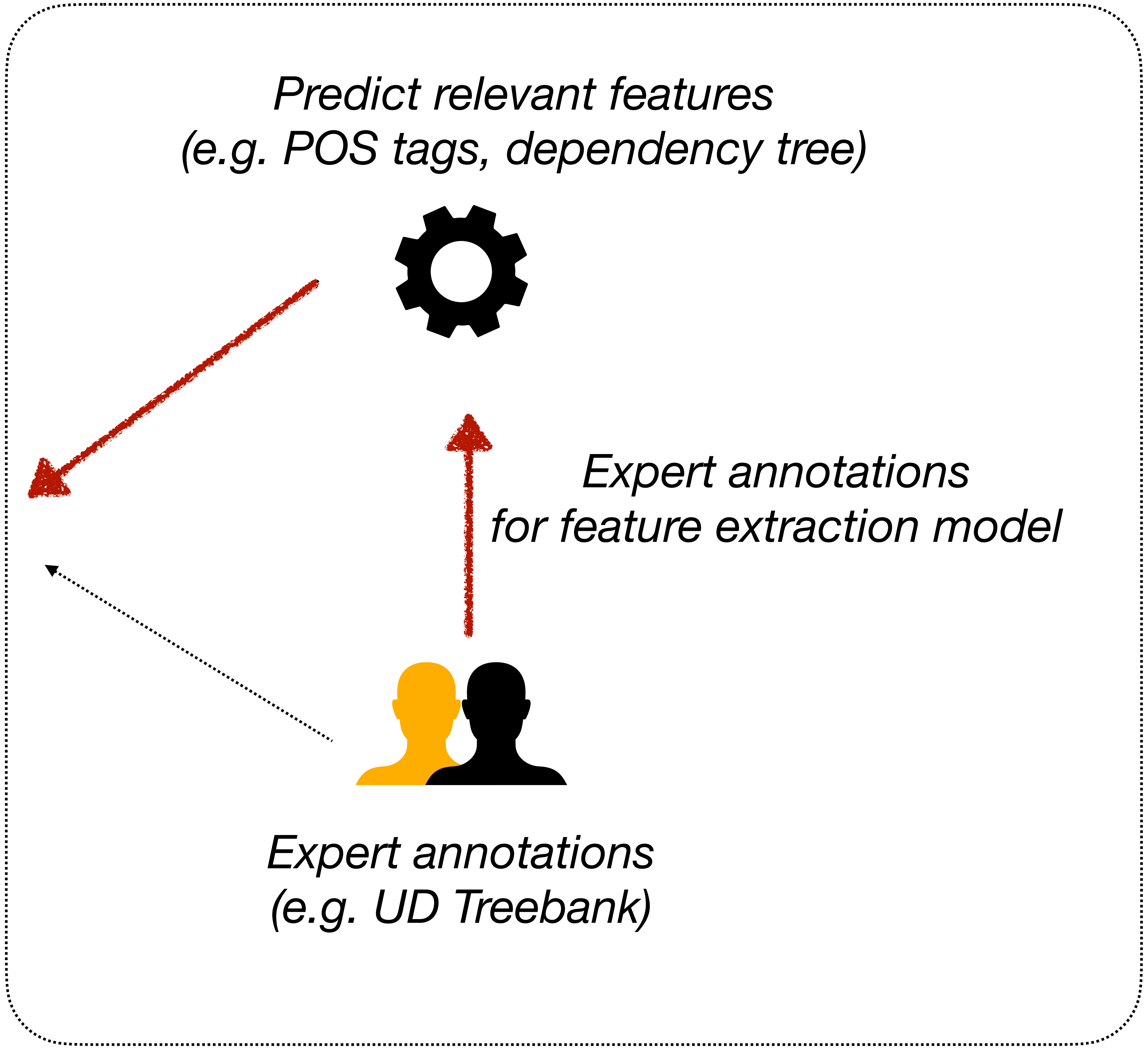
How are subjects ordered in Marathi?

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules



Predict relevant features (e.g. POS tags, dependency tree)

Expert annotations for feature extraction model

Expert annotations (e.g. UD Treebank)

Carnegie Mellon University

Types of Grammar Aspects covered in AutoLEX

Types of Grammar Aspects covered in AutoLEX

Agreement

Types of Grammar Aspects covered in AutoLEX

Agreement

Word Order

Types of Grammar Aspects covered in AutoLEX

Agreement

Word Order

Affix Usage

Types of Grammar Aspects covered in AutoLEX

Agreement

Word Order

Affix Usage

Case Marking

Types of Grammar Aspects covered in AutoLEX

Agreement

Word Order

Affix Usage

Case Marking

Word Usage

Types of Grammar Aspects covered in AutoLEX

Agreement

Word Order

Affix Usage

Case Marking

Word Usage

Types of Grammar Aspects covered in AutoLEX

Morpho-Syntax

- Agreement
- Word Order
- Affix Usage
- Case Marking

- Word Usage

Types of Grammar Aspects covered in AutoLEX

Morpho-Syntax

Agreement

Word Order

Affix Usage

Case Marking

Word Usage

Types of Grammar Aspects covered in AutoLEX

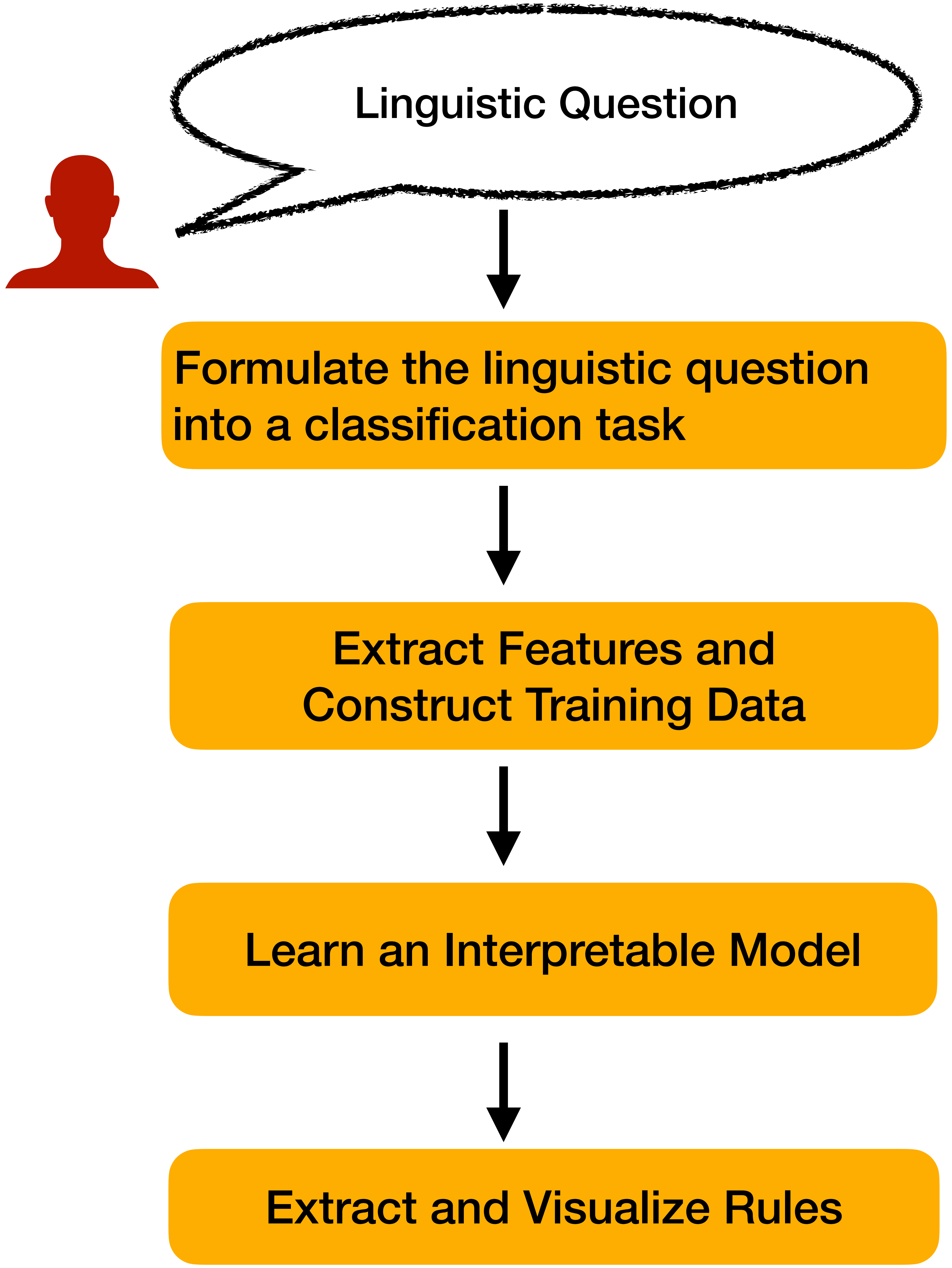
Morpho-Syntax

- Agreement
- Word Order
- Affix Usage
- Case Marking

Lexical Semantics

- Word Usage

Types of Grammar Aspects covered in AutoLEX



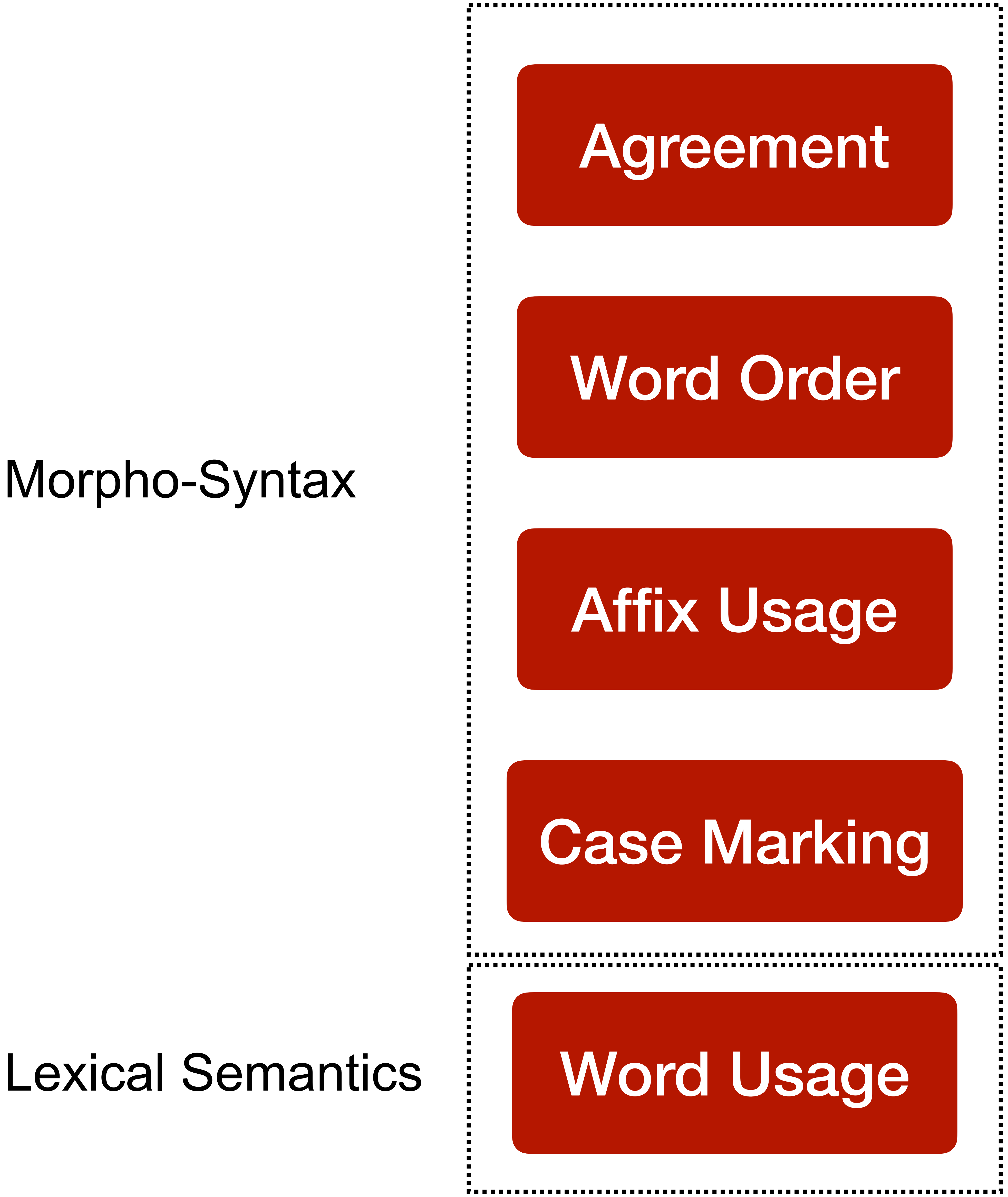
Linguistic Question

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules



Morpho-Syntax

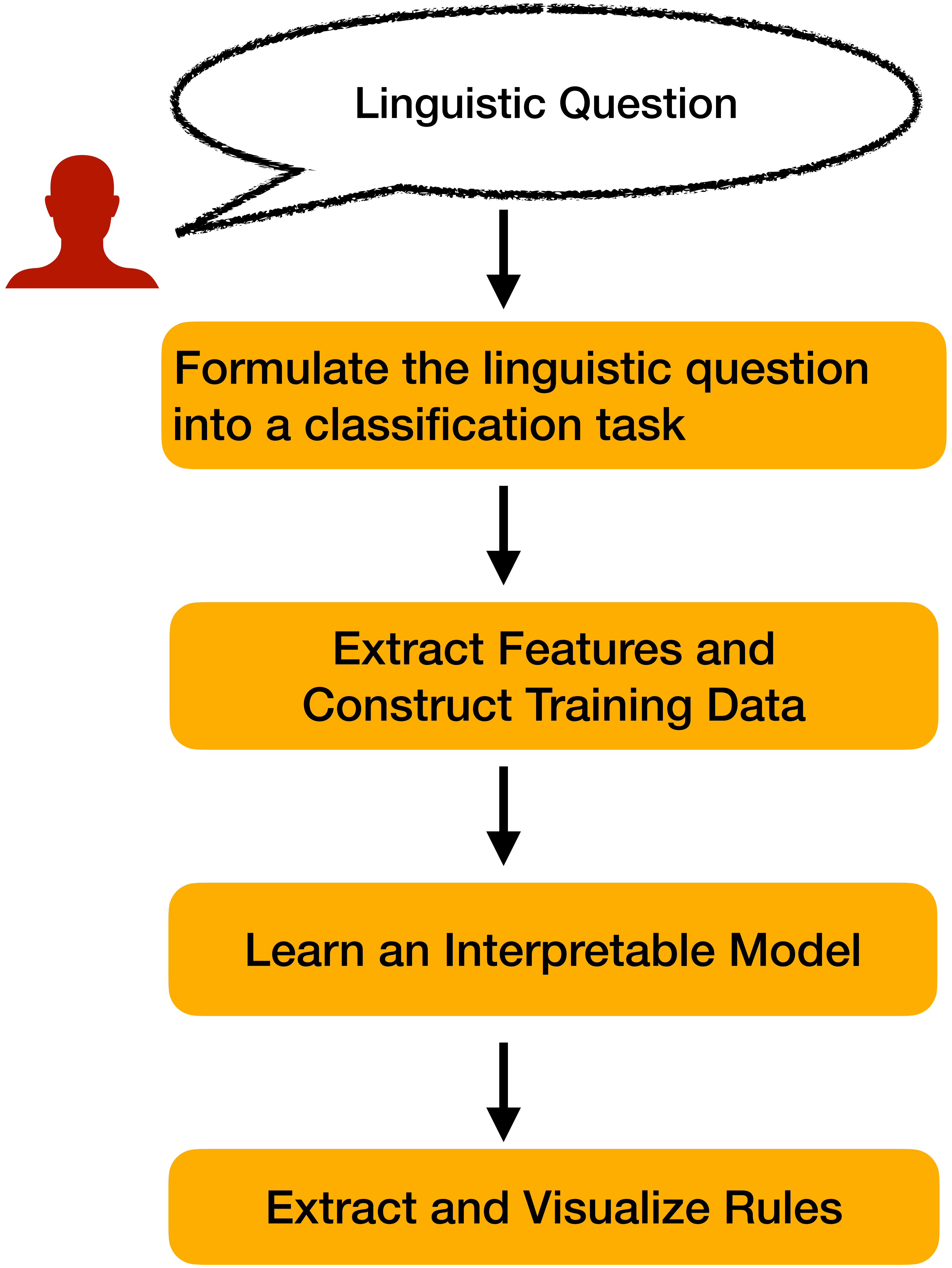
- Agreement
- Word Order
- Affix Usage
- Case Marking

Lexical Semantics

- Word Usage

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Types of Grammar Aspects covered in AutoLEX



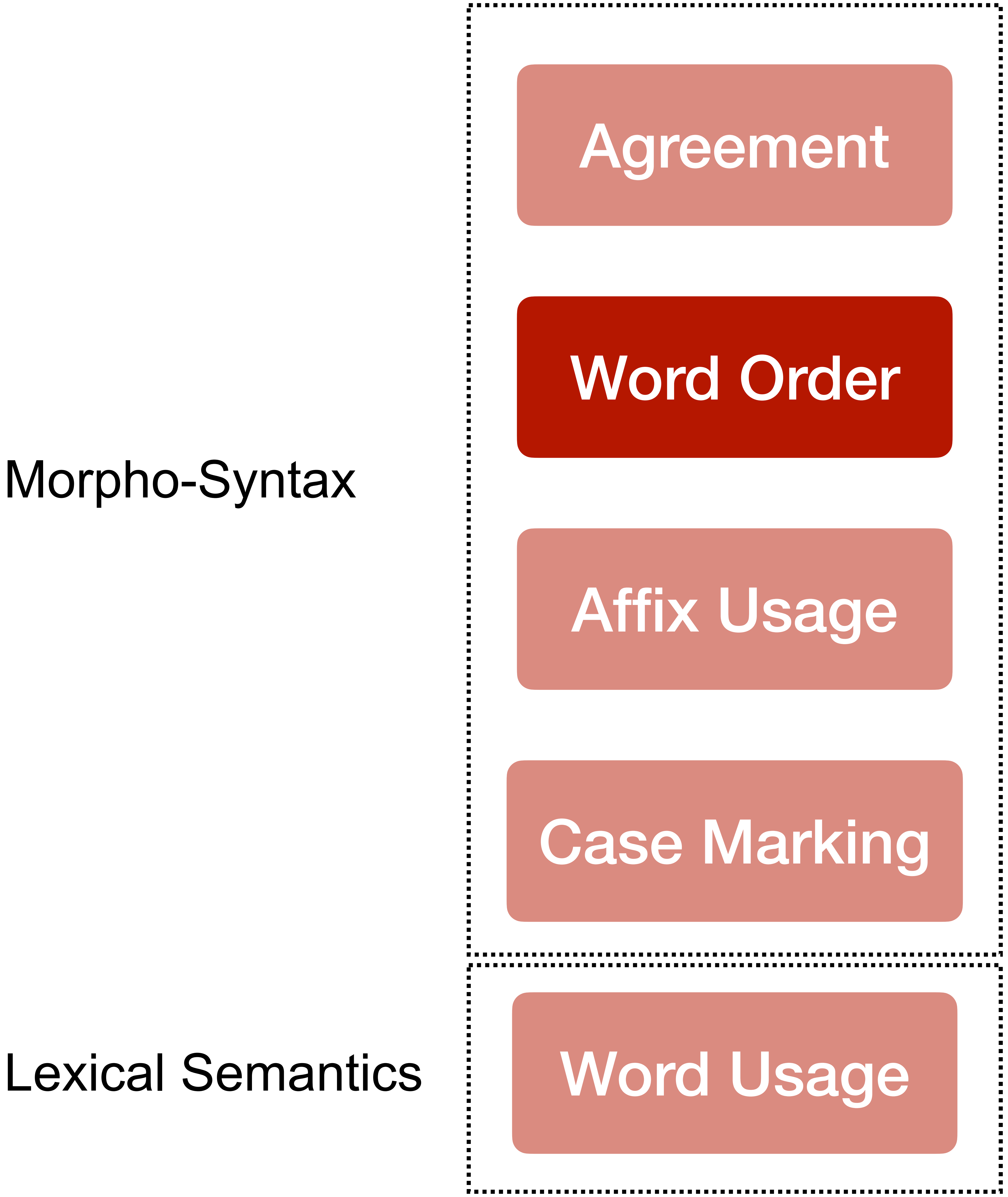
Linguistic Question

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules



Morpho-Syntax

Agreement

Word Order

Affix Usage

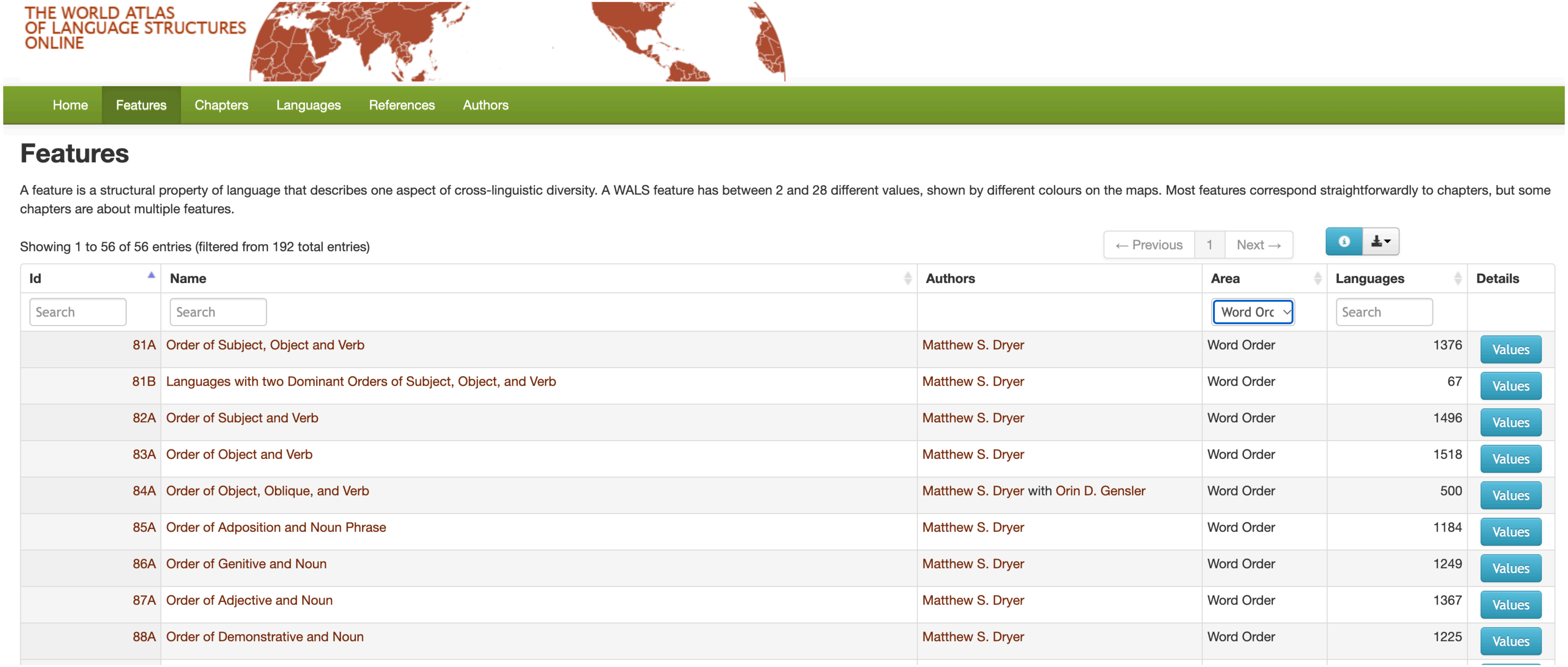
Case Marking

Lexical Semantics

Word Usage

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order



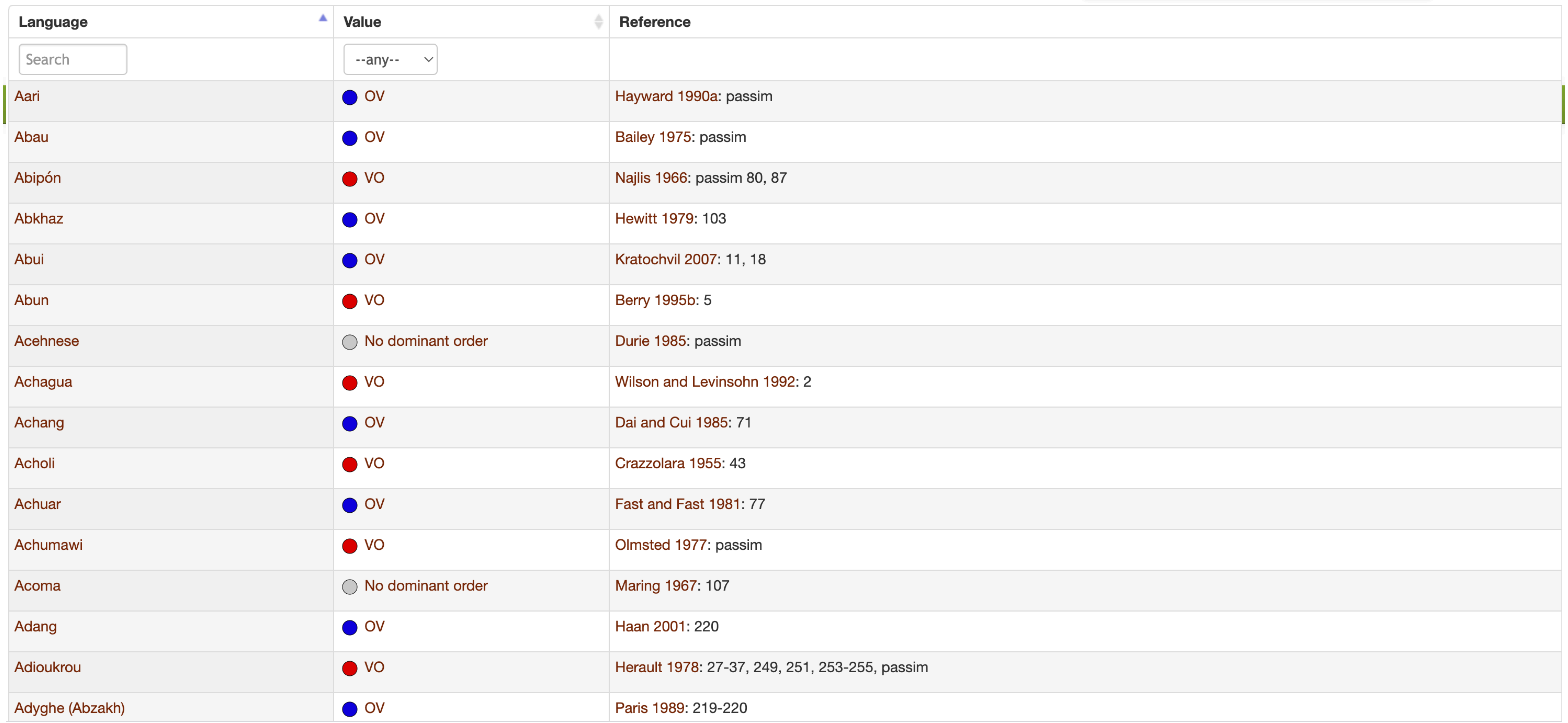
Features

A feature is a structural property of language that describes one aspect of cross-linguistic diversity. A WALS feature has between 2 and 28 different values, shown by different colours on the maps. Most features correspond straightforwardly to chapters, but some chapters are about multiple features.

Showing 1 to 56 of 56 entries (filtered from 192 total entries)

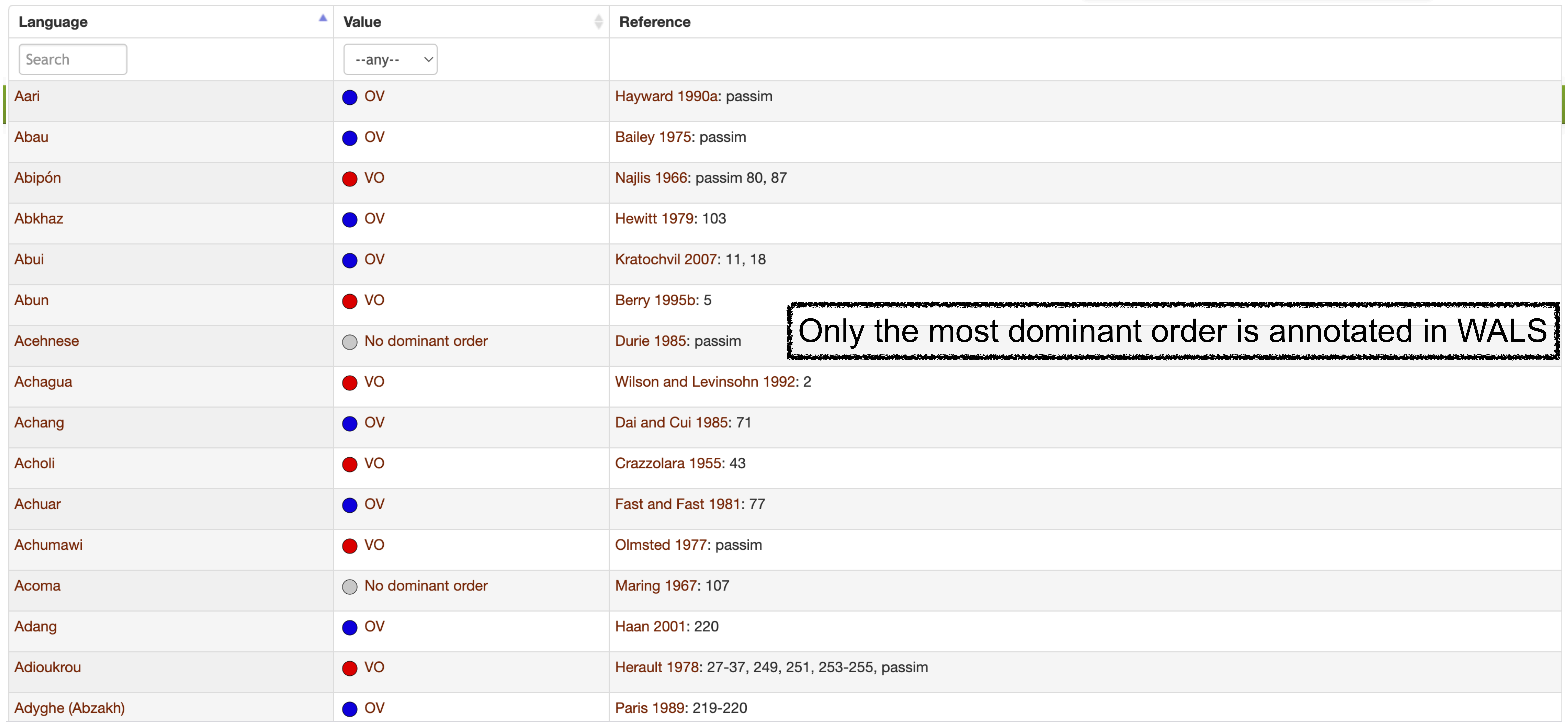
Id	Name	Authors	Area	Languages	Details
81A	Order of Subject, Object and Verb	Matthew S. Dryer	Word Order	1376	Values
81B	Languages with two Dominant Orders of Subject, Object, and Verb	Matthew S. Dryer	Word Order	67	Values
82A	Order of Subject and Verb	Matthew S. Dryer	Word Order	1496	Values
83A	Order of Object and Verb	Matthew S. Dryer	Word Order	1518	Values
84A	Order of Object, Oblique, and Verb	Matthew S. Dryer with Orin D. Gensler	Word Order	500	Values
85A	Order of Adposition and Noun Phrase	Matthew S. Dryer	Word Order	1184	Values
86A	Order of Genitive and Noun	Matthew S. Dryer	Word Order	1249	Values
87A	Order of Adjective and Noun	Matthew S. Dryer	Word Order	1367	Values
88A	Order of Demonstrative and Noun	Matthew S. Dryer	Word Order	1225	Values

Word Order



Language	Value	Reference
Search	--any--	
Aari	🔵 OV	Hayward 1990a: passim
Abau	🔵 OV	Bailey 1975: passim
Abipón	🔴 VO	Najlis 1966: passim 80, 87
Abkhaz	🔵 OV	Hewitt 1979: 103
Abui	🔵 OV	Kratochvil 2007: 11, 18
Abun	🔴 VO	Berry 1995b: 5
Acehnese	⚪ No dominant order	Durie 1985: passim
Achagua	🔴 VO	Wilson and Levinsohn 1992: 2
Achang	🔵 OV	Dai and Cui 1985: 71
Acholi	🔴 VO	Crazzolara 1955: 43
Achuar	🔵 OV	Fast and Fast 1981: 77
Achumawi	🔴 VO	Olmsted 1977: passim
Acoma	⚪ No dominant order	Maring 1967: 107
Adang	🔵 OV	Haan 2001: 220
Adioukrou	🔴 VO	Herault 1978: 27-37, 249, 251, 253-255, passim
Adyghe (Abzakh)	🔵 OV	Paris 1989: 219-220

Word Order



Language	Value	Reference
Search	--any--	
Aari	🔵 OV	Hayward 1990a: passim
Abau	🔵 OV	Bailey 1975: passim
Abipón	🔴 VO	Najlis 1966: passim 80, 87
Abkhaz	🔵 OV	Hewitt 1979: 103
Abui	🔵 OV	Kratochvil 2007: 11, 18
Abun	🔴 VO	Berry 1995b: 5
Acehnese	⚪ No dominant order	Durie 1985: passim
Achagua	🔴 VO	Wilson and Levinsohn 1992: 2
Achang	🔵 OV	Dai and Cui 1985: 71
Acholi	🔴 VO	Crazzolara 1955: 43
Achuar	🔵 OV	Fast and Fast 1981: 77
Achumawi	🔴 VO	Olmsted 1977: passim
Acoma	⚪ No dominant order	Maring 1967: 107
Adang	🔵 OV	Haan 2001: 220
Adioukrou	🔴 VO	Herault 1978: 27-37, 249, 251, 253-255, passim
Adyghe (Abzakh)	🔵 OV	Paris 1989: 219-220

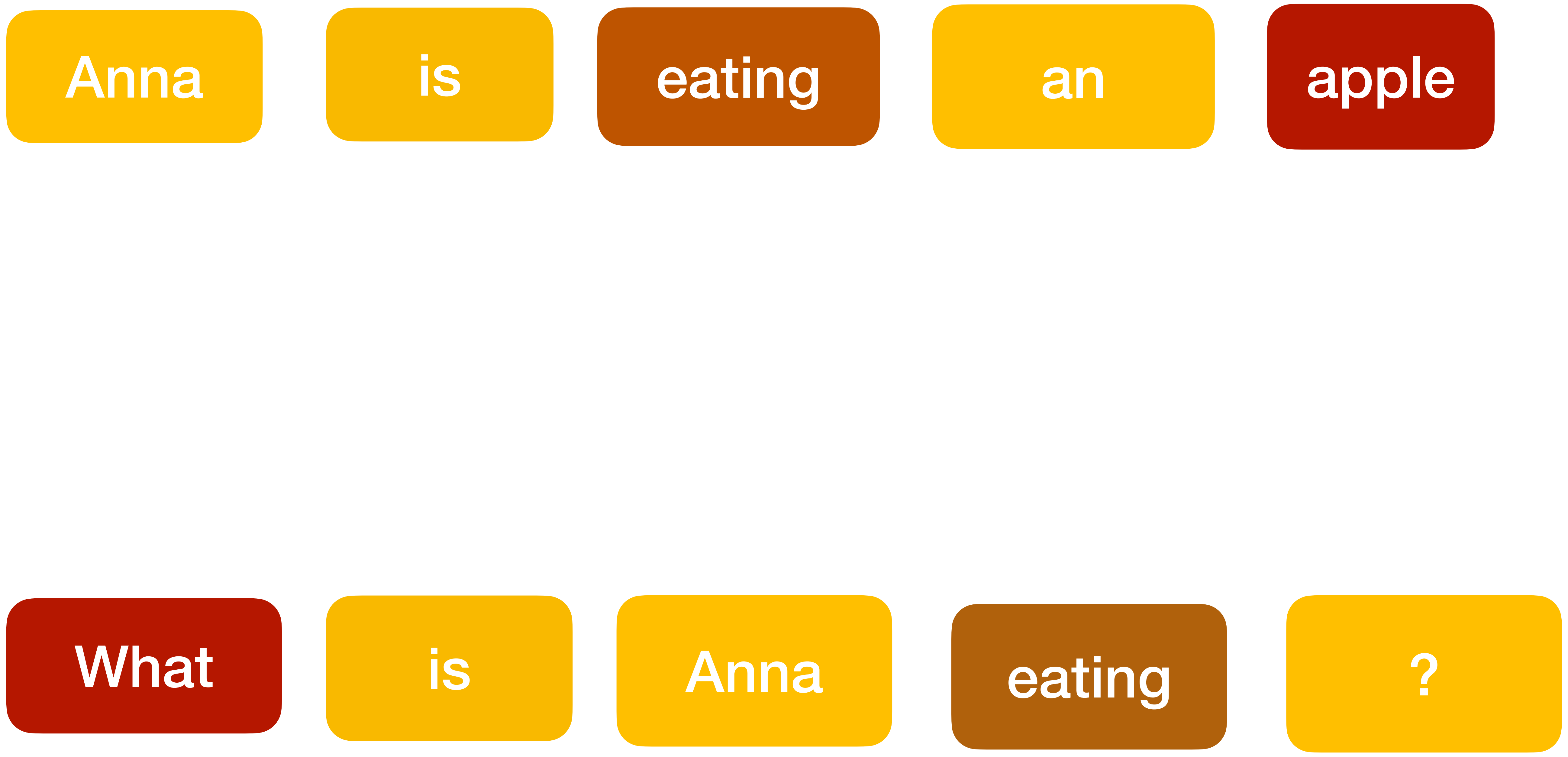
Only the most dominant order is annotated in WALS

Word Order

Word Order

Anna is eating an apple

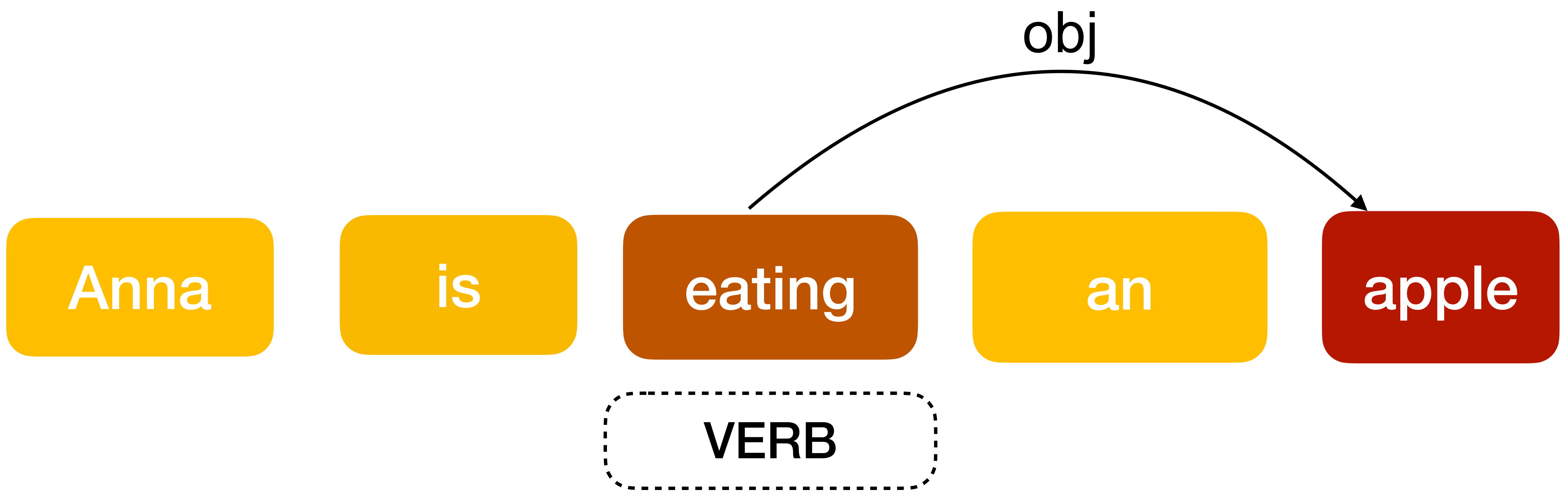
Word Order



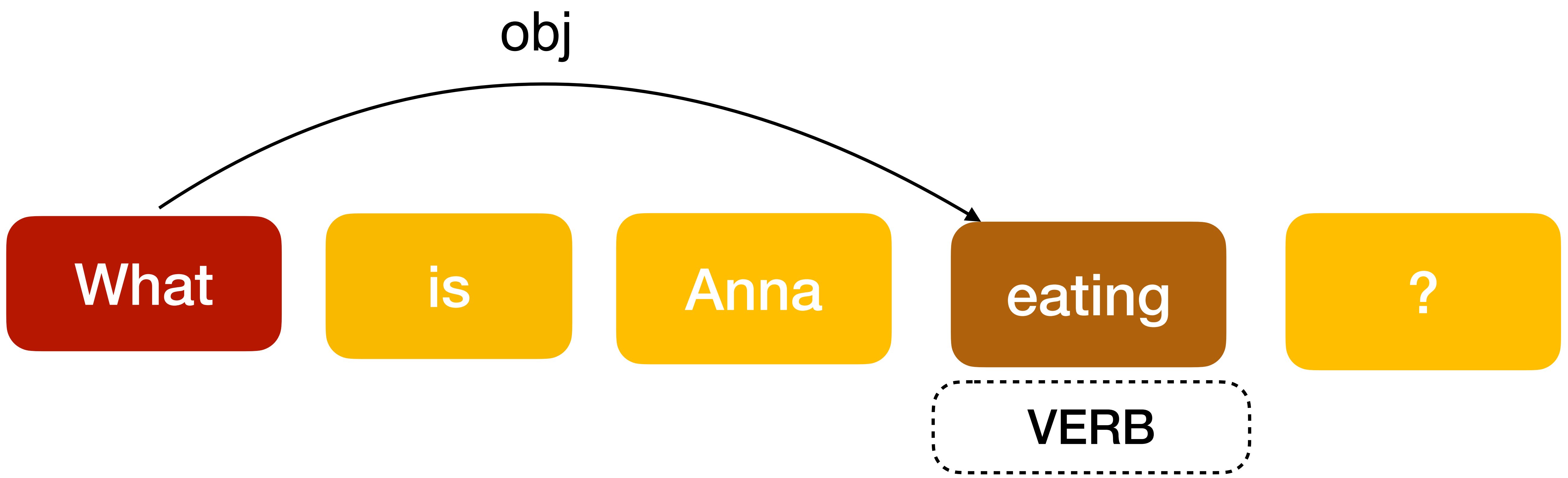
Anna is eating an apple

What is Anna eating ?

Word Order

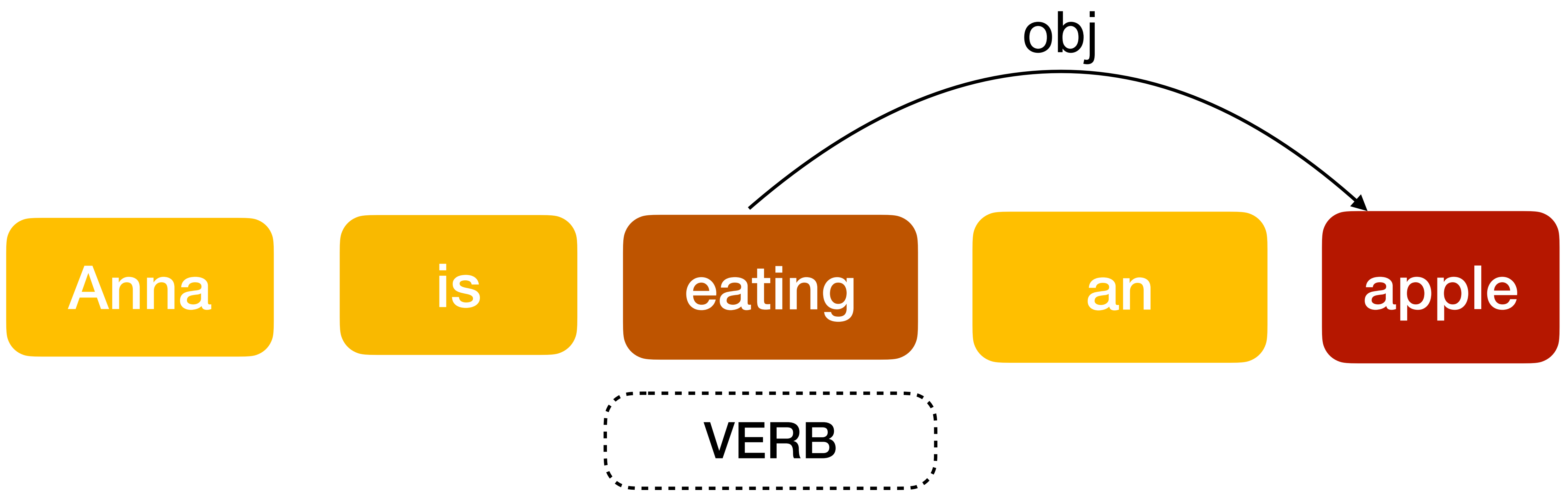


Verb-Object

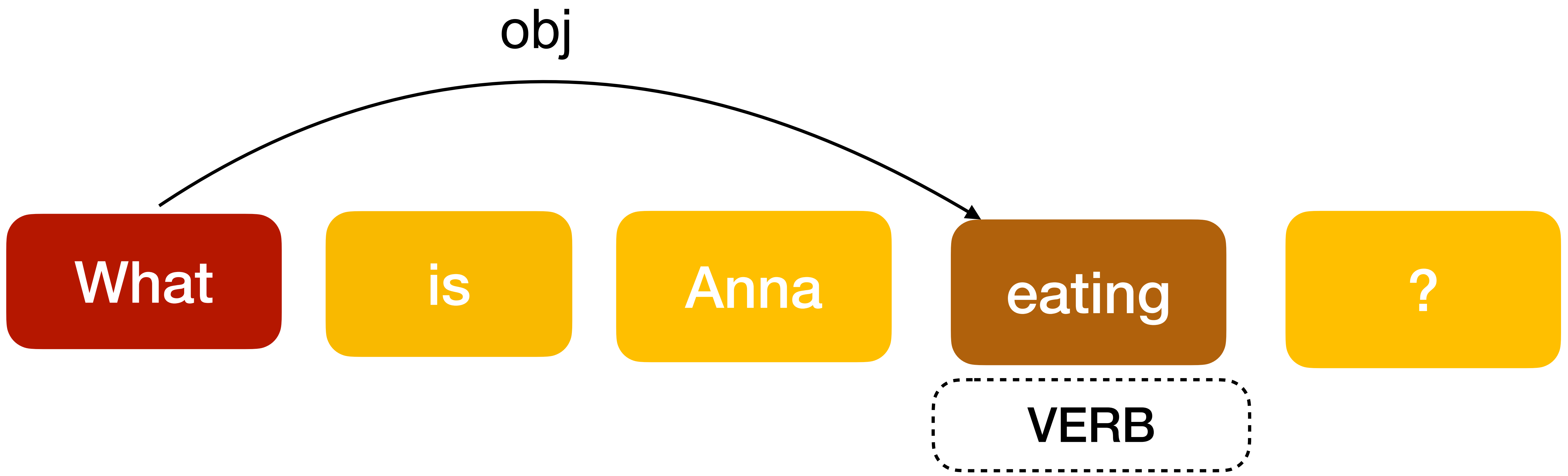


Object-Verb

Word Order



Verb-Object

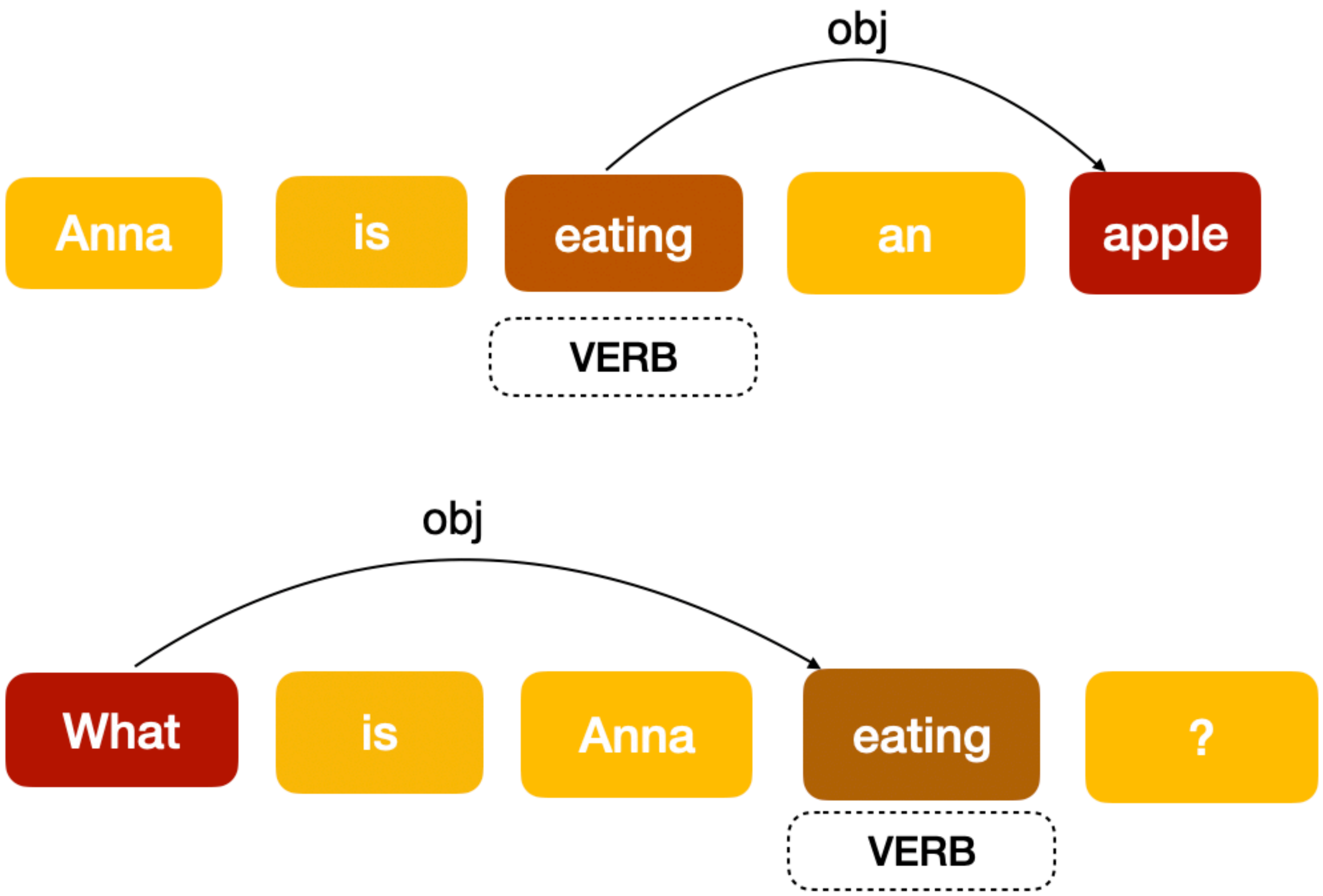


Object-Verb

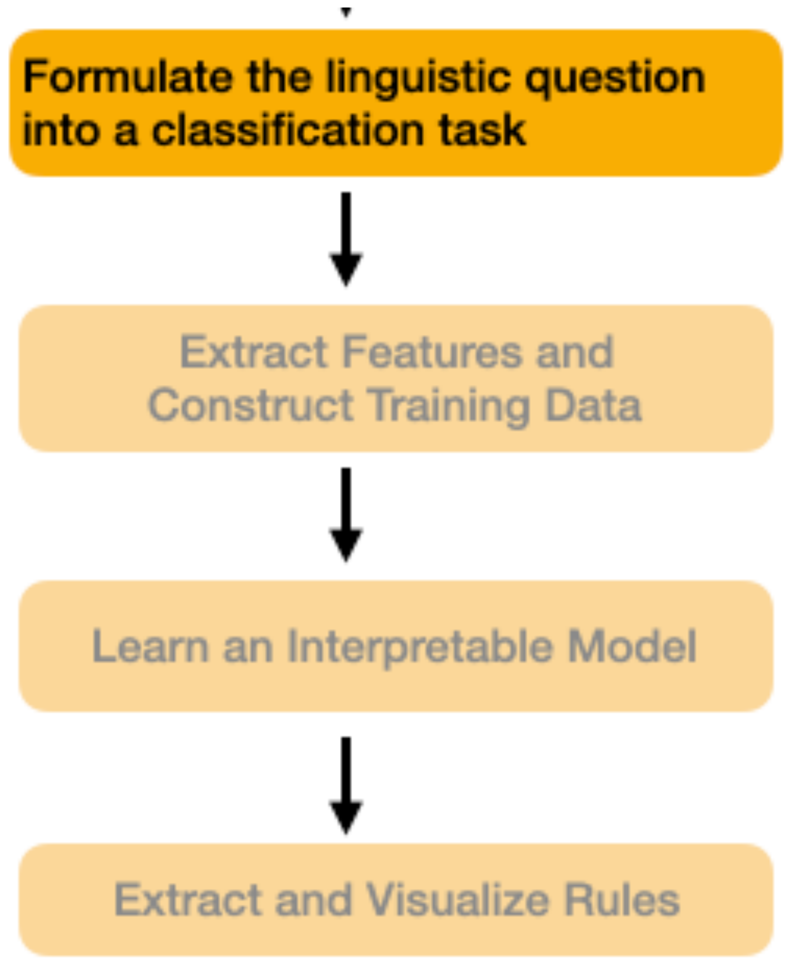
When are **objects** before or after the **verbs**?

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Formulate Linguistic Question



When are **objects** before or after the **verbs**?

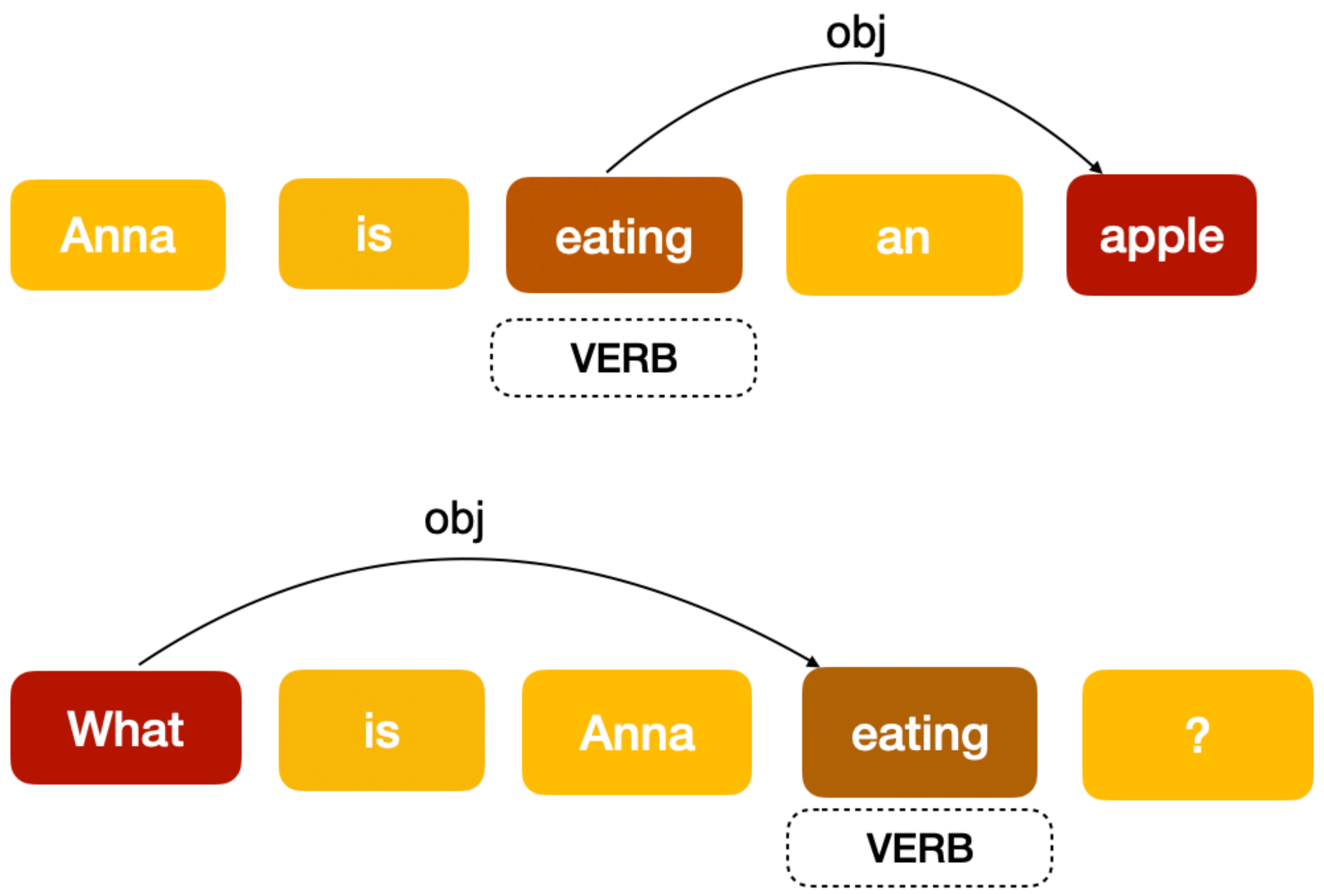

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

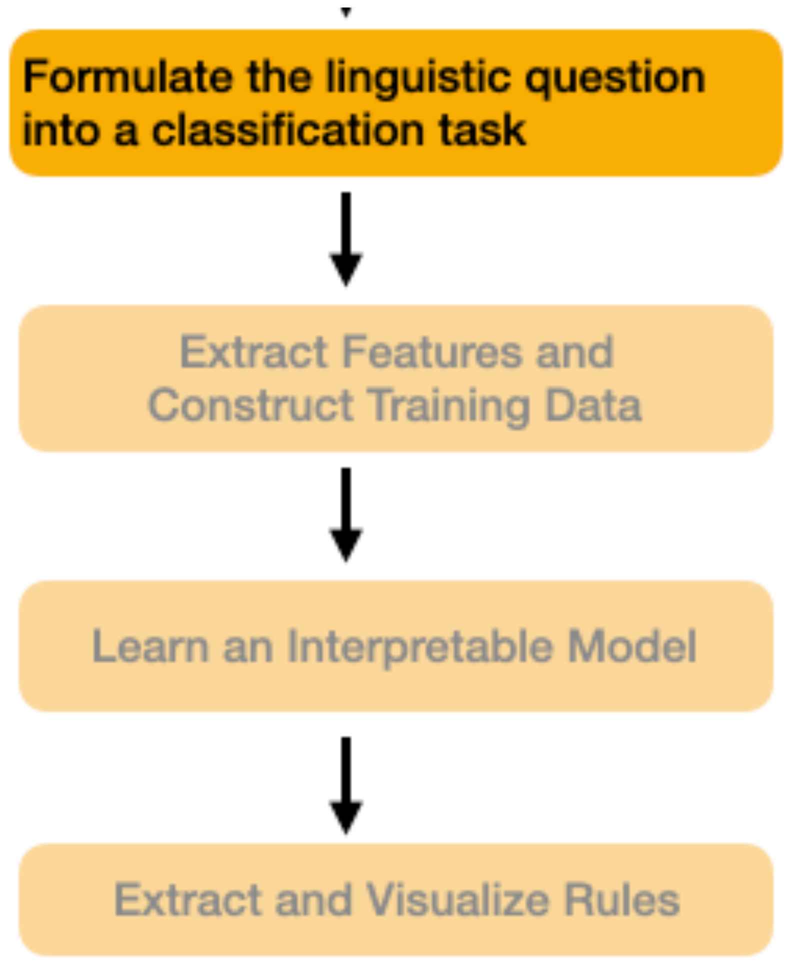
Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Formulate Linguistic Question



When are **objects** before or after the **verbs**?


Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

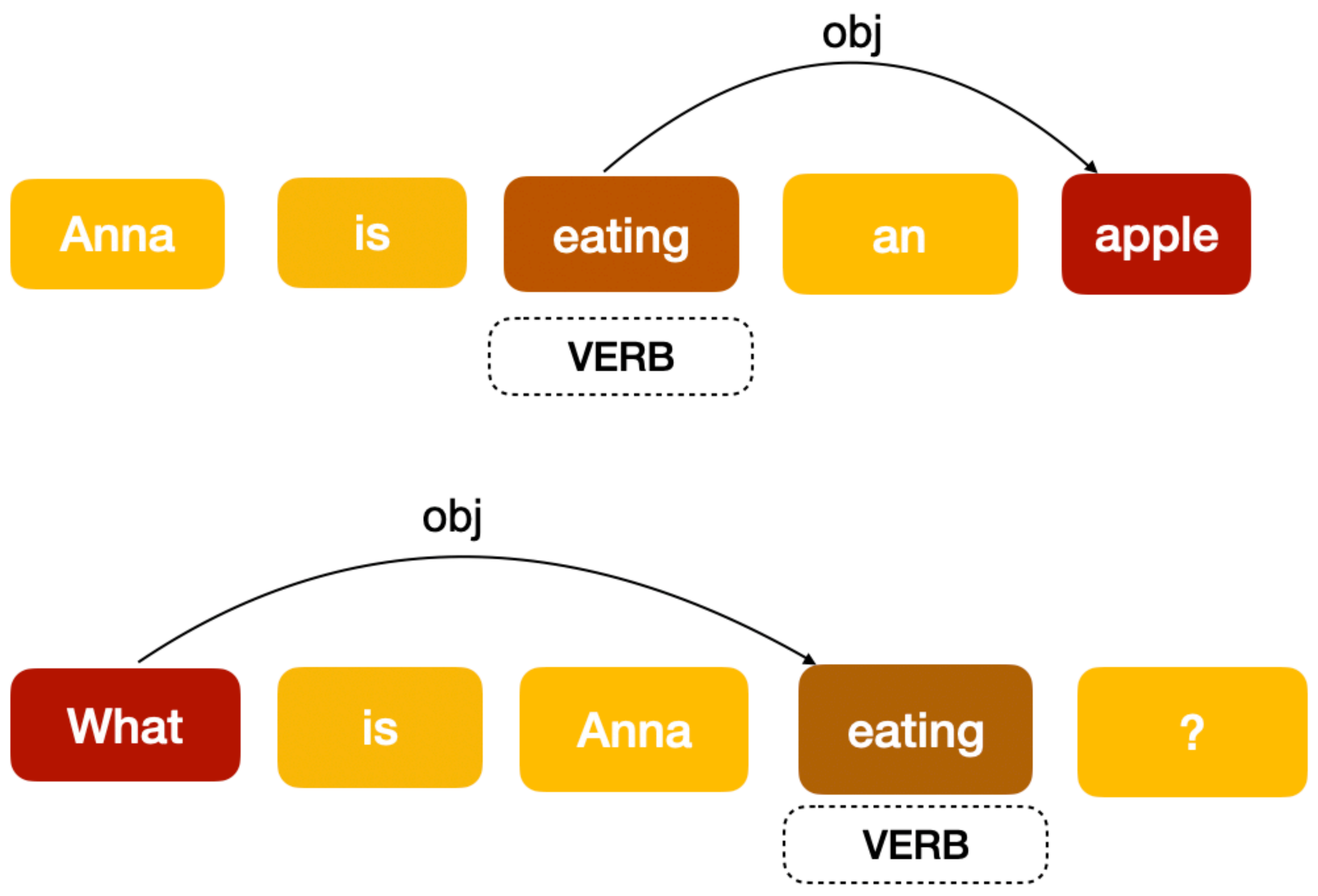
Learn an Interpretable Model

Extract and Visualize Rules

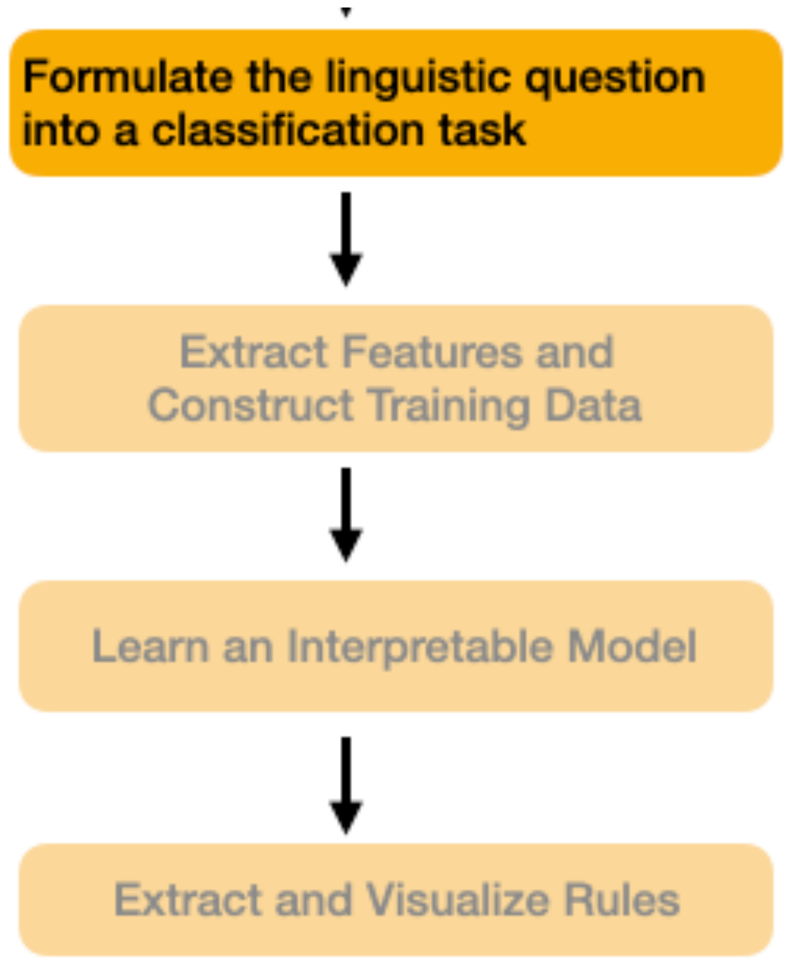
AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Carnegie Mellon University

Word Order: Formulate Linguistic Question



When are **objects** before or after the **verbs**?



Formulate the linguistic question into a classification task

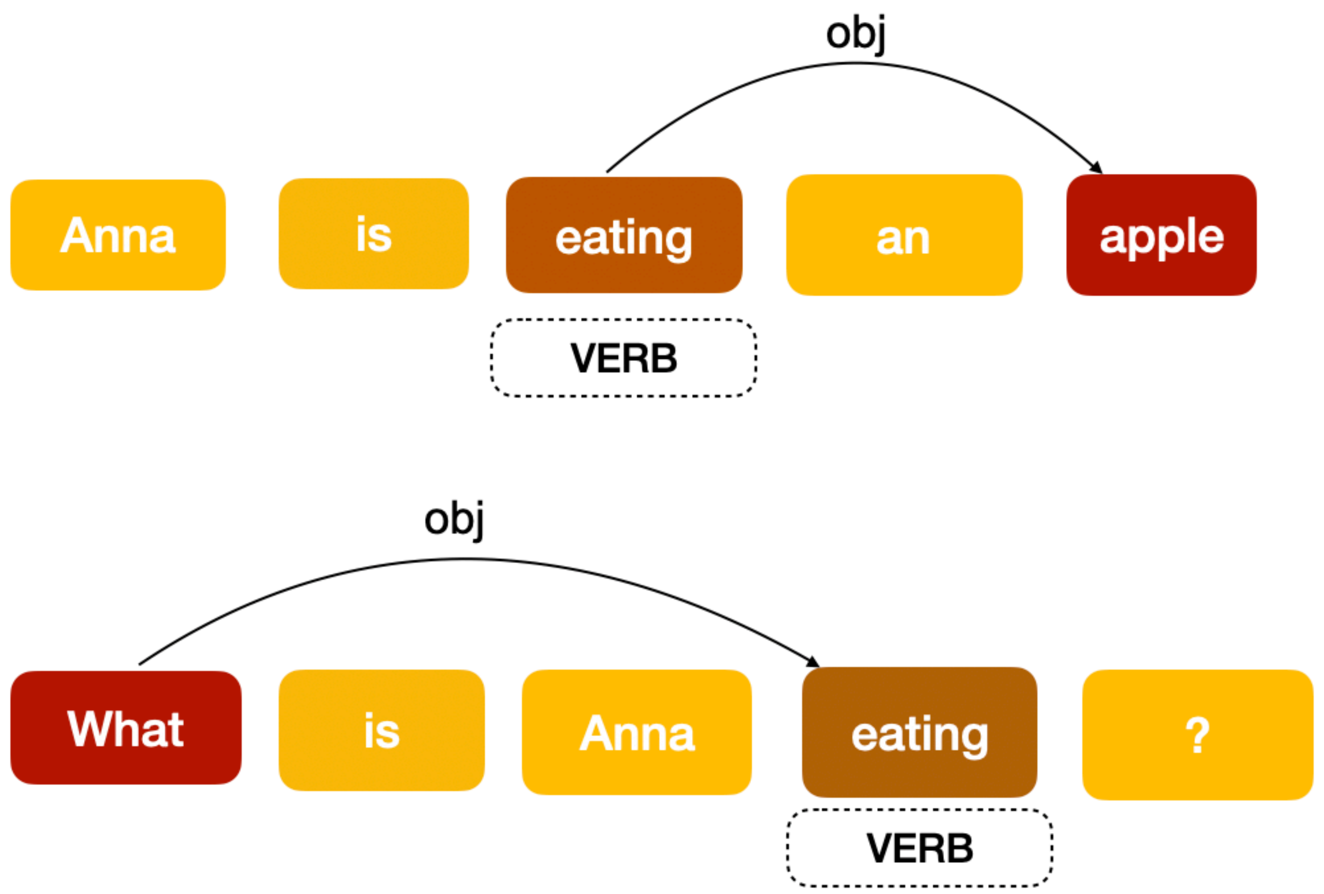
Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

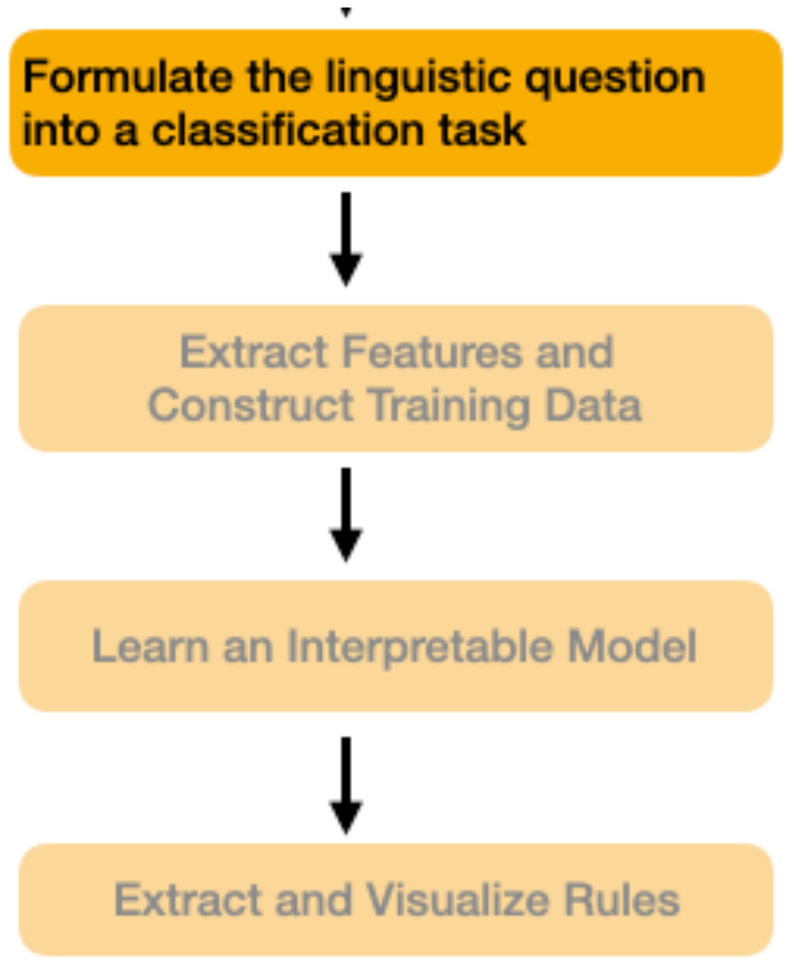
AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Formulate Linguistic Question

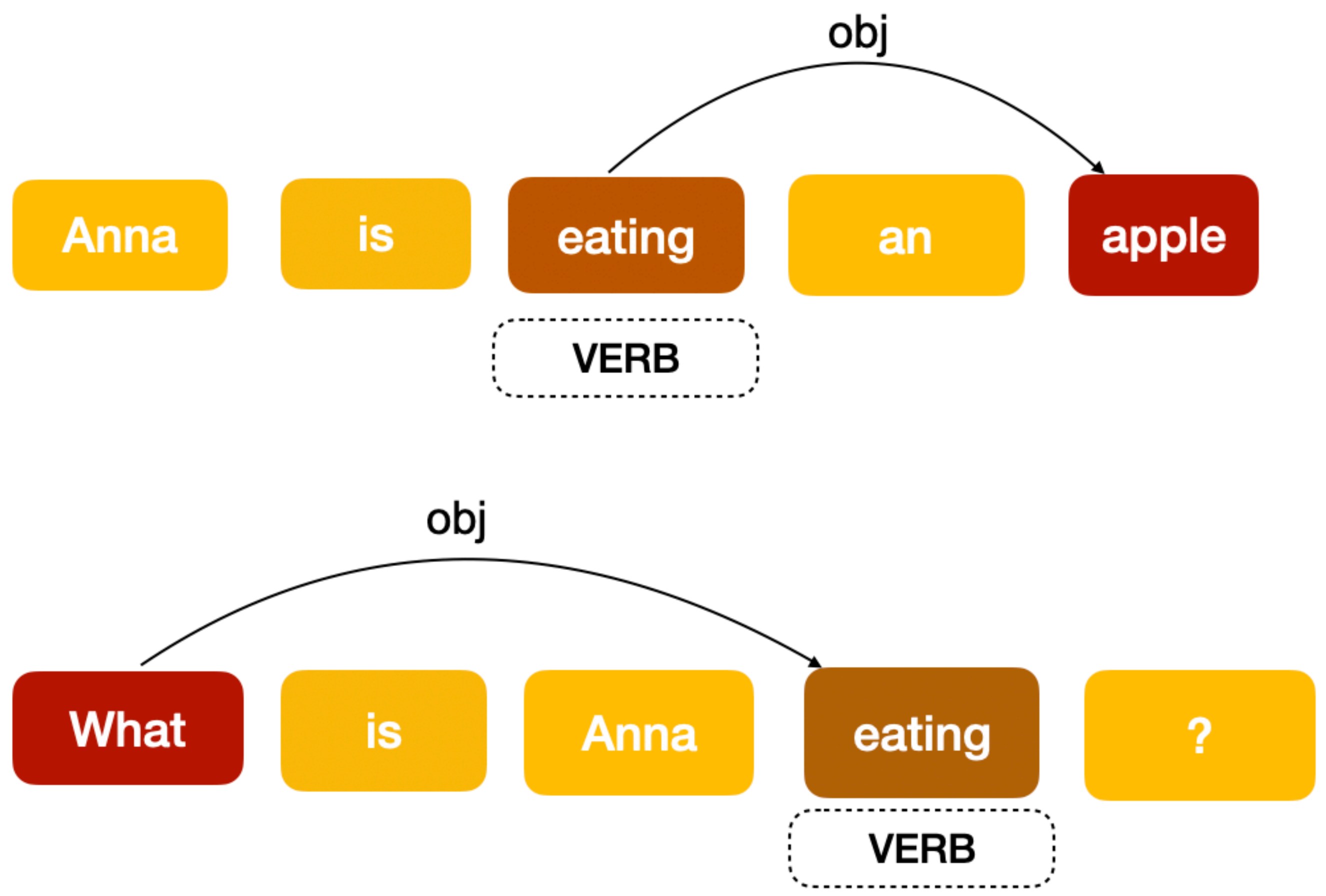


When are **objects** before or after the **verbs**?

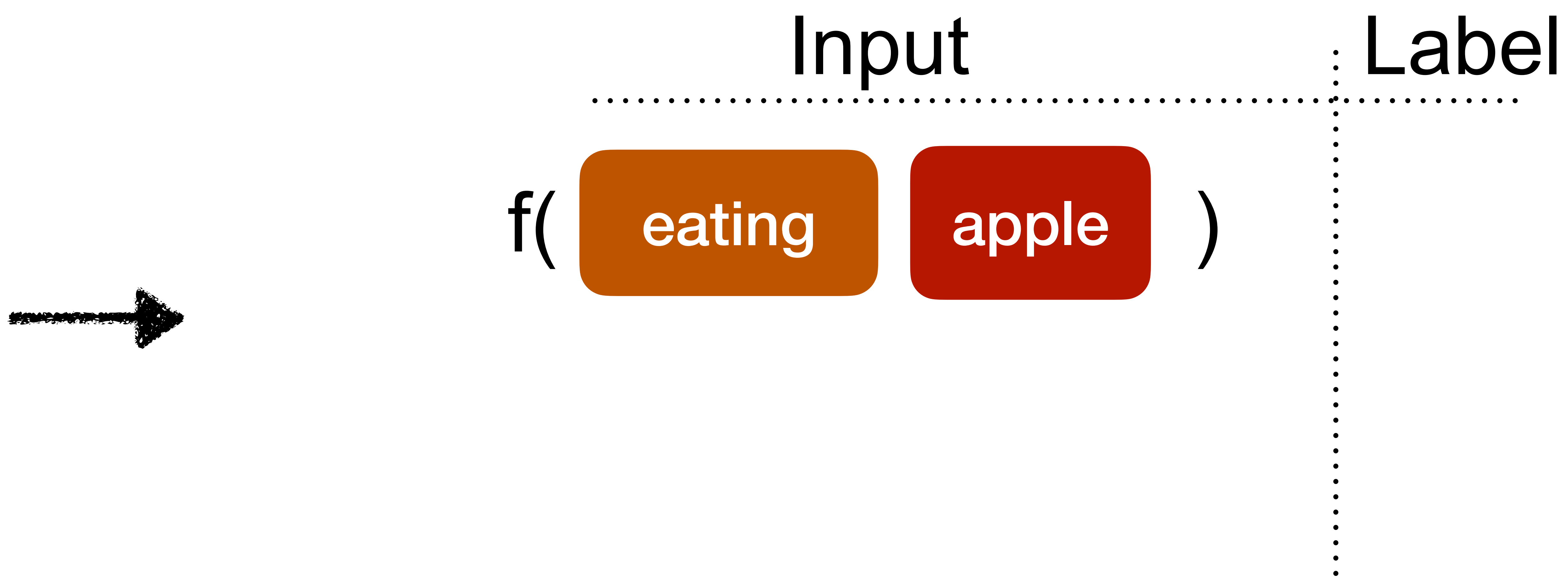
Input	Label

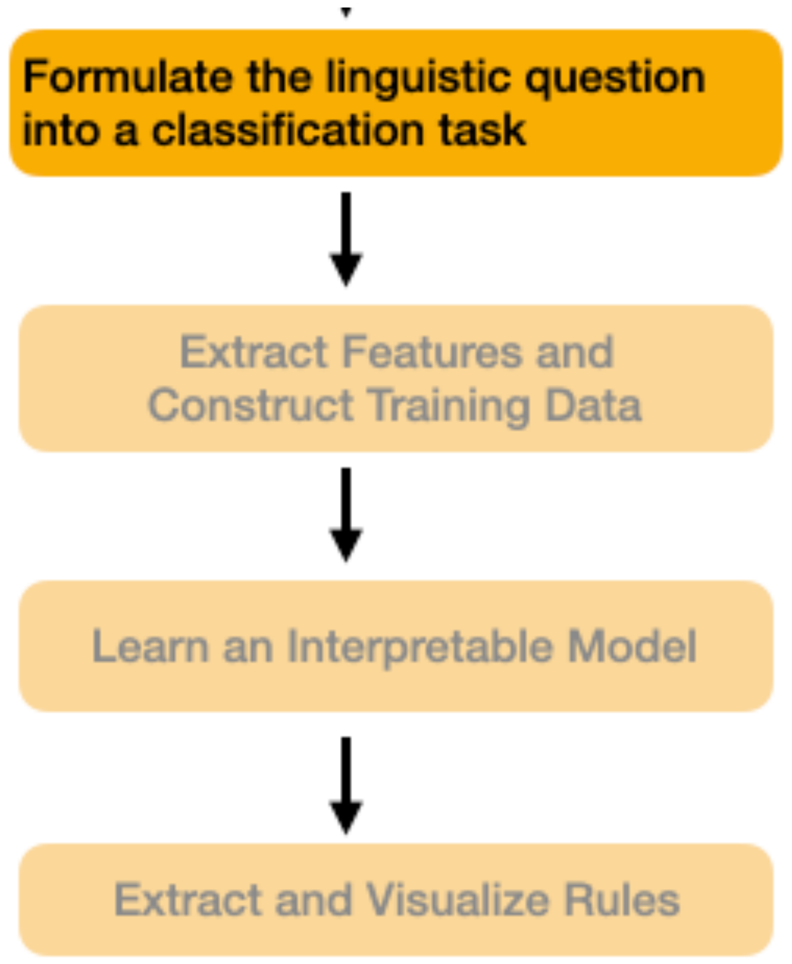


Word Order: Formulate Linguistic Question

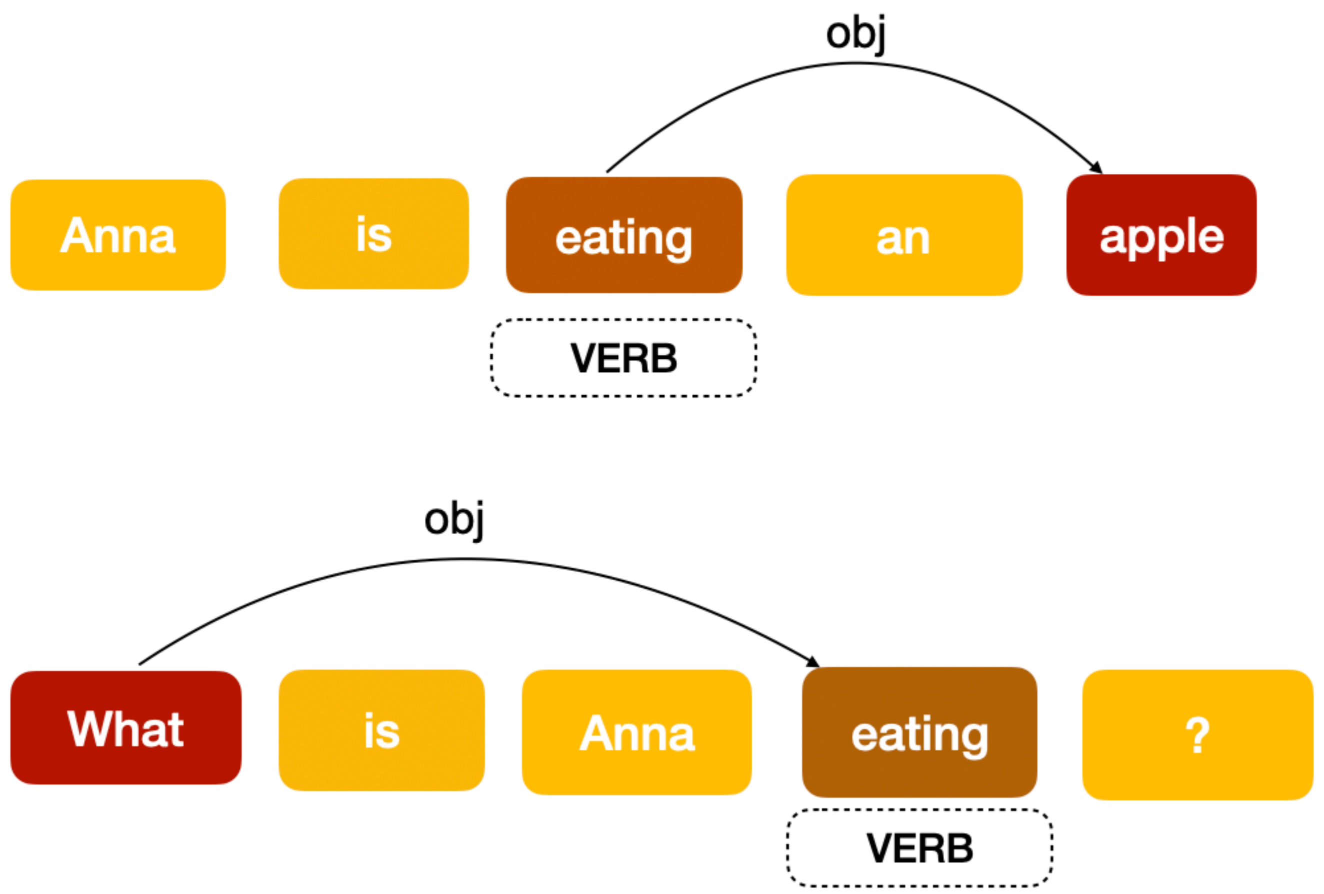


When are **objects** before or after the **verbs**?

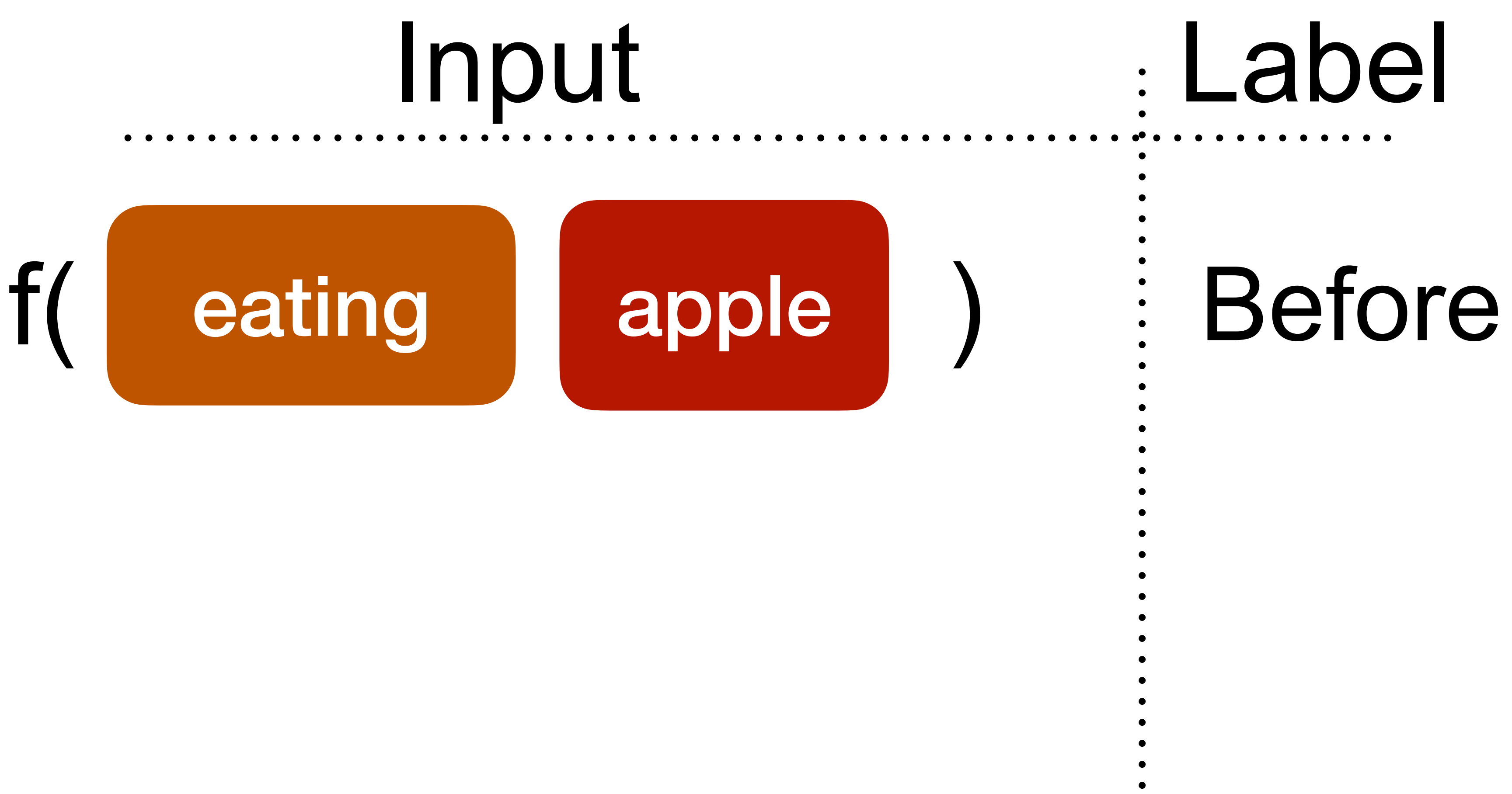


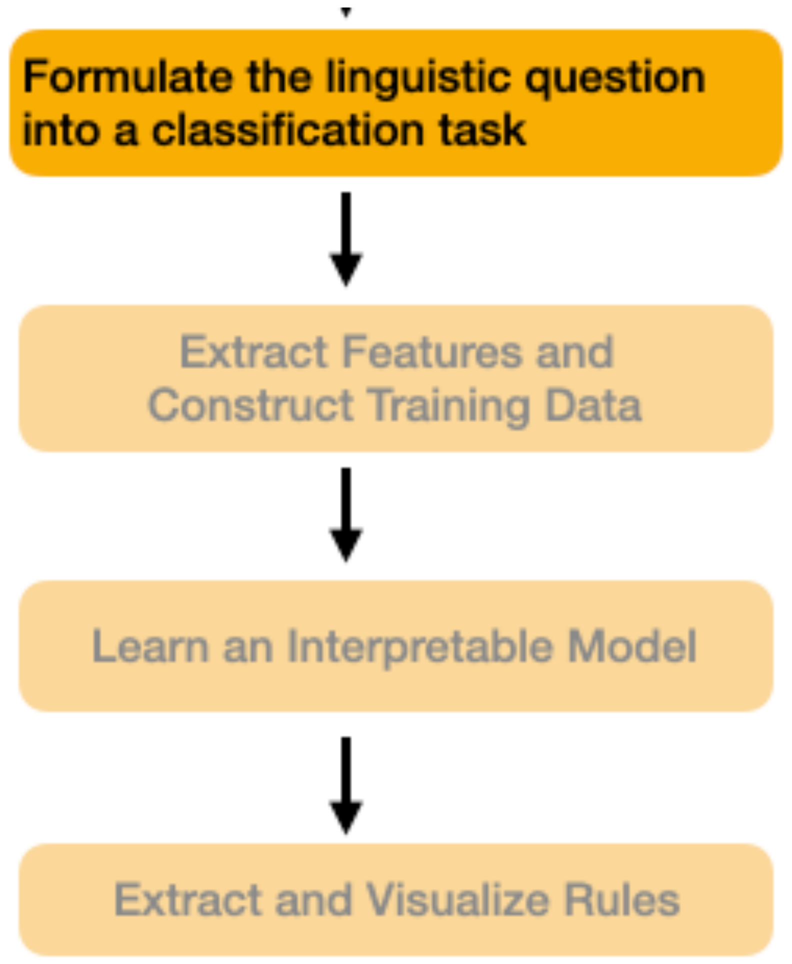


Word Order: Formulate Linguistic Question



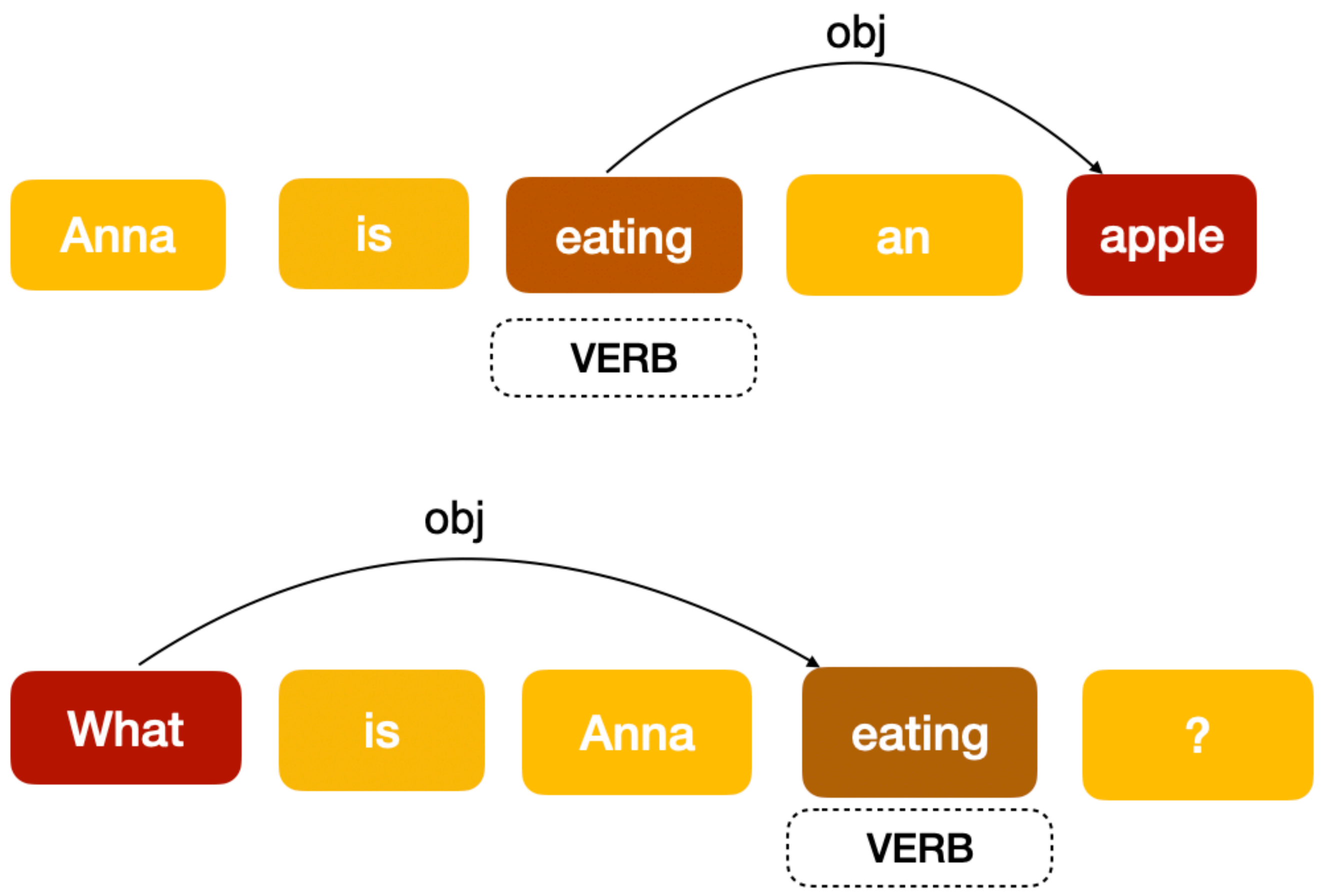
When are **objects** before or after the **verbs**?



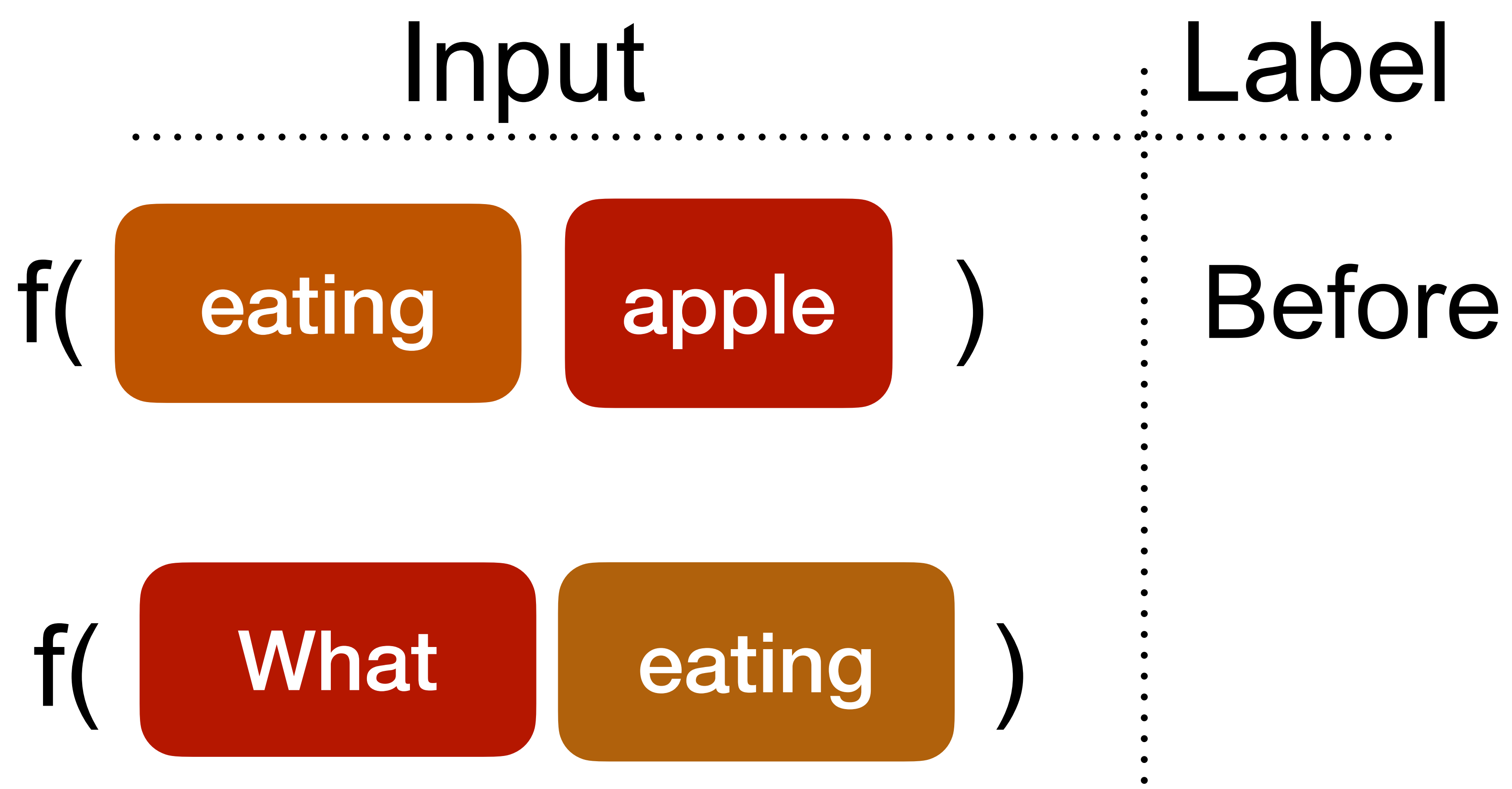


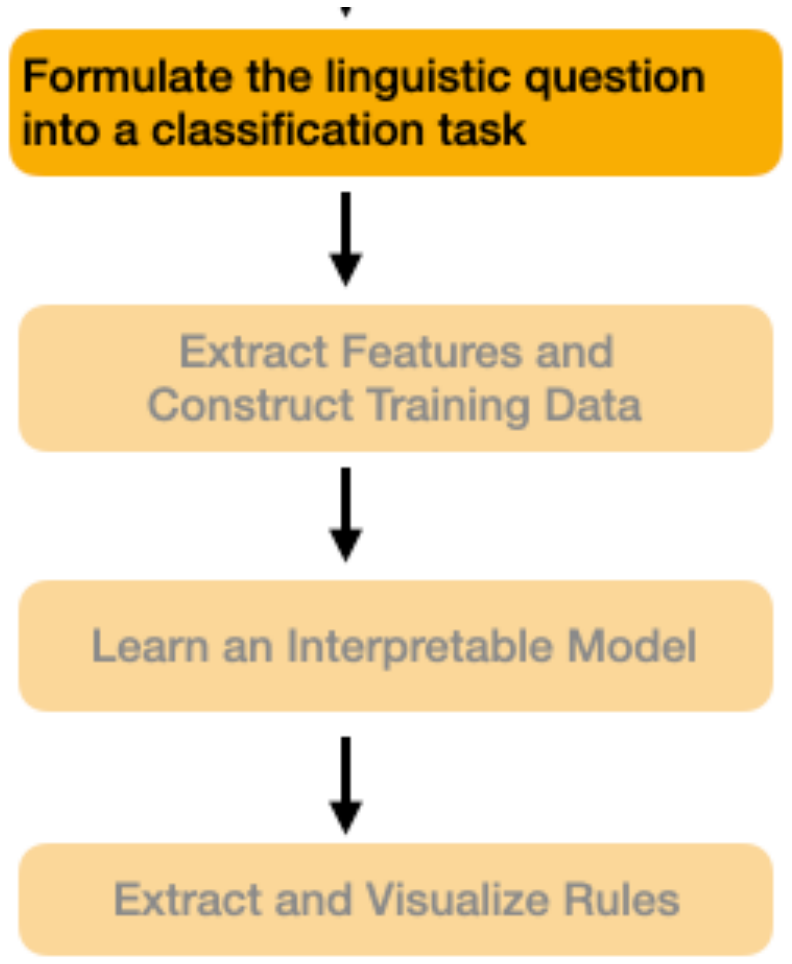
AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Formulate Linguistic Question



When are **objects** before or after the **verbs**?





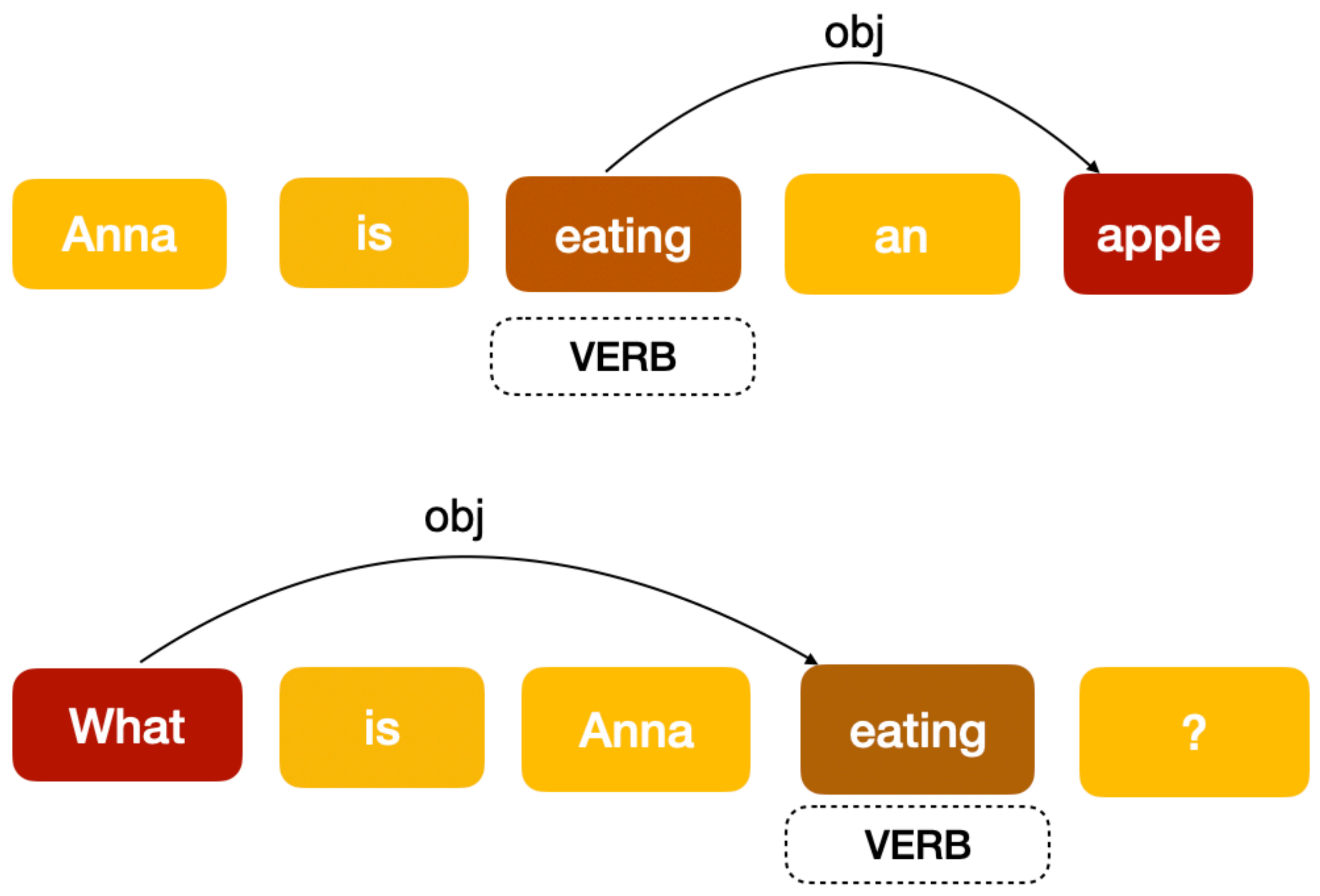
Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

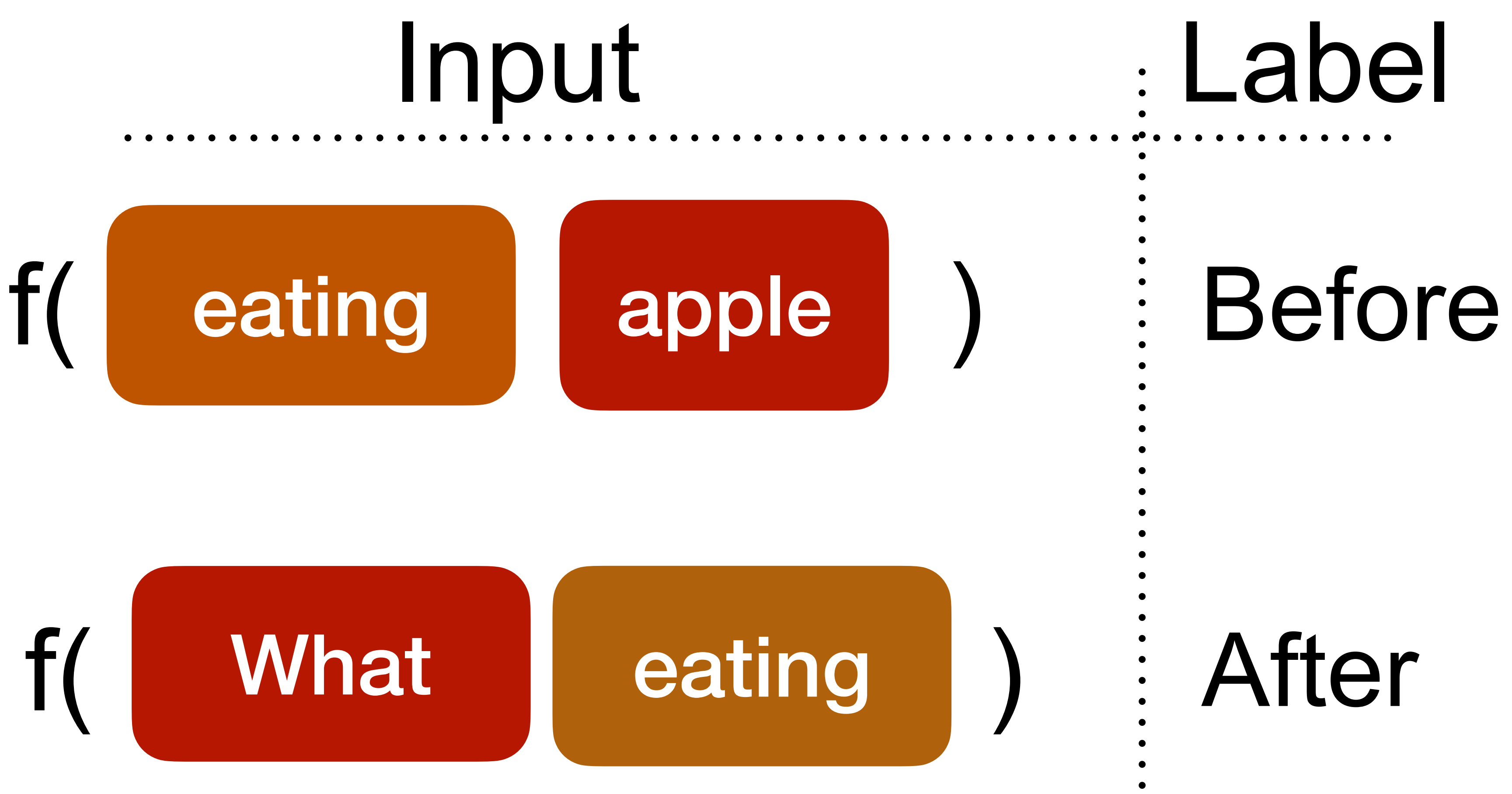
Learn an Interpretable Model

Extract and Visualize Rules

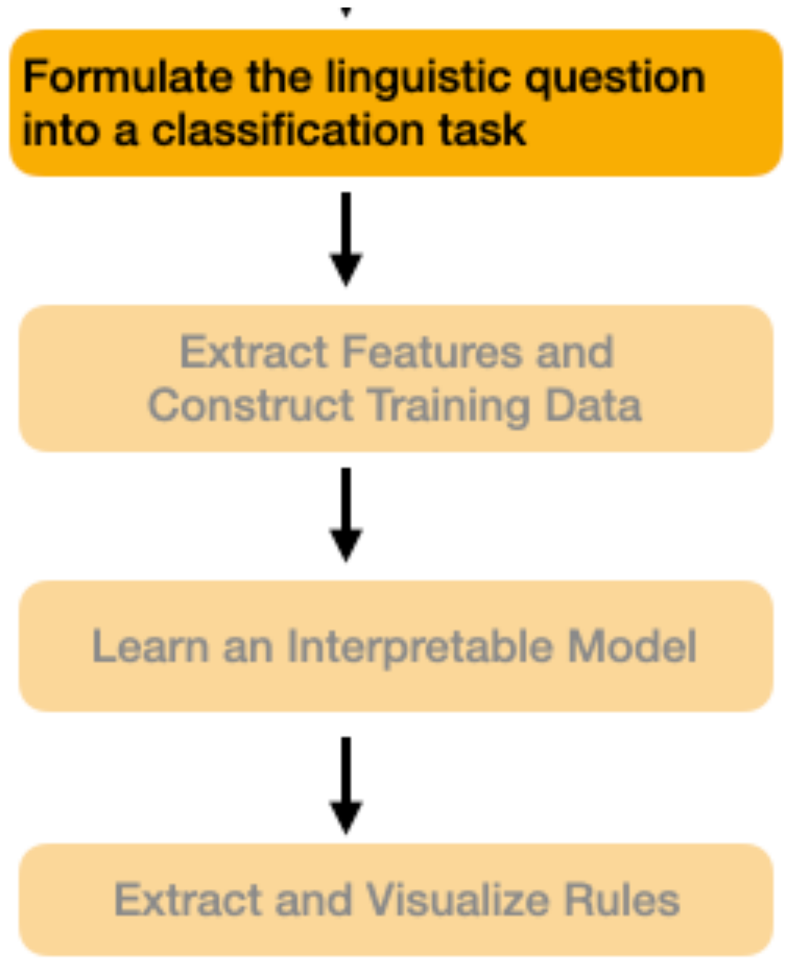
Word Order: Formulate Linguistic Question



When are **objects** before or after the **verbs**?



Input	Label
f(eating apple)	Before
f(What eating)	After



Formulate the linguistic question into a classification task

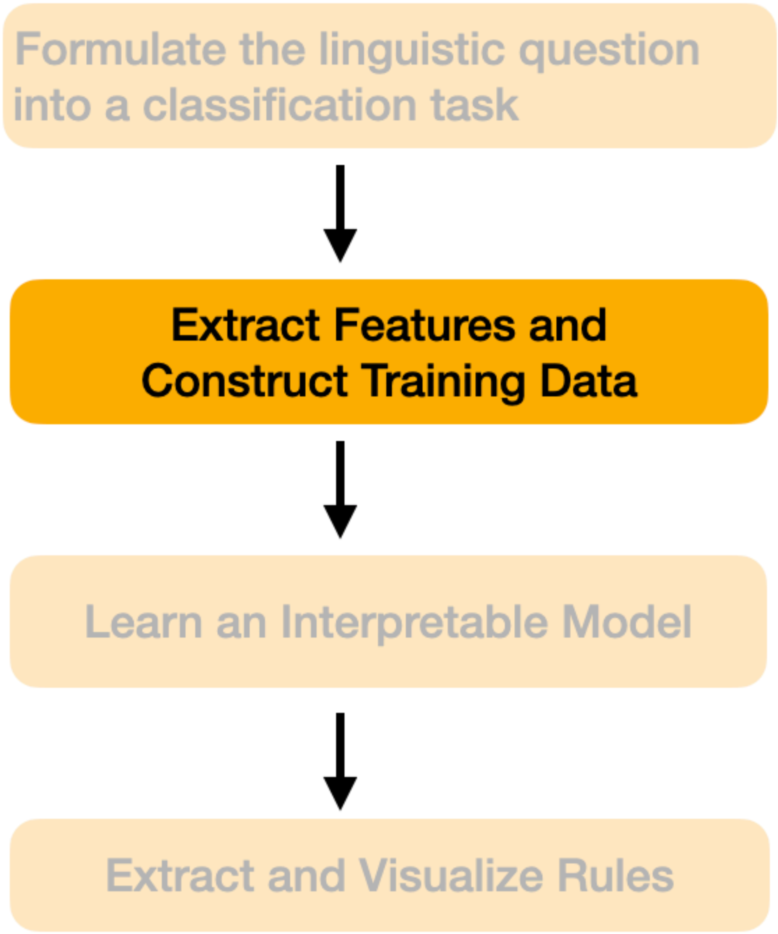
Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Formulate Linguistic Question

When are **objects** before or after the **verbs** in English?



Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

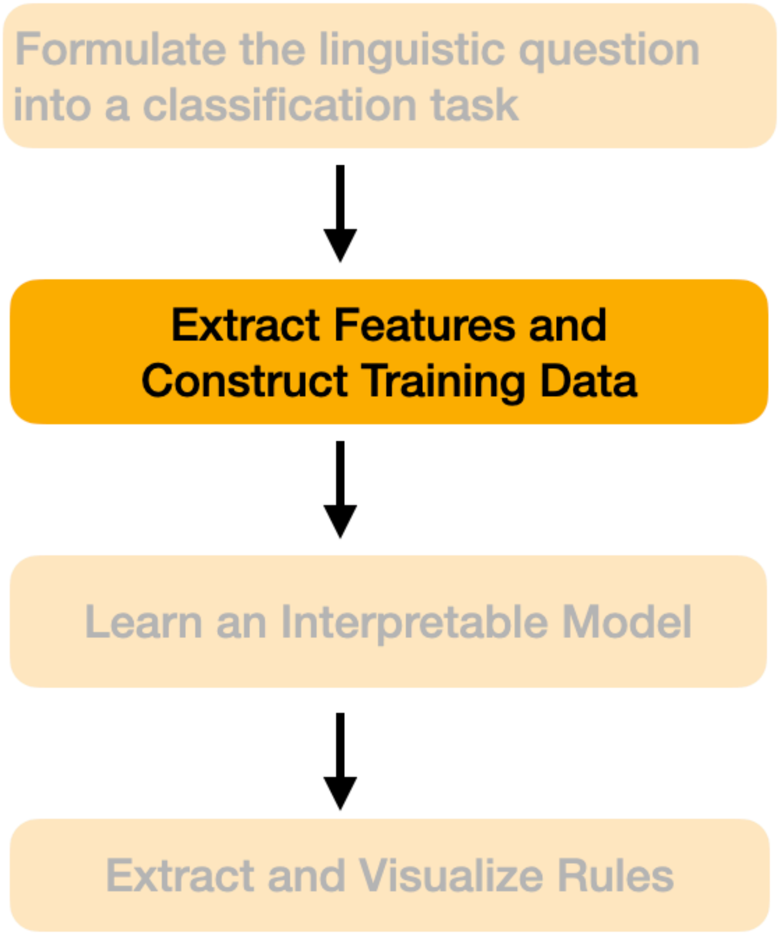
Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Formulate Linguistic Question

When are **objects** before or after the **verbs** in English?

Anna is eating an apple



Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

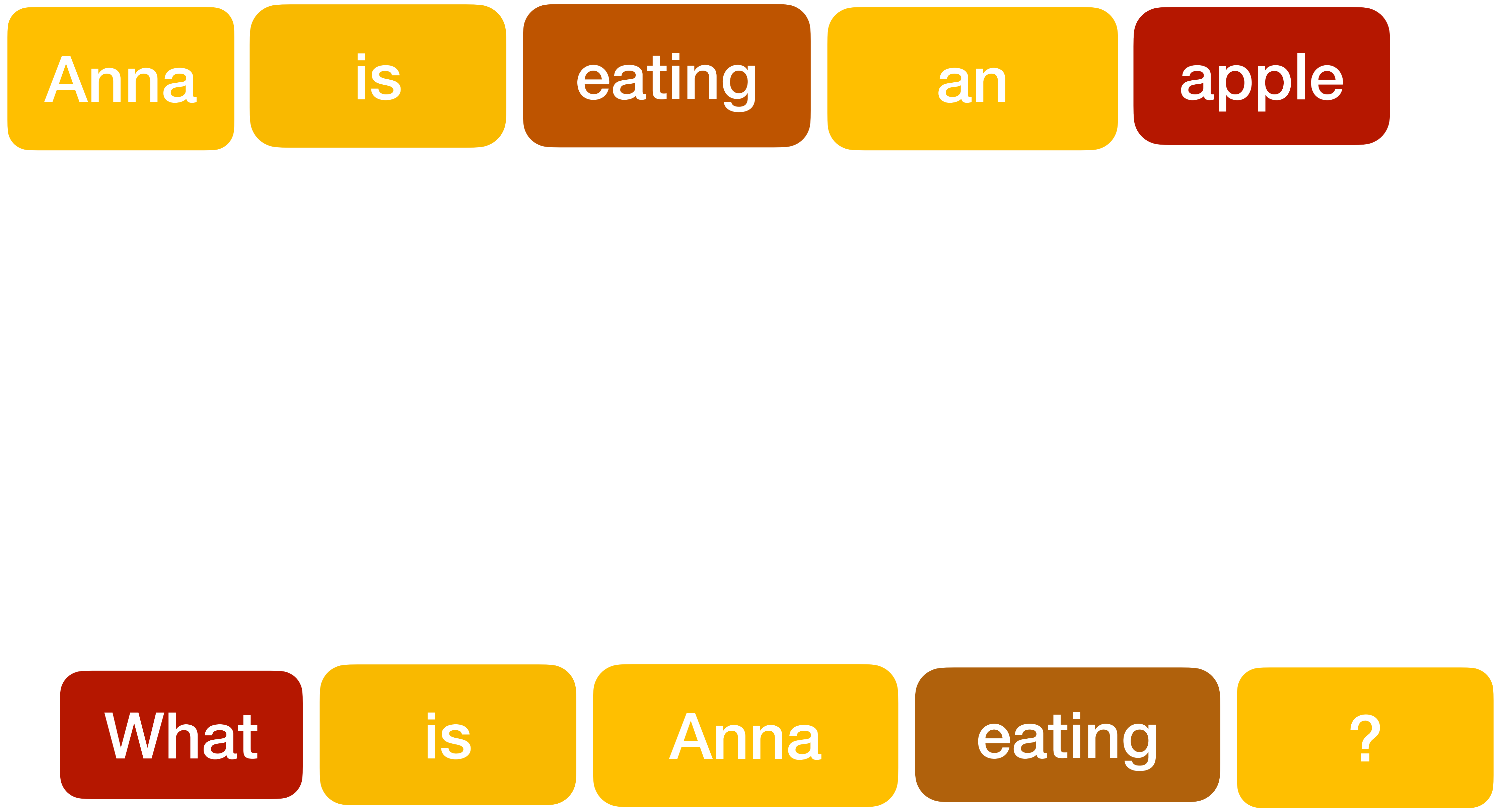
Learn an Interpretable Model

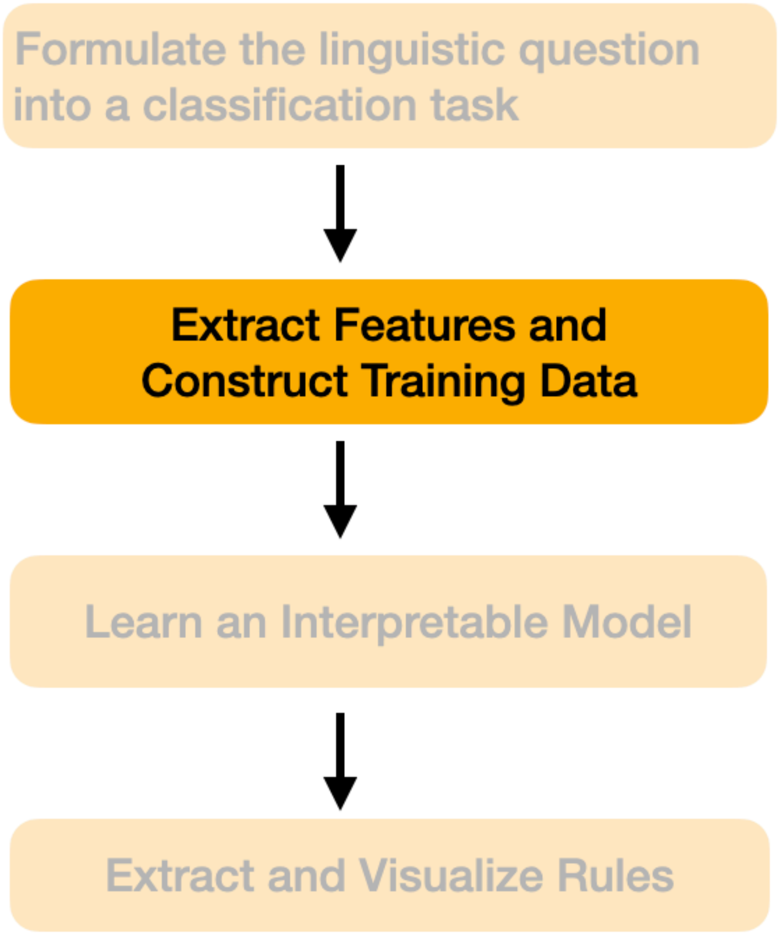
Extract and Visualize Rules

Carnegie Mellon University

Word Order: Formulate Linguistic Question

When are **objects** before or after the **verbs** in English?





Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

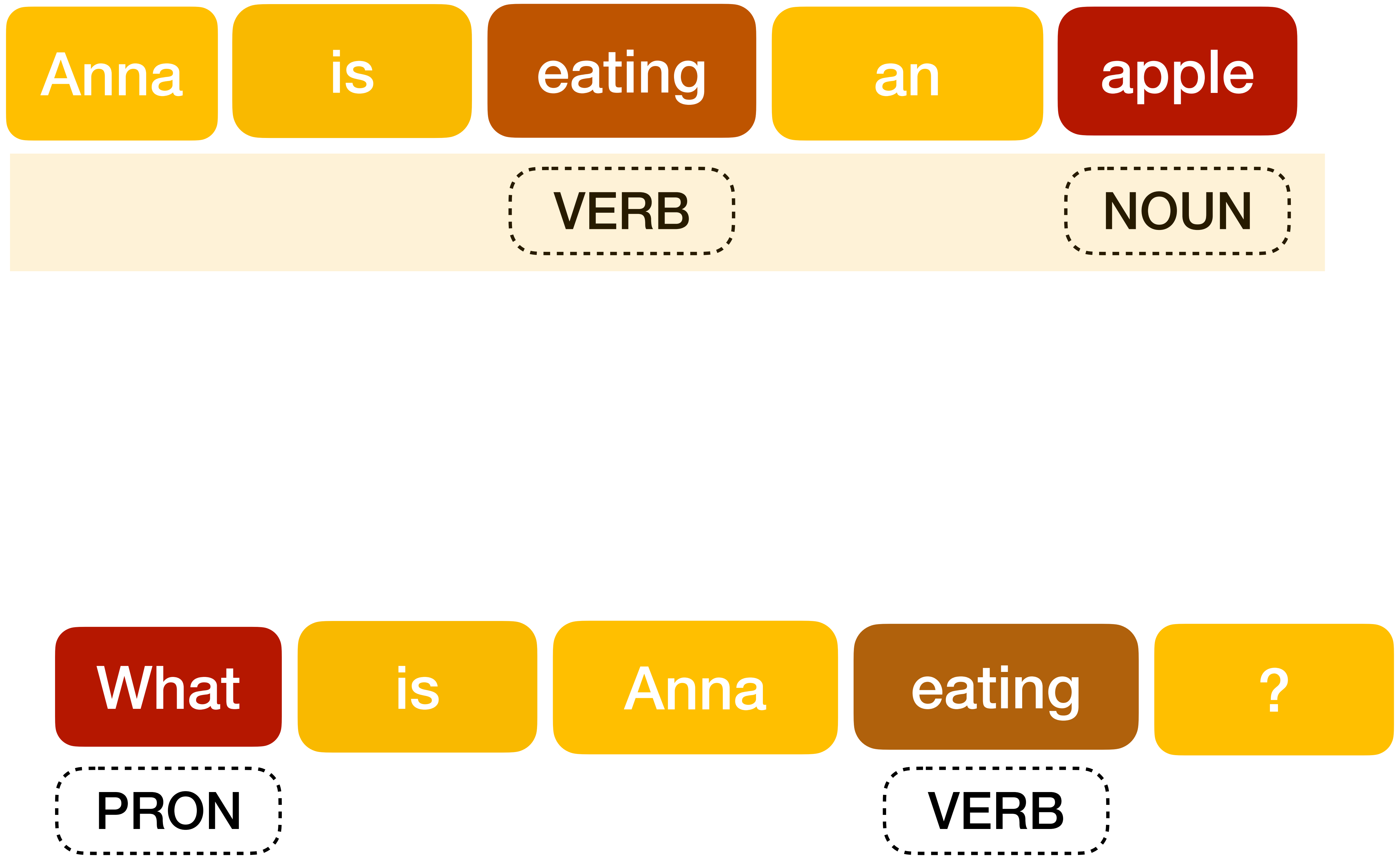
Learn an Interpretable Model

Extract and Visualize Rules

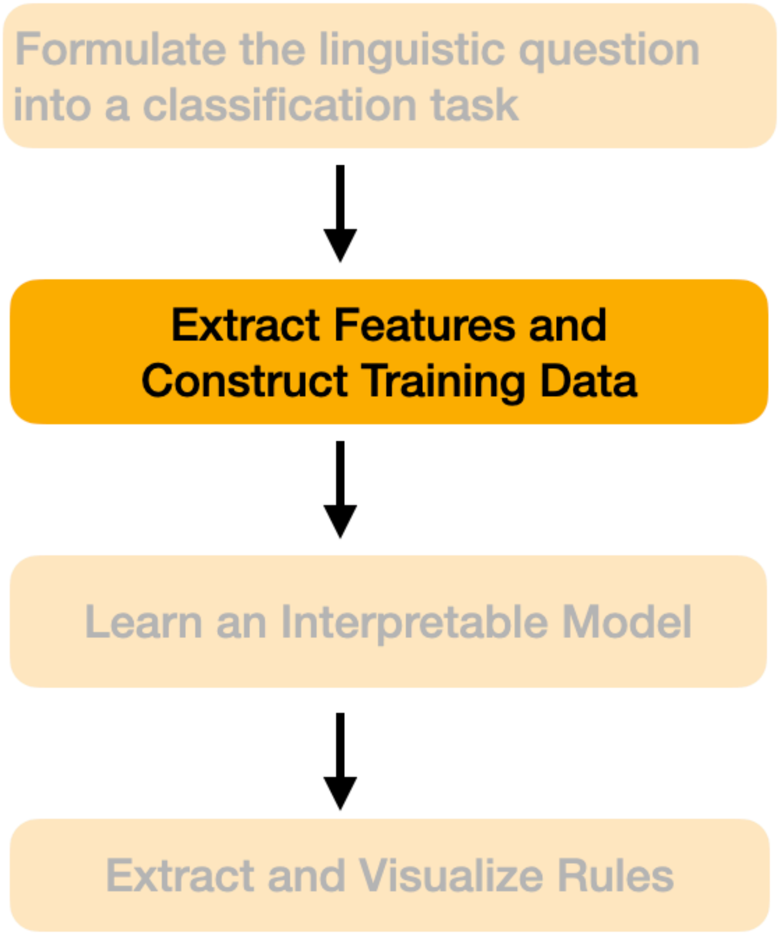
AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Formulate Linguistic Question

When are **objects** before or after the **verbs** in English?

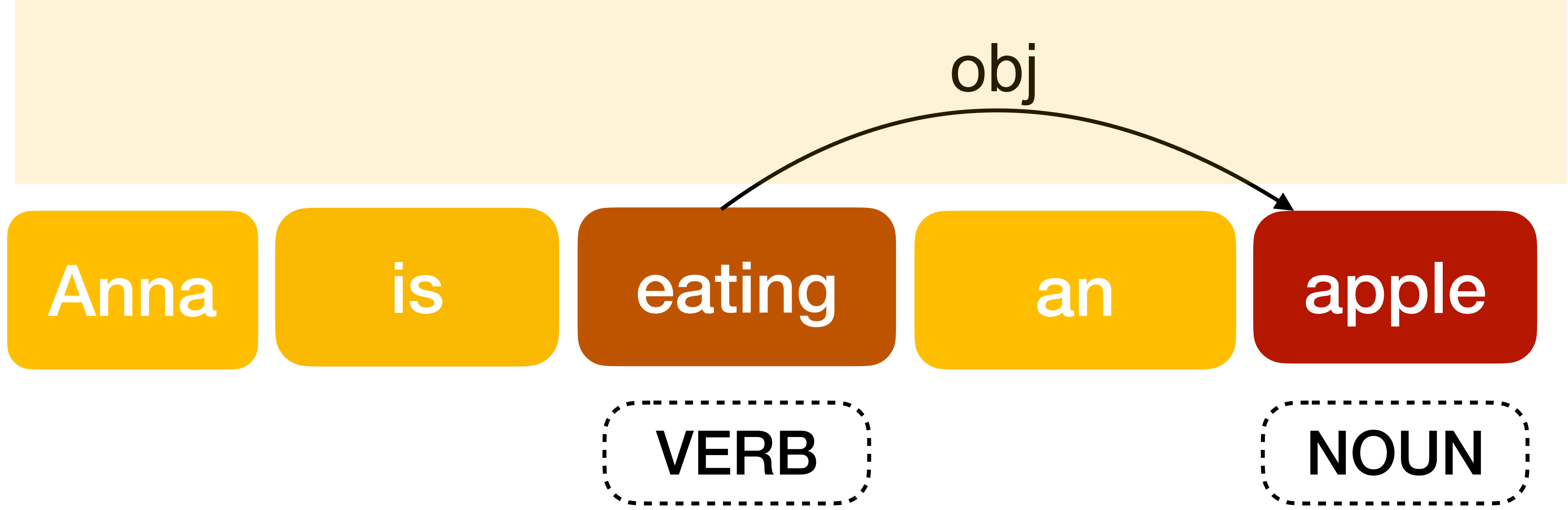


POS Tagging



Word Order: Formulate Linguistic Question

When are **objects** before or after the **verbs** in English?



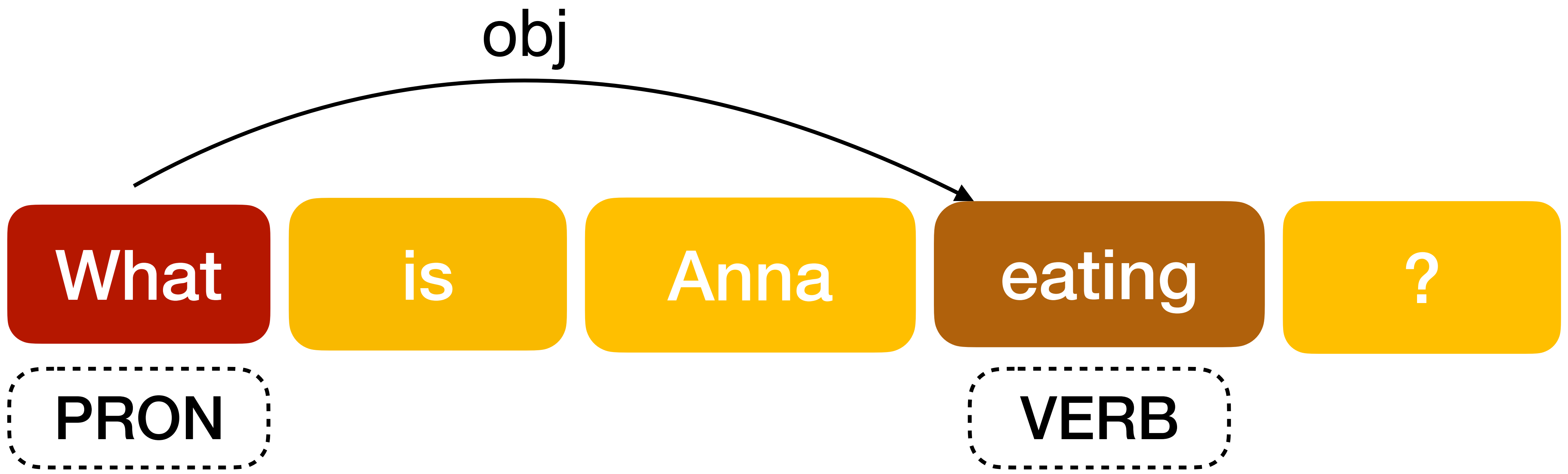
obj

Anna is eating an apple

VERB NOUN

Dependency Parsing

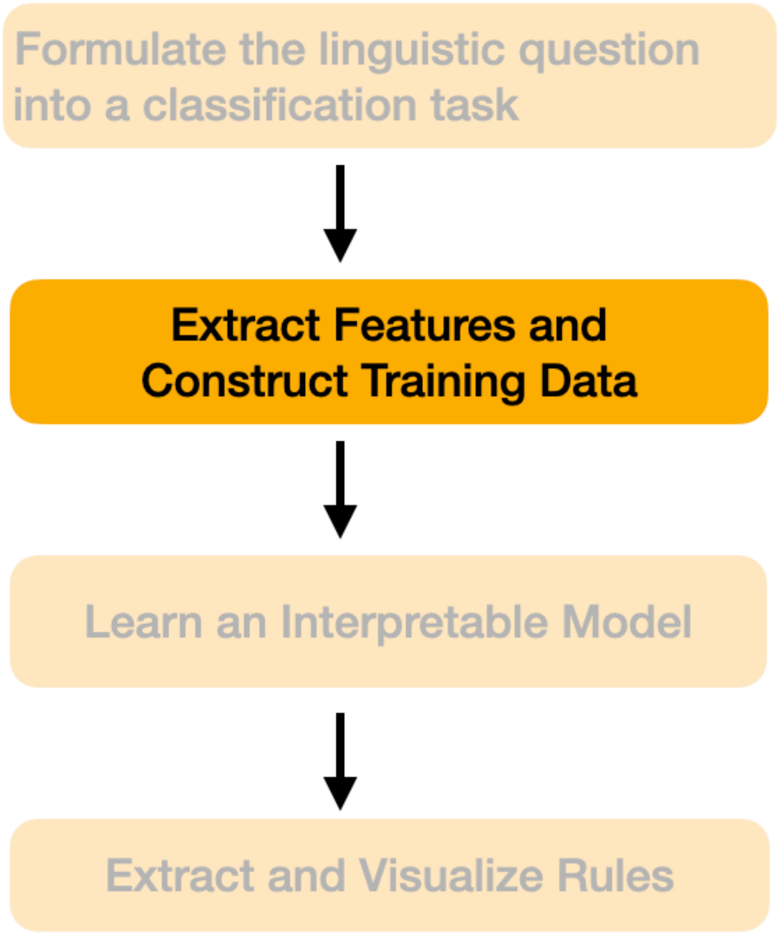
POS Tagging



obj

What is Anna eating ?

PRON VERB



Formulate the linguistic question into a classification task

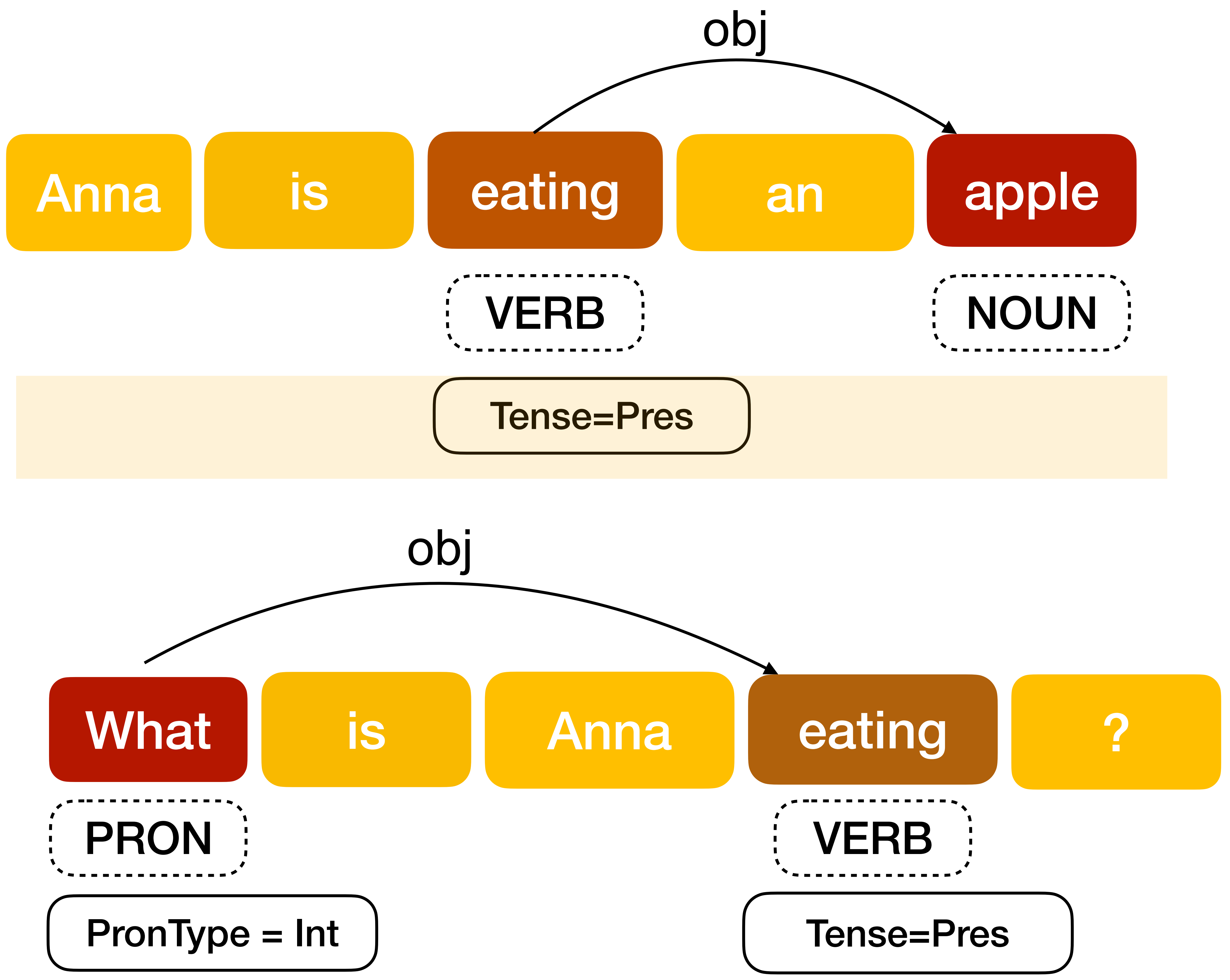
Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Formulate Linguistic Question

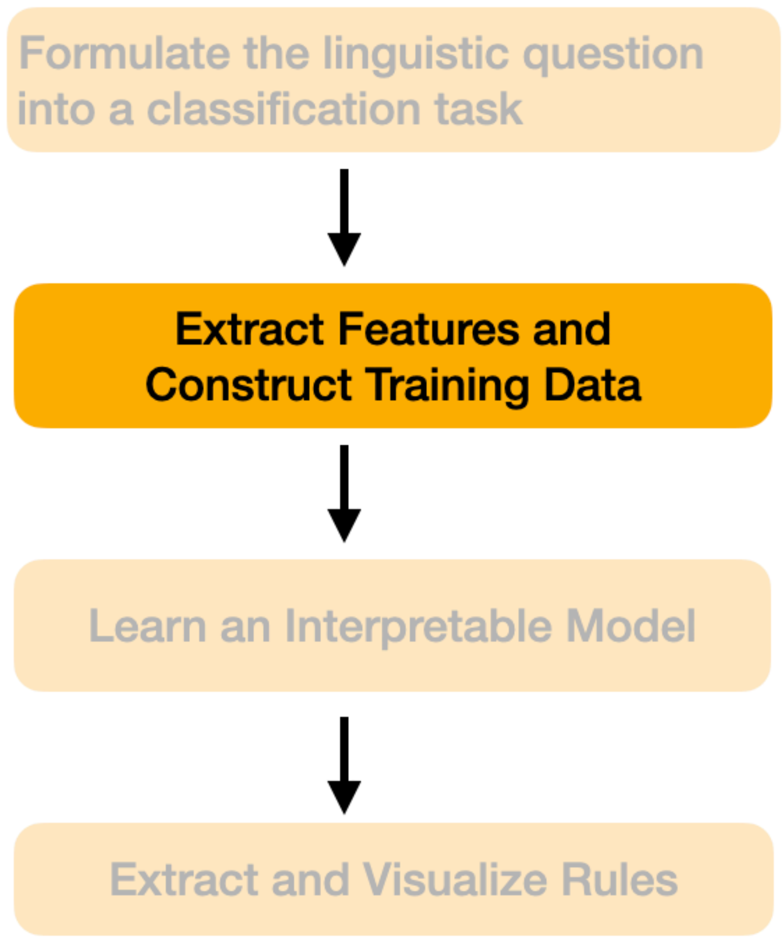
When are **objects** before or after the **verbs** in English?



Dependency Parsing

POS Tagging

Morphological Analysis

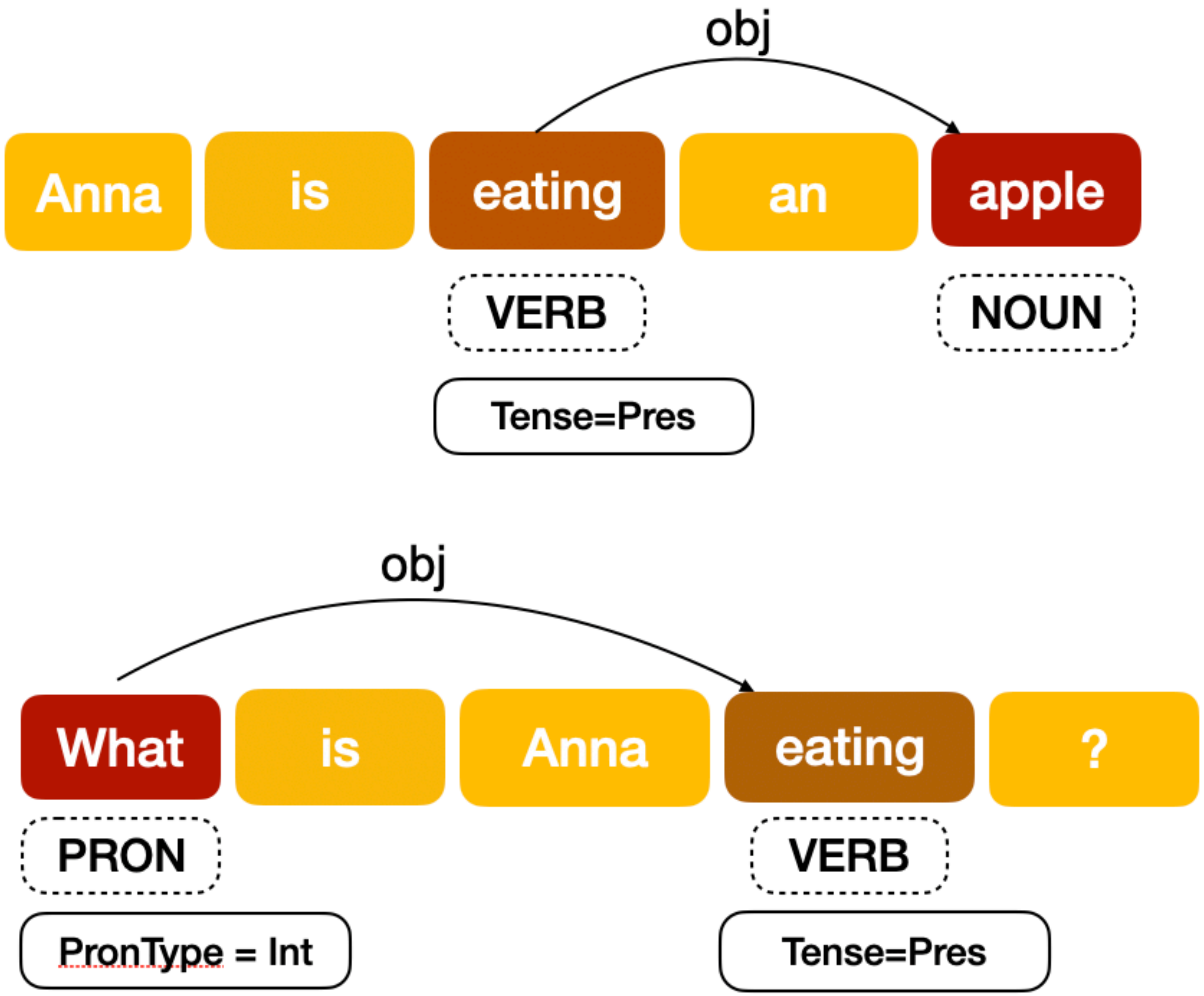


AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Carnegie Mellon University

Word Order: Extract Features and Training Data

When are **objects** before or after the **verbs** in English?



Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

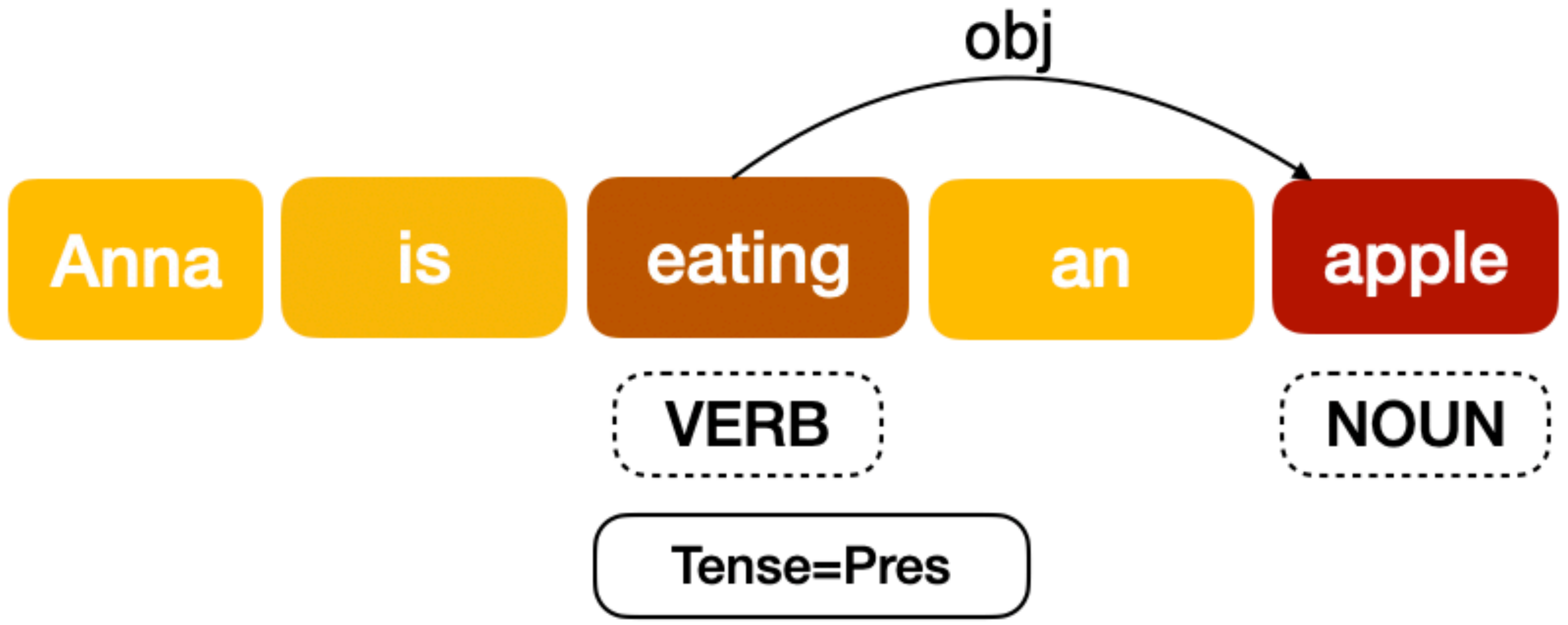
Learn an Interpretable Model

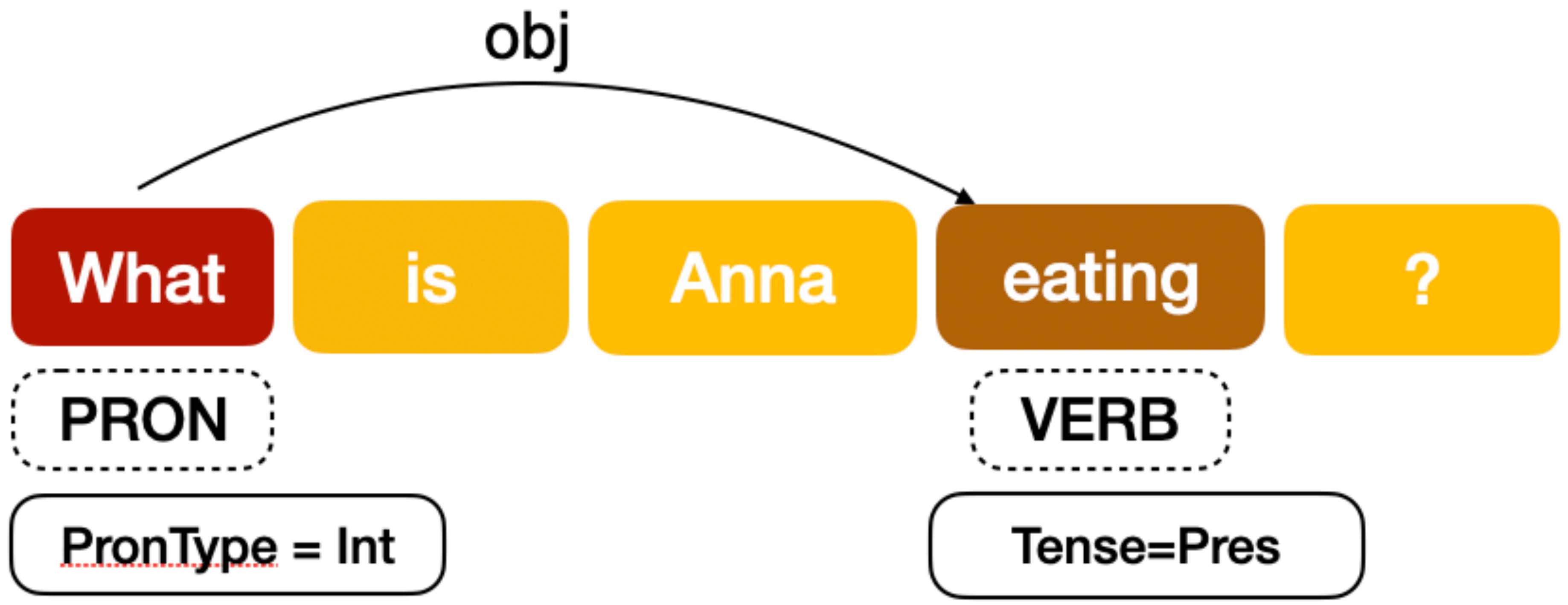
Extract and Visualize Rules

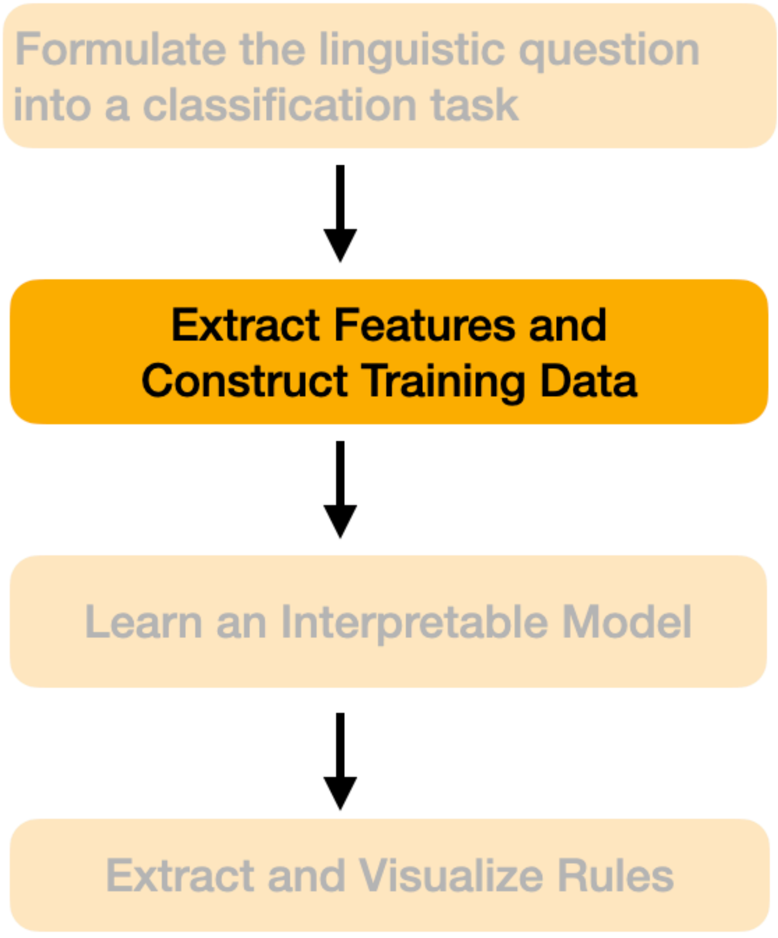
Carnegie Mellon University

Word Order: Extract Features and Training Data

When are **objects** before or after the **verbs** in English?







Formulate the linguistic question into a classification task

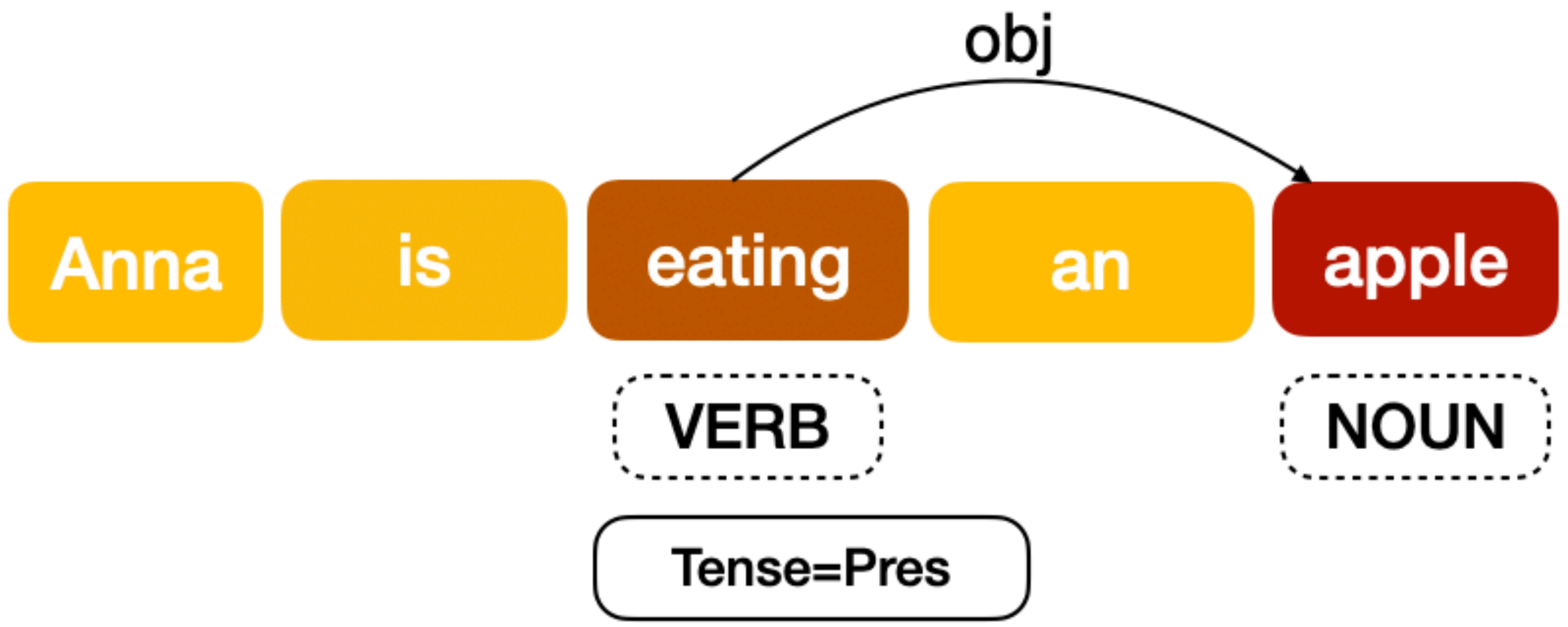
Extract Features and Construct Training Data

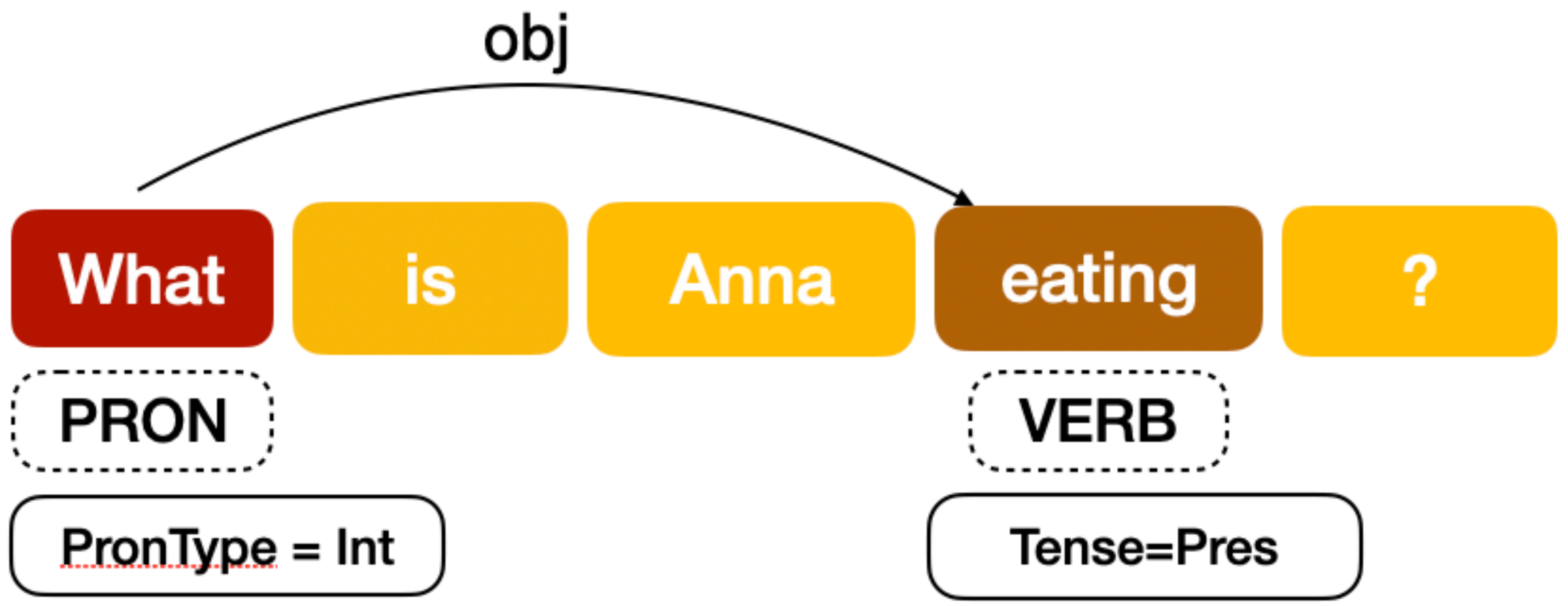
Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Extract Features and Training Data

When are **objects** before or after the **verbs** in English?





Input	Features	Label

Formulate the linguistic question into a classification task

↓

Extract Features and Construct Training Data

↓

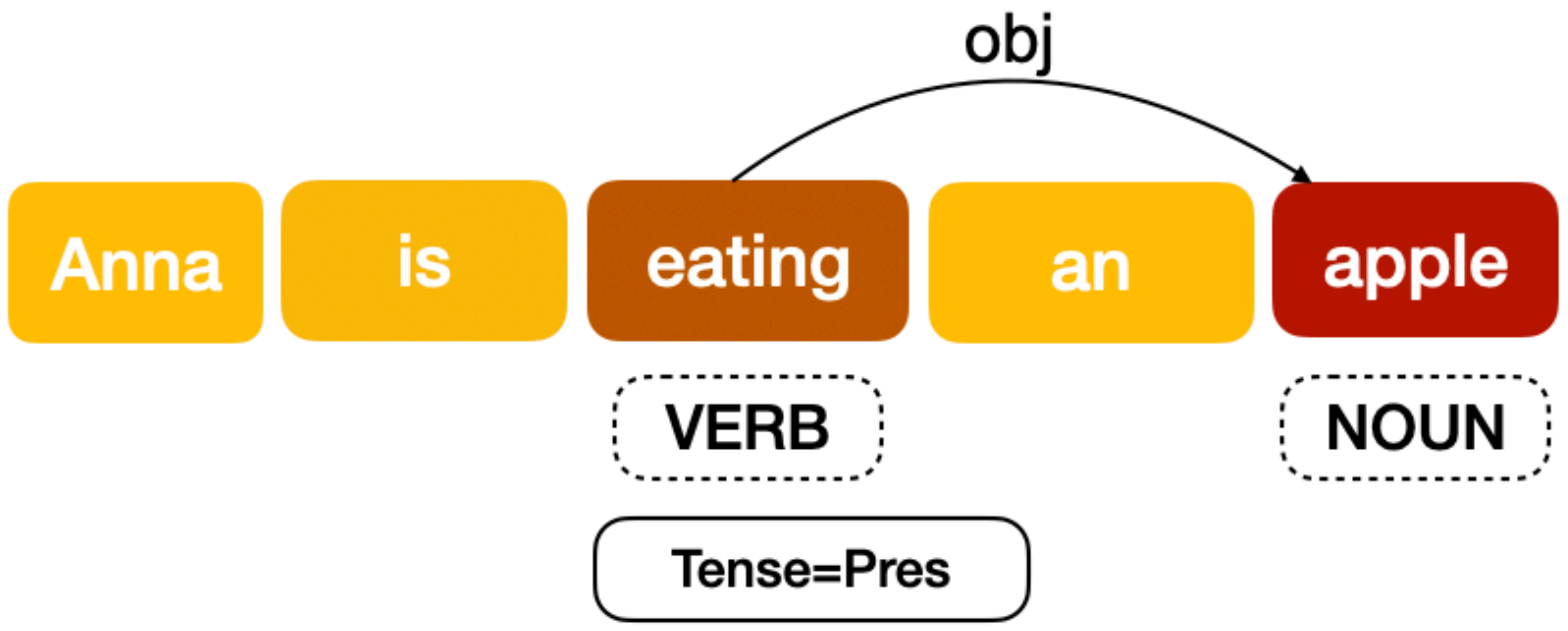
Learn an Interpretable Model

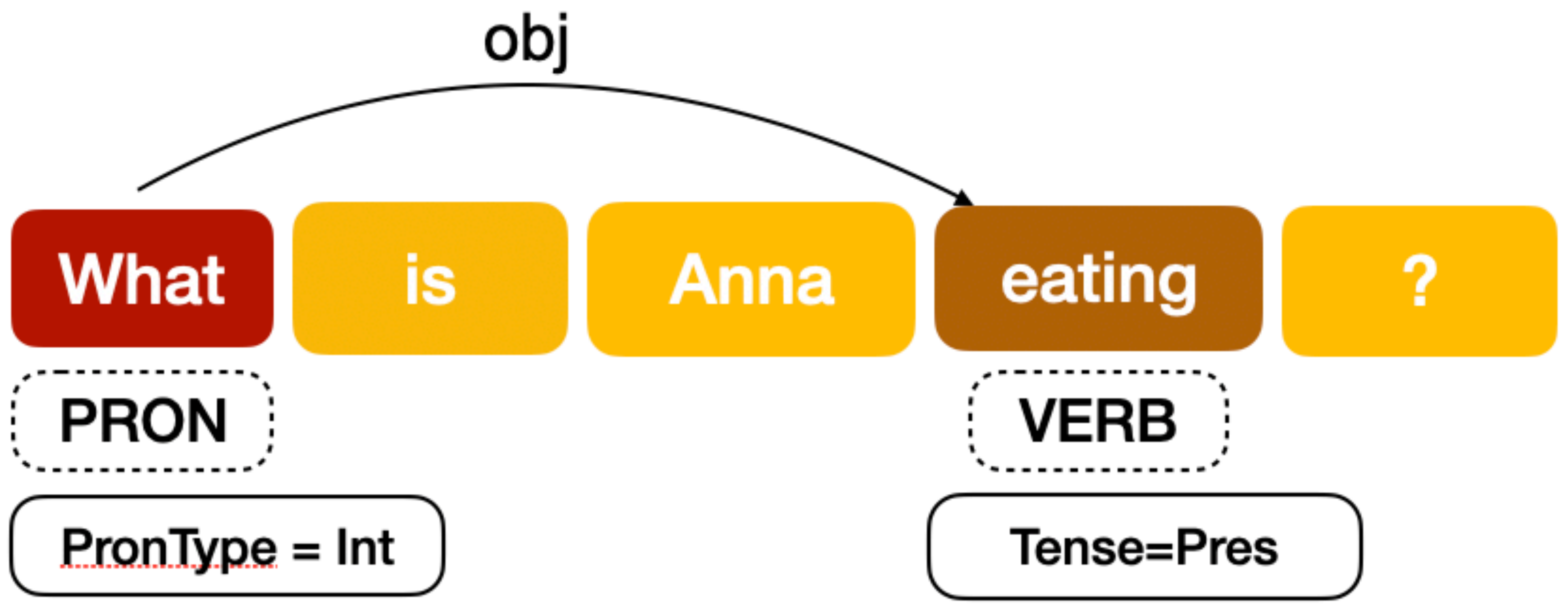
↓

Extract and Visualize Rules

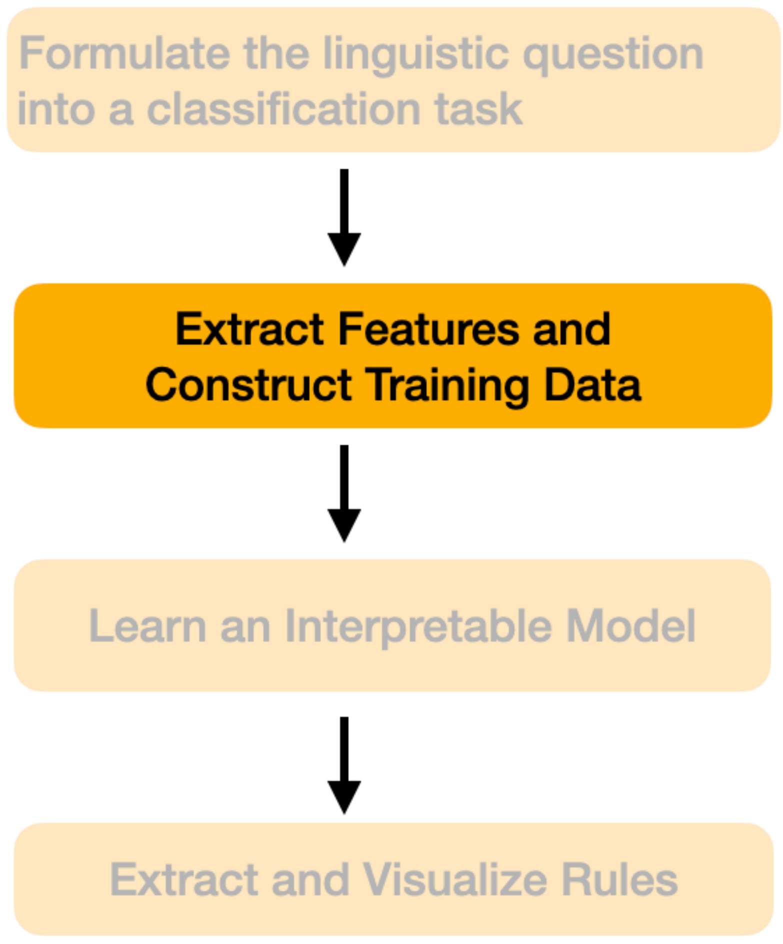
Word Order: Extract Features and Training Data

When are **objects** before or after the **verbs** in English?





Input	Features	Label
f(eating apple)		Before



Formulate the linguistic question into a classification task

↓

Extract Features and Construct Training Data

↓

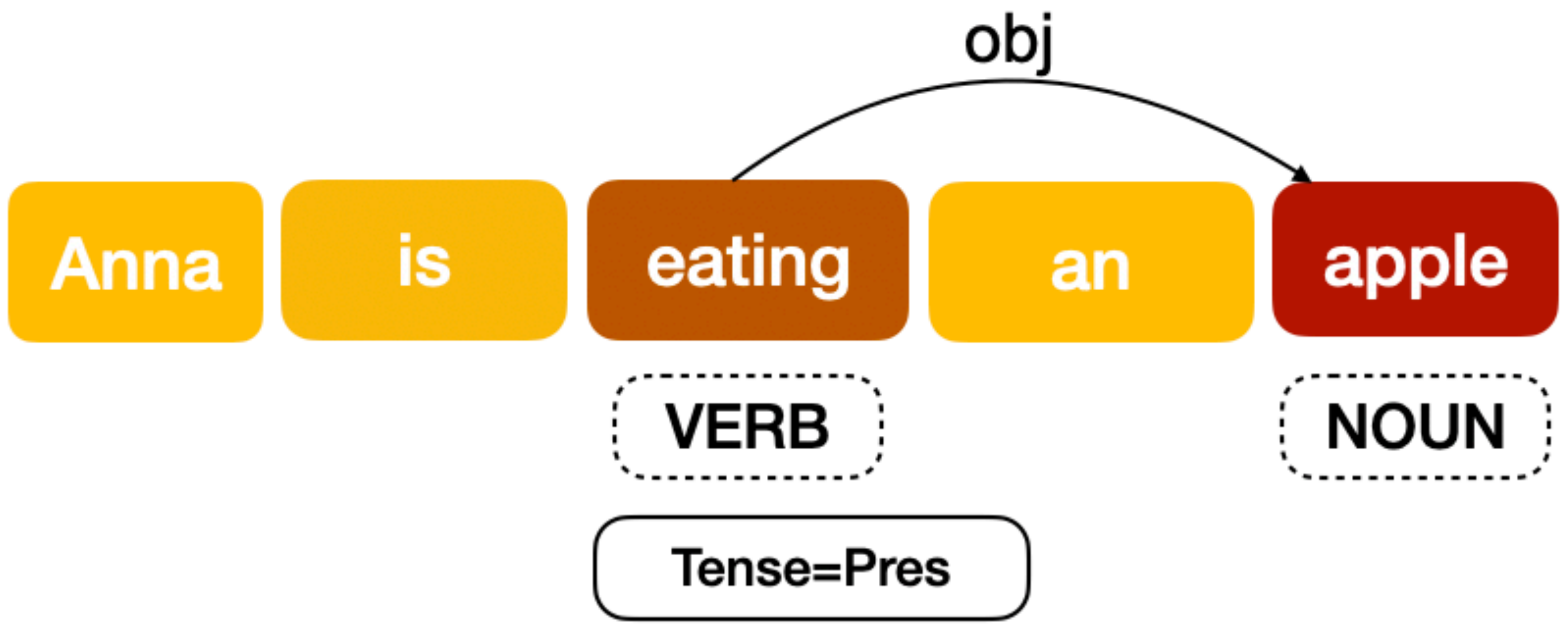
Learn an Interpretable Model

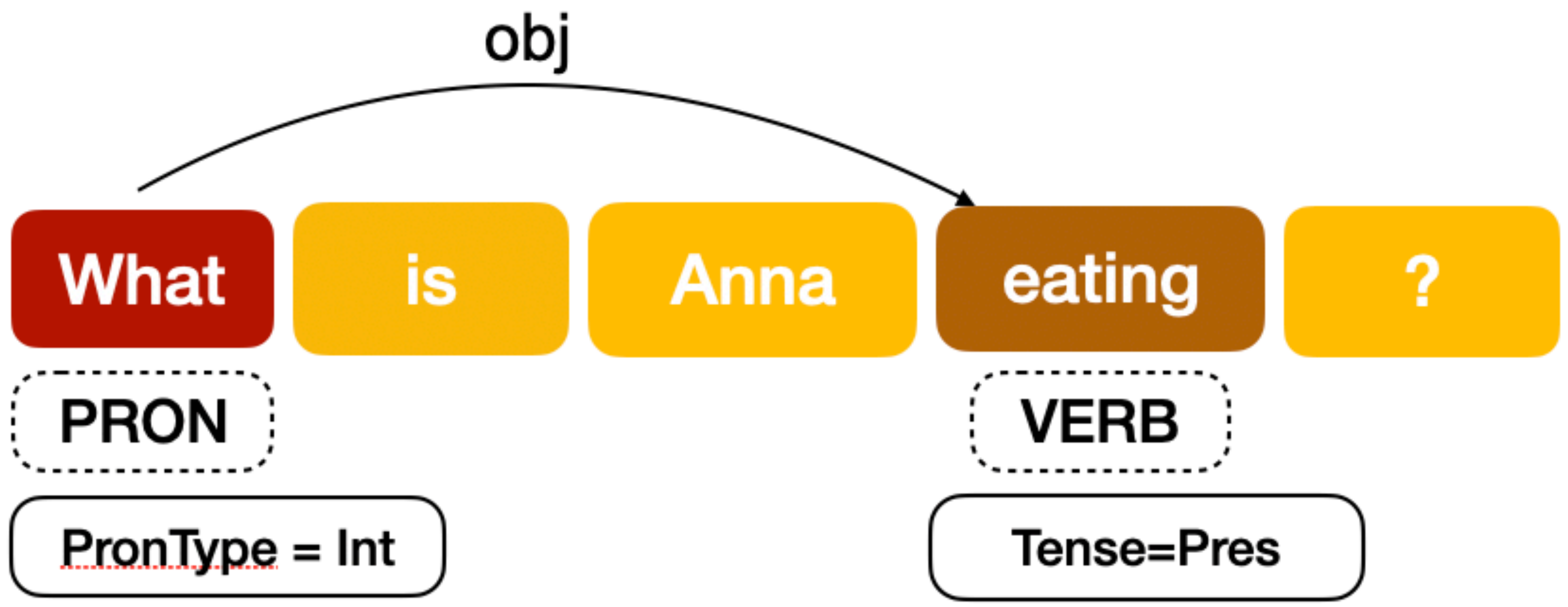
↓

Extract and Visualize Rules

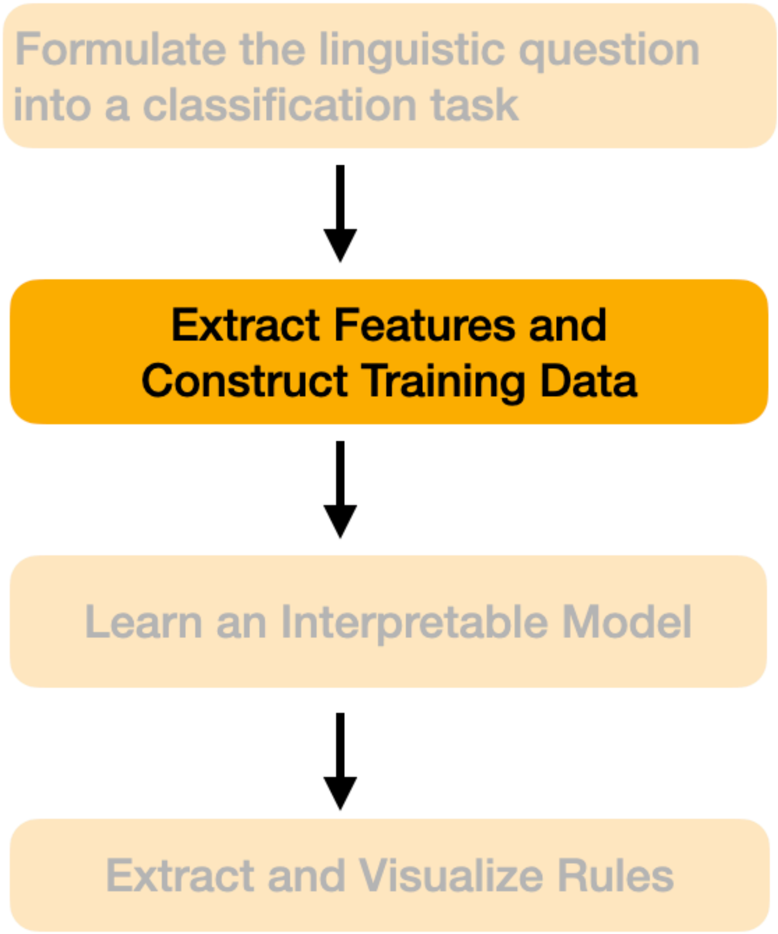
Word Order: Extract Features and Training Data

When are **objects** before or after the **verbs** in English?





Input	Features	Label
f(eating apple)		Before
f(What eating)		After


Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

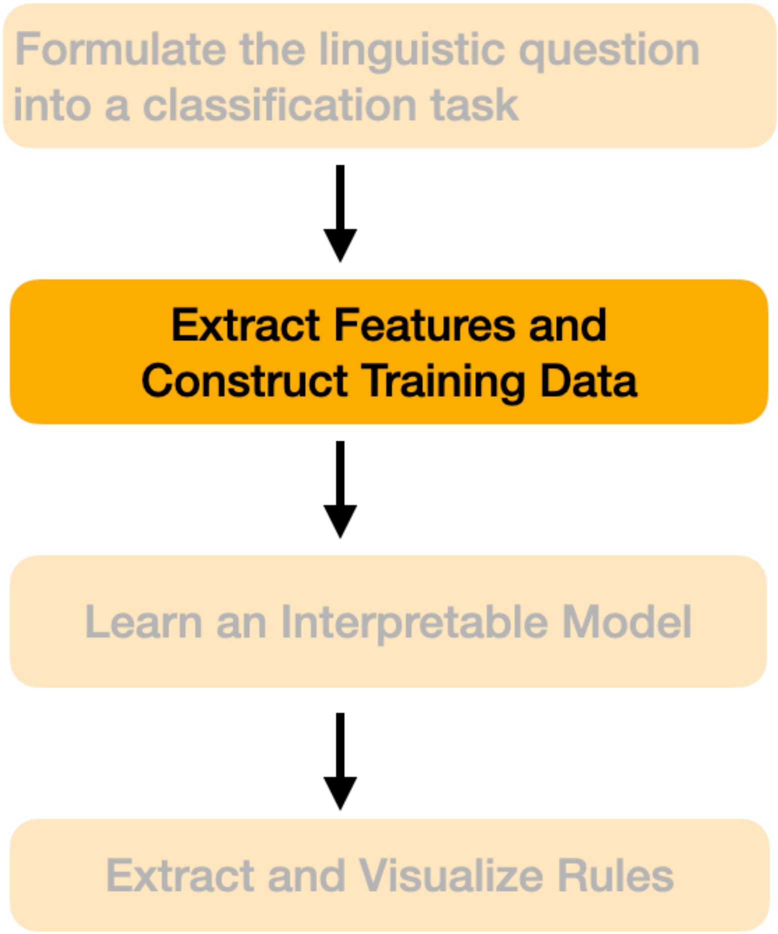
Extract and Visualize Rules

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Extract Features and Training Data

When are **objects** before or after the **verbs** in English?

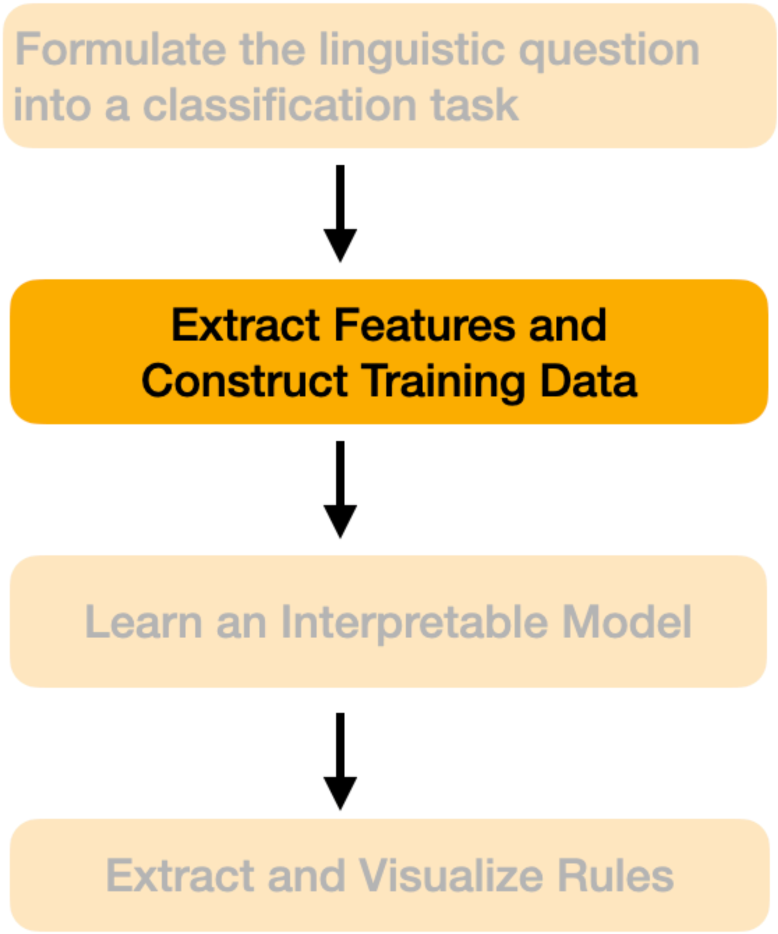
Input	Features	Label
f(eating apple)	is-object-NOUN	Before
f(What eating)	is-object-PRON	After



Word Order: Extract Features and Training Data

When are **objects** before or after the **verbs** in English?

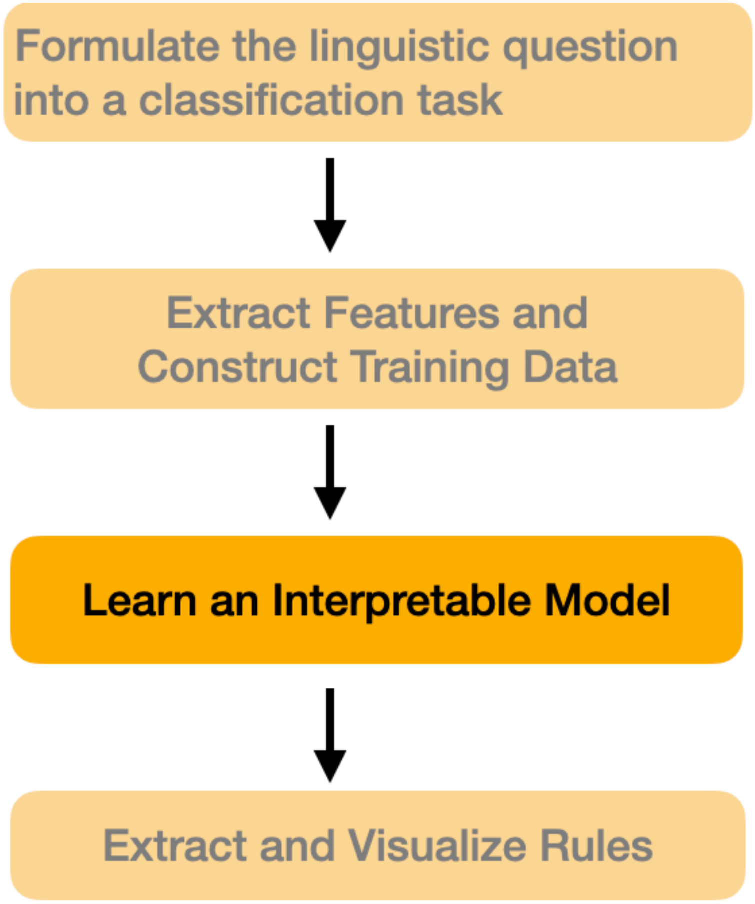
Input	Features	Label
f(eating apple)	is-object-NOUN is-verb-tense-present	Before
f(What eating)	is-object-PRON is-object-Interrogative is-verb-tense-present	After



AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Learn a Model

When are **objects** before or after the **verbs** in English?



Formulate the linguistic question into a classification task

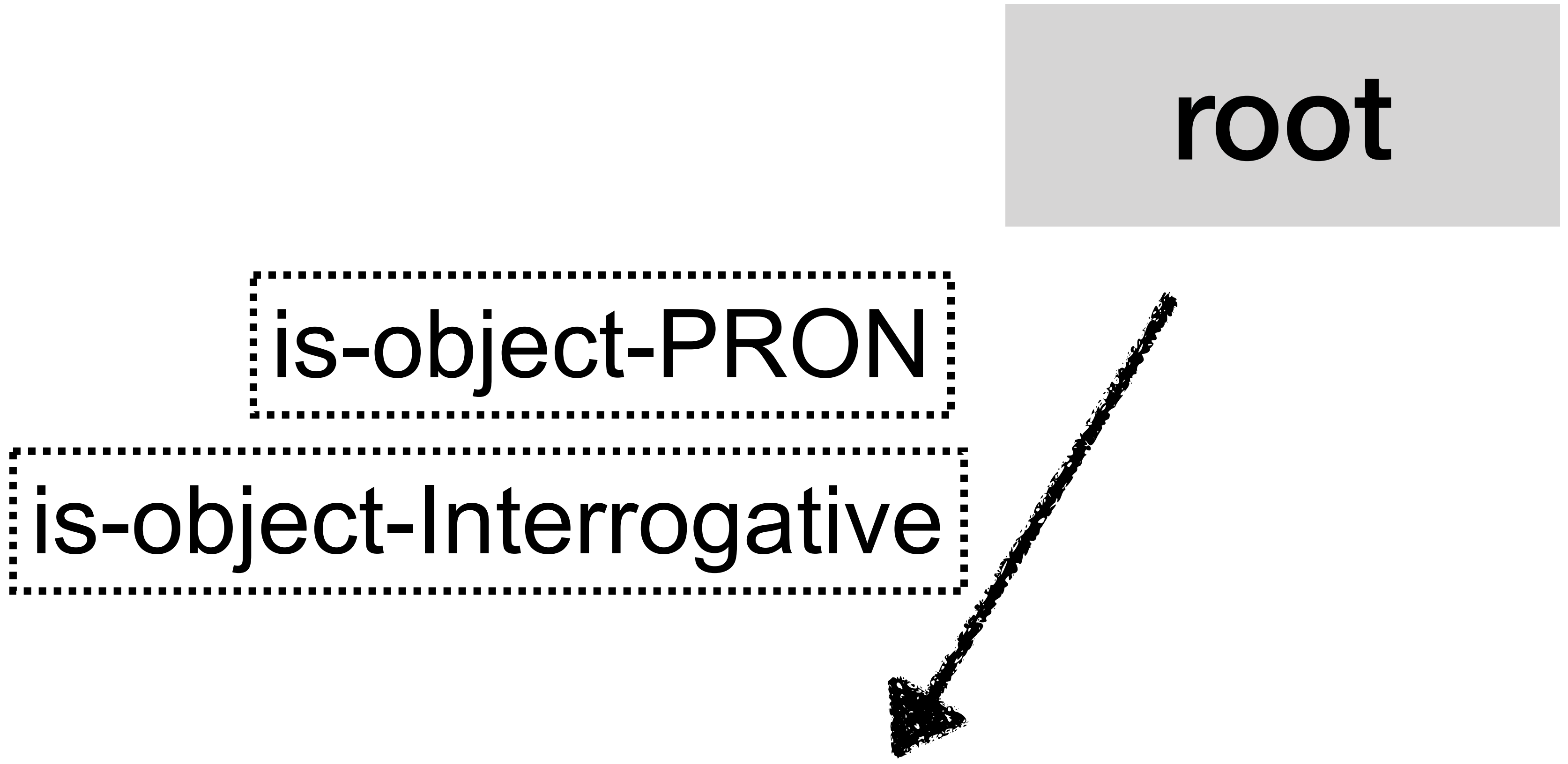
Extract Features and Construct Training Data

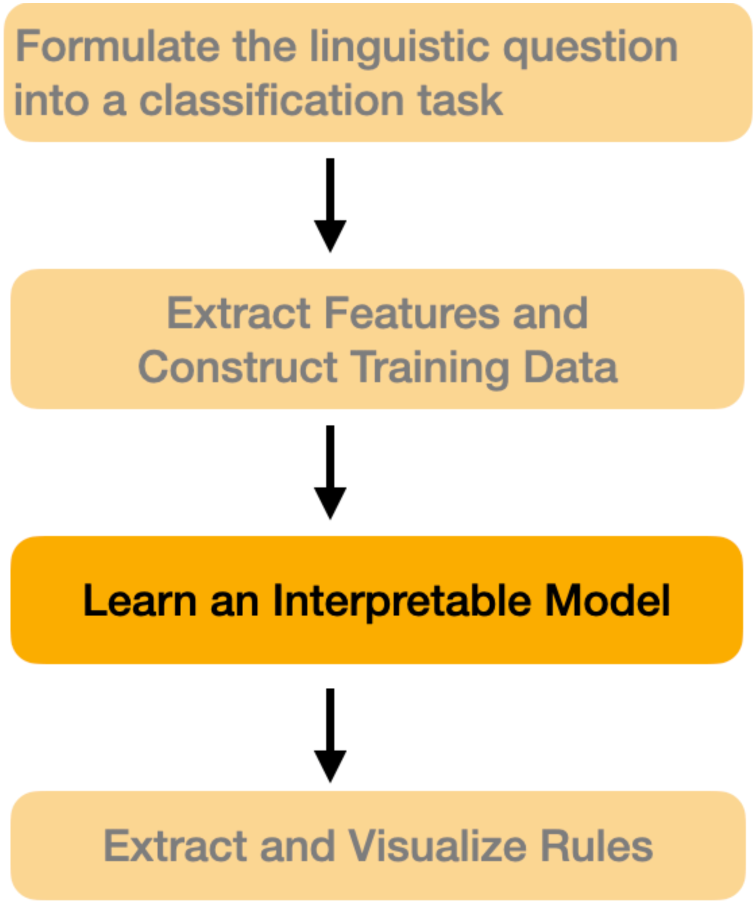
Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Learn a Model

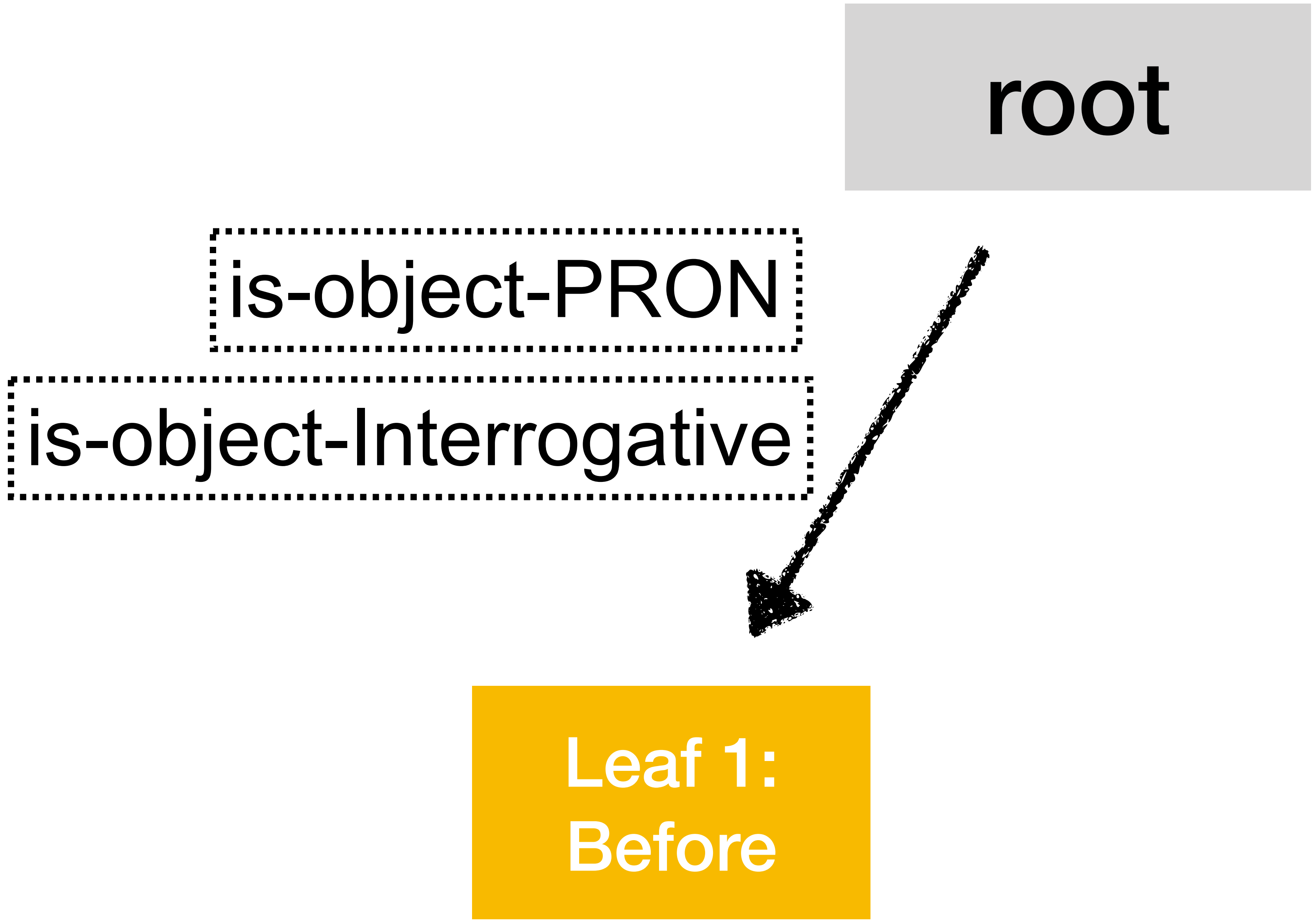
When are **objects** before or after the **verbs** in English?





Word Order: Learn a Model

When are **objects** before or after the **verbs** in English?



root

is-object-PRON

is-object-Interrogative

Leaf 1: Before

Formulate the linguistic question into a classification task

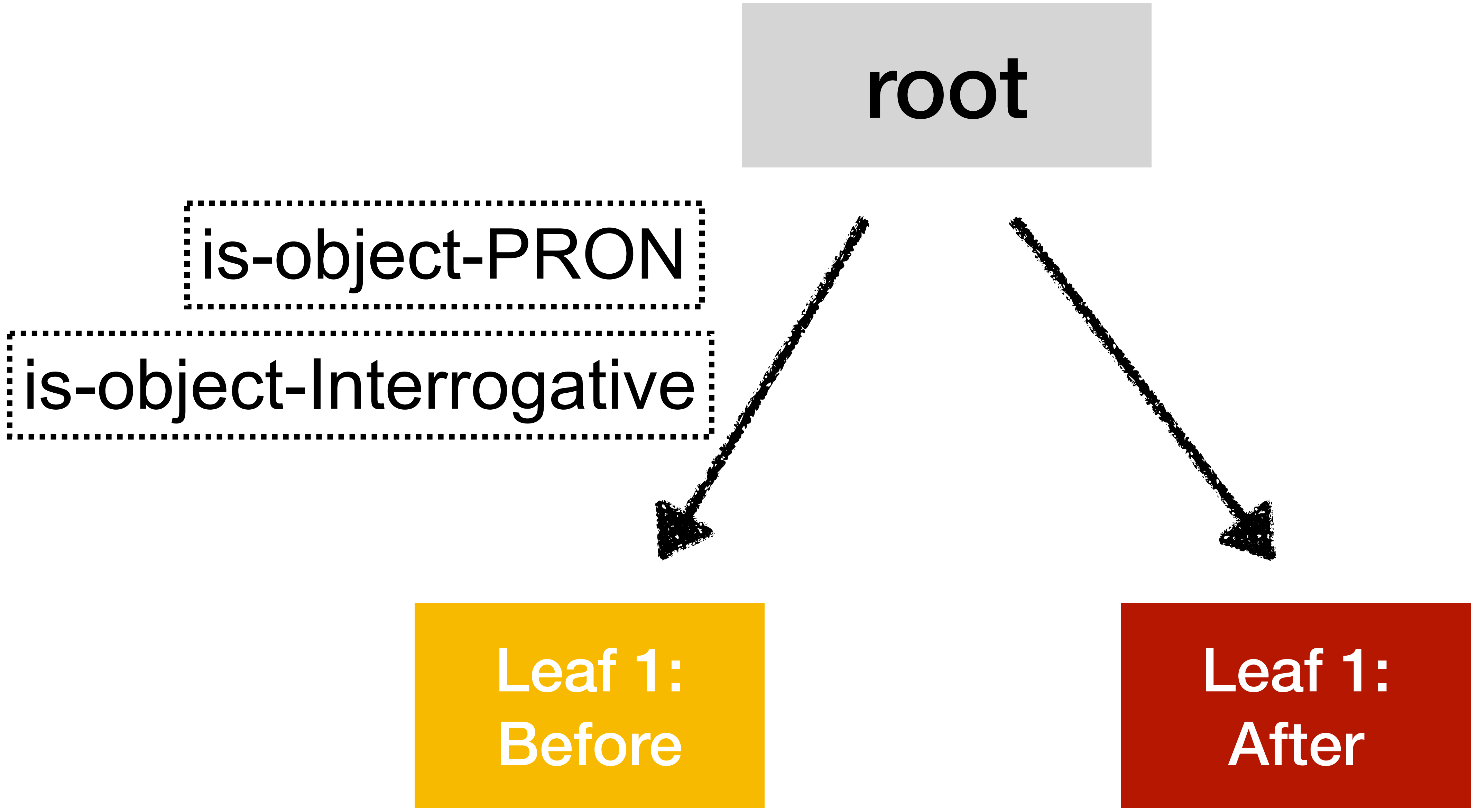
Extract Features and Construct Training Data

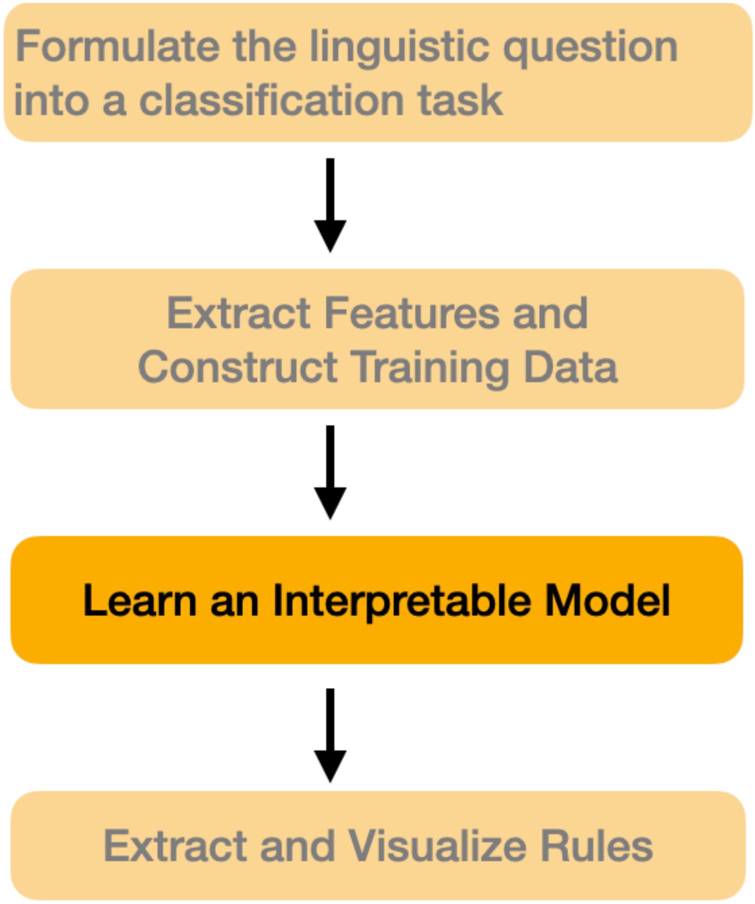
Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Learn a Model

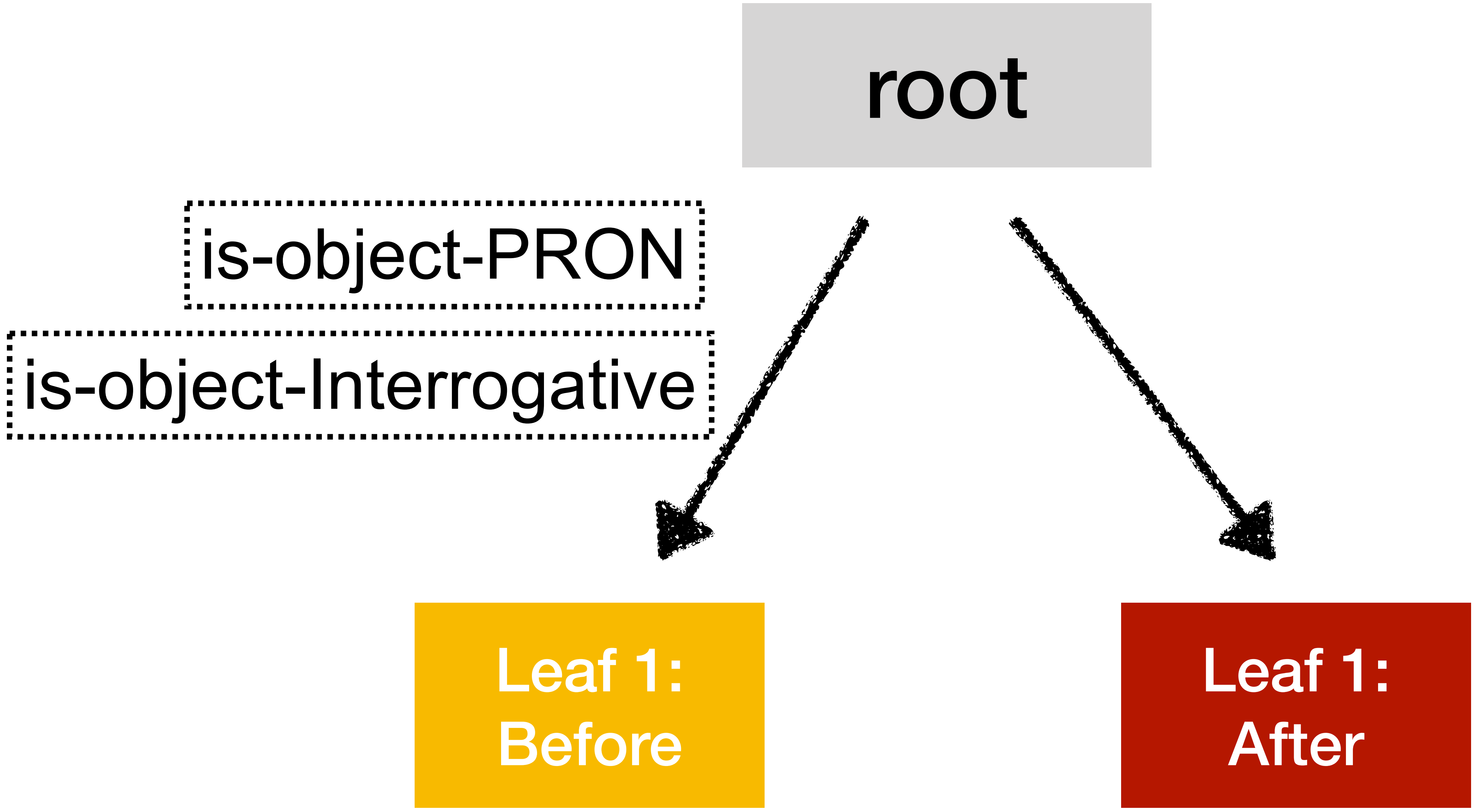
When are **objects** before or after the **verbs** in English?



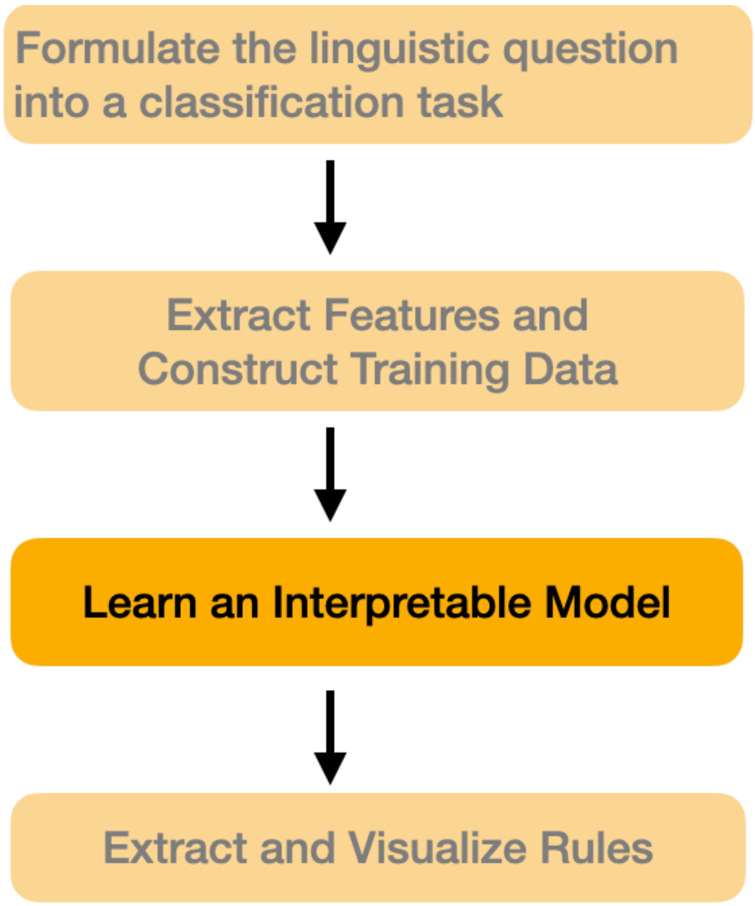


Word Order: Learn a Model

When are **objects** before or after the **verbs** in English?



Rule: When object is an interrogative pronoun, it comes **BEFORE** the verb.



Word Order: Data

When are **objects** before or after the **verbs** in English?

Formulate the linguistic question into a classification task

↓

Extract Features and Construct Training Data

↓

Learn an Interpretable Model

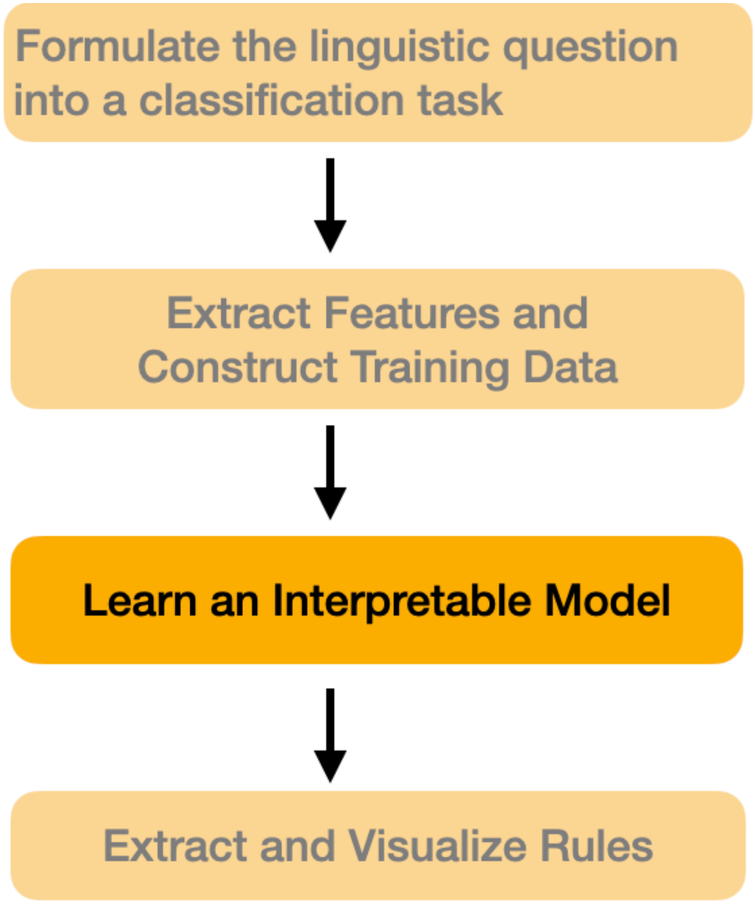
↓

Extract and Visualize Rules

Word Order: Data

When are **objects** before or after the **verbs** in English?

- Compare the model in a clean setting ⟶ **Syntactic Universal Dependencies (SUD)**



Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

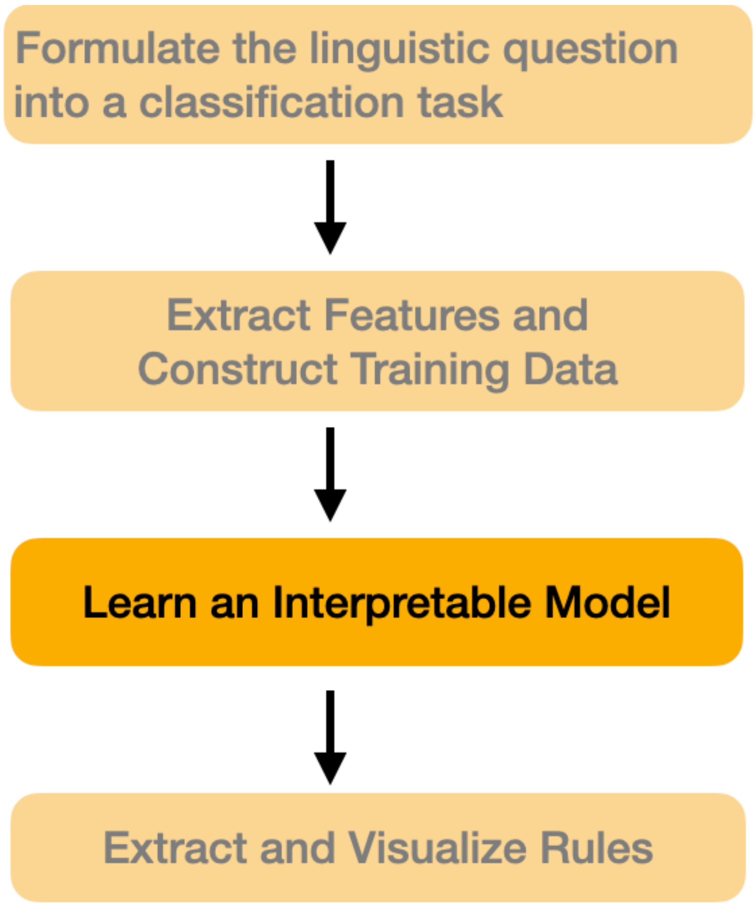
Extract and Visualize Rules

Word Order: Data

When are **objects** before or after the **verbs** in English?

- Compare the model in a clean setting ⟶ **Syntactic Universal Dependencies (SUD)**

- **Prefers syntactic heads** over content heads ⟶ more conducive to our goal of rule extraction



Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

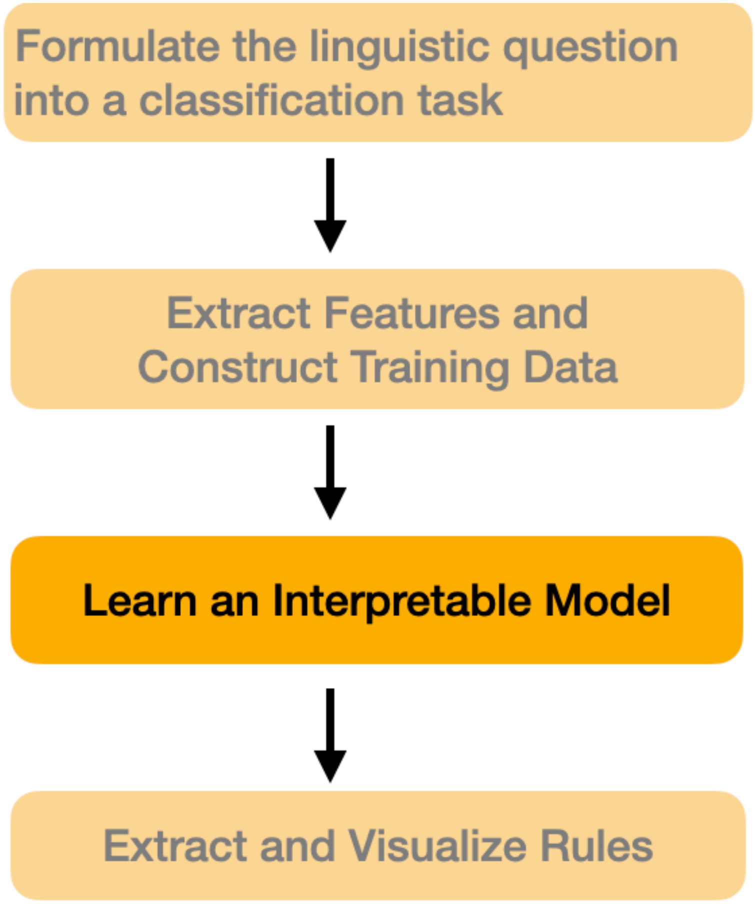
Word Order: Data

When are **objects** before or after the **verbs** in English?

- Compare the model in a clean setting ⟶ **Syntactic Universal Dependencies (SUD)**

- **Prefers syntactic heads** over content heads ⟶ more conducive to our goal of rule extraction

- Expert-annotated syntactic analysis for **>60 languages**



Formulate the linguistic question into a classification task

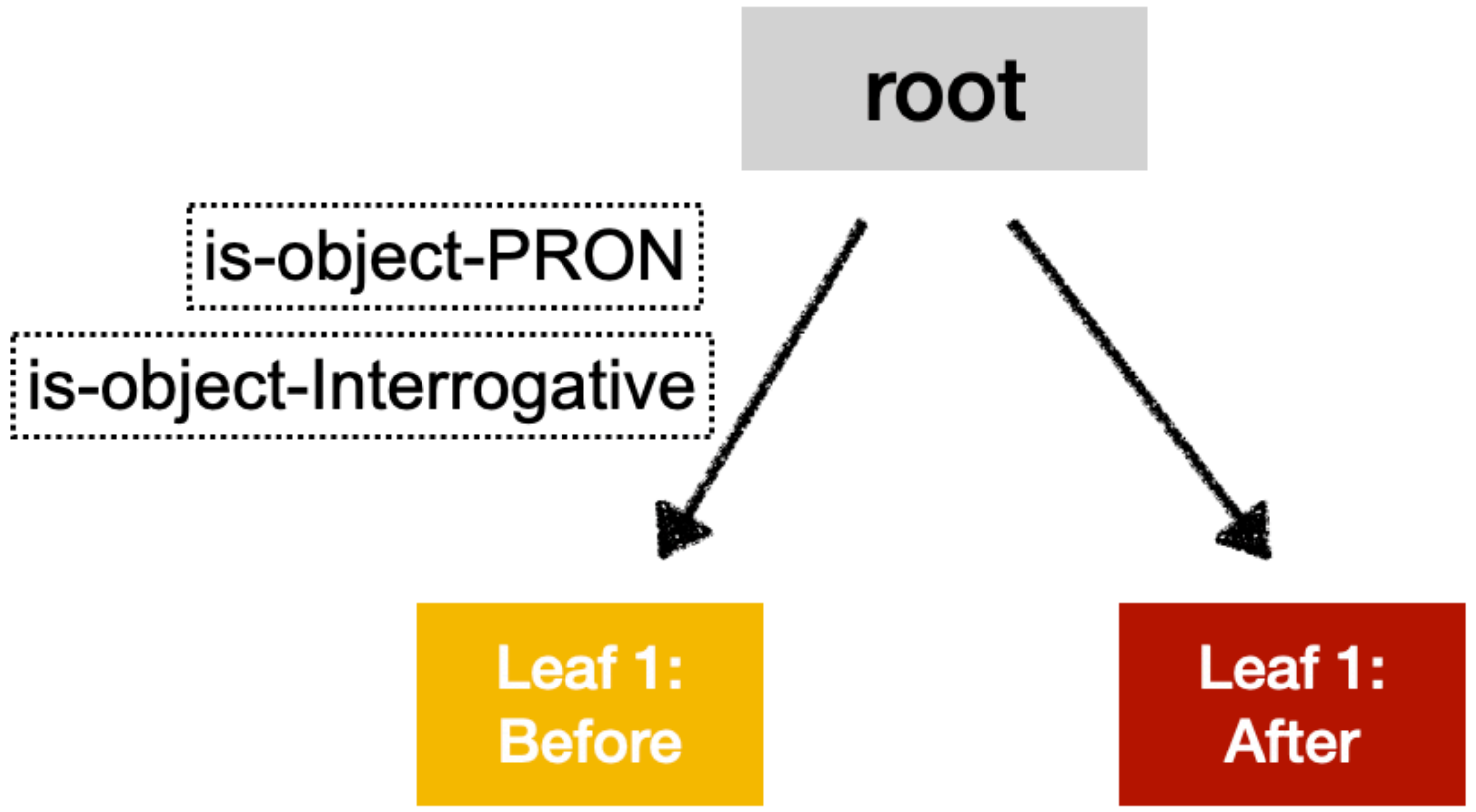
Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Automated Evaluation

When are **objects** before or after the **verbs** in English?



root

is-object-PRON

is-object-Interrogative

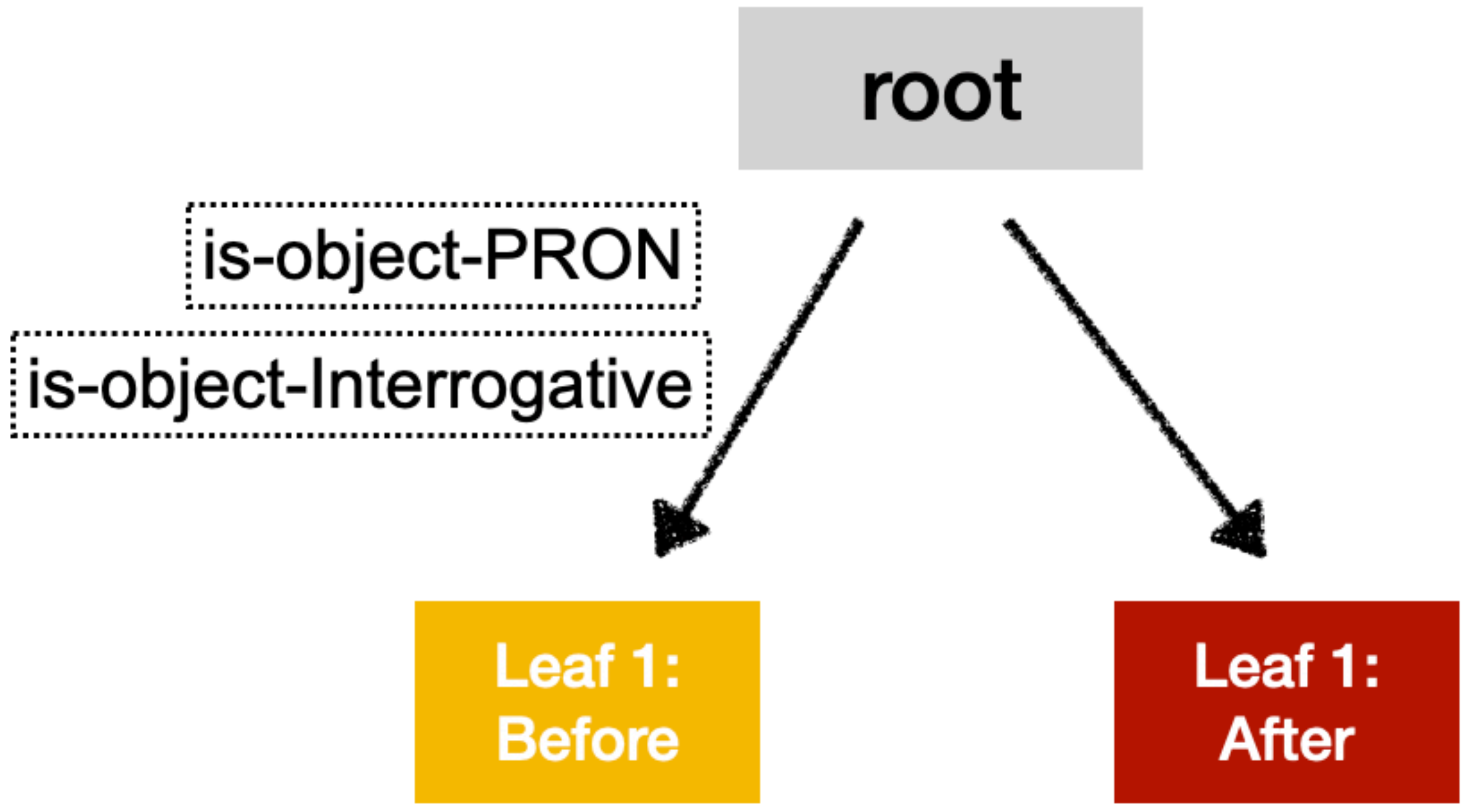
Leaf 1: Before

Leaf 1: After

Apply model
on held-out sentences

Word Order: Automated Evaluation

When are **objects** before or after the **verbs** in English?



root

is-object-PRON

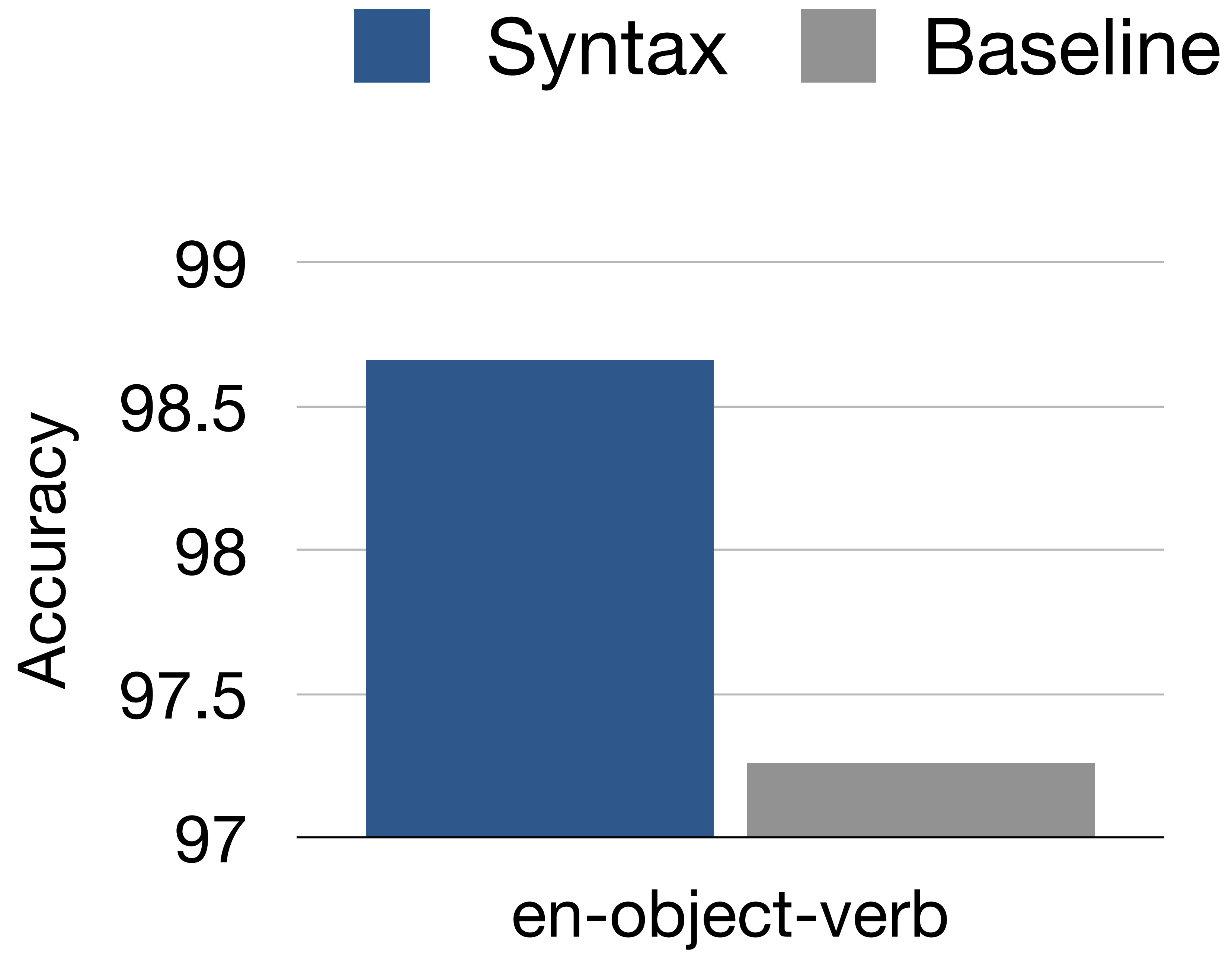
is-object-Interrogative

Leaf 1: Before

Leaf 1: After

Apply model
on held-out sentences

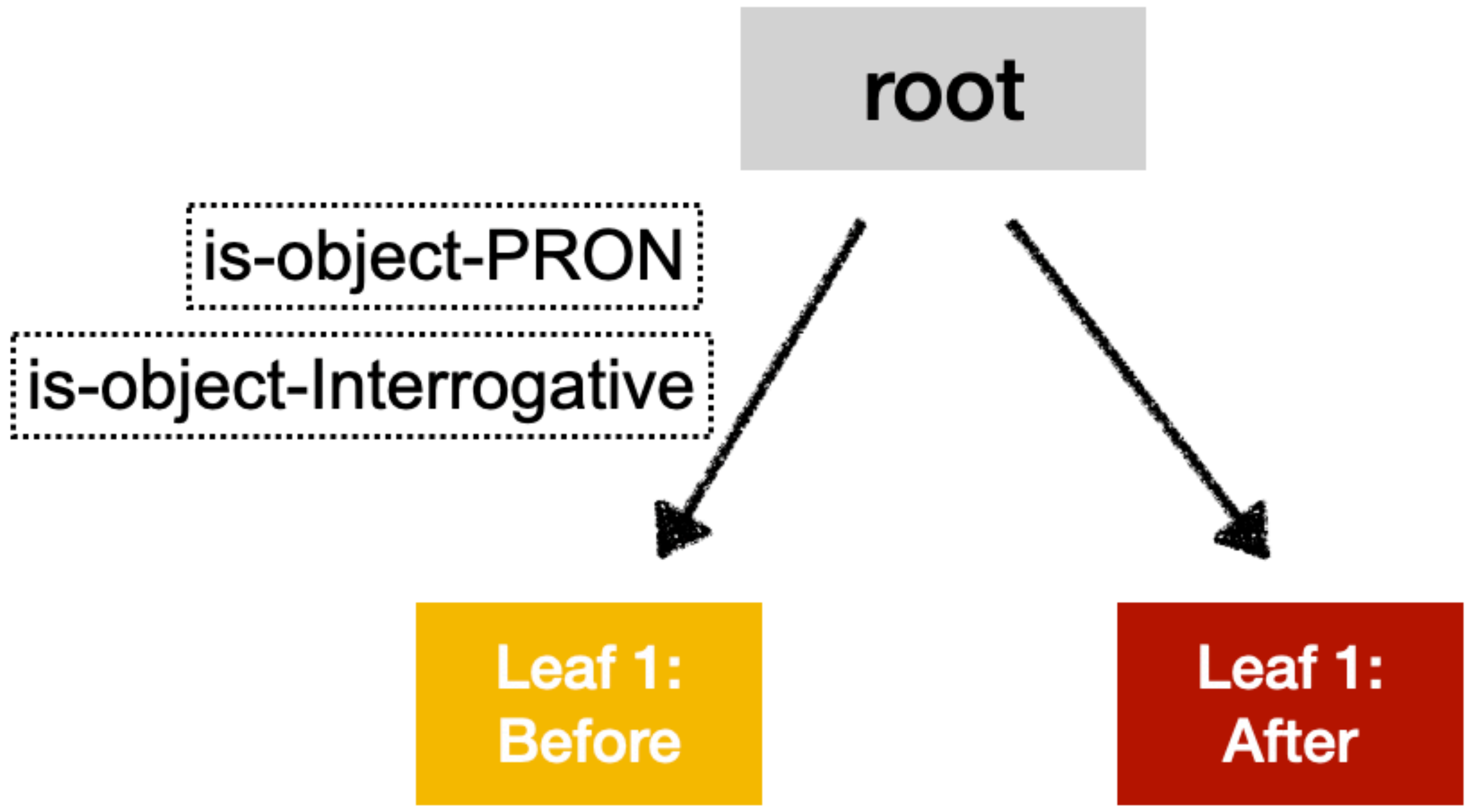
Baseline: most frequent
label in test data



AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Automated Evaluation

When are **objects** before or after the **verbs** in English?



root

is-object-PRON

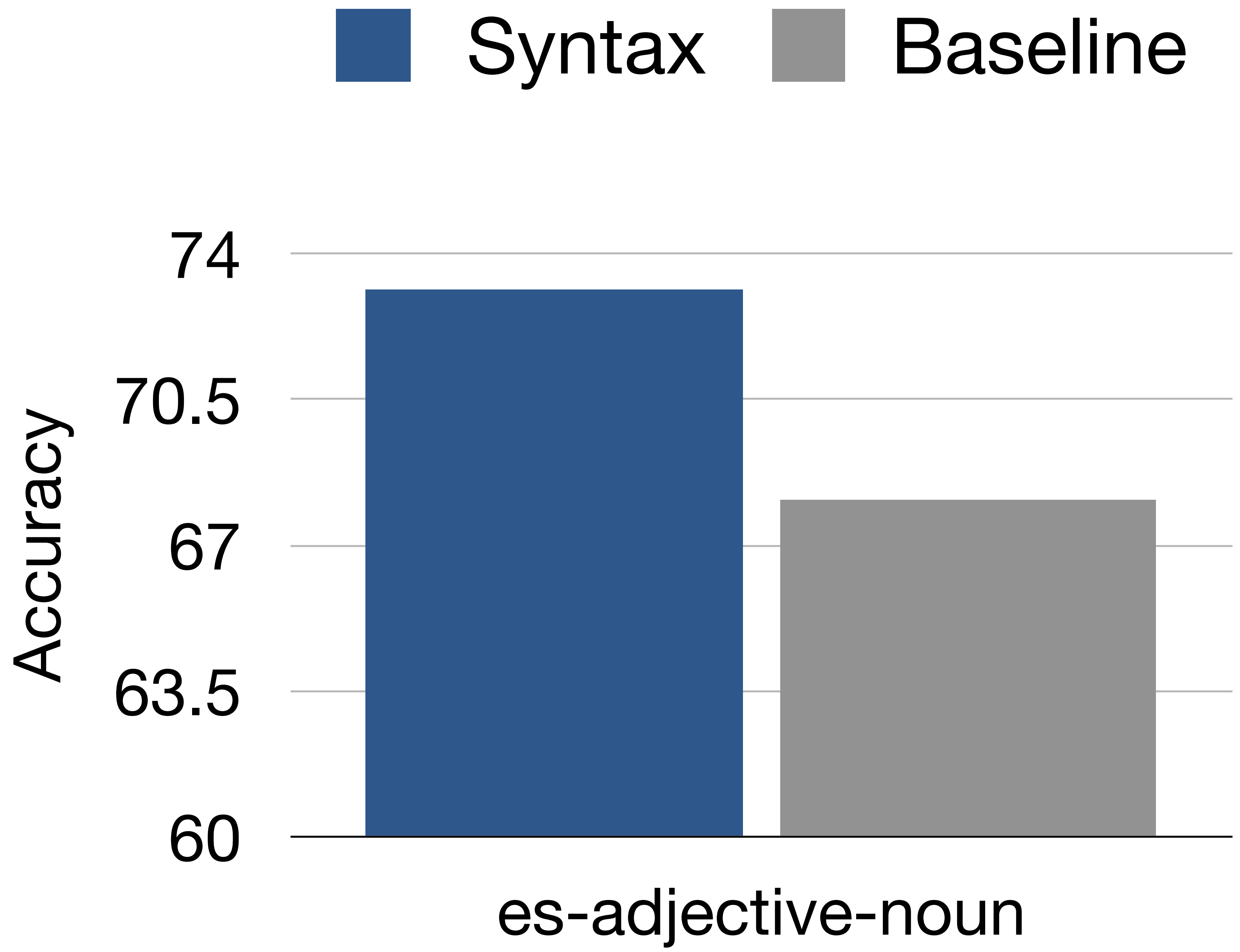
is-object-Interrogative

Leaf 1: Before

Leaf 1: After

Apply model on held-out sentences →

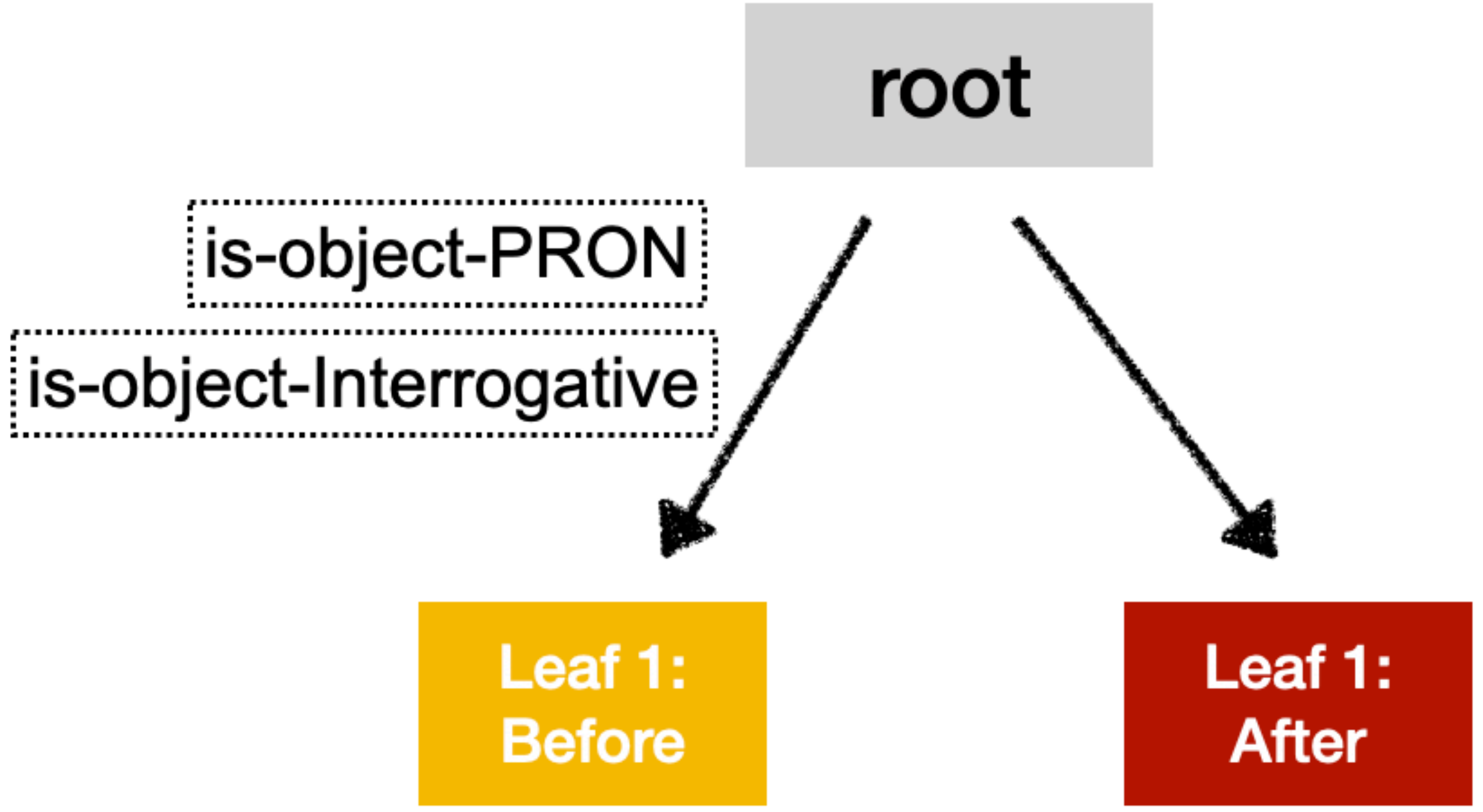
Baseline: most frequent label in test data



AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

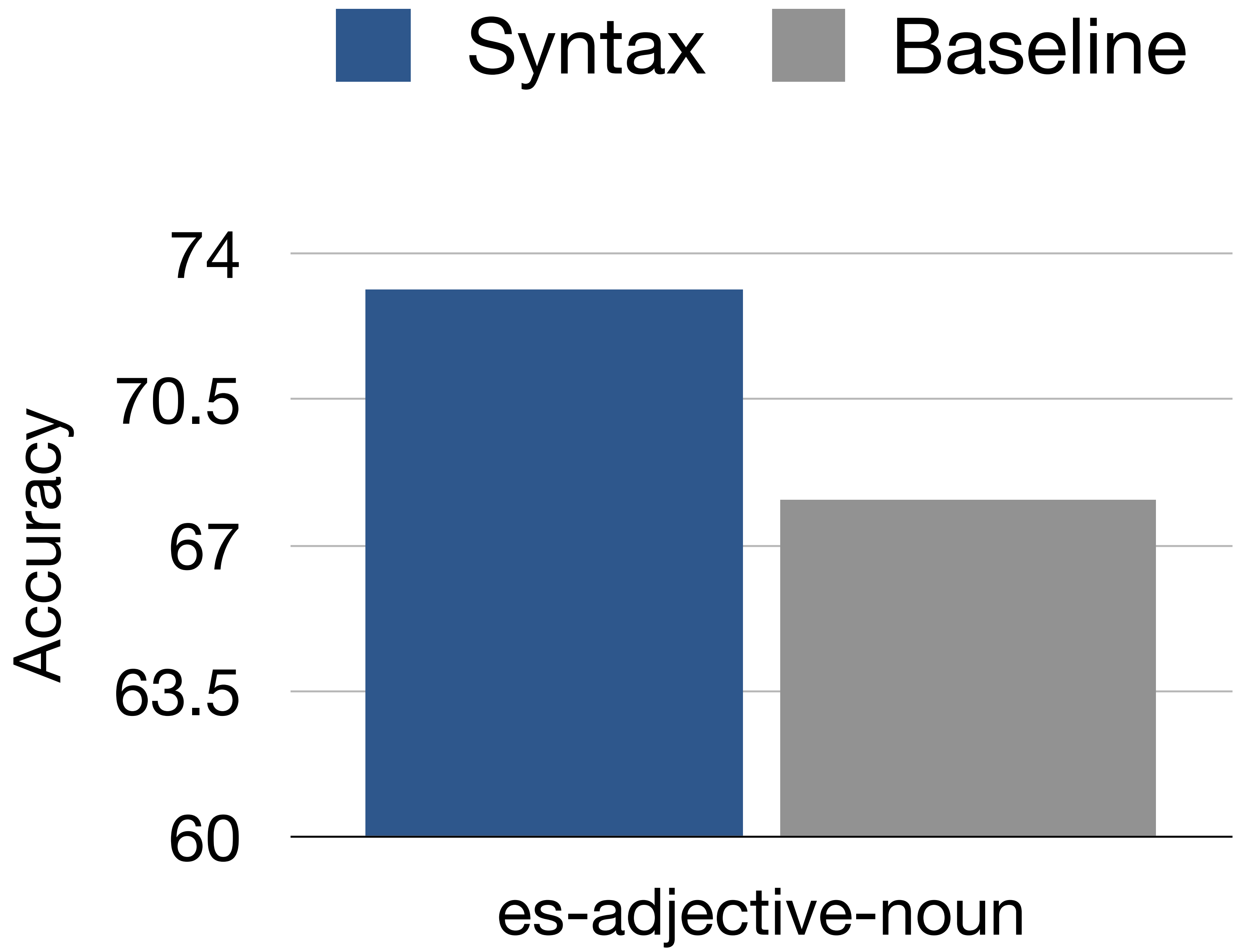
Word Order: Automated Evaluation

When are **objects** before or after the **verbs** in English?



Apply model
on held-out sentences

Baseline: most frequent
label in test data



Using syntactic signals **insufficient**!

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary, *Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission***

Word Order: Types of Features

grandes	libros	fueron	comprados	por	la	niña	pequeña
great	*books*	*were*	*bought*	*by*	*the*	*girl*	*small*

Word Order: Types of Features

grandes	libros	fueron	comprados	por	la	niña	pequeña
great	*books*	*were*	*bought*	*by*	*the*	*girl*	*small*
ADJ	NOUN					NOUN	ADJ

Word Order: Types of Features

mod

mod

grandes *great* ADJ

libros *books* NOUN

fueron *were*

comprados *bought*

por *by*

la *the*

niña *girl* NOUN

pequeña *small* ADJ

Word Order: Types of Features

syntactic

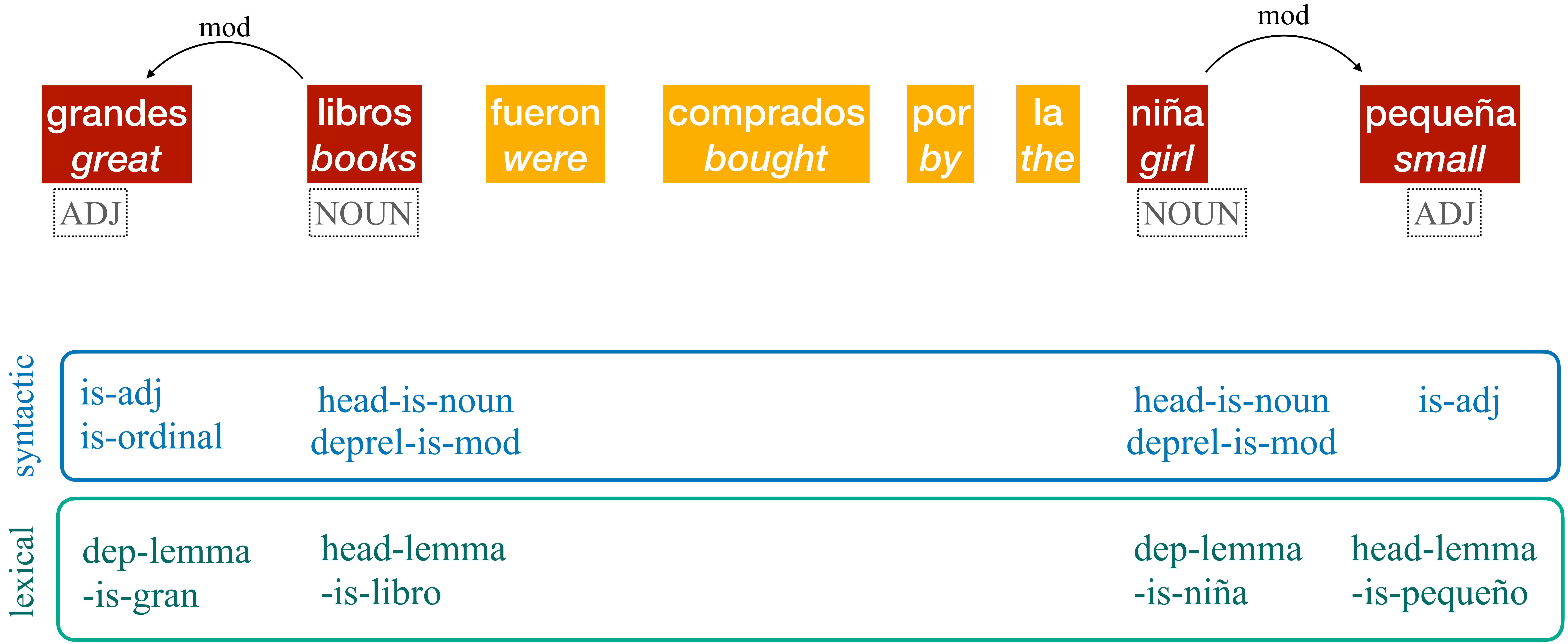
is-adj
is-ordinal

head-is-noun
deprel-is-mod

head-is-noun
deprel-is-mod

is-adj

Word Order: Types of Features



mod

mod

grandes *great* ADJ

libros *books* NOUN

fueron *were*

comprados *bought*

por *by*

la *the*

niña *girl* NOUN

pequeña *small* ADJ

syntactic

is-adj
is-ordinal

head-is-noun
deprel-is-mod

head-is-noun
deprel-is-mod

is-adj

lexical

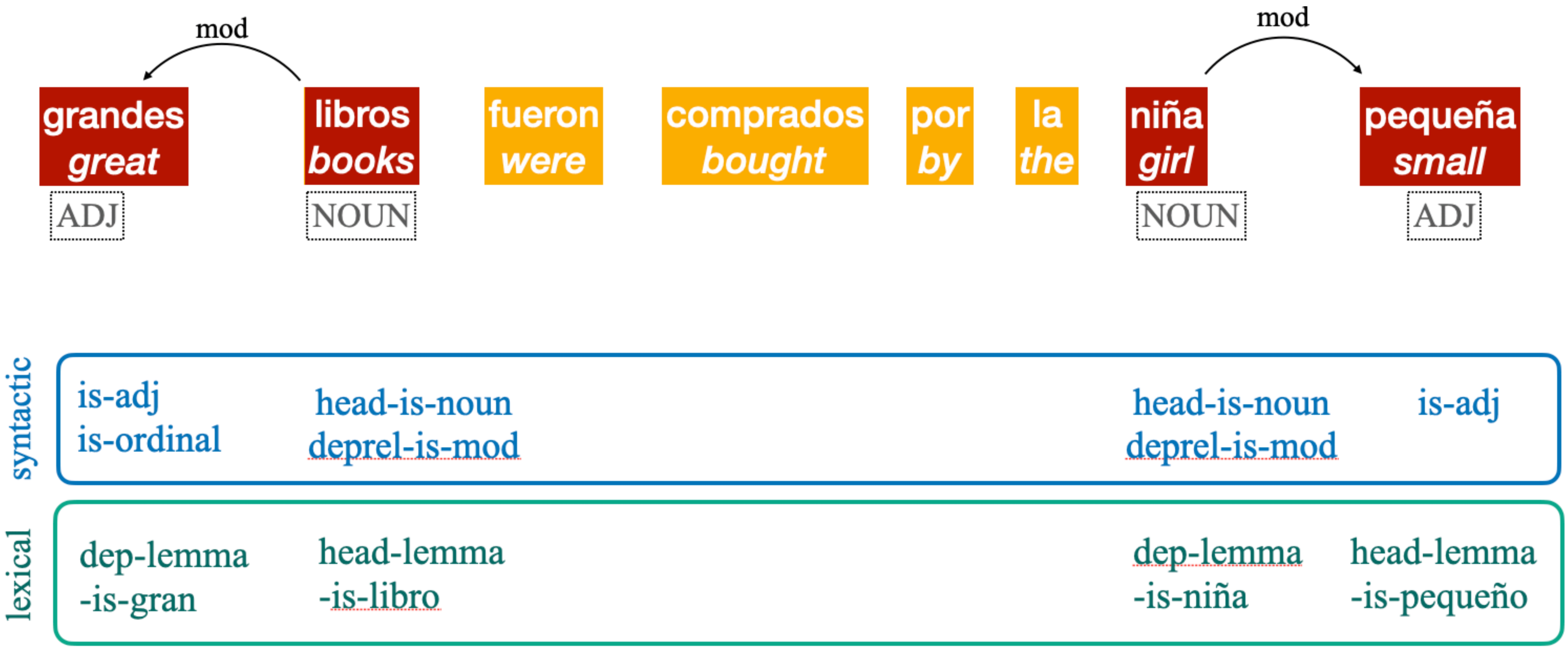
dep-lemma
-is-gran

head-lemma
-is-libro

dep-lemma
-is-niña

head-lemma
-is-pequeño

Word Order: Types of Features



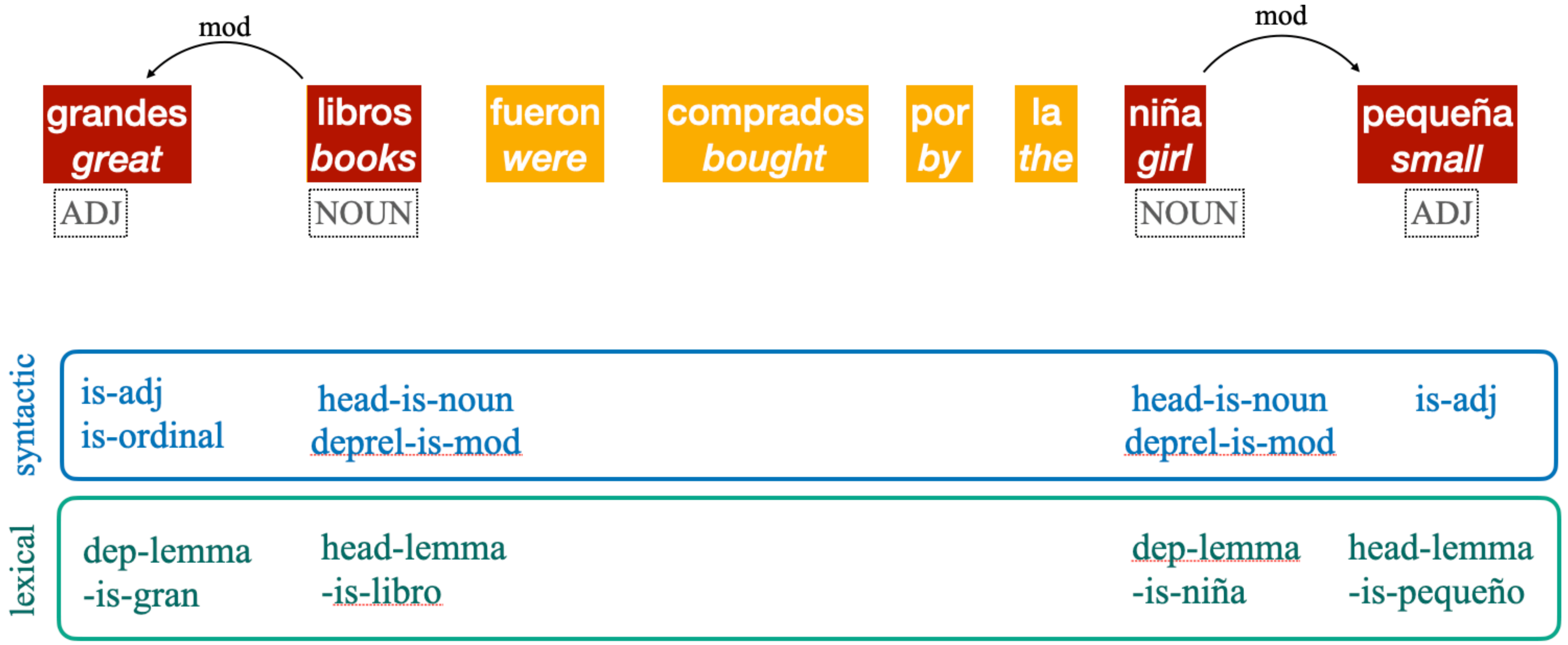
mod

mod

grandes *great*	libros *books*	fueron *were*	comprados *bought*	por *by*	la *the*	niña *girl*	pequeña *small*
ADJ	NOUN					NOUN	ADJ

syntactic	is-adj is-ordinal	head-is-noun deprel-is-mod	head-is-noun deprel-is-mod	is-adj
lexical	dep-lemma -is-gran	head-lemma -is-libro	dep-lemma -is-niña	head-lemma -is-pequeño

Word Order: Types of Features



mod

mod

grandes *great* ADJ

libros *books* NOUN

fueron *were*

comprados *bought*

por *by*

la *the*

niña *girl* NOUN

pequeña *small* ADJ

syntactic

is-adj
is-ordinal

head-is-noun
deprel-is-mod

head-is-noun
deprel-is-mod

is-adj

lexical

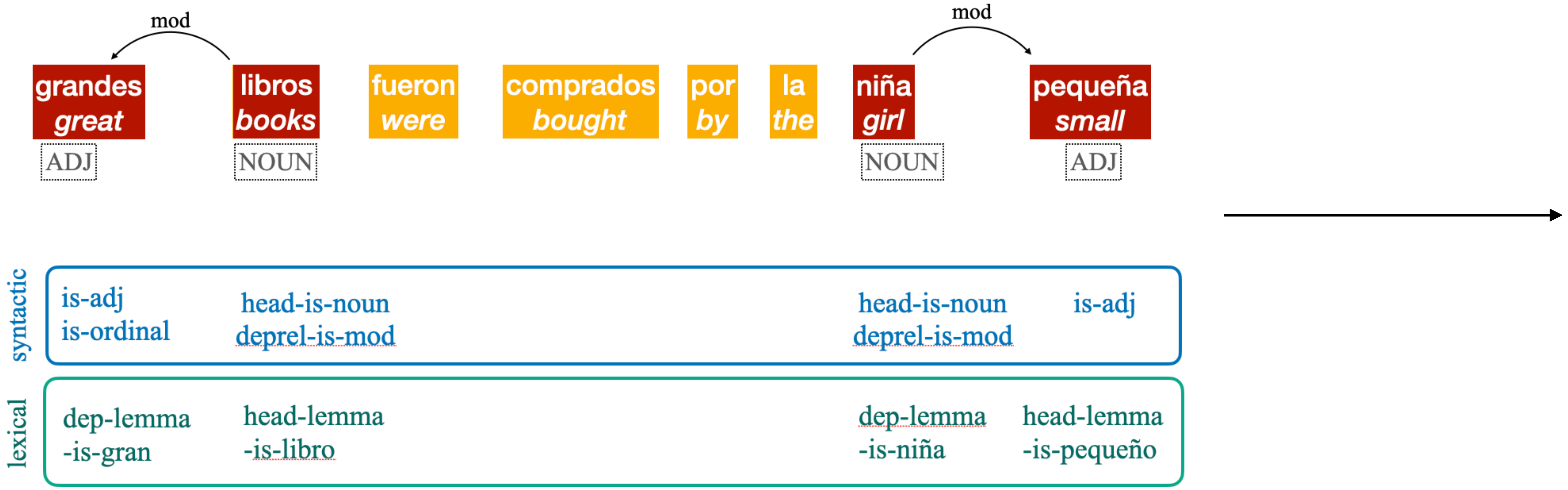
dep-lemma
-is-gran

head-lemma
-is-libro

dep-lemma
-is-niña

head-lemma
-is-pequeño

Word Order: Types of Features

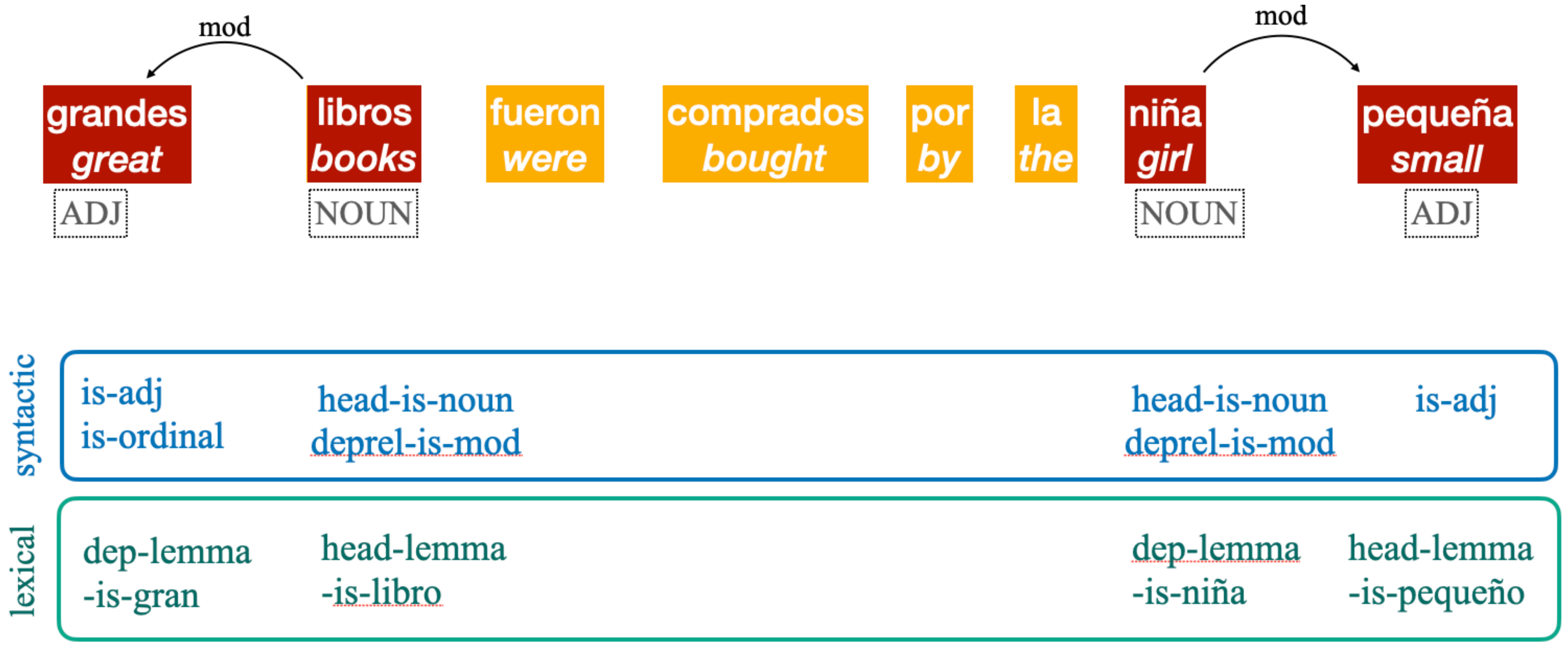


syntactic: is-adj is-ordinal | head-is-noun deprel-is-mod | head-is-noun deprel-is-mod | is-adj

lexical: dep-lemma-is-gran | head-lemma-is-libro | dep-lemma-is-niña | head-lemma-is-pequeño

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Types of Features



mod

grandes *great* ADJ

libros *books* NOUN

fueron *were*

comprados *bought*

por *by*

la *the*

mod

niña *girl* NOUN

pequeña *small* ADJ

syntactic

is-adj
is-ordinal

head-is-noun
deprel-is-mod

head-is-noun
deprel-is-mod

is-adj

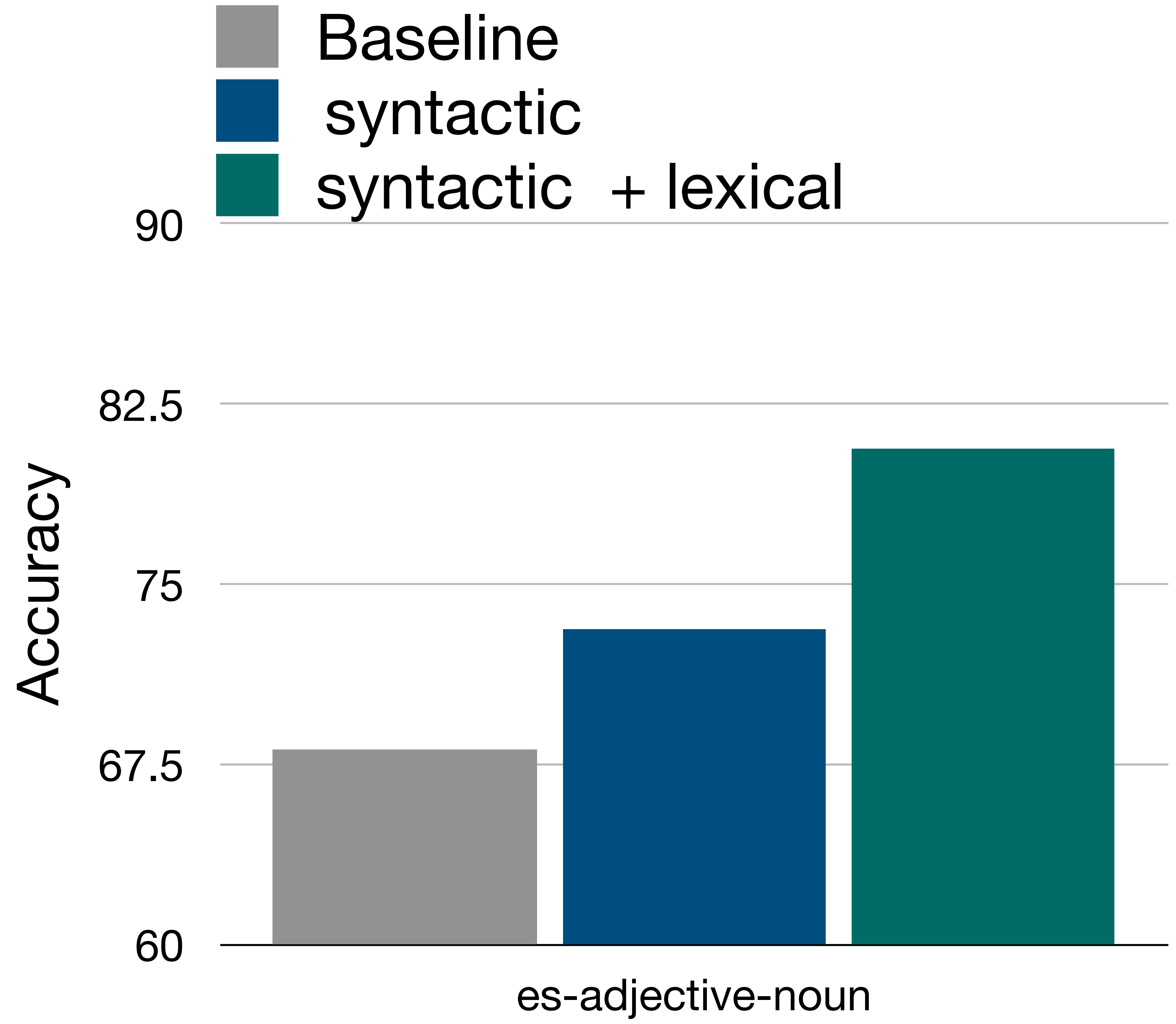
lexical

dep-lemma
-is-gran

head-lemma
-is-libro

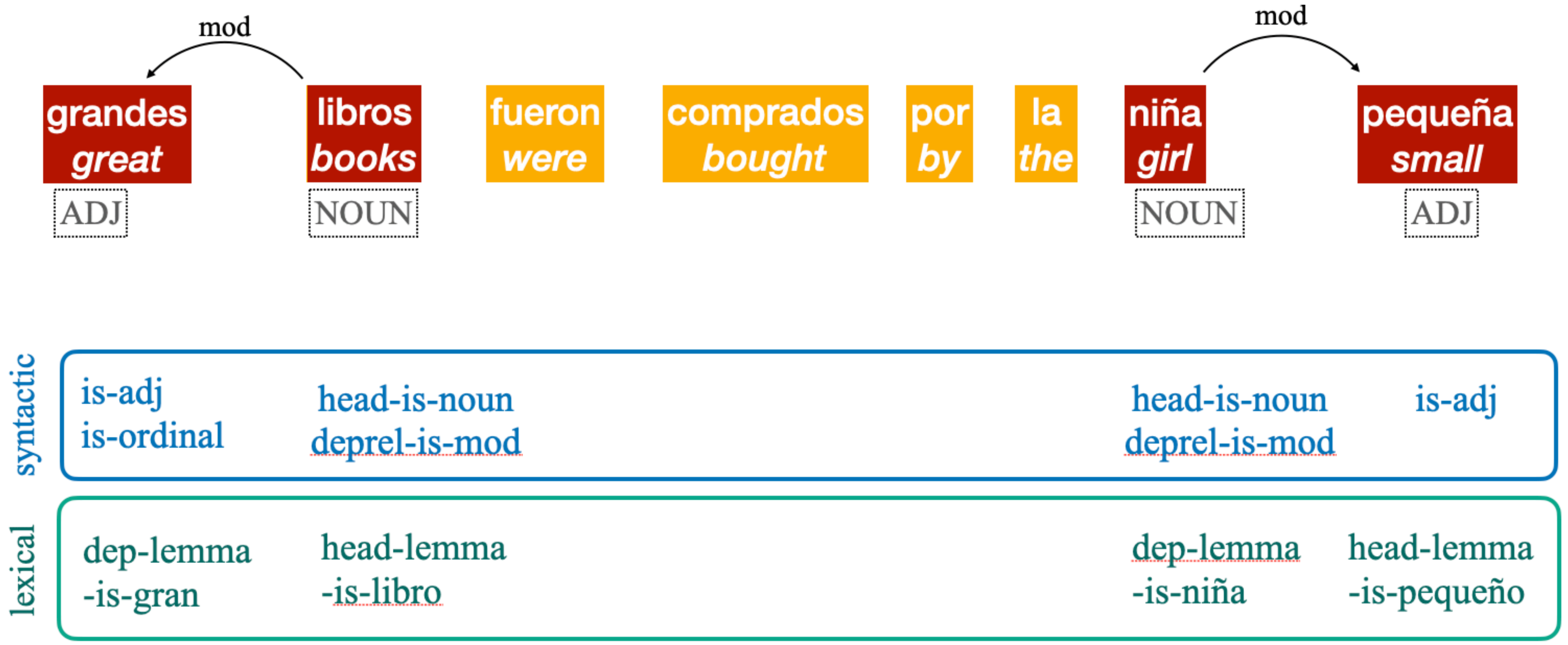
dep-lemma
-is-niña

head-lemma
-is-pequeño



AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Types of Features



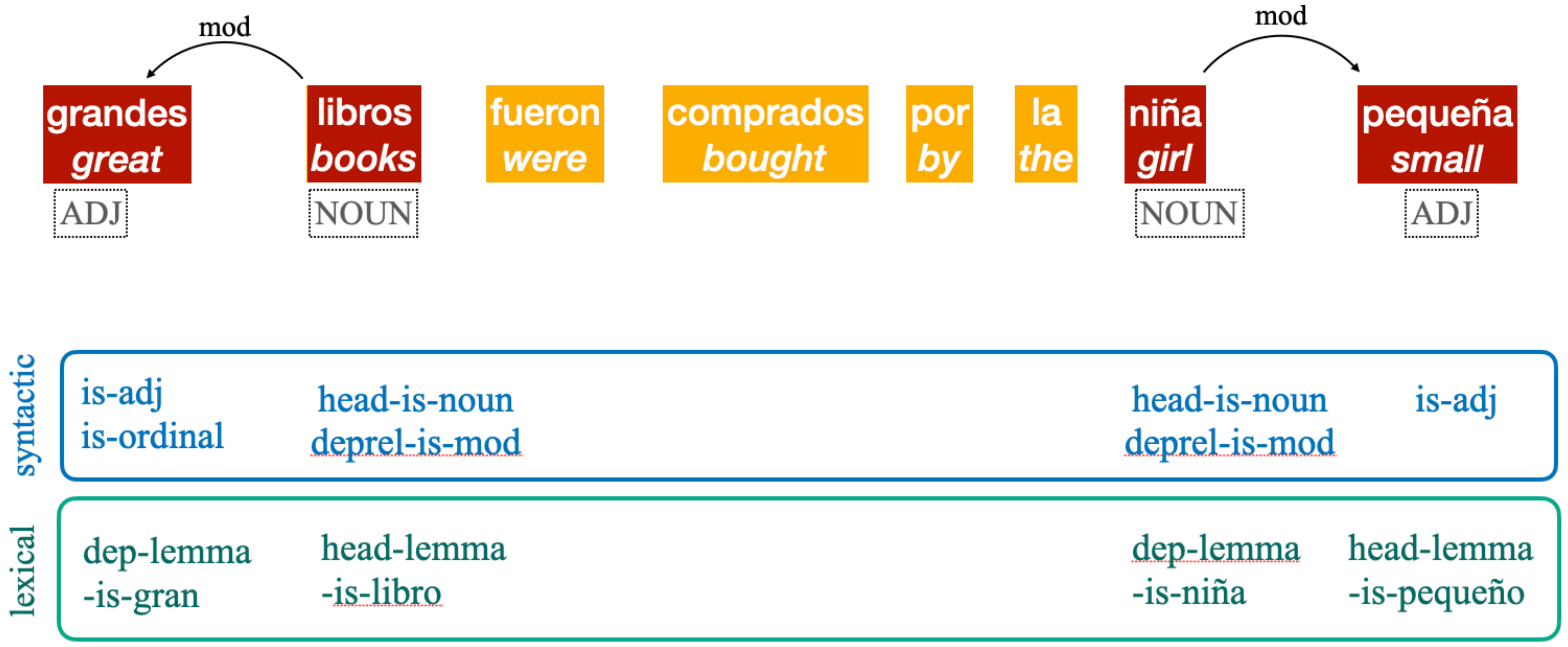
mod
mod

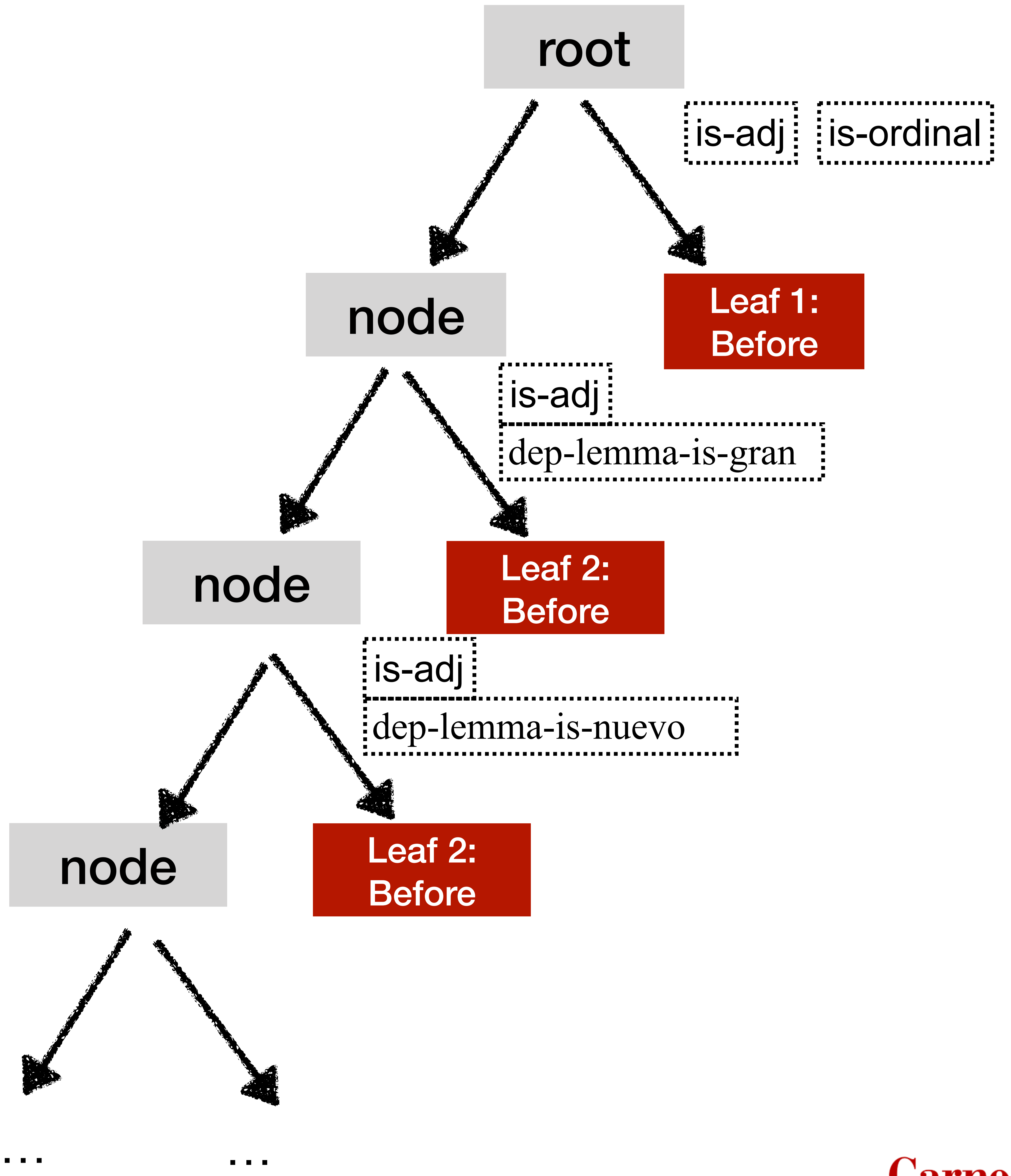
grandes *great* ADJ
libros *books* NOUN
fueron *were*
comprados *bought*
por *by*
la *the*
niña *girl* NOUN
pequeña *small* ADJ

syntactic
is-adj
is-ordinal
head-is-noun
deprel-is-mod
head-is-noun
deprel-is-mod
is-adj

lexical
dep-lemma -is-gran
head-lemma -is-libro
dep-lemma -is-niña
head-lemma -is-pequeño

Word Order: Types of Features





AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Types of Features

Word Order: Types of Features

- **Word Embeddings** ⟶ capture semantic/syntactic similarity

Word Order: Types of Features

- **Word Embeddings** ⟶ capture semantic/syntactic similarity
- **BUT** how to interpret what each dimension means?

Word Order: Types of Features

- **Word Embeddings** ⟶ capture semantic/syntactic similarity

- **BUT** how to interpret what each dimension means?

- **SPINE**[1] ⟶ sparse and non-negative transformation over pretrained embeddings

---

[1] SPINE: SParse Interpretable Neural Embeddings
Anant Subramanian, Danish Pruthi, Harsh Jhamtani, Taylor Berg-Kirkpatrick, Eduard Hovy

Word Order: Types of Features

- **Word Embeddings** ⟶ capture semantic/syntactic similarity
- **BUT** how to interpret what each dimension means?
- **SPINE**[1] ⟶ sparse and non-negative transformation over pretrained embeddings

dim-0: wrist, shoulder, ligament, ankle, thigh

Semantic similarity

---

[1] SPINE: SParse Interpretable Neural Embeddings
Anant Subramanian, Danish Pruthi, Harsh Jhamtani, Taylor Berg-Kirkpatrick, Eduard Hovy

Word Order: Types of Features

- **Word Embeddings** ⟶ capture semantic/syntactic similarity
- **BUT** how to interpret what each dimension means?
- **SPINE**[1] ⟶ sparse and non-negative transformation over pretrained embeddings

dim-0: wrist, shoulder, ligament, ankle, thigh

Semantic similarity

dim-1: extensively, adequately, traditionally, royally, fully

Syntactic similarity

---

[1] SPINE: SParse Interpretable Neural Embeddings
Anant Subramanian, Danish Pruthi, Harsh Jhamtani, Taylor Berg-Kirkpatrick, Eduard Hovy

Word Order: Types of Features

- **Word Embeddings** ⟶ capture semantic/syntactic similarity

- **BUT** how to interpret what each dimension means?

- **SPINE**[1] ⟶ sparse and non-negative transformation over pretrained embeddings

dim-0: wrist, shoulder, ligament, ankle, thigh — Semantic similarity

dim-1: extensively, adequately, traditionally, royally, fully — Syntactic similarity

dim-2: torque, joystick, grip, wrist, swinging

---

[1] SPINE: SParse Interpretable Neural Embeddings
Anant Subramanian, Danish Pruthi, Harsh Jhamtani, Taylor Berg-Kirkpatrick, Eduard Hovy

Word Order: Types of Features

- **Word Embeddings** ⟶ capture semantic/syntactic similarity
- **BUT** how to interpret what each dimension means?
- **SPINE**[1] ⟶ sparse and non-negative transformation over pretrained embeddings

dim-0: wrist, shoulder, ligament, ankle, thigh — Semantic similarity

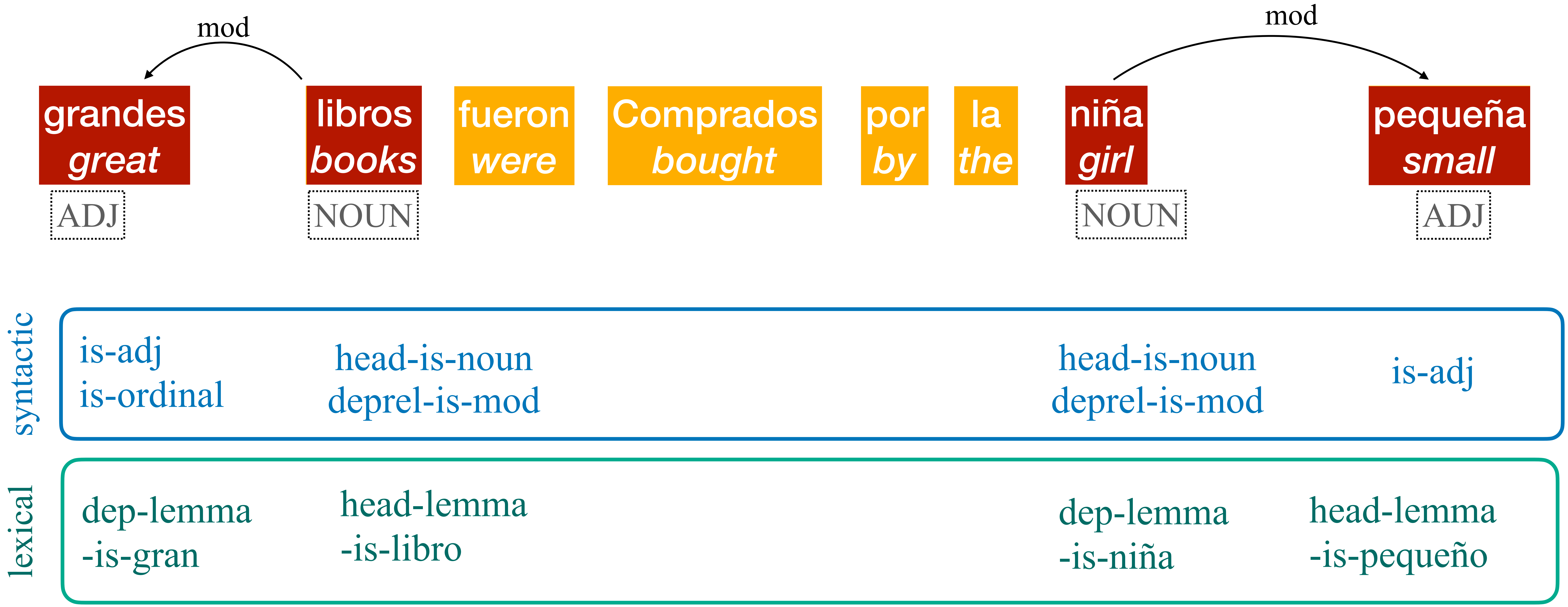
dim-1: extensively, adequately, traditionally, royally, fully — Syntactic similarity

dim-2: torque, joystick, grip, wrist, swinging — Multi-Aspect Information

---

[1] SPINE: SParse Interpretable Neural Embeddings
Anant Subramanian, Danish Pruthi, Harsh Jhamtani, Taylor Berg-Kirkpatrick, Eduard Hovy

Word Order: Types of Features



mod

mod

grandes *great*	libros *books*	fueron *were*	Comprados *bought*	por *by*	la *the*	niña *girl*	pequeña *small*
ADJ	NOUN					NOUN	ADJ

syntactic	is-adj is-ordinal	head-is-noun deprel-is-mod	head-is-noun deprel-is-mod	is-adj
lexical	dep-lemma -is-gran	head-lemma -is-libro	dep-lemma -is-niña	head-lemma -is-pequeño

Word Order: Types of Features

grandes	libros	fueron	Comprados	por	la	niña	pequeña
great	*books*	*were*	*bought*	*by*	*the*	*girl*	*small*
ADJ	NOUN					NOUN	ADJ

syntactic	is-adj is-ordinal	head-is-noun deprel-is-mod	head-is-noun deprel-is-mod	is-adj
lexical	dep-lemma -is-gran	head-lemma -is-libro	dep-lemma -is-niña	head-lemma -is-pequeño
semantic	word-is-like-primera	head-word-is-like-págin, publicación	word-is-like-niño	head-word-is-like-nueva, vuelta

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Types of Features

mod

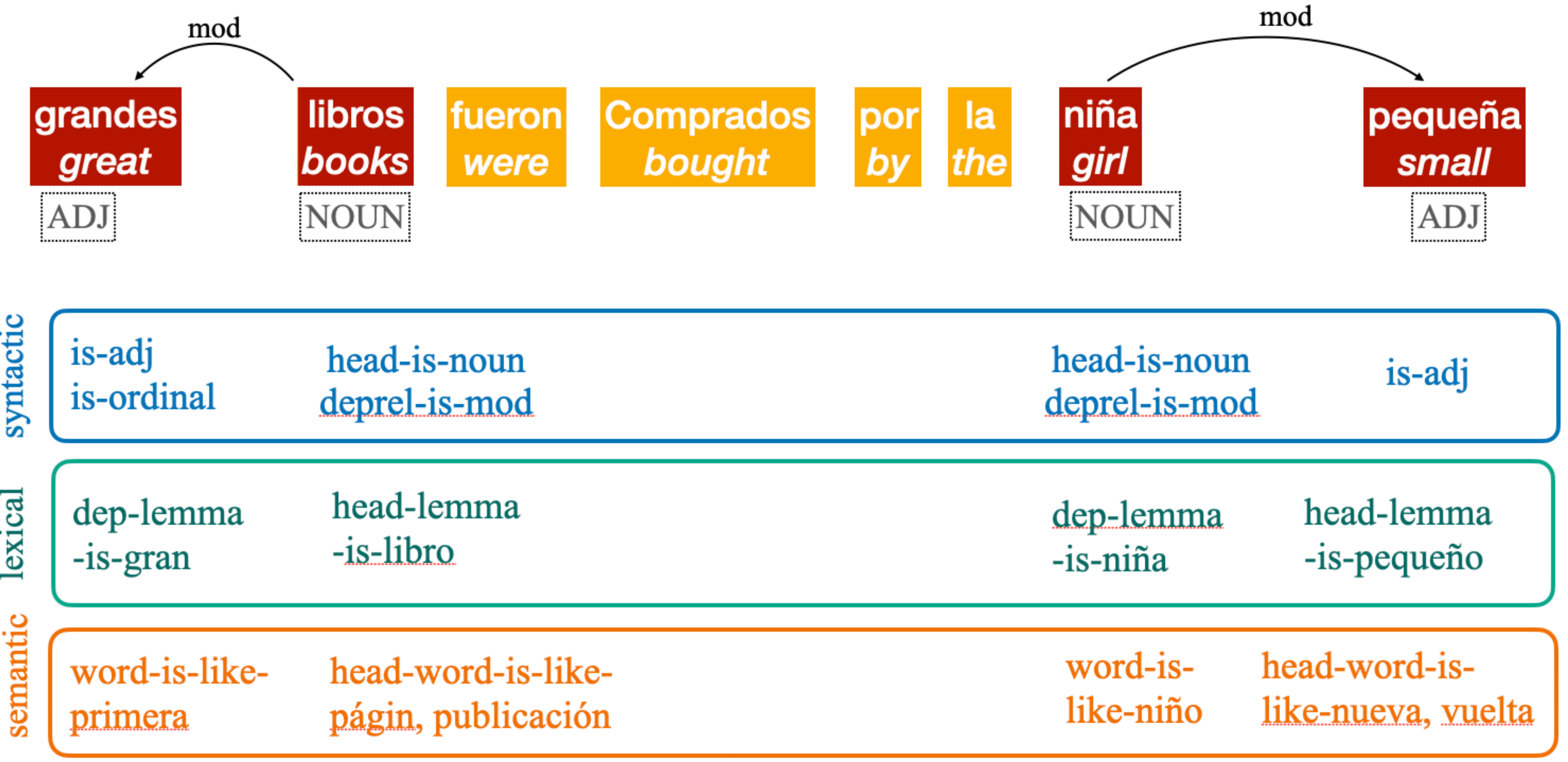
mod

grandes *great*	libros *books*	fueron *were*	Comprados *bought*	por *by*	la *the*	niña *girl*	pequeña *small*
ADJ	NOUN					NOUN	ADJ

syntactic	is-adj is-ordinal	head-is-noun deprel-is-mod	head-is-noun deprel-is-mod	is-adj
lexical	dep-lemma -is-gran	head-lemma -is-libro	dep-lemma -is-niña	head-lemma -is-pequeño
semantic	word-is-like-primera	head-word-is-like-págin, publicación	word-is-like-niño	head-word-is-like-nueva, vuelta

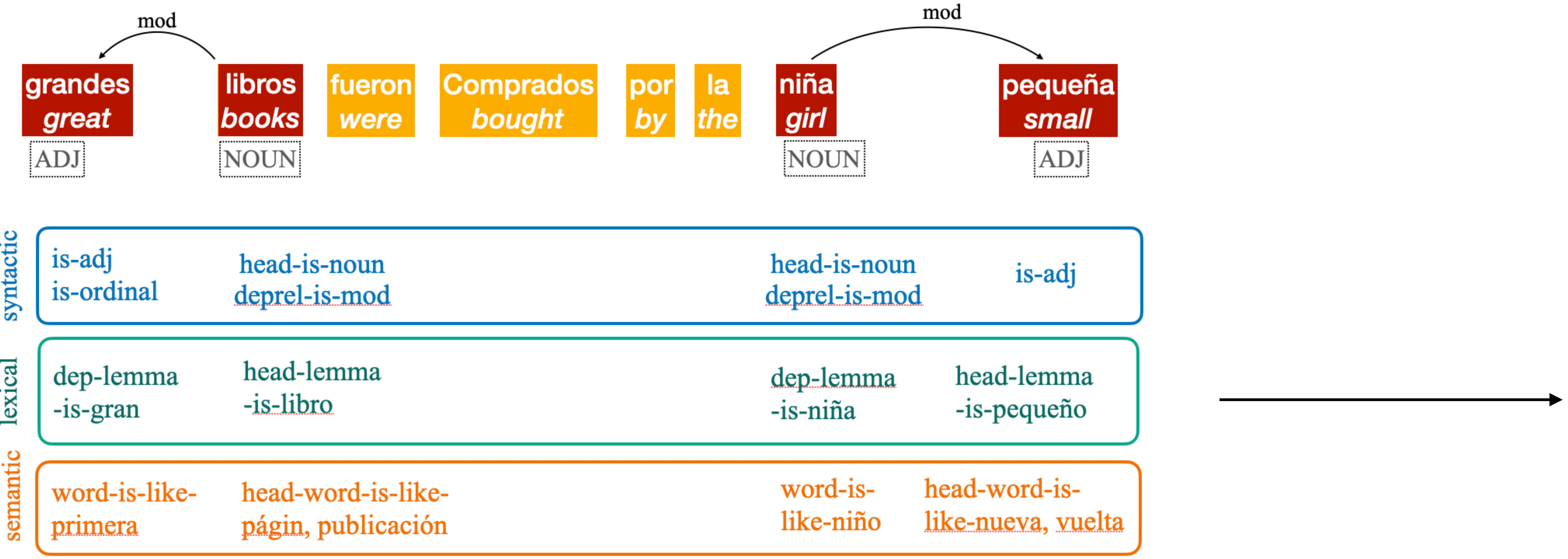
AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Types of Features



AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Types of Features



is-adj
is-ordinal

head-is-noun
deprel-is-mod

head-is-noun
deprel-is-mod

is-adj

dep-lemma -is-gran

head-lemma -is-libro

dep-lemma -is-niña

head-lemma -is-pequeño

word-is-like-primera

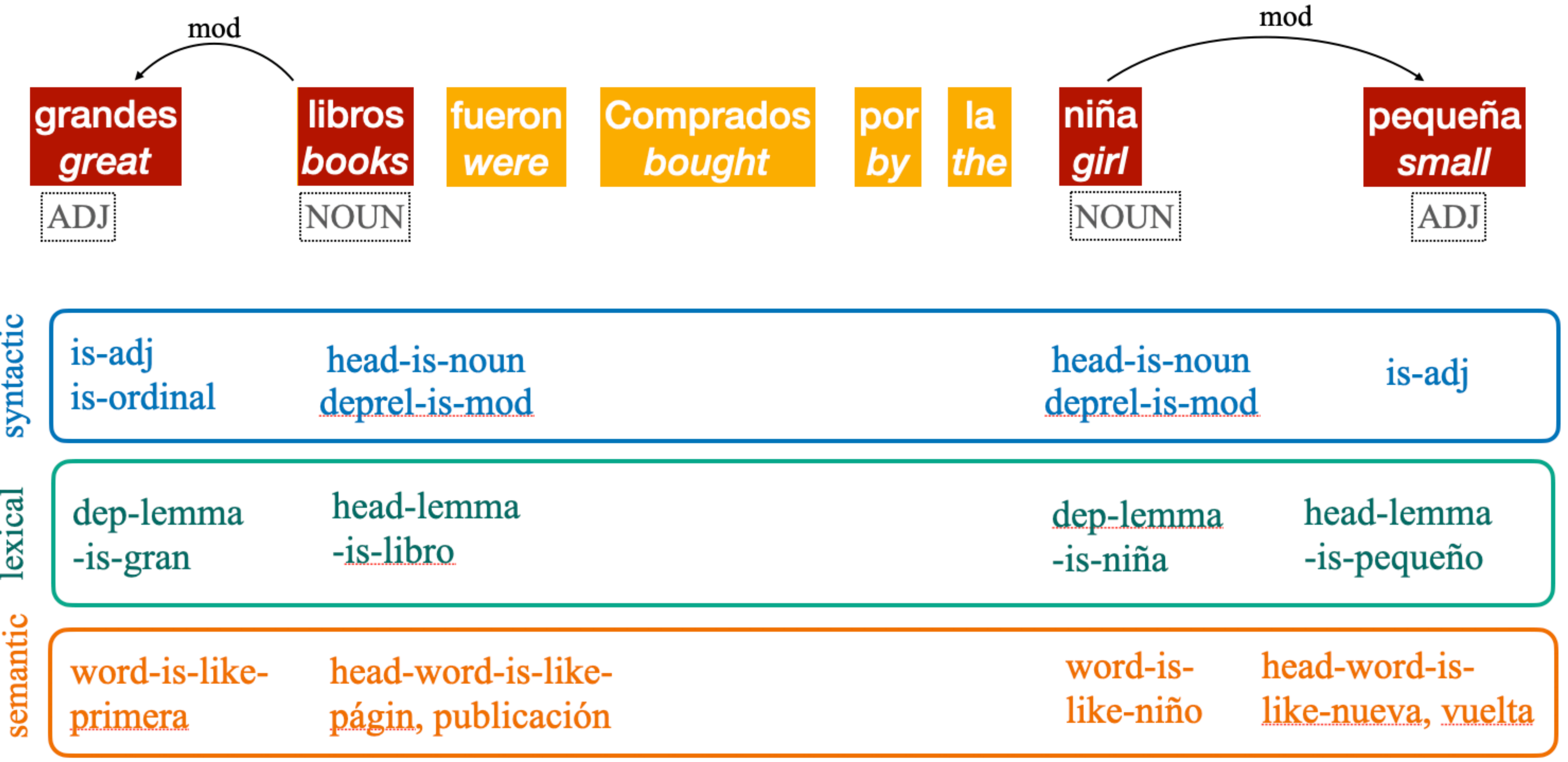
head-word-is-like-págin, publicación

word-is-like-niño

head-word-is-like-nueva, vuelta

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Types of Features



mod

mod

grandes *great* ADJ

libros *books* NOUN

fueron *were*

Comprados *bought*

por *by*

la *the*

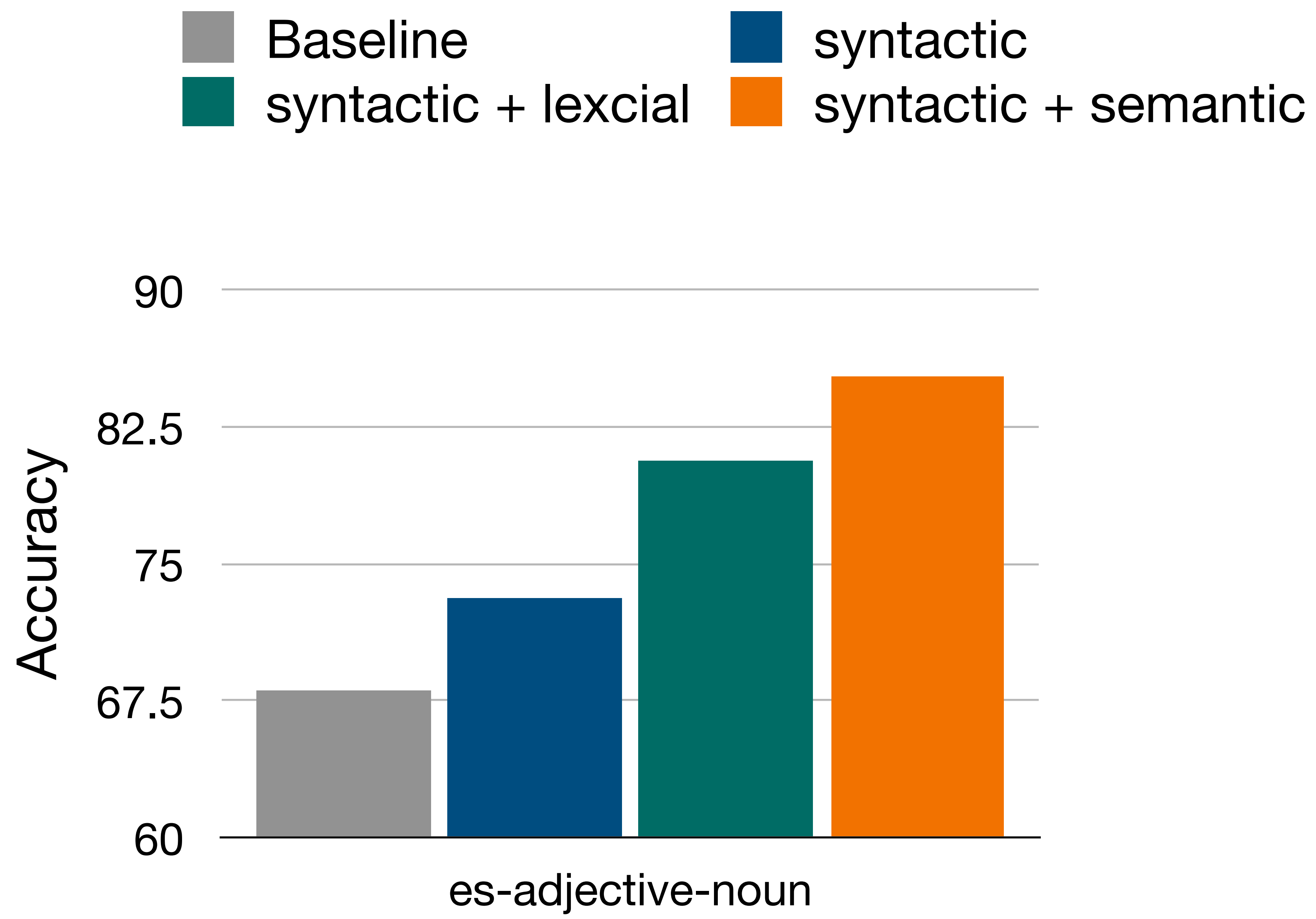
niña *girl* NOUN

pequeña *small* ADJ

syntactic: is-adj is-ordinal | head-is-noun deprel-is-mod | head-is-noun deprel-is-mod | is-adj

lexical: dep-lemma -is-gran | head-lemma -is-libro | dep-lemma -is-niña | head-lemma -is-pequeño

semantic: word-is-like-primera | head-word-is-like-págin, publicación | word-is-like-niño | head-word-is-like-nueva, vuelta



AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Extract rules

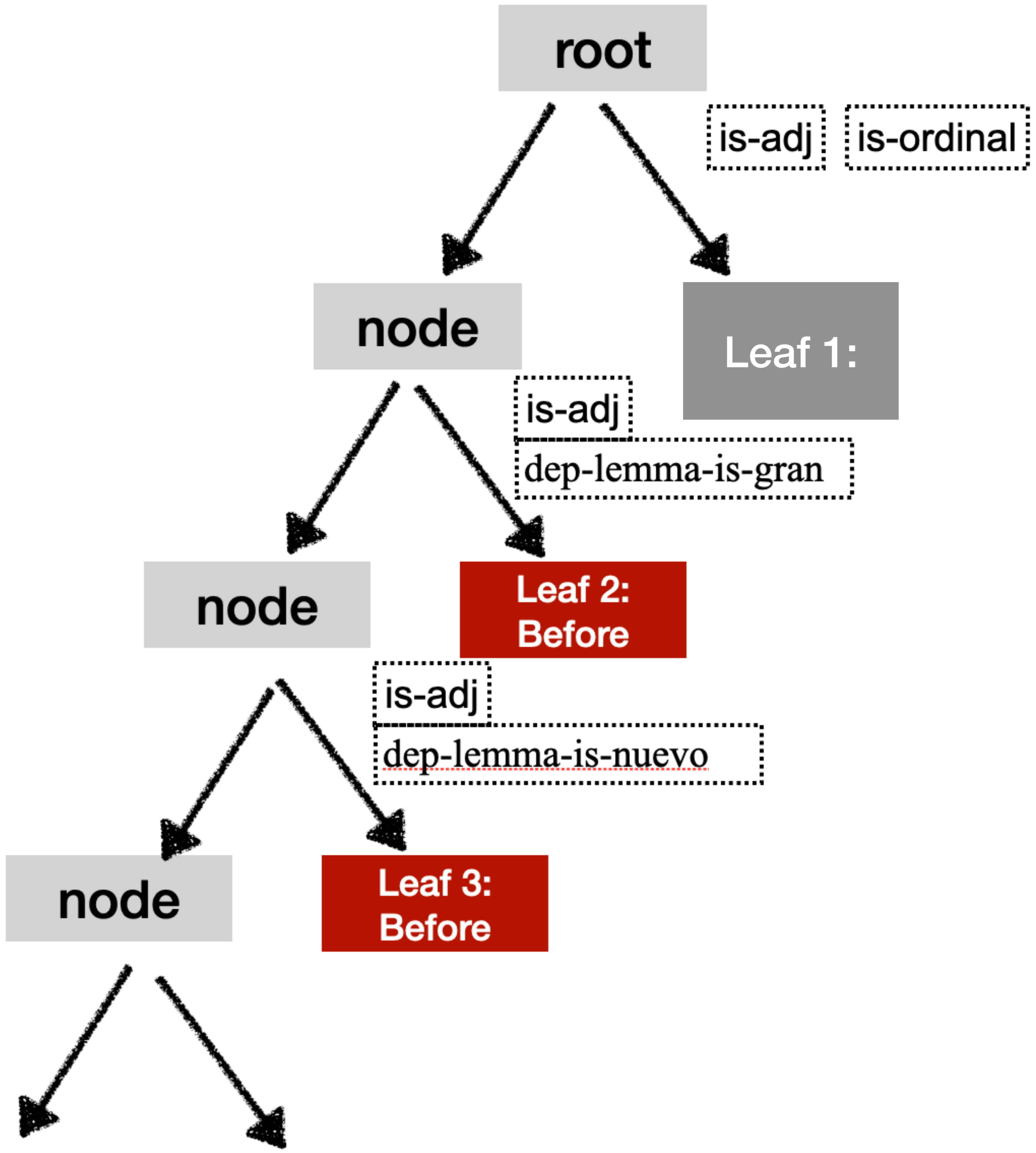
Formulate the linguistic question into a classification task

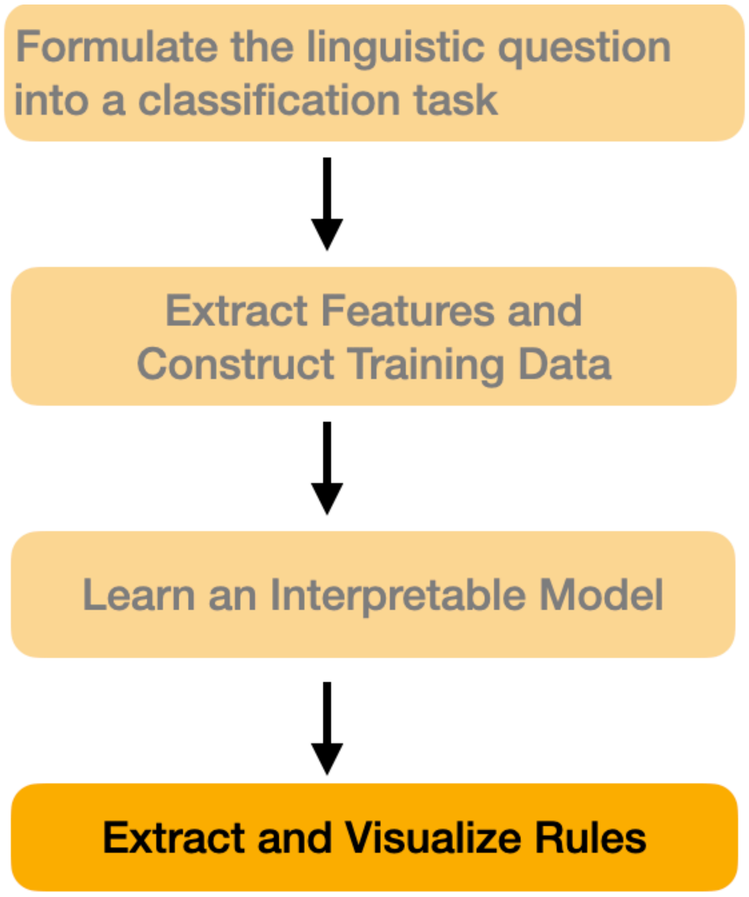
Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Extract rules





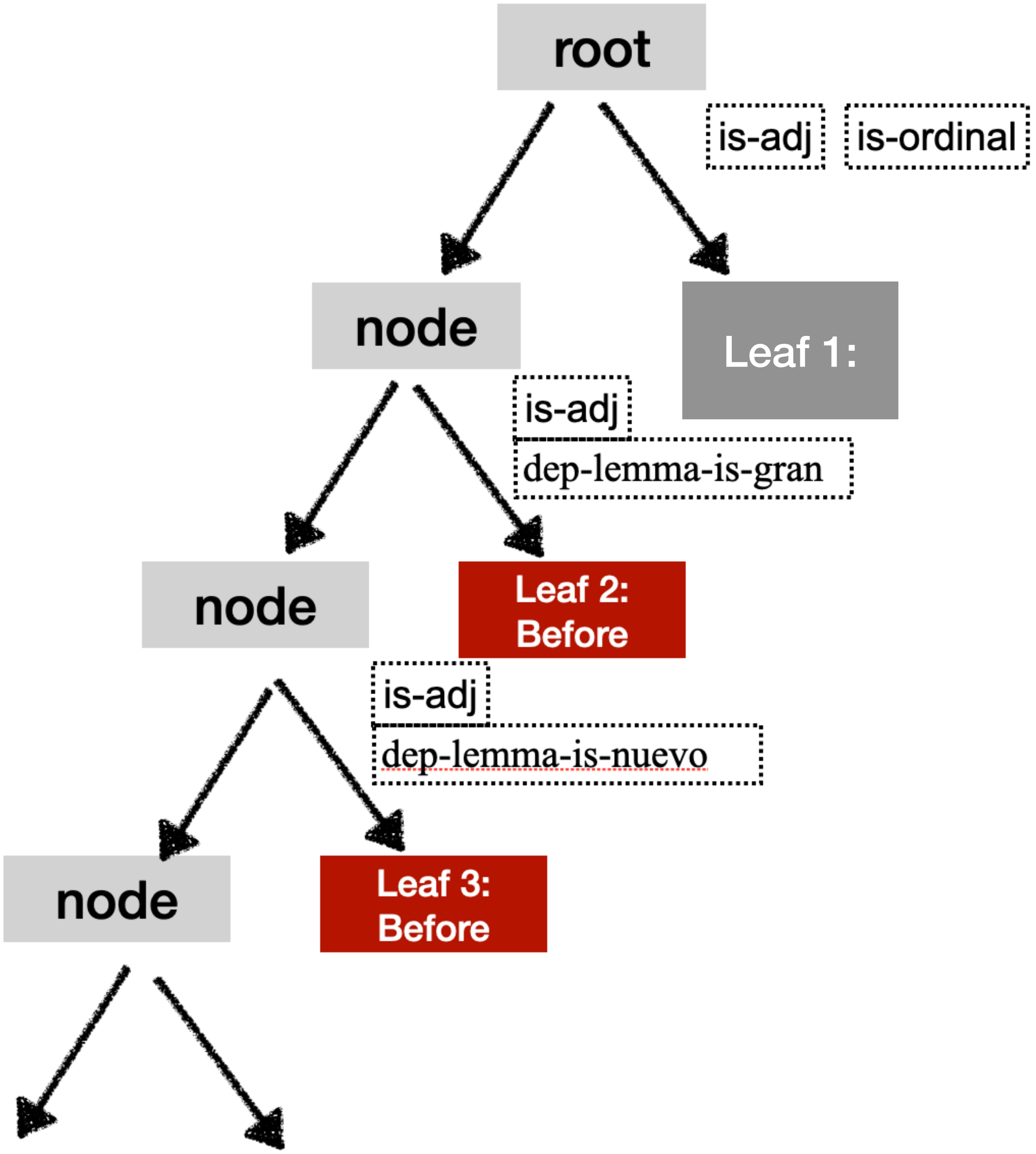
Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

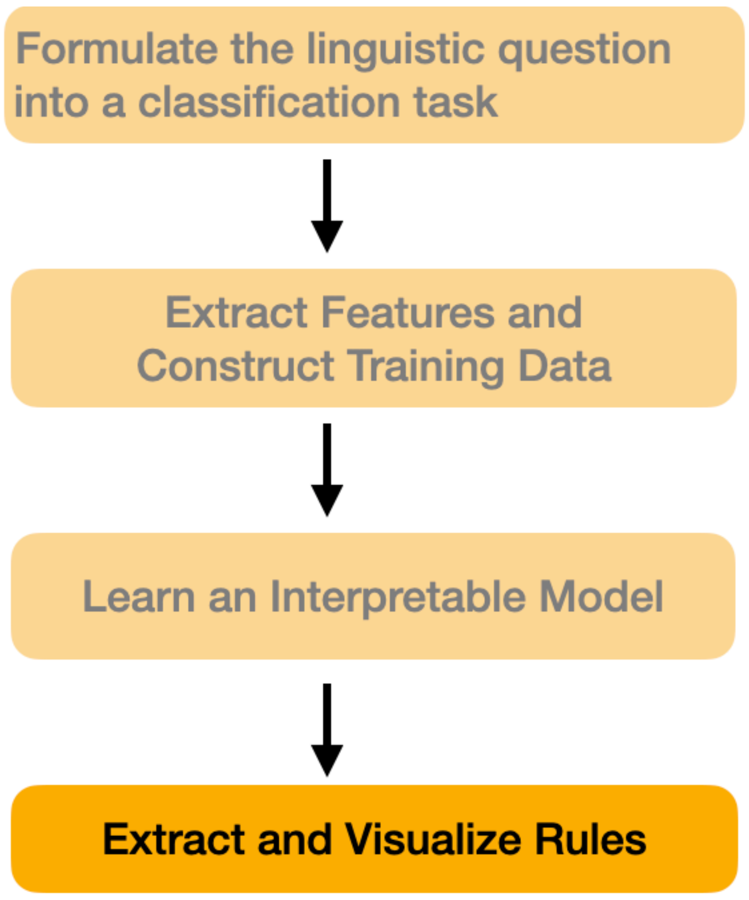
Extract and Visualize Rules

Word Order: Extract rules


root
is-adj
is-ordinal
node
Leaf 1:
is-adj
dep-lemma-is-gran
node
Leaf 2: Before
is-adj
dep-lemma-is-nuevo
node
Leaf 3: Before

Leaf 1:

After: 1000 **Before**: 8000


Formulate the linguistic question into a classification task

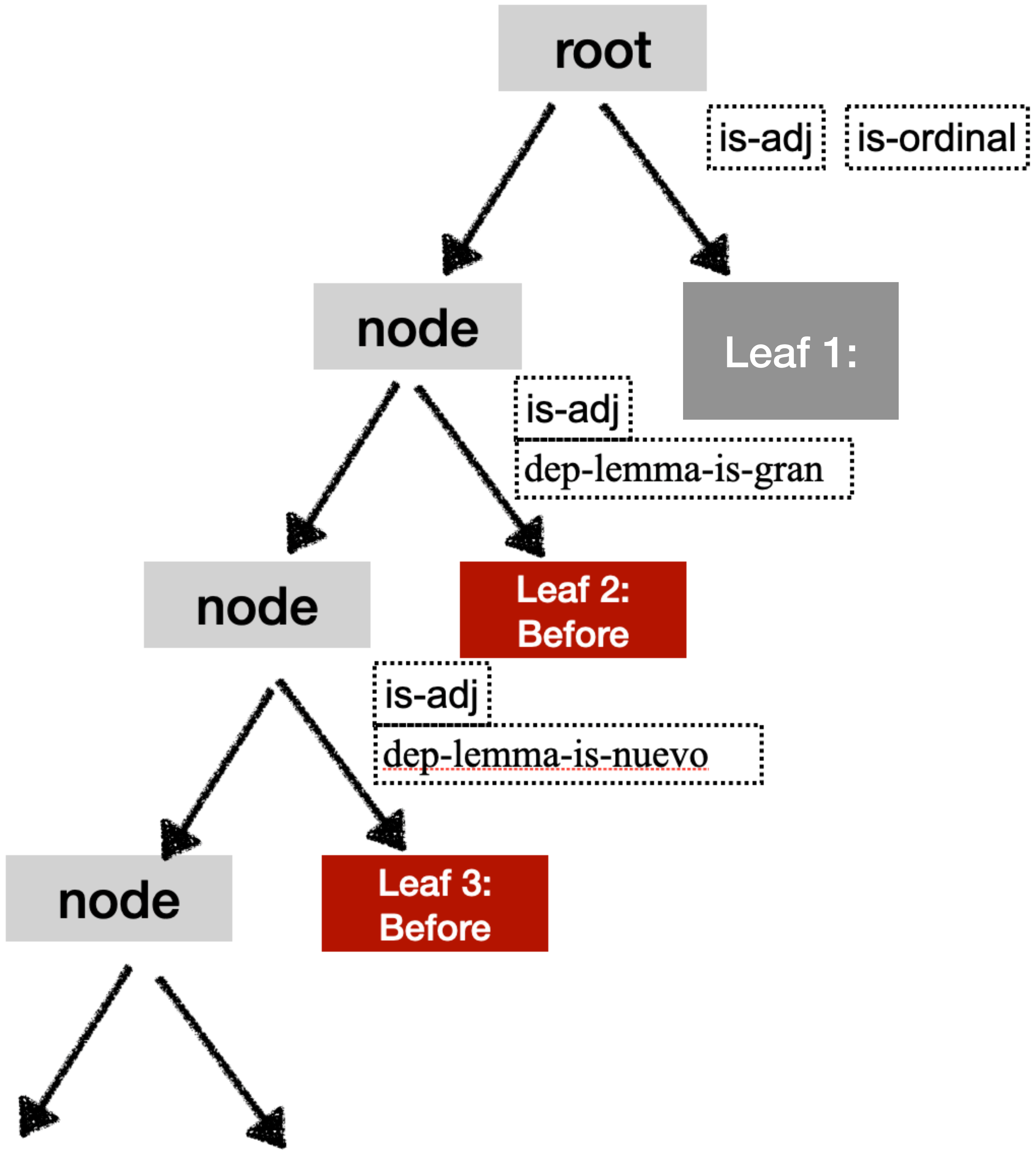
Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

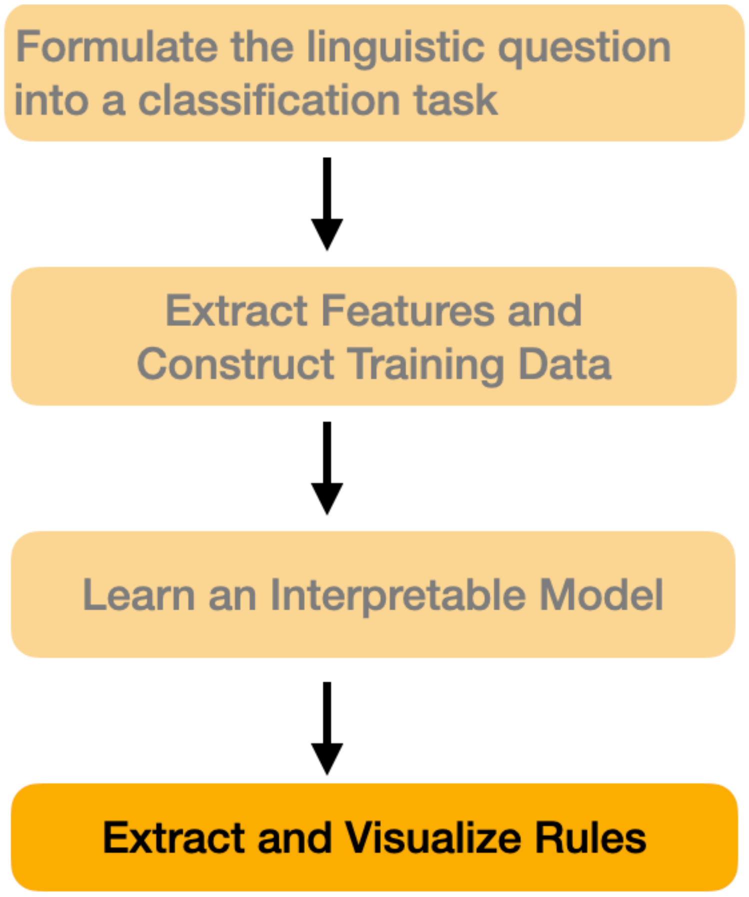
AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Extract rules



Leaf 1:

After: 1000 **Before**: 1800



Formulate the linguistic question into a classification task

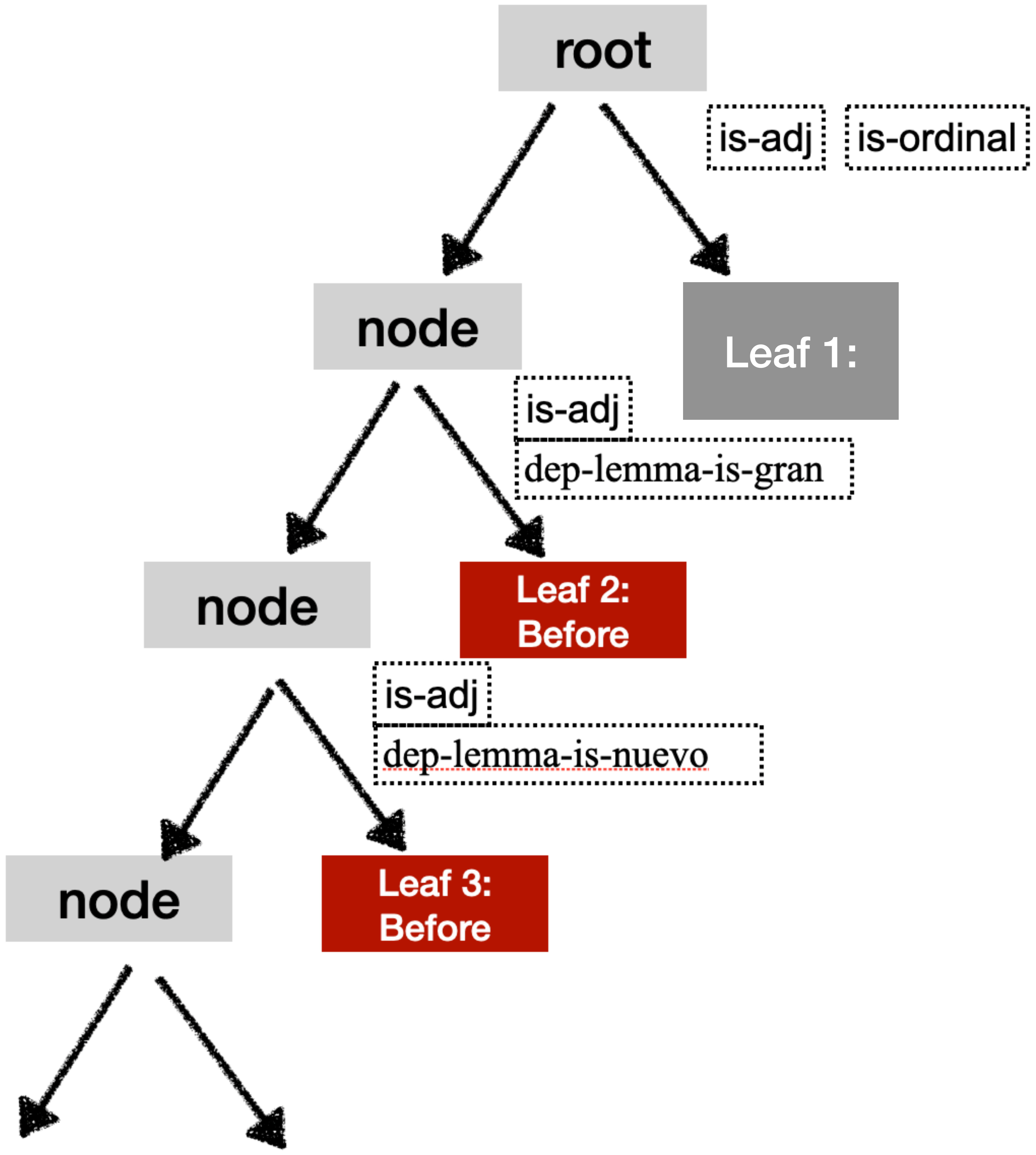
Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Extract rules



Leaf 1:

After: 1000 **Before**: 1800

↓ Statistical threshold

Significance test χ^2 + **Effect Size**

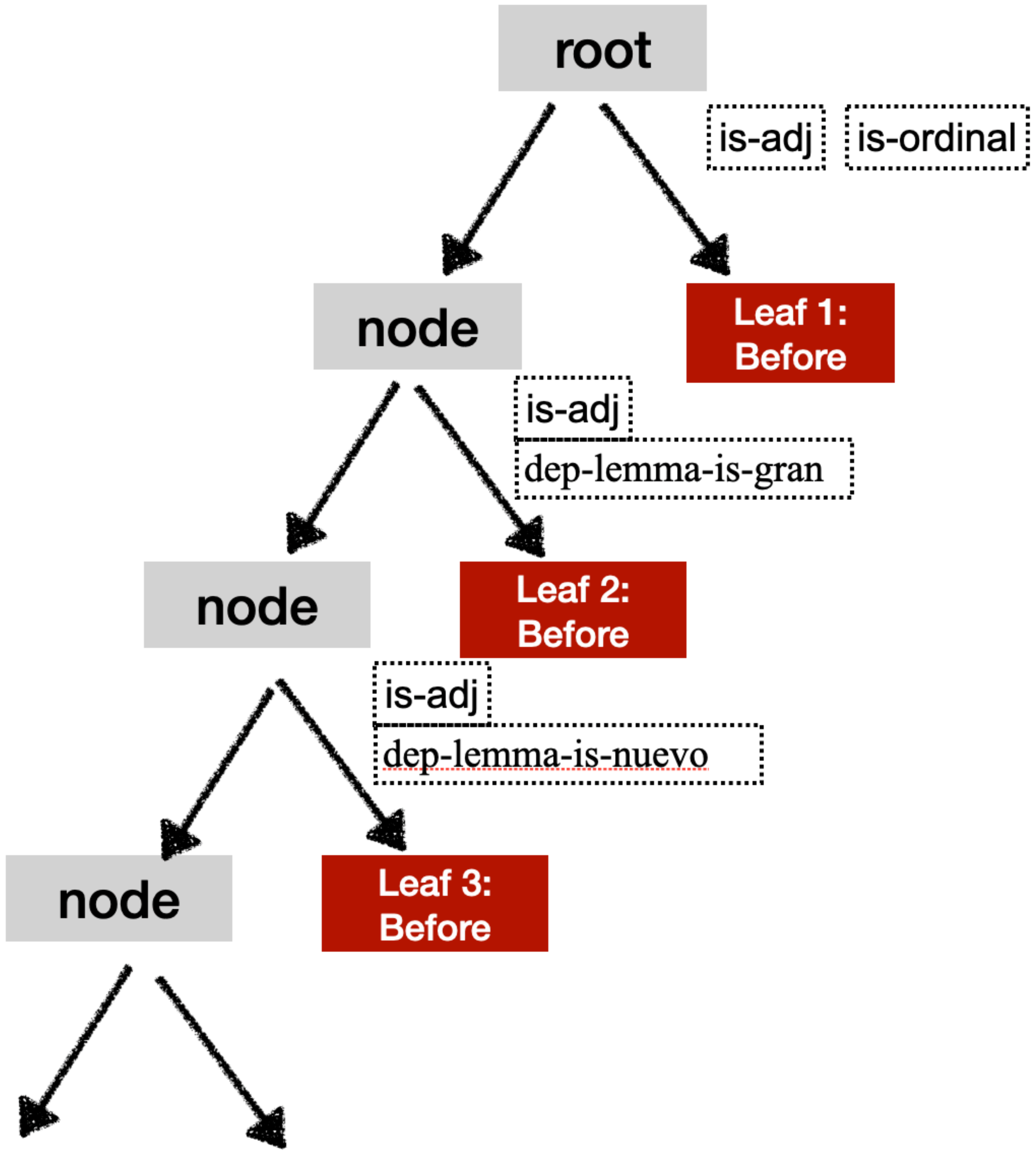
Observed distribution is significant

Magnitude of significance is large

Formulate the linguistic question into a classification task
↓
Extract Features and Construct Training Data
↓
Learn an Interpretable Model
↓
Extract and Visualize Rules

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Extract rules



Leaf 1:

After: 1000 **Before**: 1800

Label: **Before**

↓ Statistical threshold

Significance test χ^2 + **Effect Size**

Observed distribution is significant

Magnitude of significance is large

Formulate the linguistic question into a classification task

↓

Extract Features and Construct Training Data

↓

Learn an Interpretable Model

↓

Extract and Visualize Rules

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary**, Sheikh, Mortensen, Anastasopoulos, Neubig.* ***In Submission

Word Order: Visualize rules

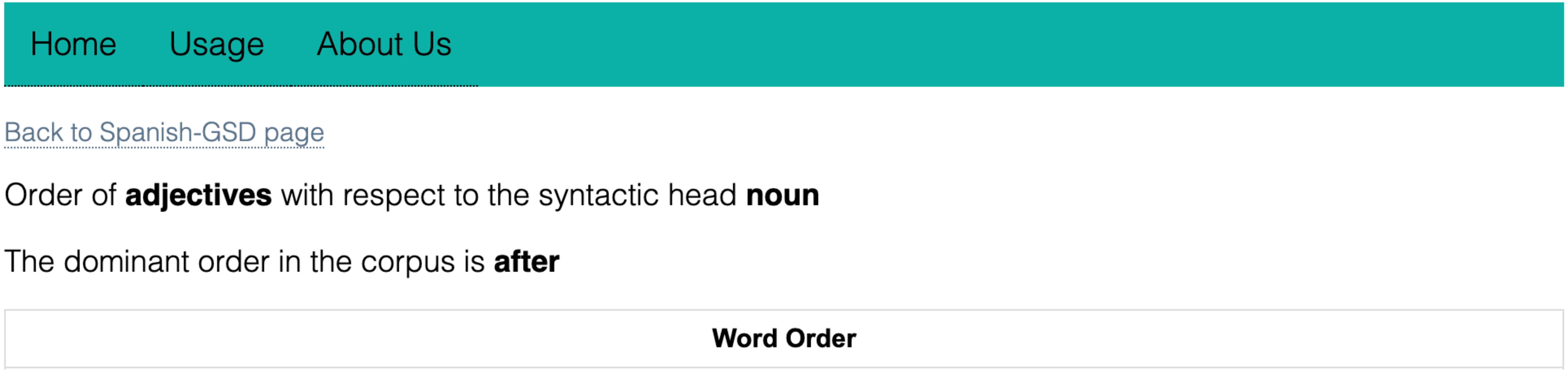
Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Visualize rules

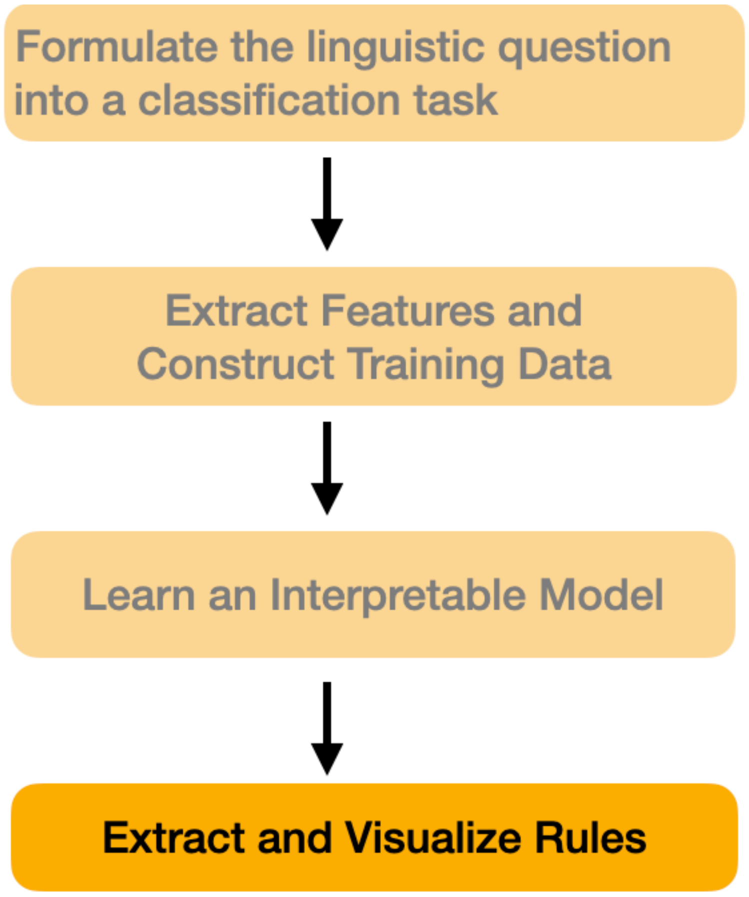


Back to Spanish-GSD page

Order of **adjectives** with respect to the syntactic head **noun**

The dominant order in the corpus is **after**

Word Order



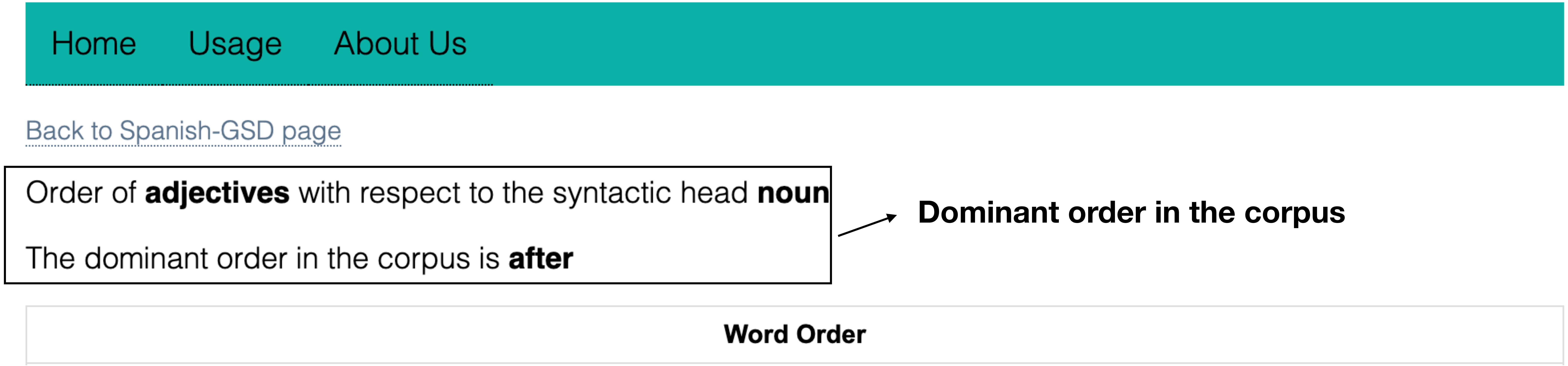
Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Visualize rules



Home Usage About Us

Back to Spanish-GSD page

Order of **adjectives** with respect to the syntactic head **noun**

The dominant order in the corpus is **after**

Word Order

Dominant order in the corpus

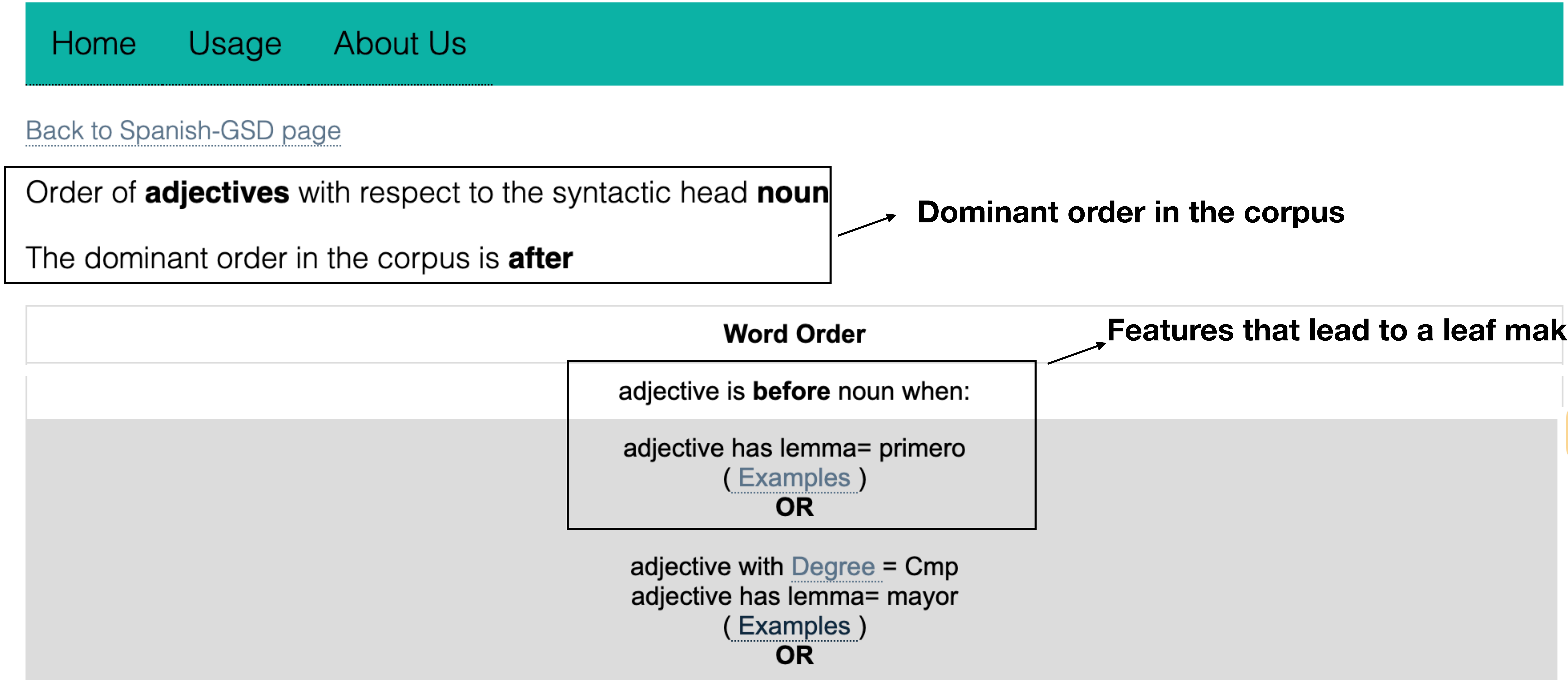
Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Word Order: Visualize rules



Back to Spanish-GSD page

Order of **adjectives** with respect to the syntactic head **noun**

The dominant order in the corpus is **after**

Word Order

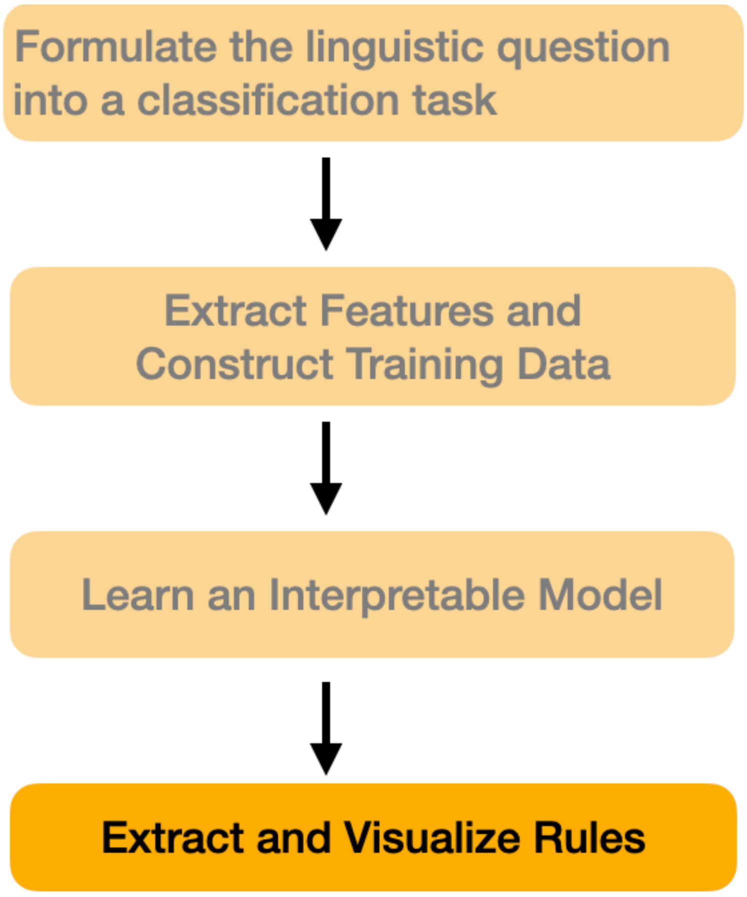
adjective is **before** noun when:

adjective has lemma= primero
(Examples)
OR

adjective with Degree = Cmp
adjective has lemma= mayor
(Examples)
OR

Dominant order in the corpus

Features that lead to a leaf makes a rule

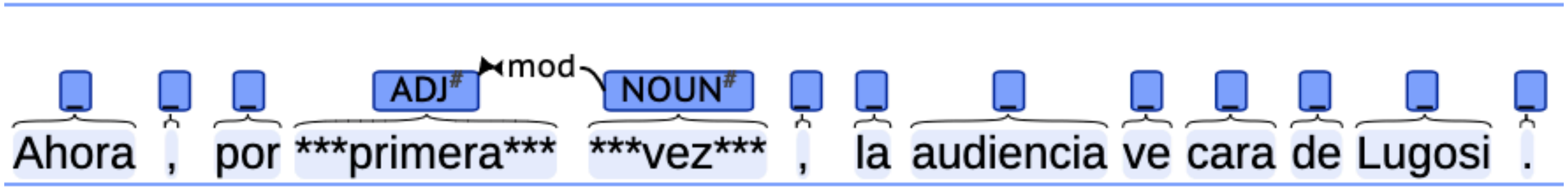


Word Order: Visualize rules

Rule: Adjective like “Primera” come **before** noun

Word Order: Visualize rules

Rule: Adjective like “Primera” come **before** noun



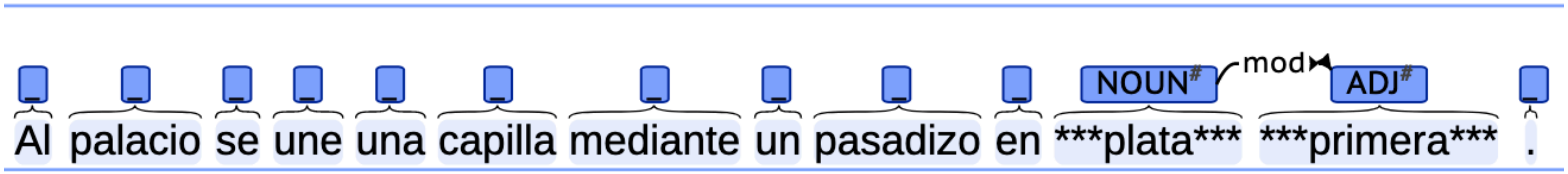
Ahora , por ***primera*** ***vez*** , la audiencia ve cara de Lugosi .

Word Order: Visualize rules

Rule: Adjective like “Primera” come **before** noun

Ahora , por ***primera*** ***vez*** , la audiencia ve cara de Lugosi .

Exceptions!



Al palacio se une una capilla mediante un pasadizo en ***plata*** ***primera*** .

Agreement

Agreement

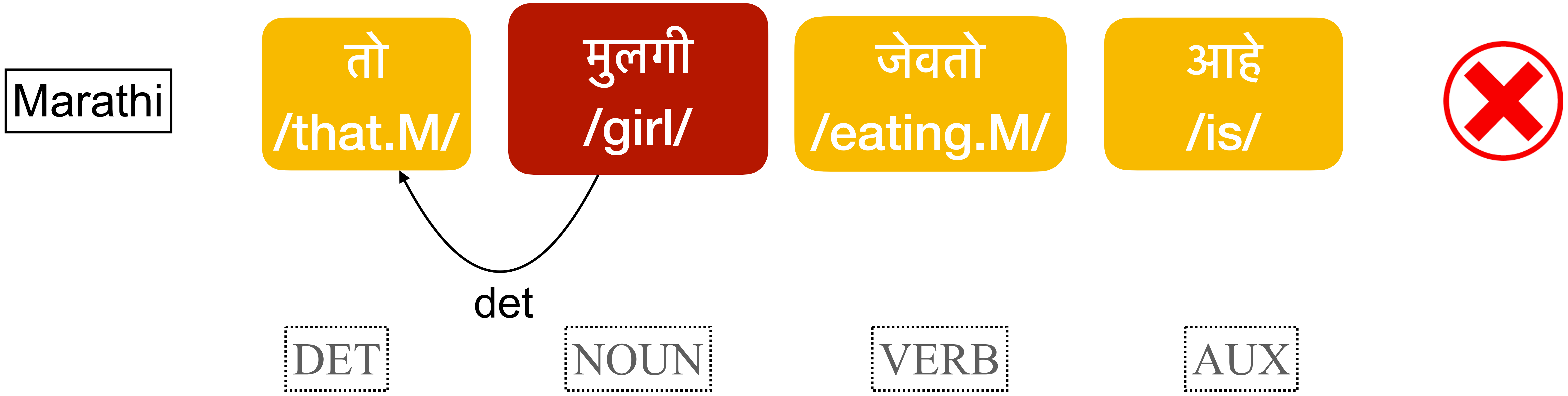
Agreement

Marathi	तो /that.M/	मुलगा /boy/	जेवतो /eating.M/	आहे /is/	✓

Agreement

Marathi	तो	मुलगी	जेवतो	आहे	✗
	/that.M/	/girl/	/eating.M/	/is/	

Agreement



Marathi

तो /that.M/

मुलगी /girl/

जेवतो /eating.M/

आहे /is/

✗

det

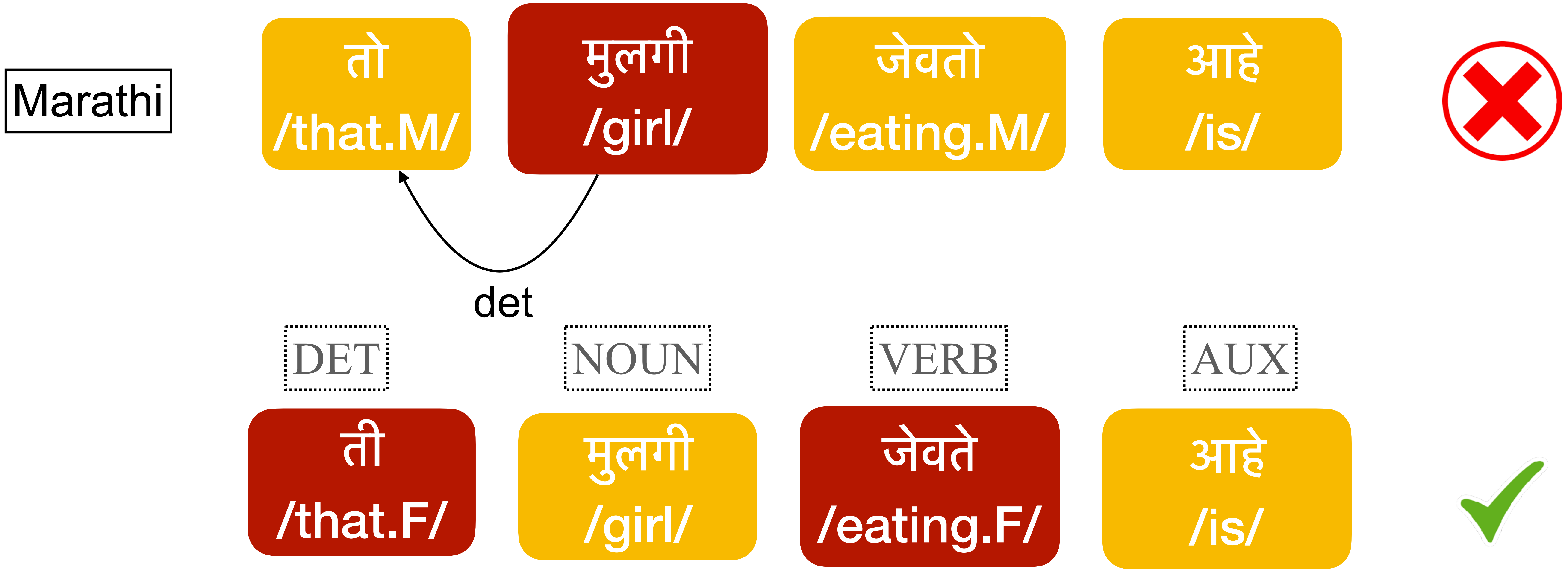
DET

NOUN

VERB

AUX

Agreement



Marathi

तो /that.M/ मुलगी /girl/ जेवतो /eating.M/ आहे /is/ ✗

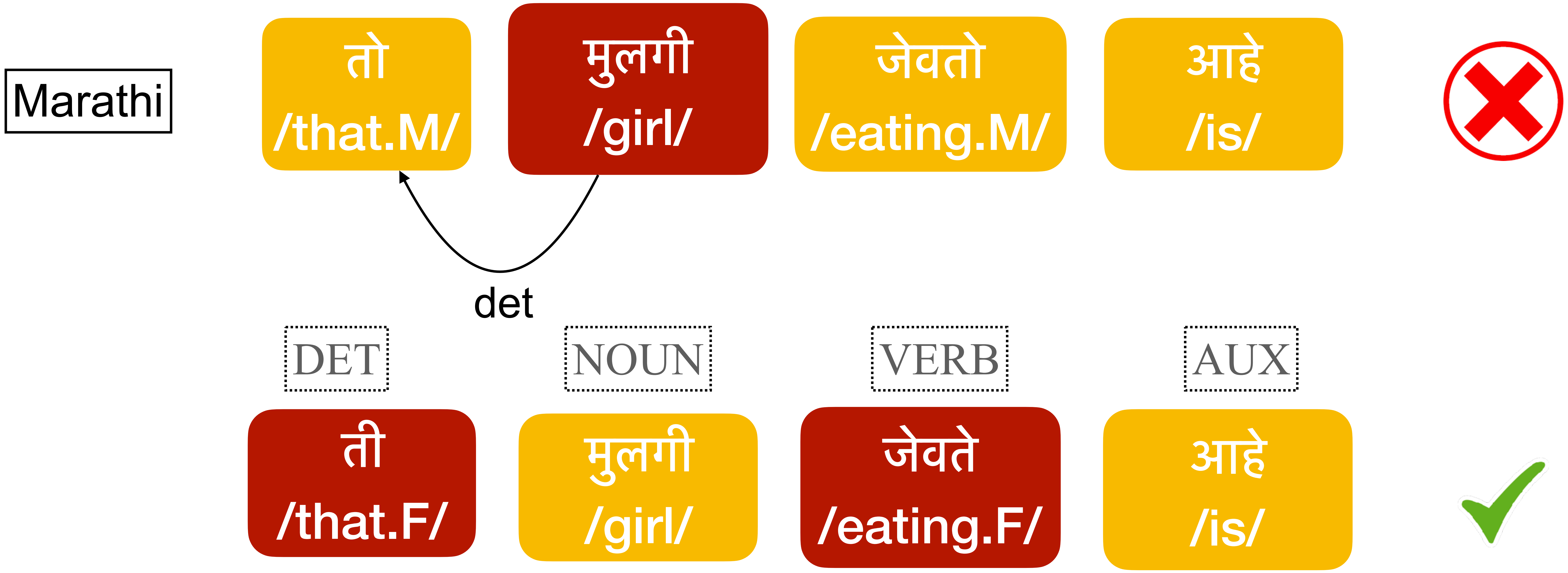
det

DET NOUN VERB AUX

ती /that.F/ मुलगी /girl/ जेवते /eating.F/ आहे /is/ ✓

Agreement

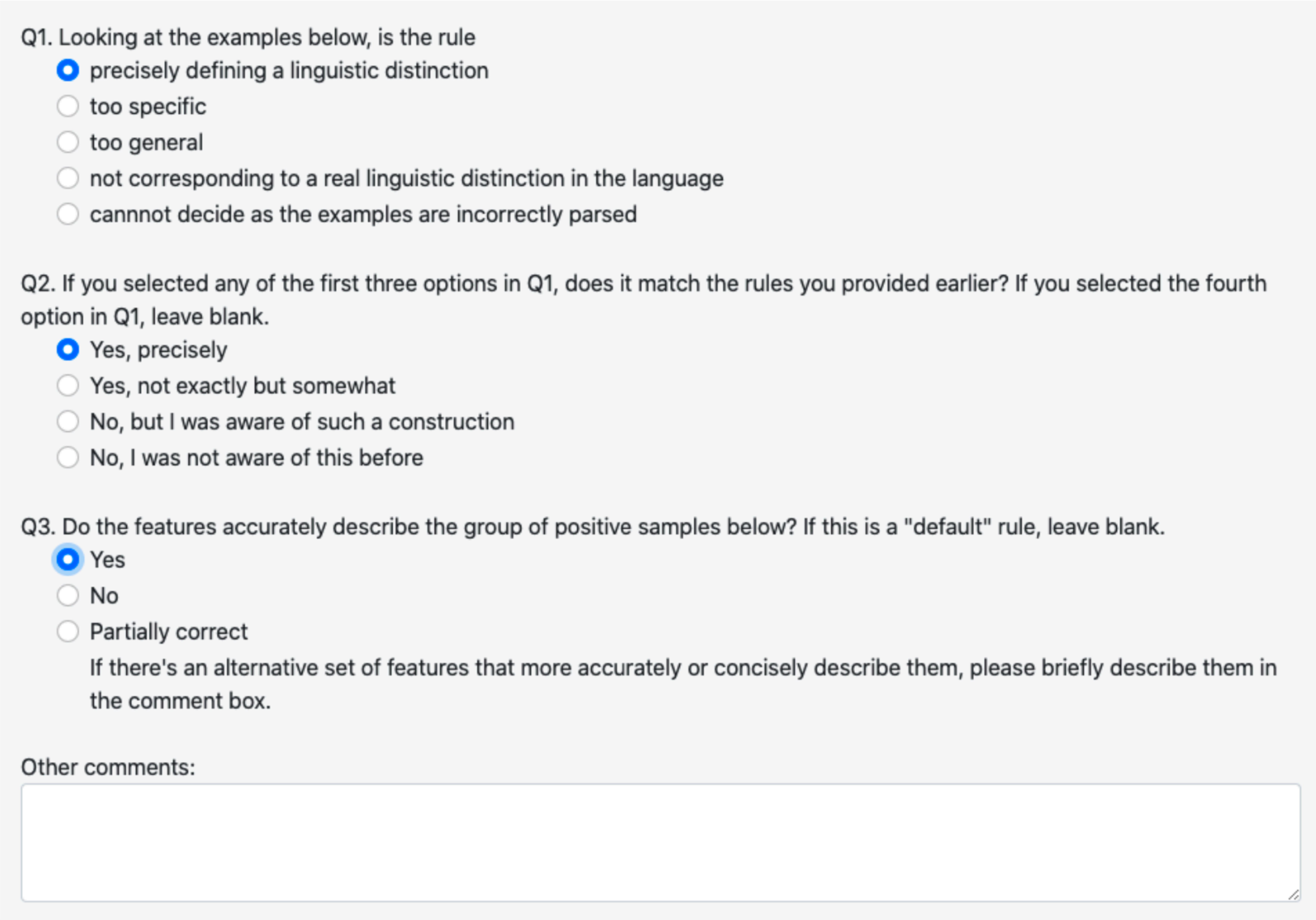
When should a head-dependent agree on gender and when it shouldn't?



Marathi

तो /that.M/	मुलगी /girl/	जेवतो /eating.M/	आहे /is/	✗
DET	NOUN	VERB	AUX	
ती /that.F/	मुलगी /girl/	जेवते /eating.F/	आहे /is/	✓

Quality Evaluation



Q1. Looking at the examples below, is the rule

- ● precisely defining a linguistic distinction
- ○ too specific
- ○ too general
- ○ not corresponding to a real linguistic distinction in the language
- ○ cannnot decide as the examples are incorrectly parsed

Q2. If you selected any of the first three options in Q1, does it match the rules you provided earlier? If you selected the fourth option in Q1, leave blank.

- ● Yes, precisely
- ○ Yes, not exactly but somewhat
- ○ No, but I was aware of such a construction
- ○ No, I was not aware of this before

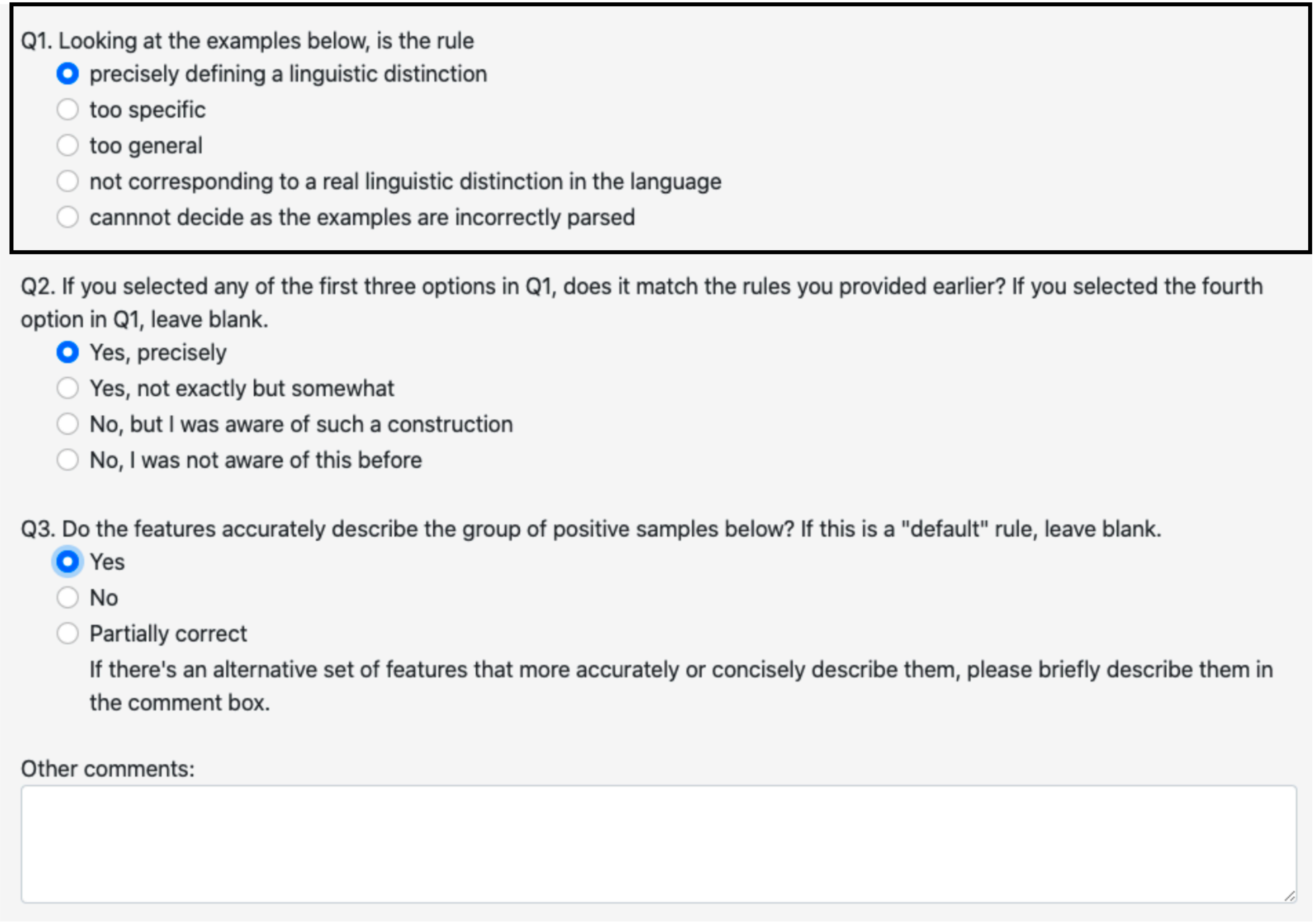
Q3. Do the features accurately describe the group of positive samples below? If this is a "default" rule, leave blank.

- ● Yes
- ○ No
- ○ Partially correct

If there's an alternative set of features that more accurately or concisely describe them, please briefly describe them in the comment box.

Other comments:

Quality Evaluation



Q1. Looking at the examples below, is the rule

- ● precisely defining a linguistic distinction
- ○ too specific
- ○ too general
- ○ not corresponding to a real linguistic distinction in the language
- ○ cannnot decide as the examples are incorrectly parsed

Q2. If you selected any of the first three options in Q1, does it match the rules you provided earlier? If you selected the fourth option in Q1, leave blank.

- ● Yes, precisely
- ○ Yes, not exactly but somewhat
- ○ No, but I was aware of such a construction
- ○ No, I was not aware of this before

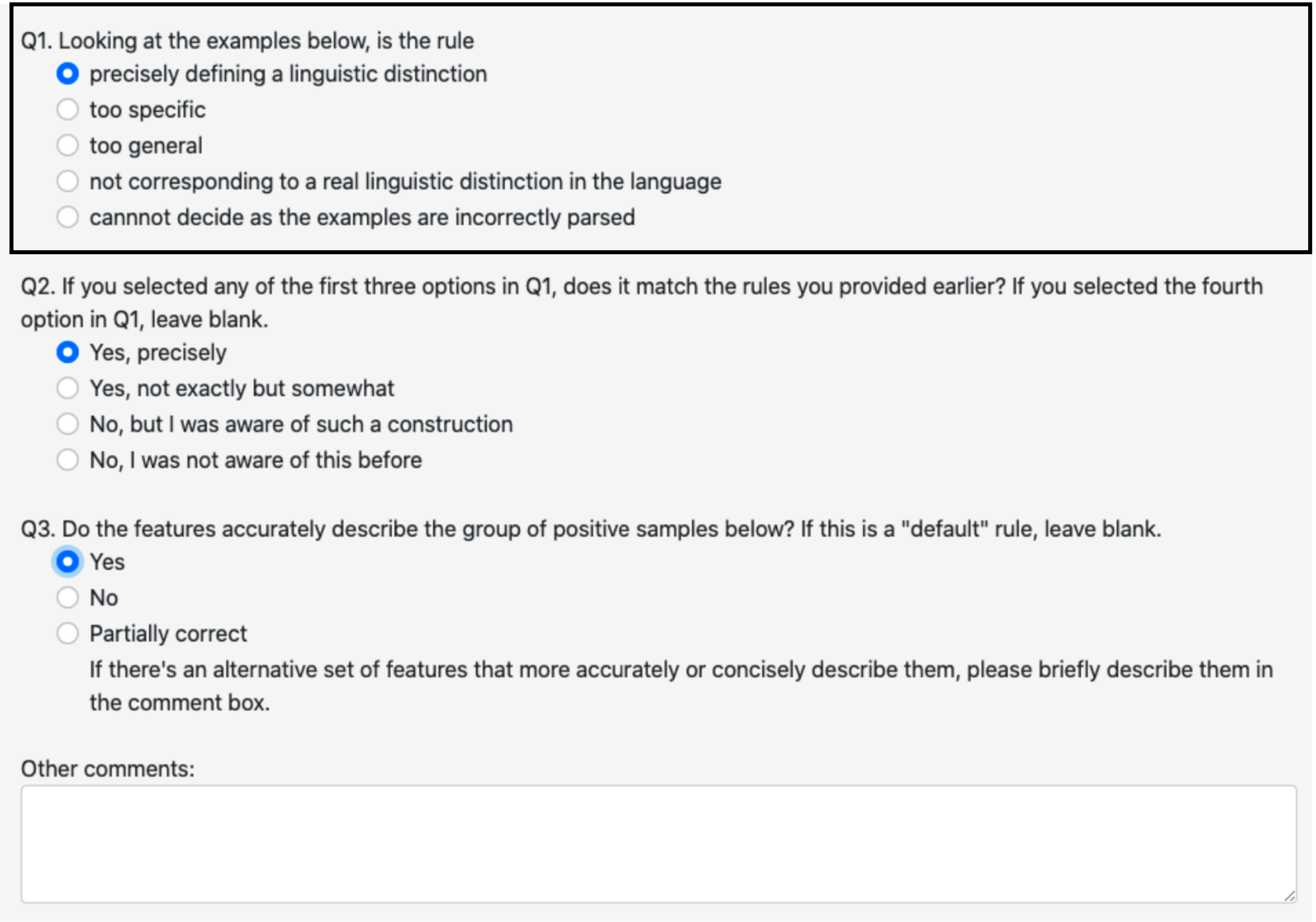
Q3. Do the features accurately describe the group of positive samples below? If this is a "default" rule, leave blank.

- ● Yes
- ○ No
- ○ Partially correct

If there's an alternative set of features that more accurately or concisely describe them, please briefly describe them in the comment box.

Other comments:

Quality Evaluation



Q1. Looking at the examples below, is the rule

- ● precisely defining a linguistic distinction
- ○ too specific
- ○ too general
- ○ not corresponding to a real linguistic distinction in the language
- ○ cannnot decide as the examples are incorrectly parsed

Q2. If you selected any of the first three options in Q1, does it match the rules you provided earlier? If you selected the fourth option in Q1, leave blank.

- ● Yes, precisely
- ○ Yes, not exactly but somewhat
- ○ No, but I was aware of such a construction
- ○ No, I was not aware of this before

Q3. Do the features accurately describe the group of positive samples below? If this is a "default" rule, leave blank.

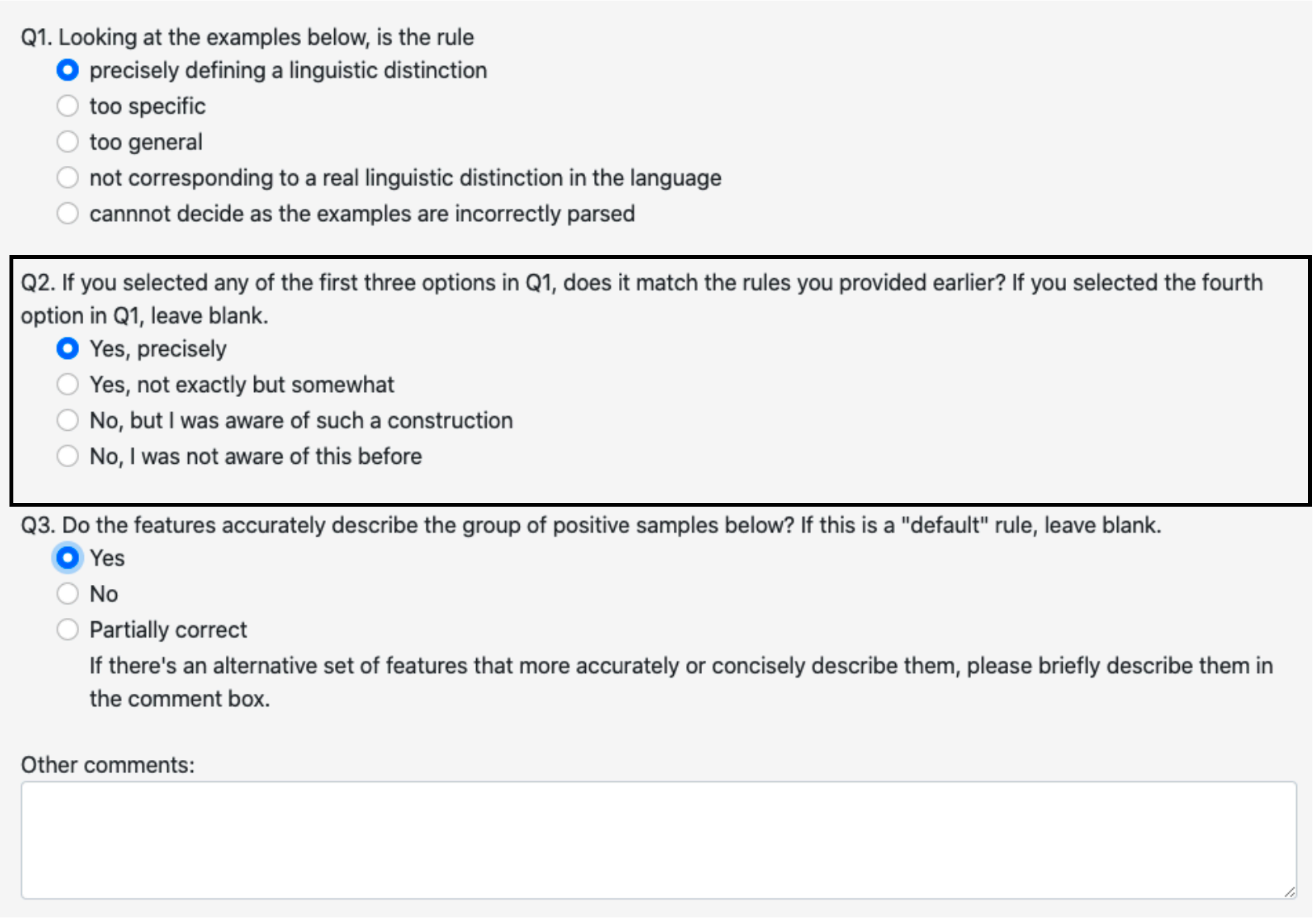
- ● Yes
- ○ No
- ○ Partially correct

If there's an alternative set of features that more accurately or concisely describe them, please briefly describe them in the comment box.

Other comments:

Correctness

Quality Evaluation



Q1. Looking at the examples below, is the rule

- ● precisely defining a linguistic distinction
- ○ too specific
- ○ too general
- ○ not corresponding to a real linguistic distinction in the language
- ○ cannnot decide as the examples are incorrectly parsed

Q2. If you selected any of the first three options in Q1, does it match the rules you provided earlier? If you selected the fourth option in Q1, leave blank.

- ● Yes, precisely
- ○ Yes, not exactly but somewhat
- ○ No, but I was aware of such a construction
- ○ No, I was not aware of this before

Q3. Do the features accurately describe the group of positive samples below? If this is a "default" rule, leave blank.

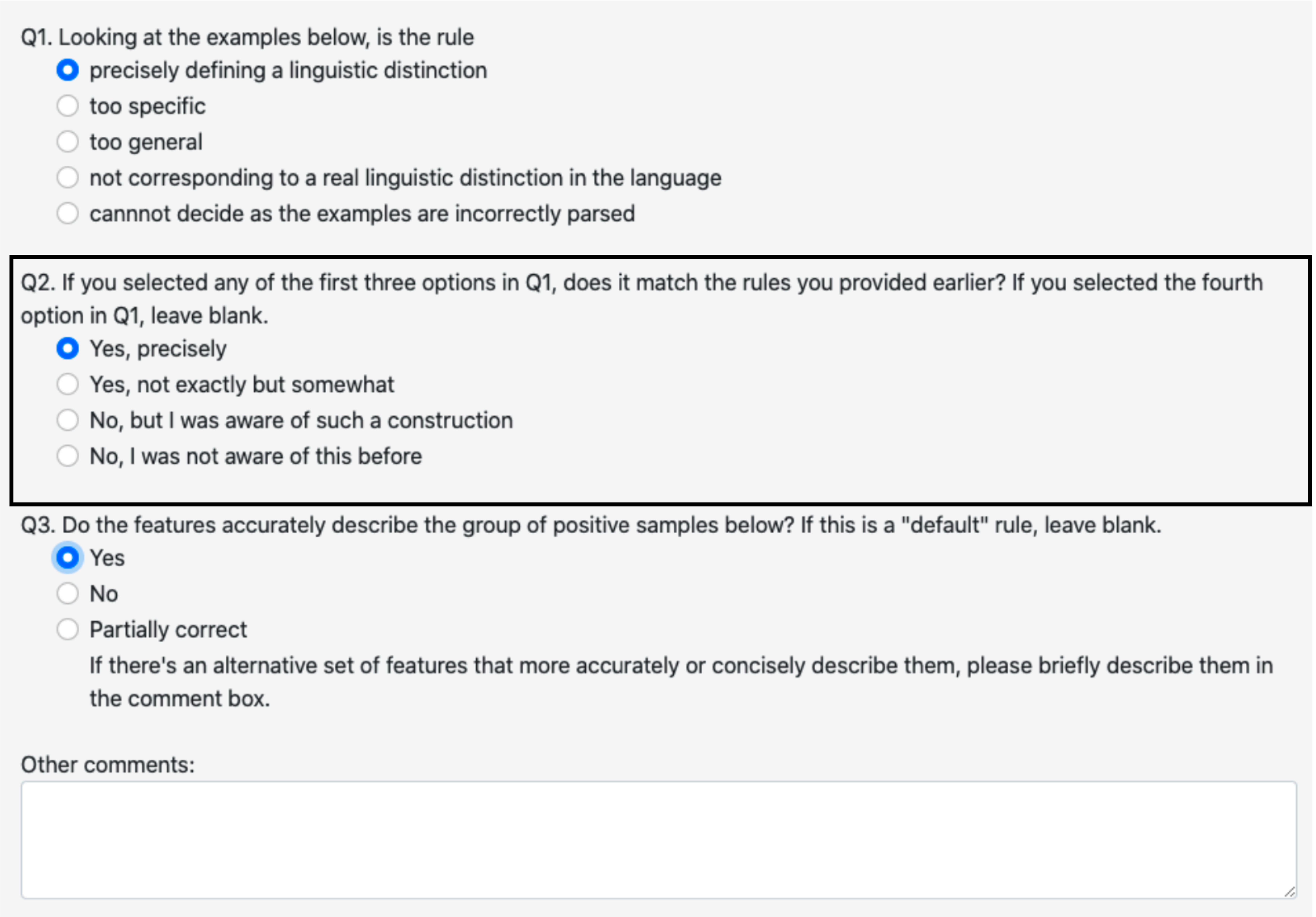
- ● Yes
- ○ No
- ○ Partially correct

If there's an alternative set of features that more accurately or concisely describe them, please briefly describe them in the comment box.

Other comments:

Correctness

Quality Evaluation



Q1. Looking at the examples below, is the rule

- ● precisely defining a linguistic distinction
- ○ too specific
- ○ too general
- ○ not corresponding to a real linguistic distinction in the language
- ○ cannnot decide as the examples are incorrectly parsed

Q2. If you selected any of the first three options in Q1, does it match the rules you provided earlier? If you selected the fourth option in Q1, leave blank.

- ● Yes, precisely
- ○ Yes, not exactly but somewhat
- ○ No, but I was aware of such a construction
- ○ No, I was not aware of this before

Q3. Do the features accurately describe the group of positive samples below? If this is a "default" rule, leave blank.

- ● Yes
- ○ No
- ○ Partially correct

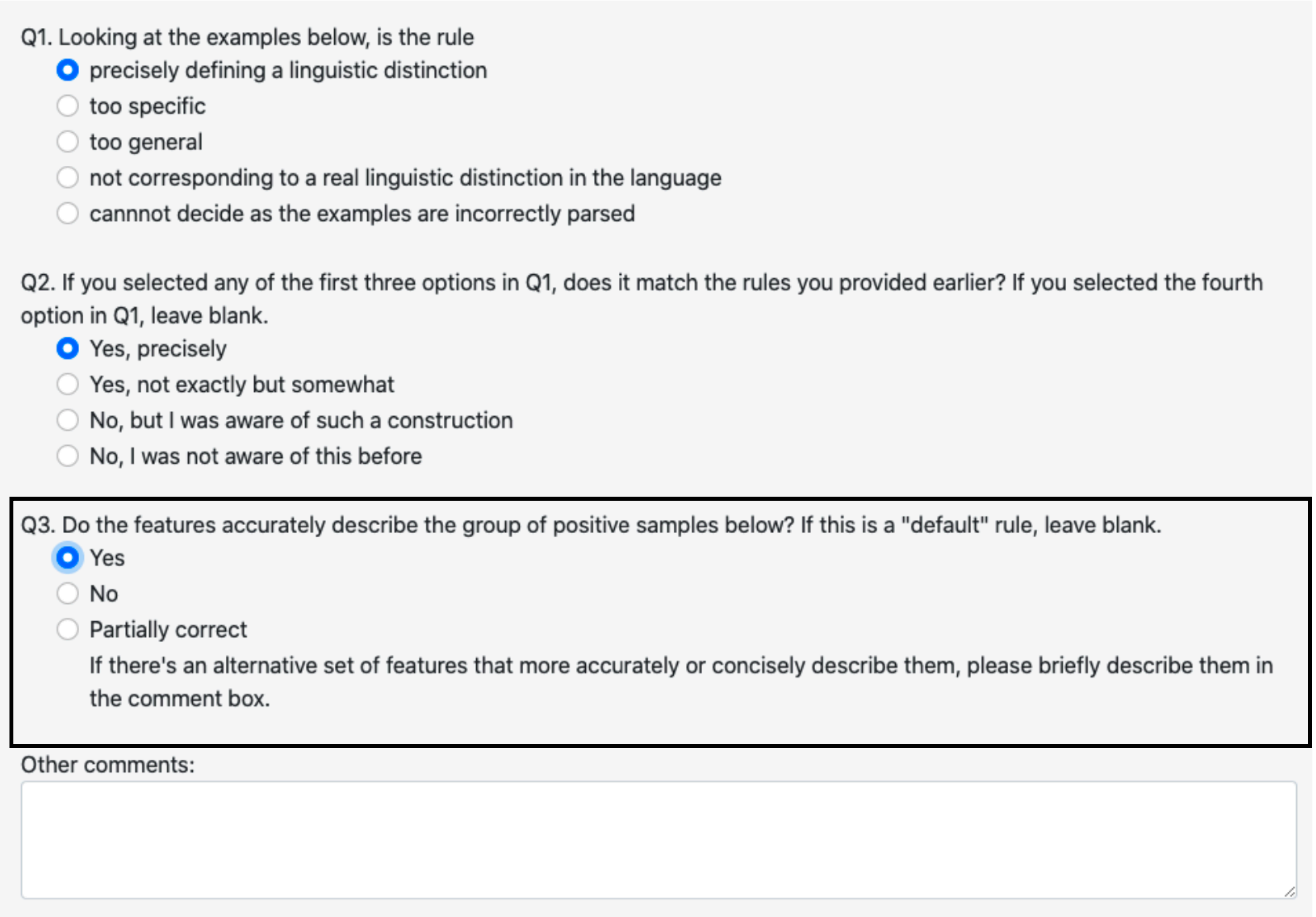
If there's an alternative set of features that more accurately or concisely describe them, please briefly describe them in the comment box.

Other comments:

Correctness

Prior knowledge

Quality Evaluation



Q1. Looking at the examples below, is the rule

- ● precisely defining a linguistic distinction
- ○ too specific
- ○ too general
- ○ not corresponding to a real linguistic distinction in the language
- ○ cannnot decide as the examples are incorrectly parsed

Q2. If you selected any of the first three options in Q1, does it match the rules you provided earlier? If you selected the fourth option in Q1, leave blank.

- ● Yes, precisely
- ○ Yes, not exactly but somewhat
- ○ No, but I was aware of such a construction
- ○ No, I was not aware of this before

Q3. Do the features accurately describe the group of positive samples below? If this is a "default" rule, leave blank.

- ● Yes
- ○ No
- ○ Partially correct

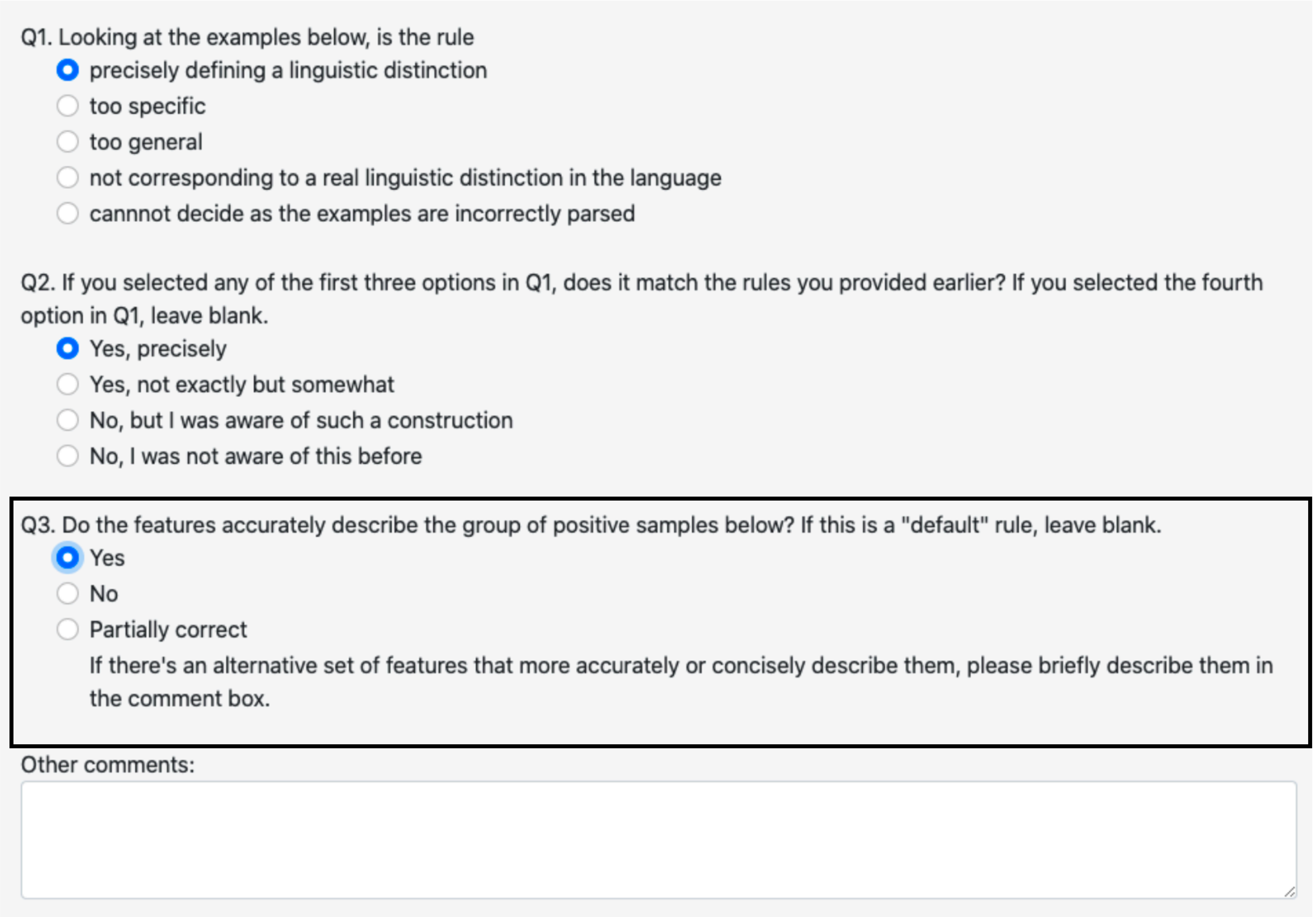
If there's an alternative set of features that more accurately or concisely describe them, please briefly describe them in the comment box.

Other comments:

Correctness

Prior knowledge

Quality Evaluation



Q1. Looking at the examples below, is the rule

- ● precisely defining a linguistic distinction
- ○ too specific
- ○ too general
- ○ not corresponding to a real linguistic distinction in the language
- ○ cannnot decide as the examples are incorrectly parsed

Q2. If you selected any of the first three options in Q1, does it match the rules you provided earlier? If you selected the fourth option in Q1, leave blank.

- ● Yes, precisely
- ○ Yes, not exactly but somewhat
- ○ No, but I was aware of such a construction
- ○ No, I was not aware of this before

Q3. Do the features accurately describe the group of positive samples below? If this is a "default" rule, leave blank.

- ● Yes
- ○ No
- ○ Partially correct

  If there's an alternative set of features that more accurately or concisely describe them, please briefly describe them in the comment box.

Other comments:

Correctness

Prior knowledge

Readability

Results: Quality Evaluation for Greek

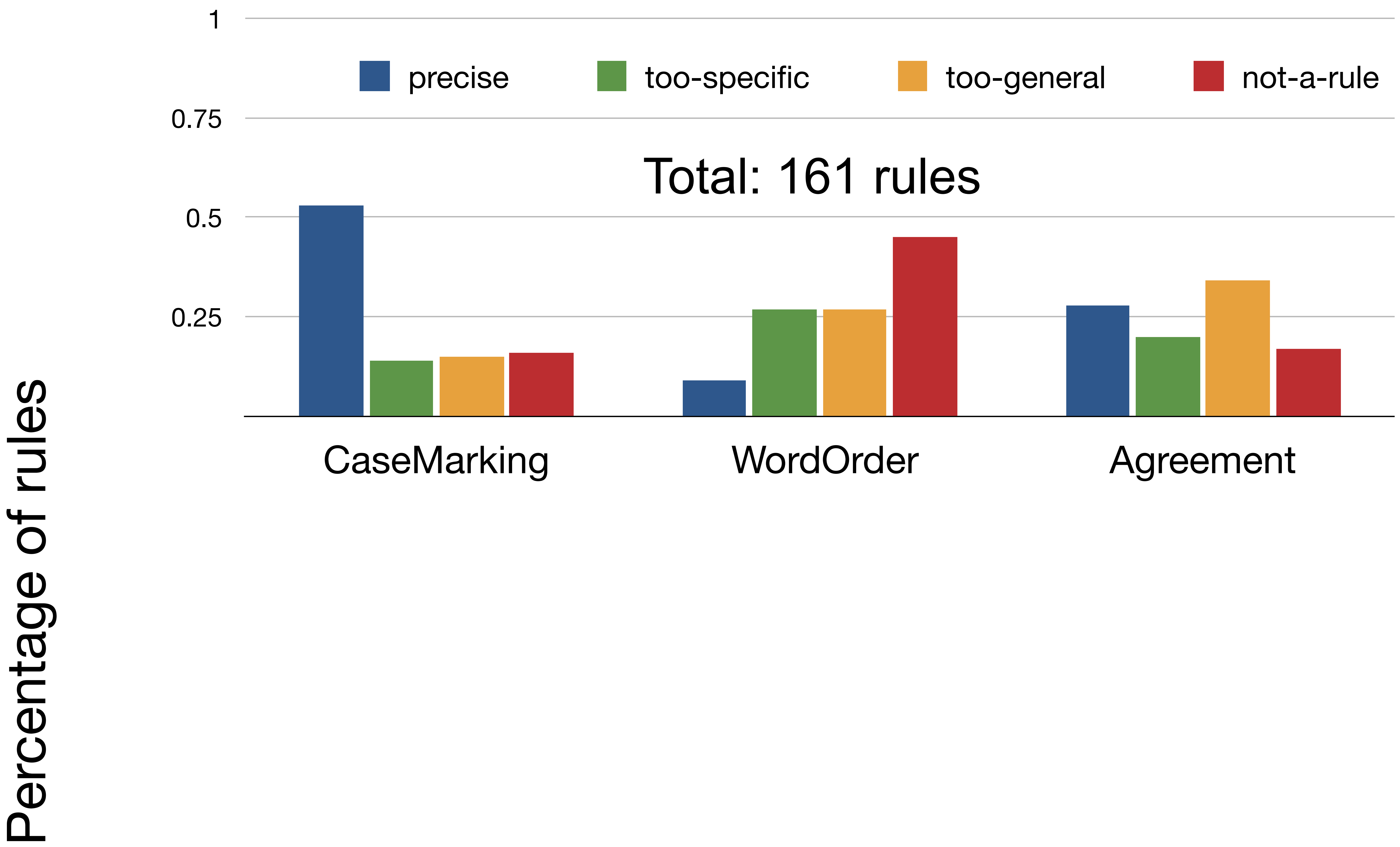
Percentage of rules

Results: Quality Evaluation for Greek

Does model discover grammatically valid rules?

Percentage of rules

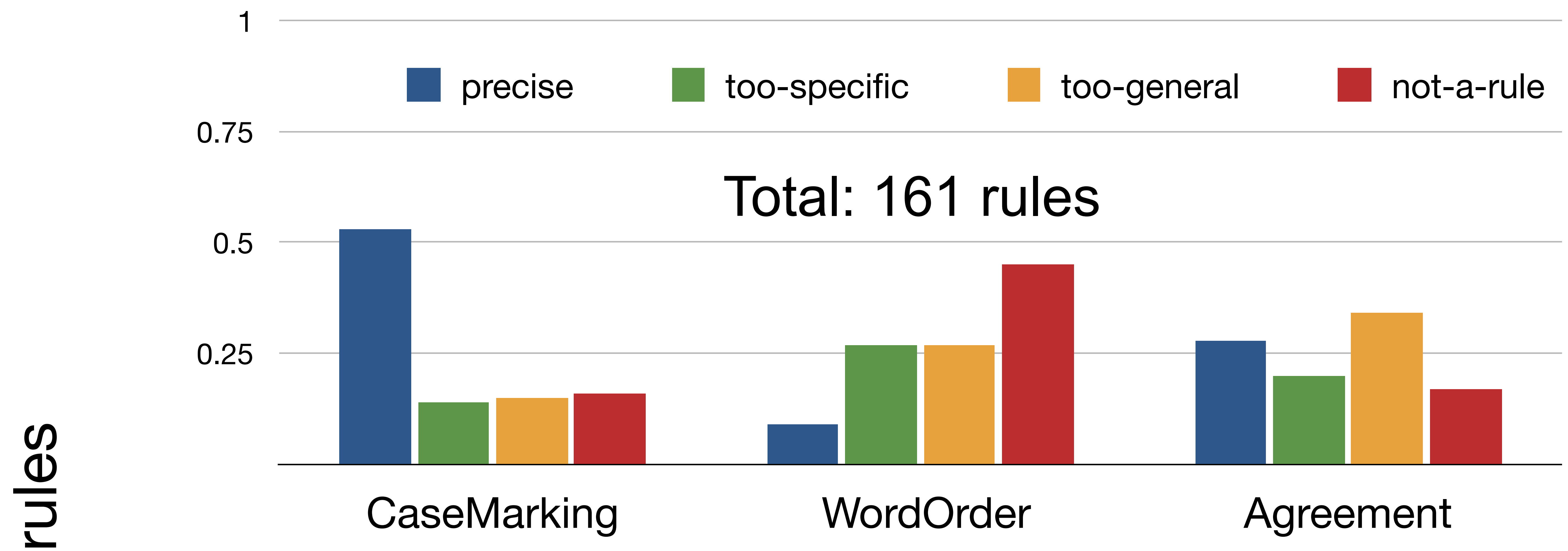
Results: Quality Evaluation for Greek



Does model discover grammatically valid rules?

80% rules are valid,
40% valid rules too specific/general

Results: Quality Evaluation for Greek

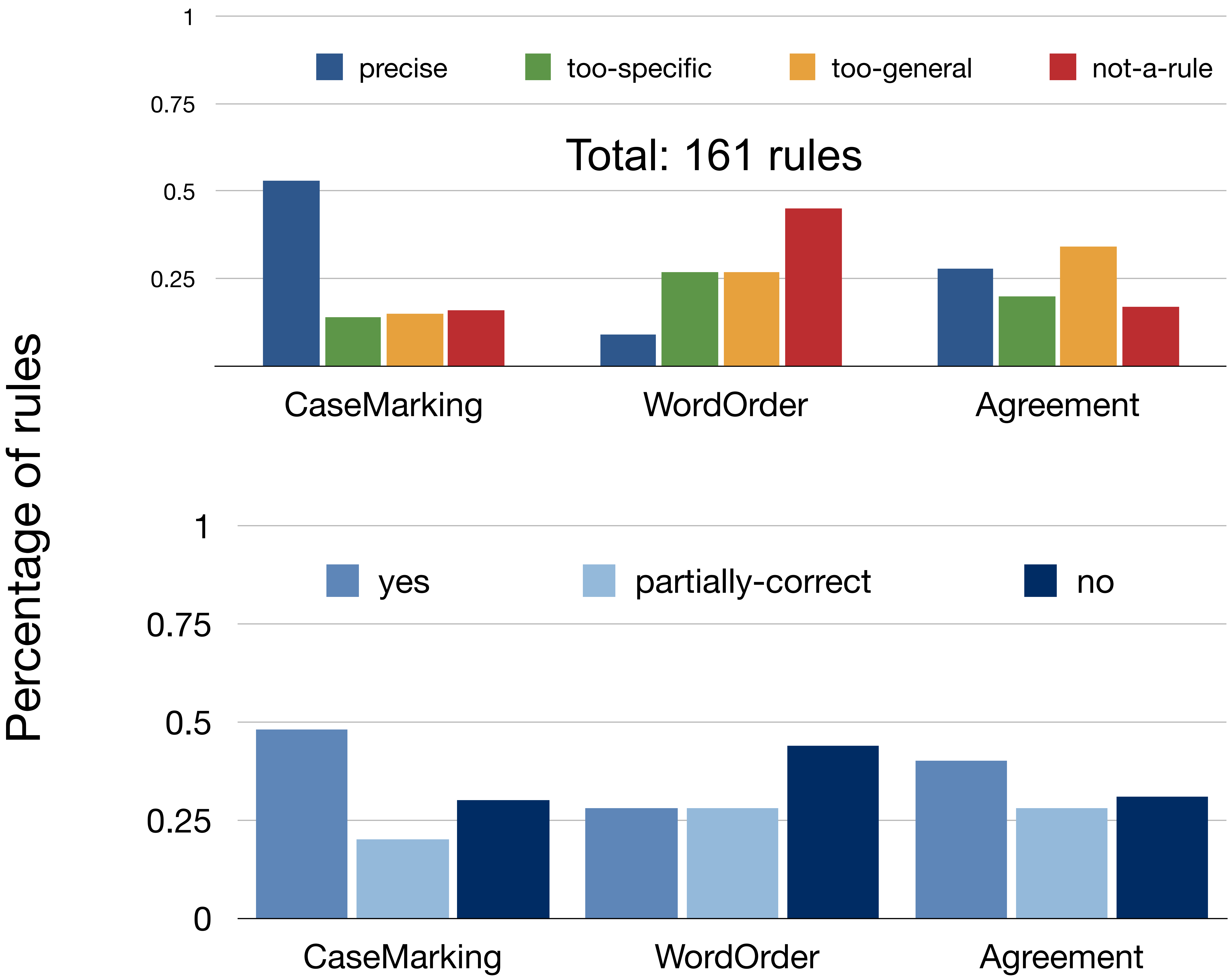


Does model discover grammatically valid rules?

80% rules are valid,
40% valid rules too specific/general

Are the rules human-readable?

Results: Quality Evaluation for Greek



Does model discover grammatically valid rules?

80% rules are valid,
40% valid rules too specific/general

Are the rules human-readable?

69% rules are informative and readable

Results: Quality Evaluation

Does the model discover *new rules?

Results: Quality Evaluation

Does the model discover *new rules?

adjective is **after** its head **noun**

Features that make up this rule	
Active Features	**Inactive Features**
adjective's head is a= PRON	-

Even for well studied languages, system discovers *new rules

Examples that agree with label: **after**: The **adjective** is denoted by ***

1 please let me know if you need anything ***else***

2 Please let us know if you need anything ***else*** .

3 No pancreatitis or anything ***abnormal*** .

4 Then take the leash off and do nothing ***else*** but that .

An aside: Parser quality matters a lot!

---

[1] 75 Languages, 1 Model: Parsing Universal Dependencies Universally, Kondratyuk and Straka, 2019

An aside: Parser quality matters a lot!

- Applied AutoLEX on **Hmong Daw (mmw)** which has **NO syntactic parser** available

---

[1] 75 Languages, 1 Model: Parsing Universal Dependencies Universally, Kondratyuk and Straka, 2019

An aside: Parser quality matters a lot!

- Applied AutoLEX on **Hmong Daw (mmw)** which has **NO syntactic parser** available

— **Zero-shot transfer** using a multilingual UDIFY[1] parser

---

[1] 75 Languages, 1 Model: Parsing Universal Dependencies Universally, Kondratyuk and Straka, 2019

An aside: Parser quality matters a lot!

- Applied AutoLEX on **Hmong Daw (mmw)** which has **NO syntactic parser** available

  — **Zero-shot transfer** using a multilingual UDIFY[1] parser

  — Discovered 30 rules for Subj-V, Obj-V, Num-N, Adj-N, Adp-N word orders **but**, only 16% showed valid patterns

[1] 75 Languages, 1 Model: Parsing Universal Dependencies Universally, Kondratyuk and Straka, 2019

Carnegie Mellon University

An aside: Parser quality matters a lot!

- Applied AutoLEX on **Hmong Daw (mmw)** which has **NO syntactic parser** available

  — **Zero-shot transfer** using a multilingual UDIFY[1] parser

  — Discovered 30 rules for Subj-V, Obj-V, Num-N, Adj-N, Adp-N word orders **but**, only 16% showed valid patterns

- Applied AutoLEX on **Kannada (kan)** which has **NO syntactic parser** but has related language data with small in-person data

---

[1] 75 Languages, 1 Model: Parsing Universal Dependencies Universally, Kondratyuk and Straka, 2019

An aside: Parser quality matters a lot!

- Applied AutoLEX on **Hmong Daw (mmw)** which has **NO syntactic parser** available

  — **Zero-shot transfer** using a multilingual UDIFY[1] parser

  — Discovered 30 rules for Subj-V, Obj-V, Num-N, Adj-N, Adp-N word orders **but**, only 16% showed valid patterns

- Applied AutoLEX on **Kannada (kan)** which has **NO syntactic parser** but has related language data with small in-person data

  — **Few-shot transfer** using a multilingual UDIFY[1] parser **with all available data**

---

[1] 75 Languages, 1 Model: Parsing Universal Dependencies Universally, Kondratyuk and Straka, 2019

An aside: Parser quality matters a lot!

- Applied AutoLEX on **Hmong Daw (mmw)** which has **NO syntactic parser** available

  — **Zero-shot transfer** using a multilingual UDIFY[1] parser

  — Discovered 30 rules for Subj-V, Obj-V, Num-N, Adj-N, Adp-N word orders **but**, only 16% showed valid patterns

- Applied AutoLEX on **Kannada (kan)** which has **NO syntactic parser** but has related language data with small in-person data

  — **Few-shot transfer** using a multilingual UDIFY[1] parser **with all available data**

  — For Subj-V, Obj-V word orders **nearly 50%** showed valid patterns

[1] 75 Languages, 1 Model: Parsing Universal Dependencies Universally, Kondratyuk and Straka, 2019

An aside: Parser quality matters a lot!

- Applied AutoLEX on **Hmong Daw (mmw)** which has **NO syntactic parser** available

  — **Zero-shot transfer** using a multilingual UDIFY[1] parser

  — Discovered 30 rules for Subj-V, Obj-V, Num-N, Adj-N, Adp-N word orders **but**, only 16% showed valid patterns

- Applied AutoLEX on **Kannada (kan)** which has **NO syntactic parser** but has related language data with small in-person data

  — **Few-shot transfer** using a multilingual UDIFY[1] parser **with all available data**

  — For Subj-V, Obj-V word orders **nearly 50%** showed valid patterns

With improvements in syntactic parser, quality of rules also improves!

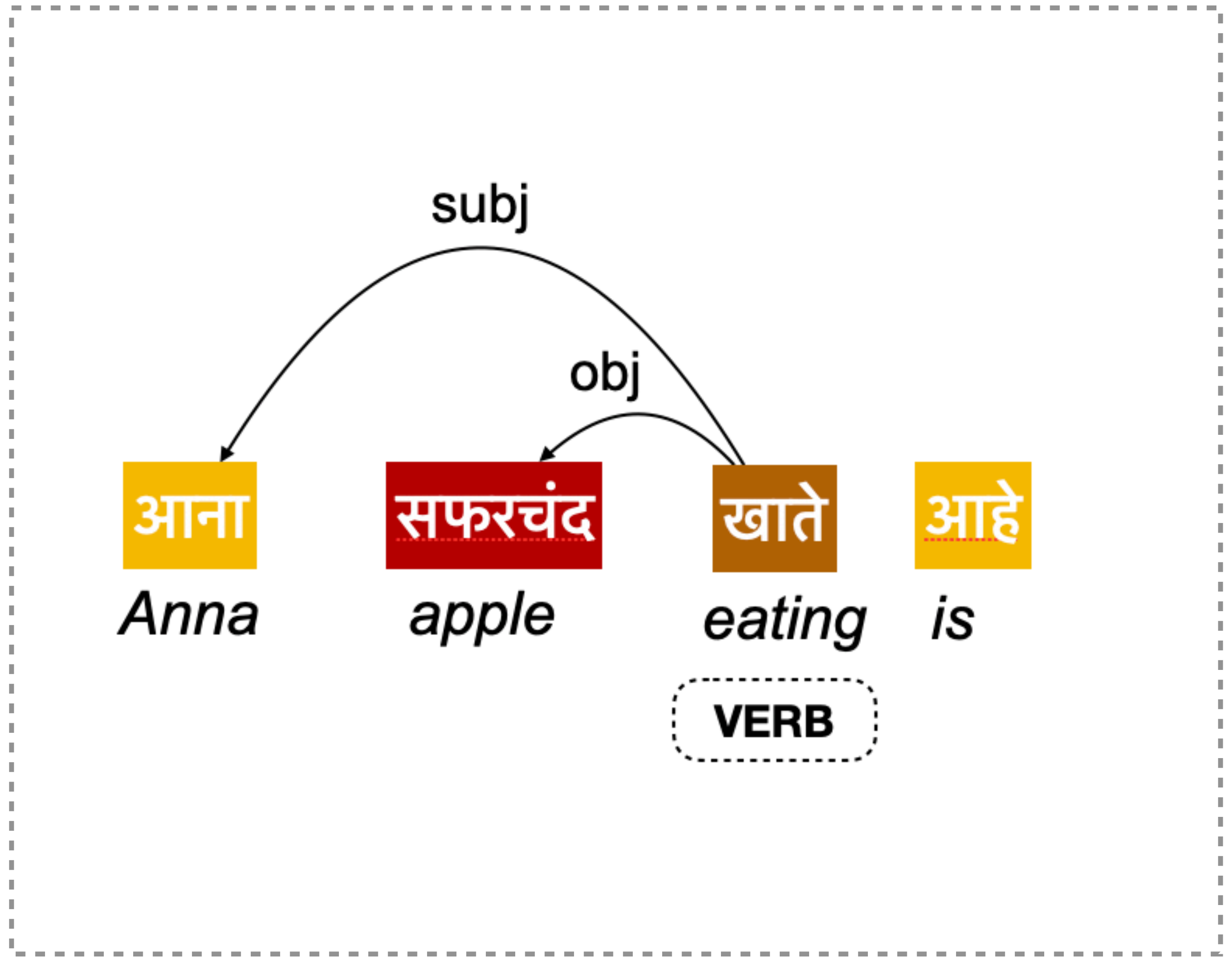
---

[1] 75 Languages, 1 Model: Parsing Universal Dependencies Universally, Kondratyuk and Straka, 2019

Revisiting our research goal

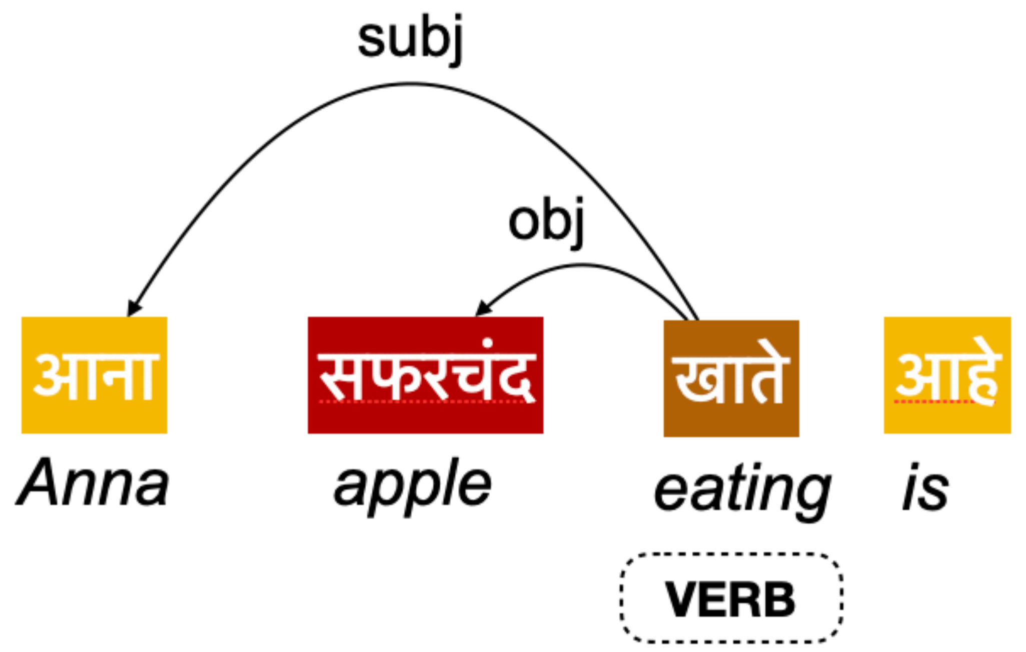
Revisiting our research goal

(Low-resource) Language Analysis



Revisiting our research goal

(Low-resource) Language Analysis



subj

obj

आना *Anna*

सफरचंद *apple*

खाते *eating*

आहे *is*

VERB

AutoLEX: Automatic Language Explorer

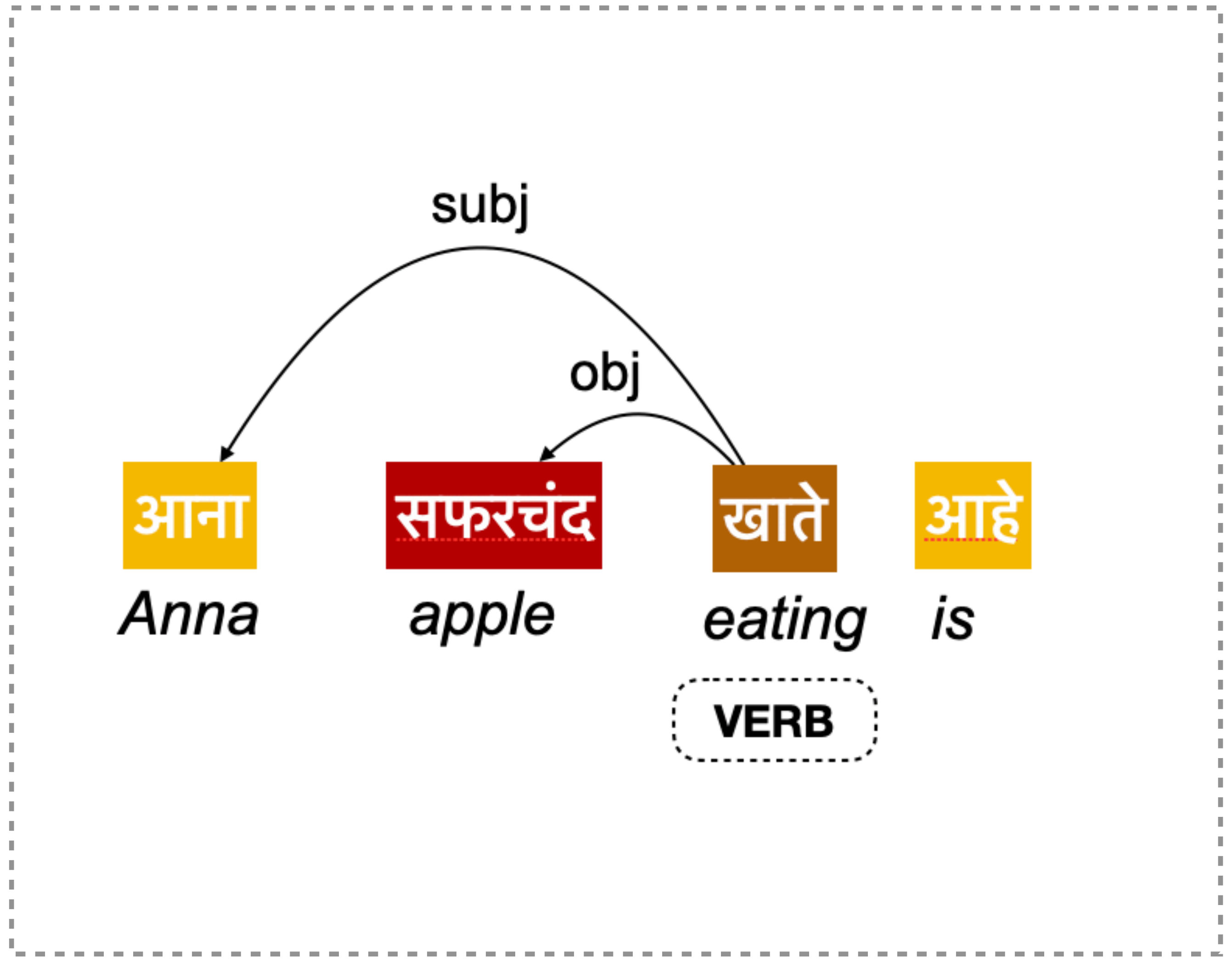
Subject Object Verb

तांदूळ
/tandul/
‘raw rice’

भात
/bhaat/
‘cooked rice’

Revisiting our research goal

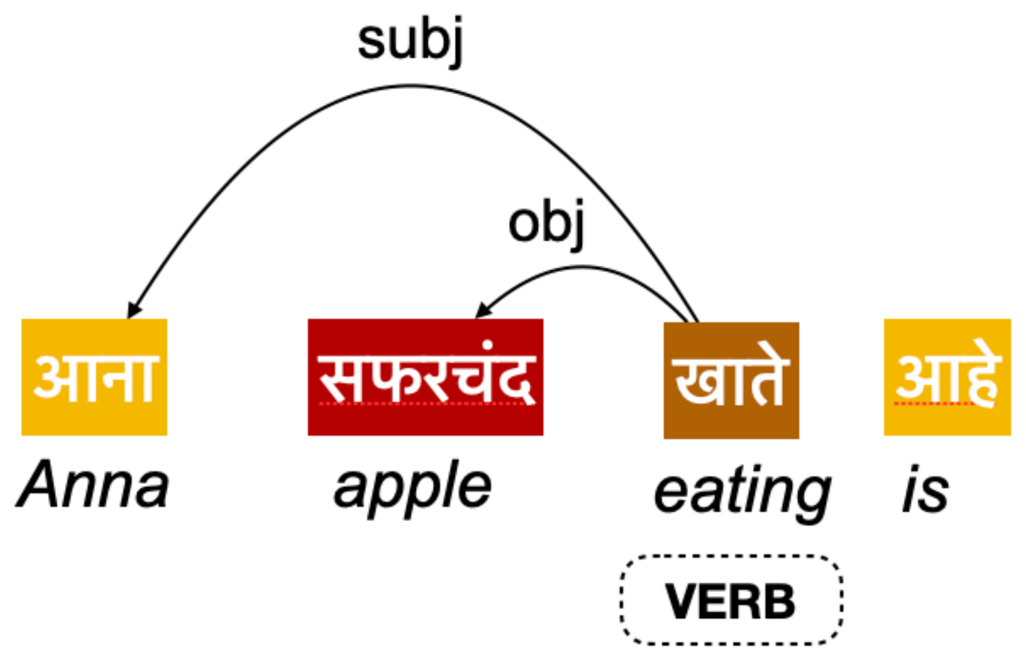
(Low-resource) Language Analysis



AutoLEX: Automatic Language Explorer

Revisiting our research goal

(Low-resource) Language Analysis



AutoLEX: Automatic Language Explorer

Subject **Object** **Verb**

तांदूळ
/tandul/
'raw rice'

भात
/bhaat/
'cooked rice'

Applications

Language Education and Documentation

Language Education

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

Language Education

- Computer-assisted language learning (CALL) systems are in **high demand**!

Language Education

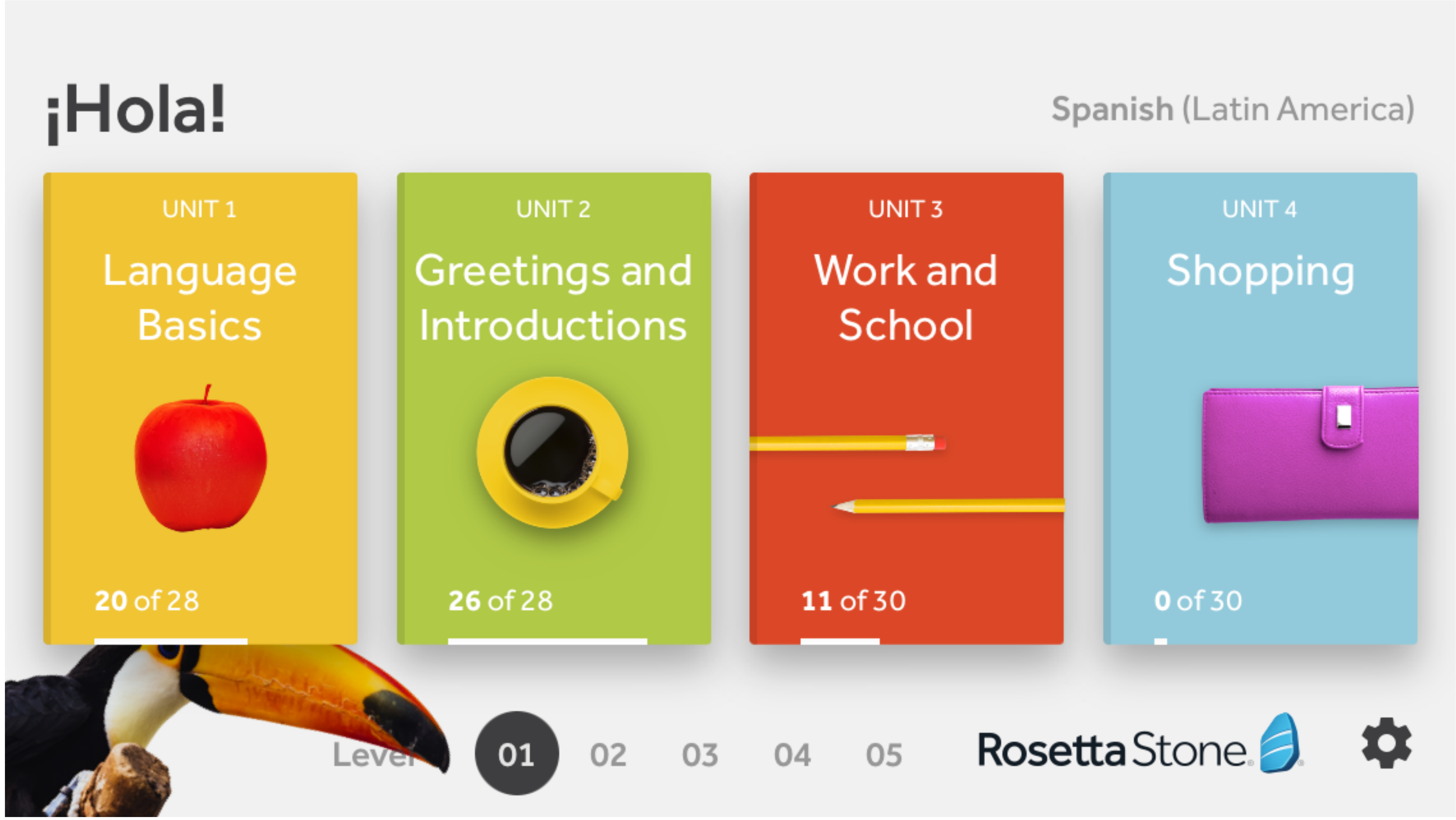
- Computer-assisted language learning (CALL) systems are in **high demand**!

Language Education

- Computer-assisted language learning (CALL) systems are in **high demand**!

Language Education

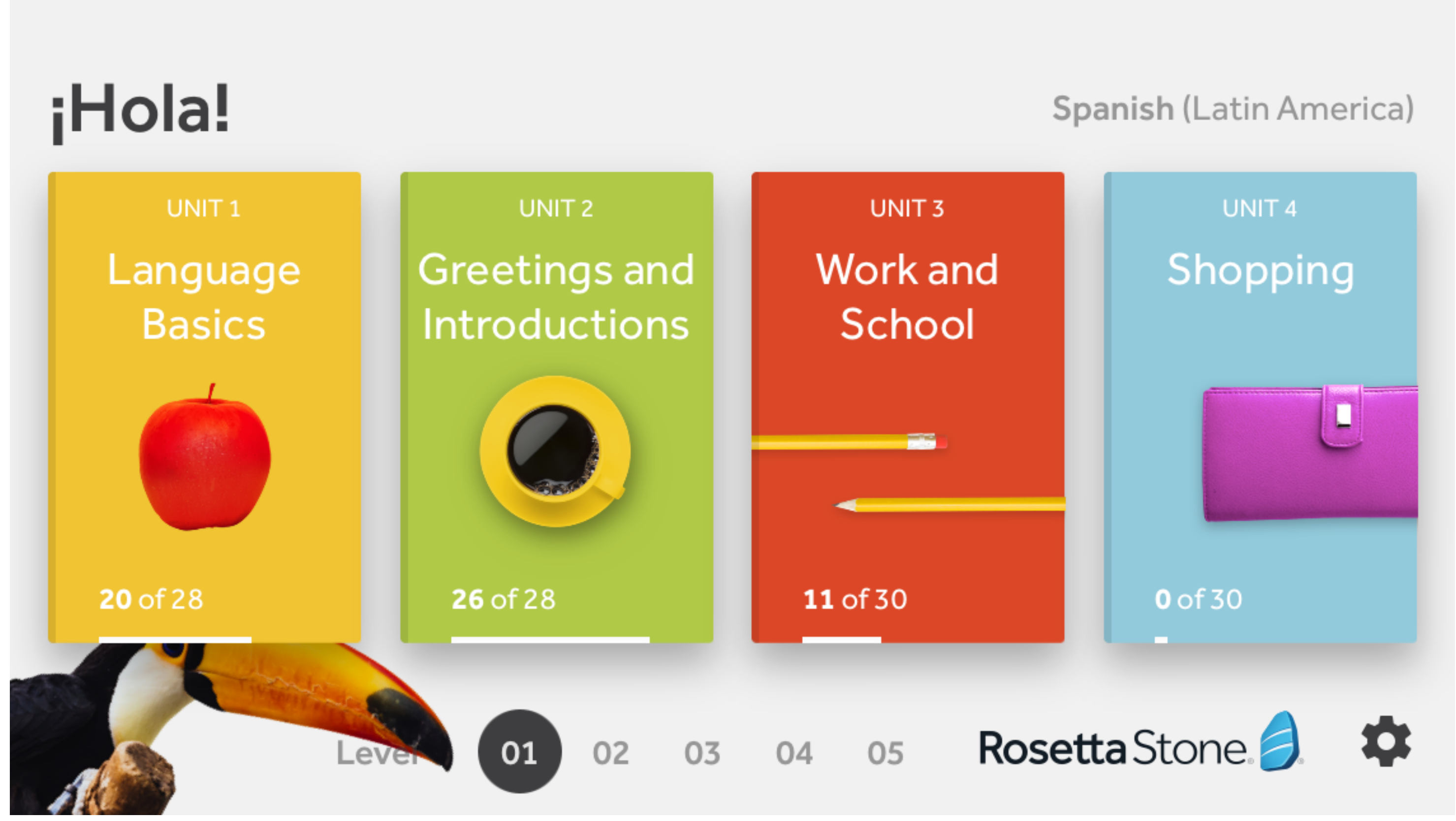
- Computer-assisted language learning (CALL) systems are in **high demand**!



Language Education

- Computer-assisted language learning (CALL) systems are in **high demand**!

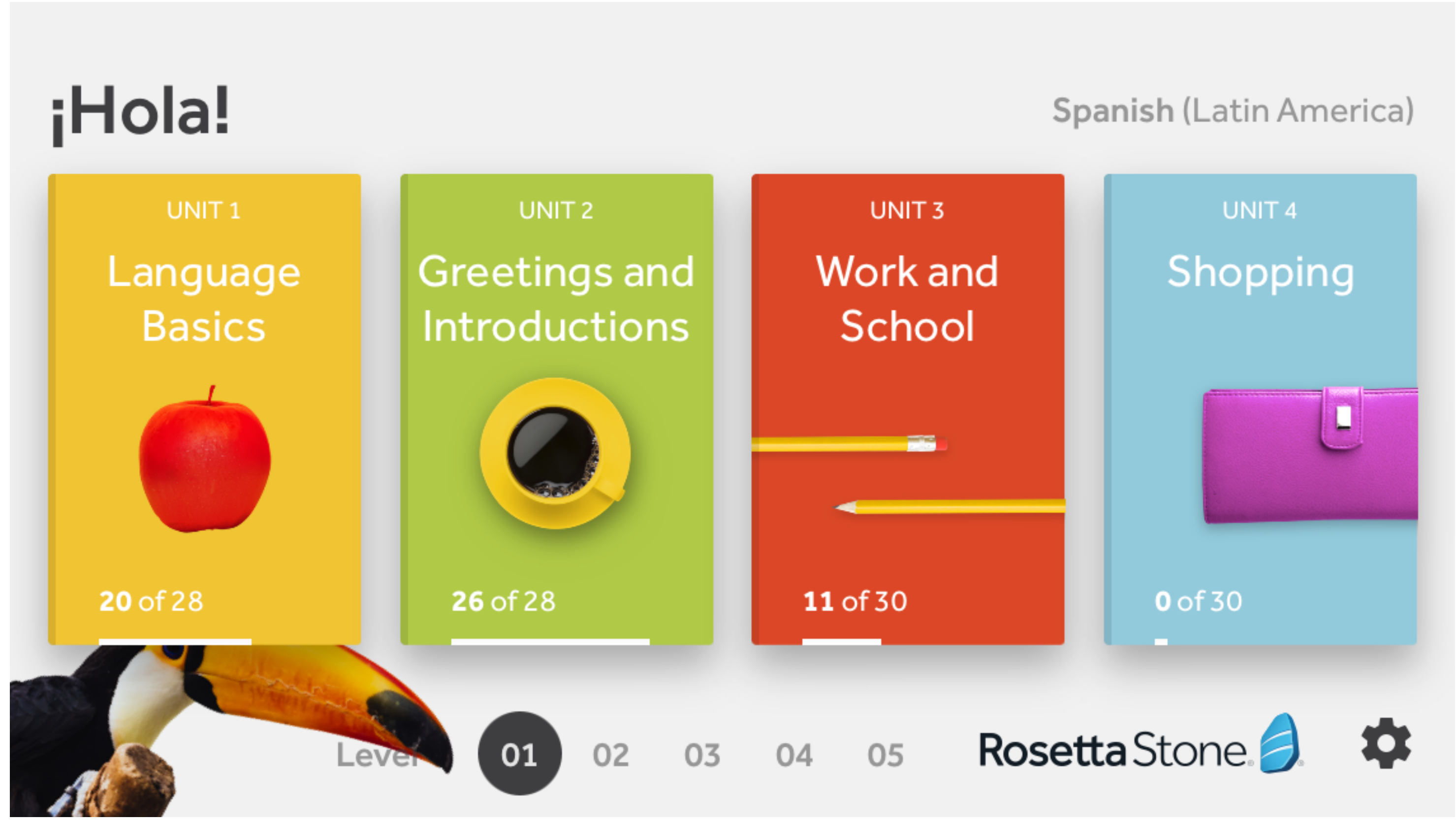
Limited in language coverage!



Language Education

- Computer-assisted language learning (CALL) systems are in **high demand**!

Limited in language coverage!

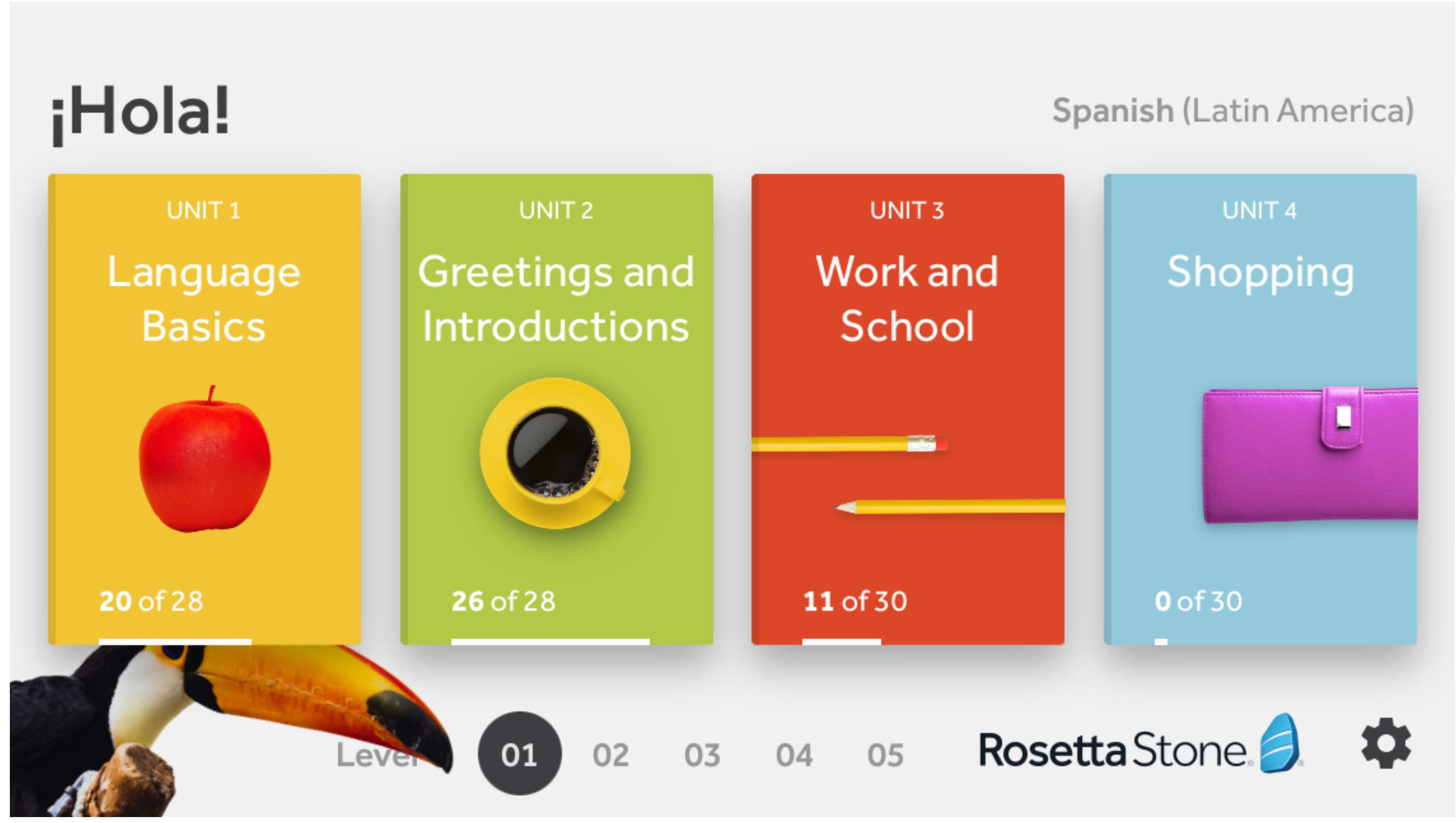


- Creating a curriculum is a **challenging process** ⟶ grammar coverage, examples, exercises …

Language Education

- Computer-assisted language learning (CALL) systems are in **high demand**!

Limited in language coverage!



- Creating a curriculum is a **challenging process** ⟶ grammar coverage, examples, exercises …

- **AutoLEX** has shown potential in doing some aspects of this **automatically**

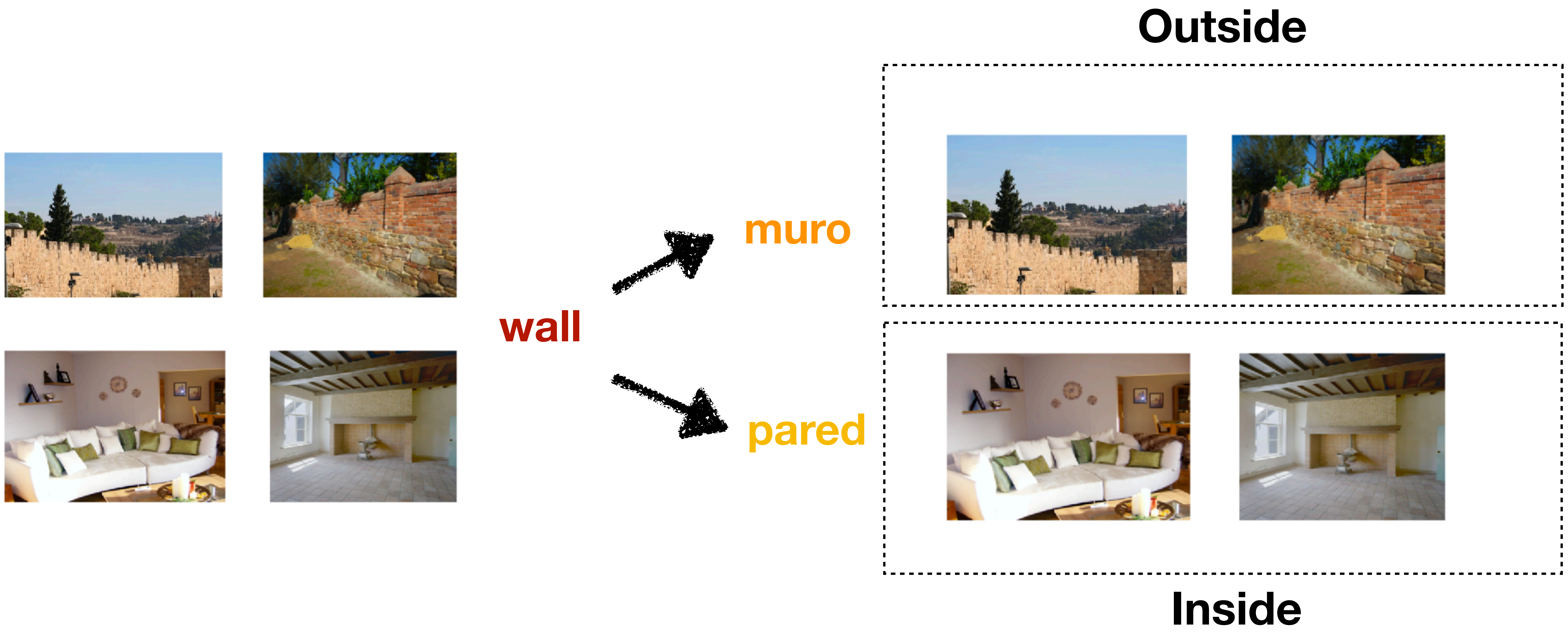
Word Usage: Semantic Subdivisions

Word Usage: Semantic Subdivisions

- Different languages carve up the semantic space differently

Word Usage: Semantic Subdivisions

- Different languages carve up the semantic space differently



Outside

wall

muro

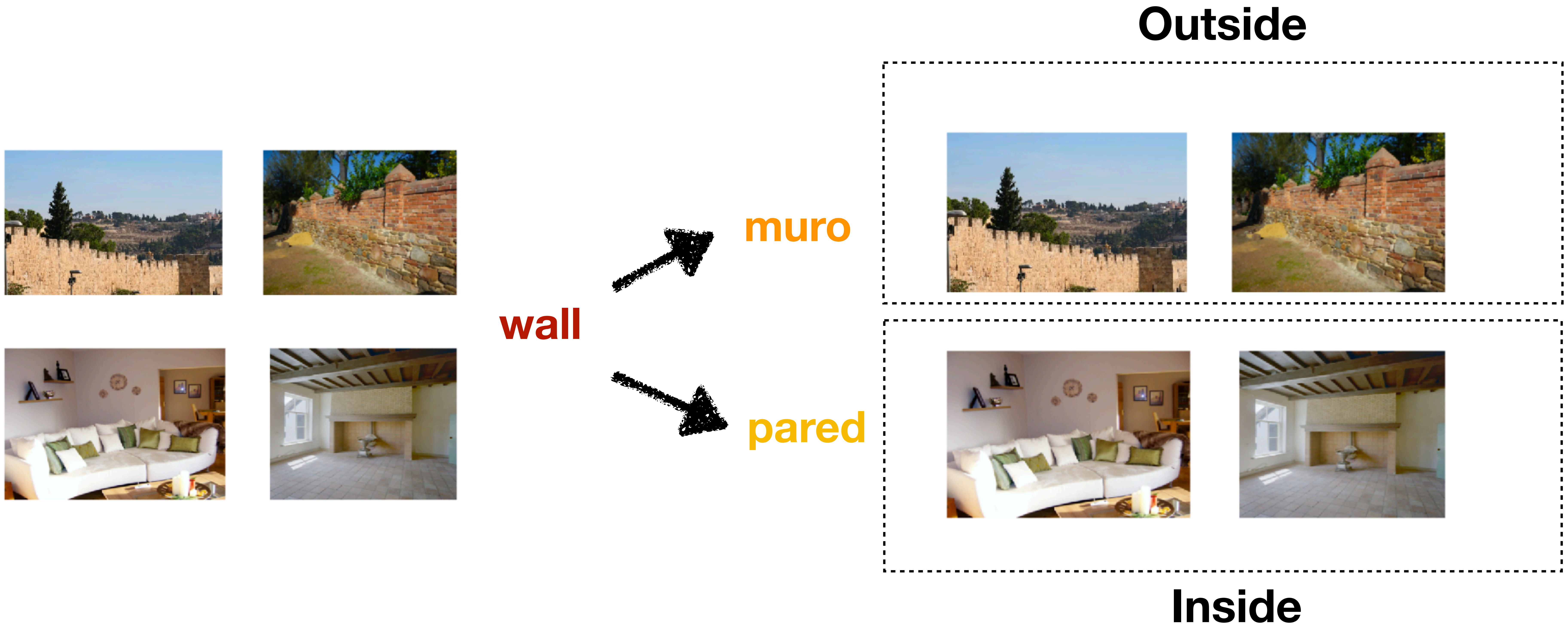
pared

Inside

When is Wall a Pared and when a Muro? Extracting Rules Governing Lexical Selection
Chaudhary**, Yin, Anastasopoulos, Neubig.* ***EMNLP 2021

Word Usage: Semantic Subdivisions

- Different languages carve up the semantic space differently



Outside

muro

wall

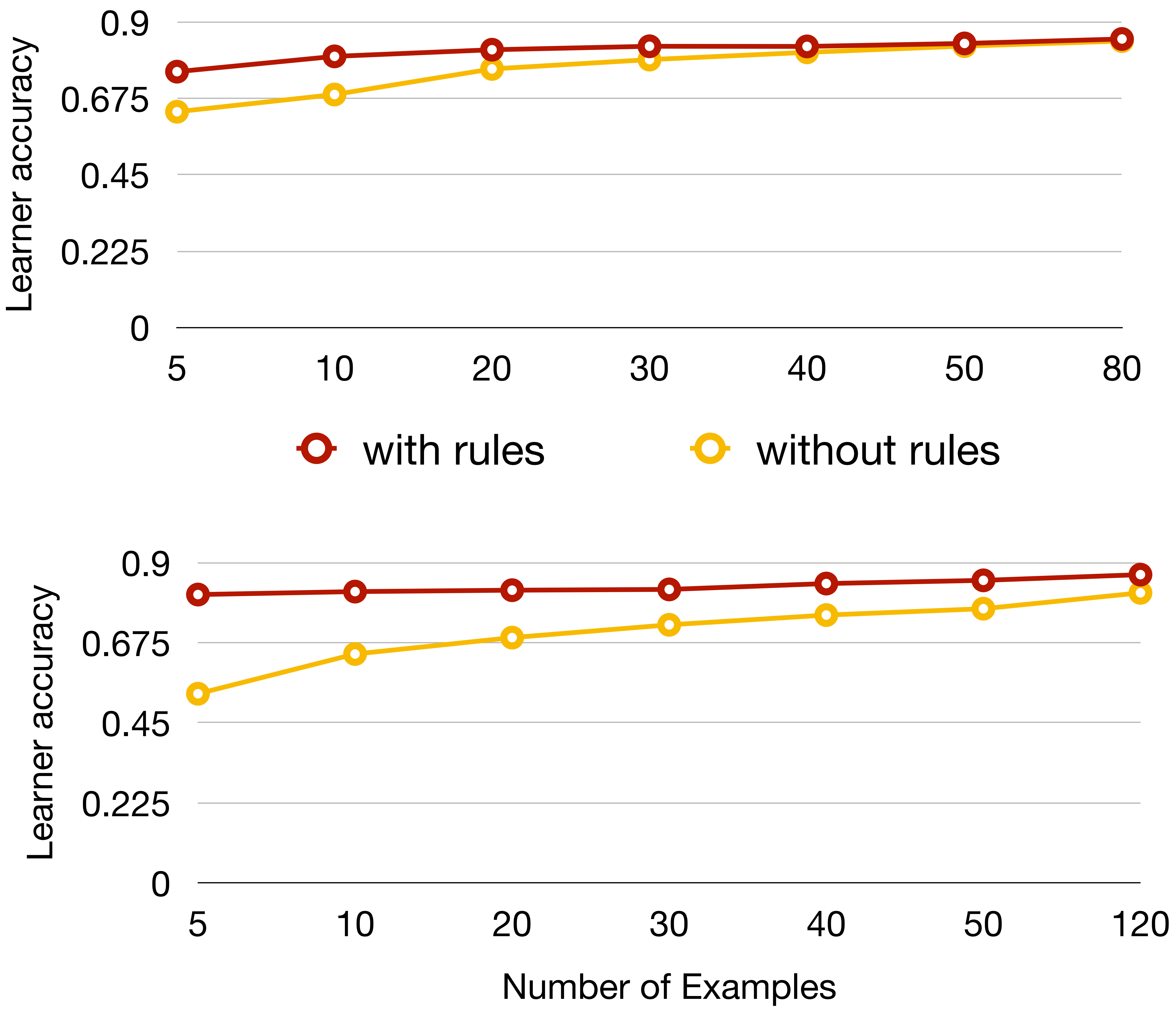
pared

Inside

- Crowdsourced study where participants recruited online had to predict correct word usage in context

When is Wall a Pared and when a Muro? Extracting Rules Governing Lexical Selection
Chaudhary**, Yin, Anastasopoulos, Neubig.* ***EMNLP 2021

Results



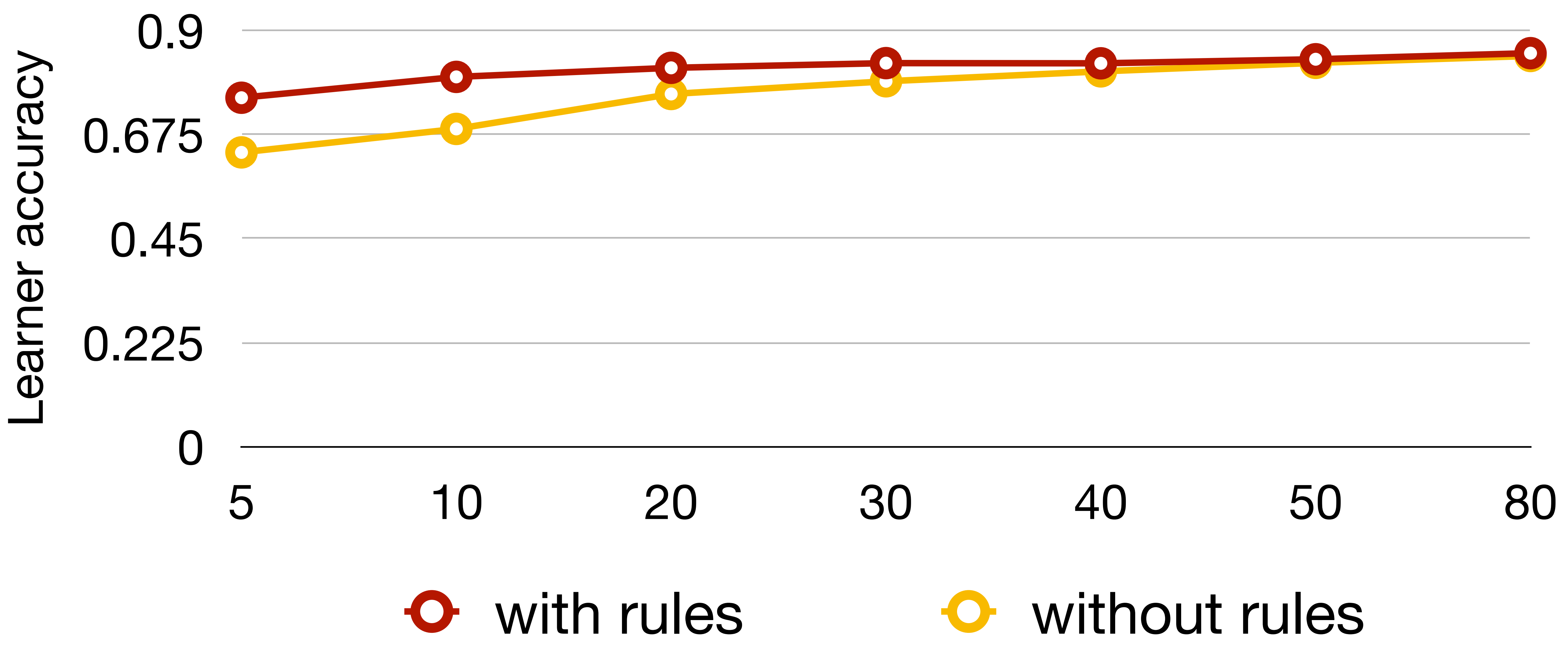
Spanish

7 learners, 9 words

Greek

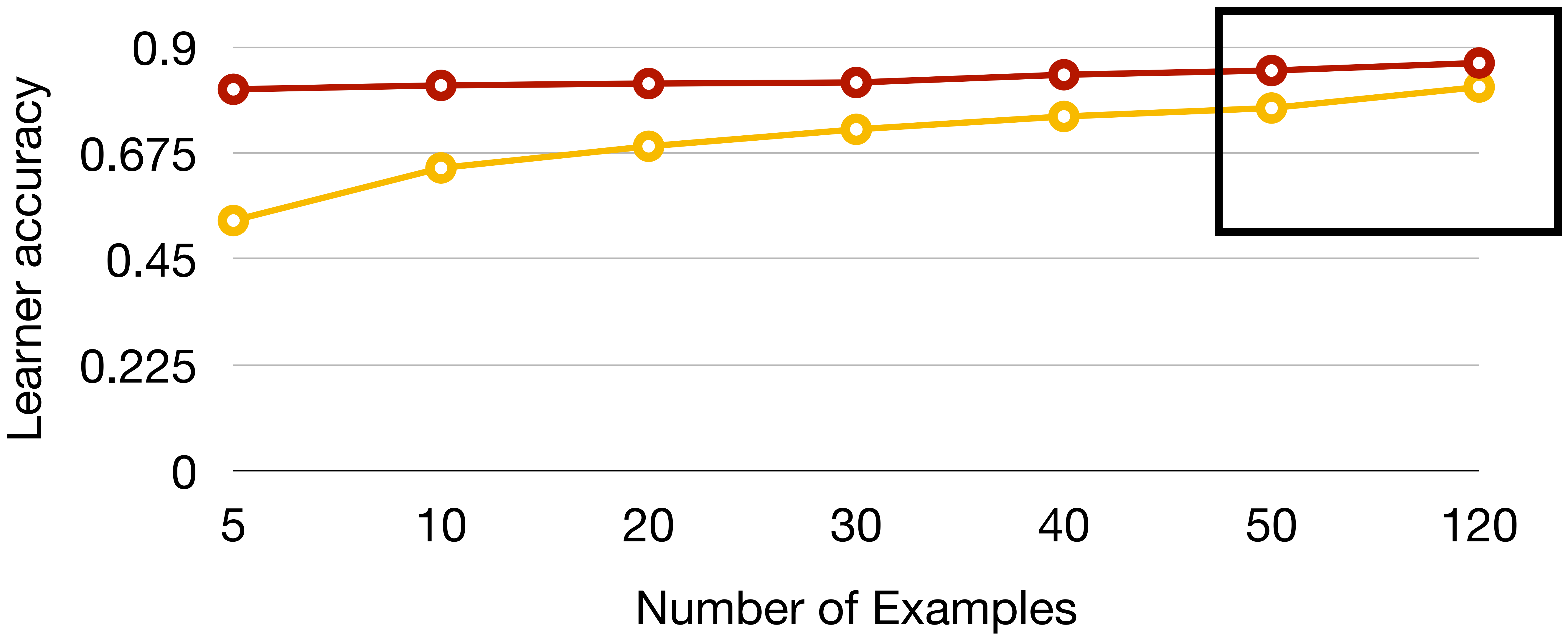
9 learners, 10 words

Results



Spanish

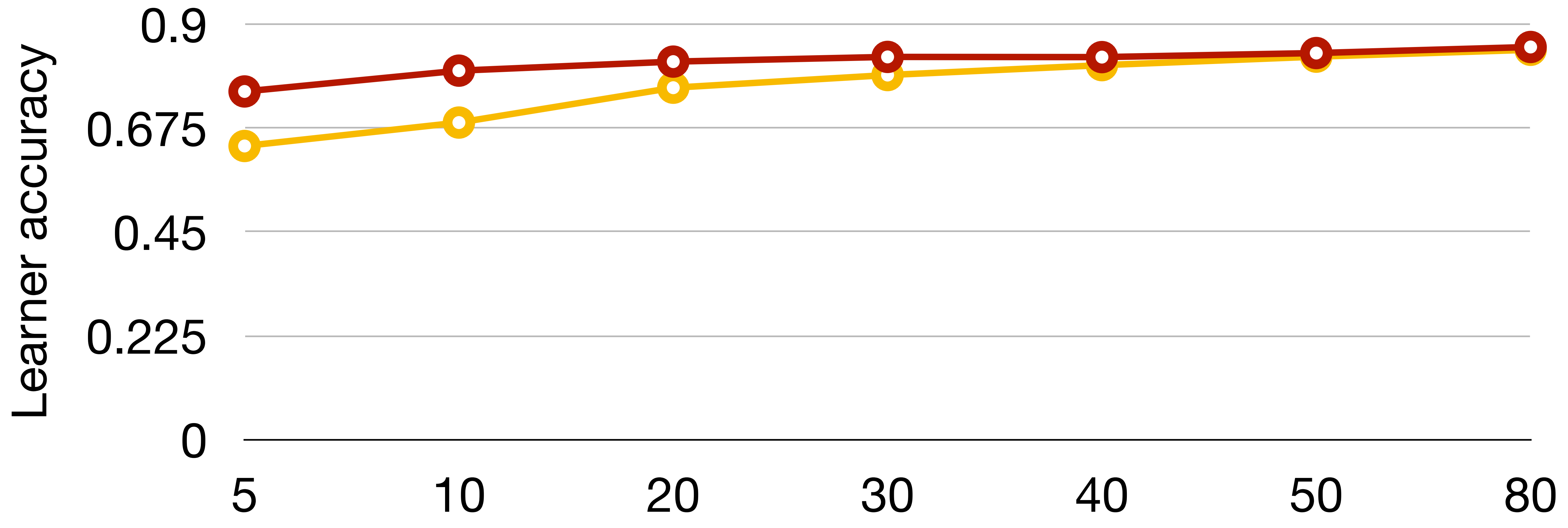
7 learners, 9 words



Greek

9 learners, 10 words

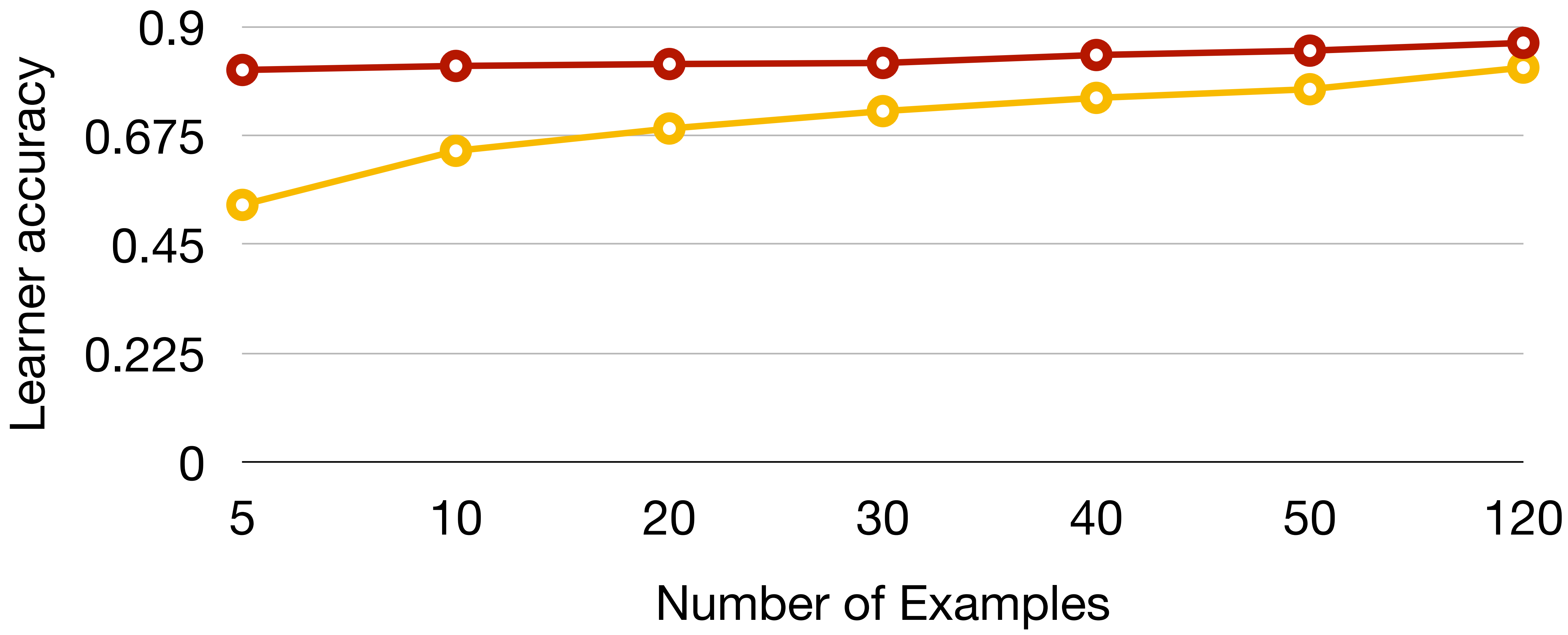
Results



Spanish

7 learners, 9 words

When shown rules, learners learn better and faster!



Greek

9 learners, 10 words

Language Education

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

Language Education

- Apply **AutoLEX** to extract grammar aspects for teaching two Indian languages

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

Language Education

- Apply **AutoLEX** to extract grammar aspects for teaching two Indian languages

Indo-Aryan language family

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

Language Education

- Apply **AutoLEX** to extract grammar aspects for teaching two Indian languages

Indo-Aryan language family

Dravidian language family

“Classical language status”

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

Language Education

- Apply **AutoLEX** to extract grammar aspects for teaching two Indian languages **outside of India**

Indo-Aryan language family

Dravidian language family

“Classical language status”

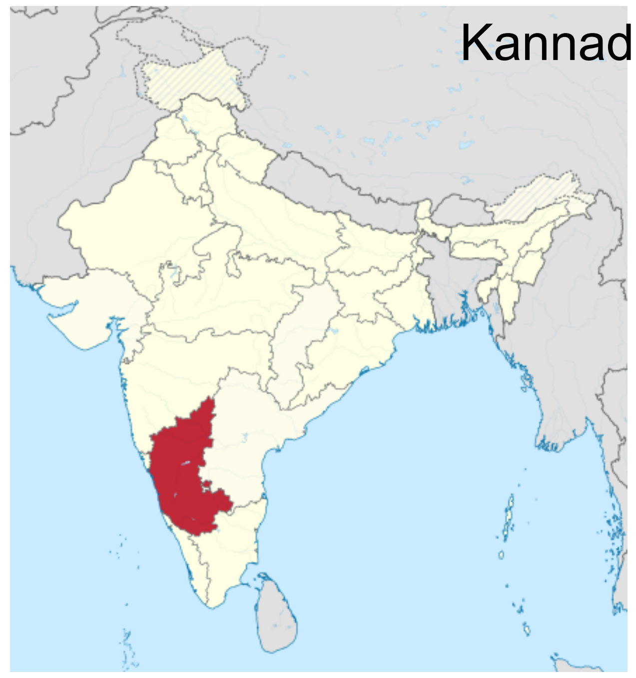
Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

Language Education

- Apply **AutoLEX** to extract grammar aspects for teaching two Indian languages **outside of India**

Marathi - मराठी

Indo-Aryan language family



Kannada - ಕನ್ನಡ

Dravidian language family

“Classical language status”

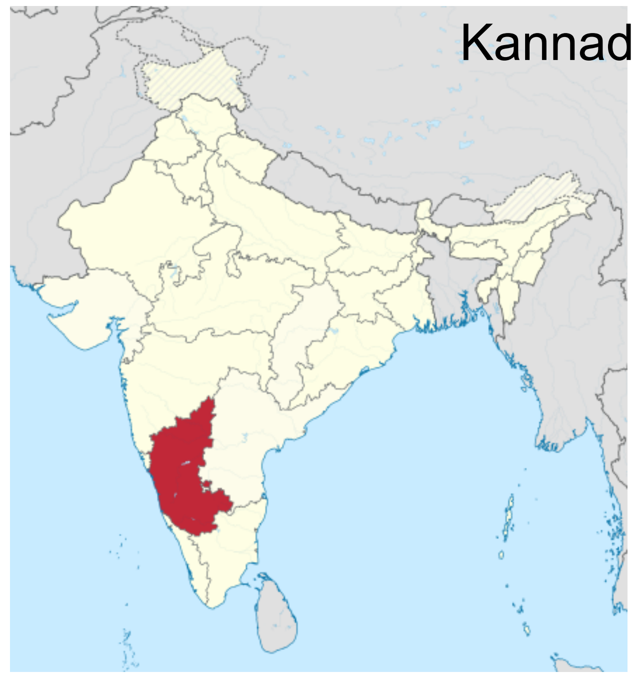
Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

Language Education

- Apply **AutoLEX** to extract grammar aspects for teaching two Indian languages **outside of India**

Marathi - मराठी

Indo-Aryan language family



Kannada - ಕನ್ನಡ

Dravidian language family

“Classical language status”

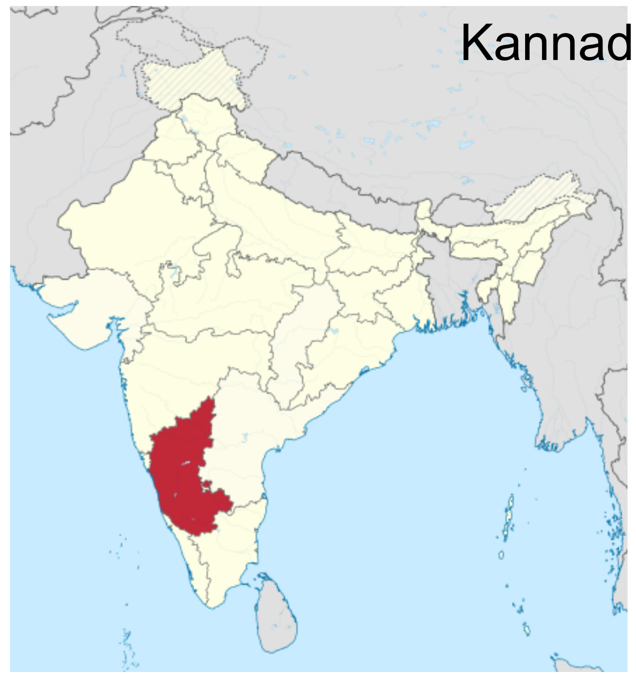
- **Under-resourced settings** w.r.t pedagogical resources as well as NLP models/resources

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

Language Education

- Apply **AutoLEX** to extract grammar aspects for teaching two Indian languages **outside of India**

Indo-Aryan language family



Dravidian language family

“Classical language status”

- **Under-resourced settings** w.r.t pedagogical resources as well as NLP models/resources

- **Access to in-service teachers** that teach these languages to English speakers

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

Communities and Consultations

Communities and Consultations

- **Kannada** ⟶ Kannada Academy largest organization in the world

  70 learning centers, 800 volunteer teachers

Arun Sampath

Director - Academics

Ashwin Sheshadri, UK

Communities and Consultations

- **Kannada** ⟶ Kannada Academy largest organization in the world

  70 learning centers, 800 volunteer teachers

Arun Sampath

Director - Academics

Ashwin Sheshadri, UK

- **Marathi** ⟶ Marathi Vidyalaya, Randolph, New Jersey

  Marathi Shala, Pittsburgh

  Small independent volunteer-run schools

AutoLEX: Selection of Grammar Aspects

AutoLEX: Selection of Grammar Aspects

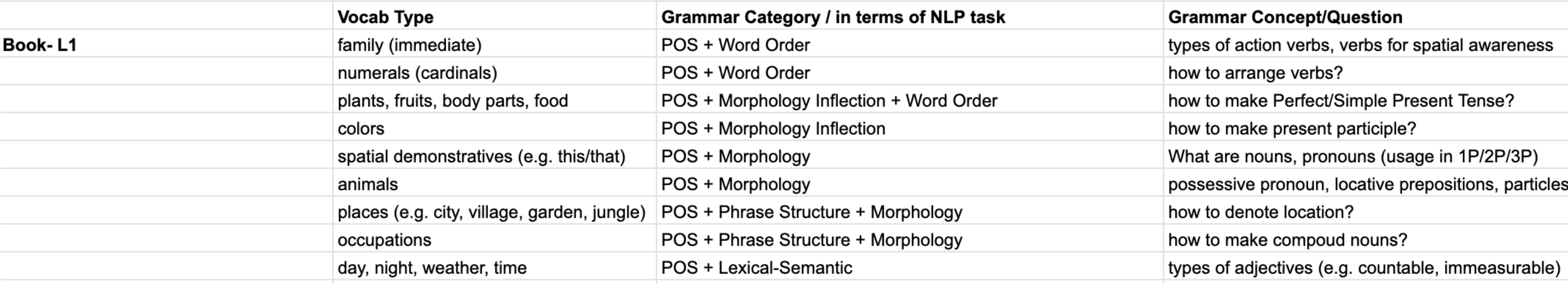
- In consultation with the curriculum designers

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points



	Vocab Type	Grammar Category / in terms of NLP task	Grammar Concept/Question
Book- L1	family (immediate)	POS + Word Order	types of action verbs, verbs for spatial awareness
	numerals (cardinals)	POS + Word Order	how to arrange verbs?
	plants, fruits, body parts, food	POS + Morphology Inflection + Word Order	how to make Perfect/Simple Present Tense?
	colors	POS + Morphology Inflection	how to make present participle?
	spatial demonstratives (e.g. this/that)	POS + Morphology	What are nouns, pronouns (usage in 1P/2P/3P)
	animals	POS + Morphology	possessive pronoun, locative prepositions, particles
	places (e.g. city, village, garden, jungle)	POS + Phrase Structure + Morphology	how to denote location?
	occupations	POS + Phrase Structure + Morphology	how to make compound nouns?
	day, night, weather, time	POS + Lexical-Semantic	types of adjectives (e.g. countable, immeasurable)

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points

	Vocab Type	**Grammar Category / in terms of NLP task**	**Grammar Concept/Question**
Book- L1	family (immediate)	POS + Word Order	types of action verbs, verbs for spatial awareness
	numerals (cardinals)	POS + Word Order	how to arrange verbs?
	plants, fruits, body parts, food	POS + Morphology Inflection + Word Order	how to make Perfect/Simple Present Tense?
	colors	POS + Morphology Inflection	how to make present participle?
	spatial demonstratives (e.g. this/that)	POS + Morphology	What are nouns, pronouns (usage in 1P/2P/3P)
	animals	POS + Morphology	possessive pronoun, locative prepositions, particles
	places (e.g. city, village, garden, jungle)	POS + Phrase Structure + Morphology	how to denote location?
	occupations	POS + Phrase Structure + Morphology	how to make compound nouns?
	day, night, weather, time	POS + Lexical-Semantic	types of adjectives (e.g. countable, immeasurable)

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points

	Vocab Type	**Grammar Category / in terms of NLP task**	**Grammar Concept/Question**
Book- L1	family (immediate)	POS + Word Order	types of action verbs, verbs for spatial awareness
	numerals (cardinals)	POS + Word Order	how to arrange verbs?
	plants, fruits, body parts, food	POS + Morphology Inflection + Word Order	how to make Perfect/Simple Present Tense?
	colors	POS + Morphology Inflection	how to make present participle?
	spatial demonstratives (e.g. this/that)	POS + Morphology	What are nouns, pronouns (usage in 1P/2P/3P)
	animals	POS + Morphology	possessive pronoun, locative prepositions, particles
	places (e.g. city, village, garden, jungle)	POS + Phrase Structure + Morphology	how to denote location?
	occupations	POS + Phrase Structure + Morphology	how to make compound nouns?
	day, night, weather, time	POS + Lexical-Semantic	types of adjectives (e.g. countable, immeasurable)

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points

	Vocab Type	**Grammar Category / in terms of NLP task**	**Grammar Concept/Question**
Book- L1	family (immediate)	POS + Word Order	types of action verbs, verbs for spatial awareness
	numerals (cardinals)	POS + Word Order	how to arrange verbs?
	plants, fruits, body parts, food	POS + Morphology Inflection + Word Order	how to make Perfect/Simple Present Tense?
	colors	POS + Morphology Inflection	how to make present participle?
	spatial demonstratives (e.g. this/that)	POS + Morphology	What are nouns, pronouns (usage in 1P/2P/3P)
	animals	POS + Morphology	possessive pronoun, locative prepositions, particles
	places (e.g. city, village, garden, jungle)	POS + Phrase Structure + Morphology	how to denote location?
	occupations	POS + Phrase Structure + Morphology	how to make compound nouns?
	day, night, weather, time	POS + Lexical-Semantic	types of adjectives (e.g. countable, immeasurable)

General Info

AutoLEX: Selection of Grammar Aspects

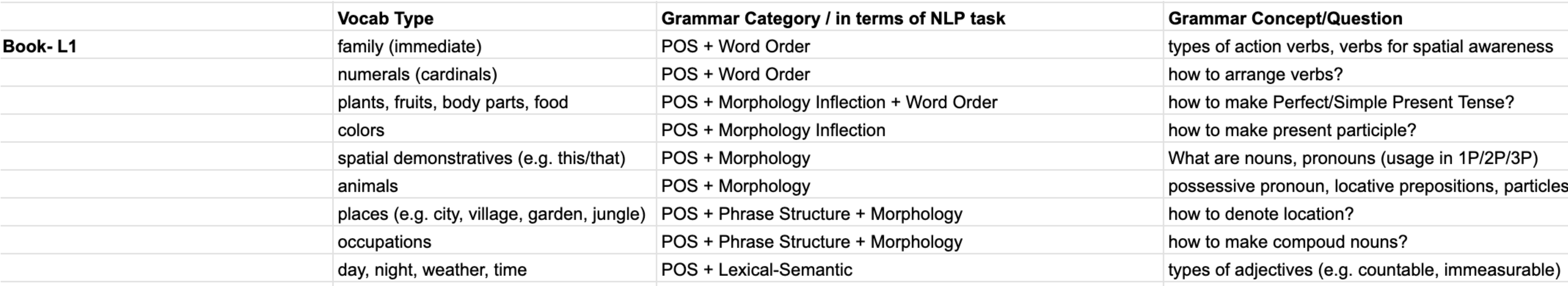
Lemma	Morphosyntactic Attributes	Gender				
		Fem	NA	Neut	Masc	
कर (kar)		-	करून (karun)	-	-	Examples
कर (kar)	2;Plur	-	करा (kara)	-	-	Examples
कर (kar)	Acc;Sing	-	-	करण्याचा (karanya)	-	Examples
कर (kar)	Nom;Sing	केली (keli)	-	करणे (karane)	केला (kelaa)	Examples
कर (kar)	3;Past;Sing	केली (keli)	-	केले (kele)	केला (kelaa)	Examples
कर (kar)	1;Plur	-	-	-	करावे (karaave)	Examples
कर (kar)	3;Sing	करावी (karavi)	-	करावे (karaave)	करावा (karawa)	Examples
कर (kar)	3;Nom;Past;Sing	-	-	-	केला (kelaa)	Examples
कर (kar)	3;Past;Plur	केल्या (kelya)	-	केली (keli)	करीत (kareet)	Examples
कर (kar)	Acc	-	केल्यामुळे (kelyamule)	-	-	Examples
कर (kar)	3;Plur;Pres	-	-	-	करतात (kartaat)	Examples

General Info

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points



	Vocab Type	Grammar Category / in terms of NLP task	Grammar Concept/Question
Book- L1	family (immediate)	POS + Word Order	types of action verbs, verbs for spatial awareness
	numerals (cardinals)	POS + Word Order	how to arrange verbs?
	plants, fruits, body parts, food	POS + Morphology Inflection + Word Order	how to make Perfect/Simple Present Tense?
	colors	POS + Morphology Inflection	how to make present participle?
	spatial demonstratives (e.g. this/that)	POS + Morphology	What are nouns, pronouns (usage in 1P/2P/3P)
	animals	POS + Morphology	possessive pronoun, locative prepositions, particles
	places (e.g. city, village, garden, jungle)	POS + Phrase Structure + Morphology	how to denote location?
	occupations	POS + Phrase Structure + Morphology	how to make compound nouns?
	day, night, weather, time	POS + Lexical-Semantic	types of adjectives (e.g. countable, immeasurable)

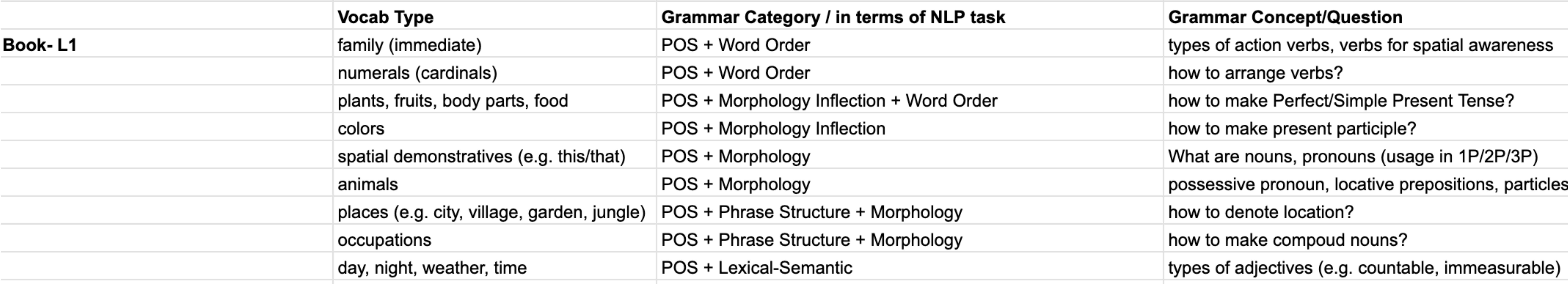
General Info

Agreement

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points



	Vocab Type	**Grammar Category / in terms of NLP task**	**Grammar Concept/Question**
Book- L1	family (immediate)	POS + Word Order	types of action verbs, verbs for spatial awareness
	numerals (cardinals)	POS + Word Order	how to arrange verbs?
	plants, fruits, body parts, food	POS + Morphology Inflection + Word Order	how to make Perfect/Simple Present Tense?
	colors	POS + Morphology Inflection	how to make present participle?
	spatial demonstratives (e.g. this/that)	POS + Morphology	What are nouns, pronouns (usage in 1P/2P/3P)
	animals	POS + Morphology	possessive pronoun, locative prepositions, particles
	places (e.g. city, village, garden, jungle)	POS + Phrase Structure + Morphology	how to denote location?
	occupations	POS + Phrase Structure + Morphology	how to make compound nouns?
	day, night, weather, time	POS + Lexical-Semantic	types of adjectives (e.g. countable, immeasurable)

General Info

Agreement

Word Order

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points

	Vocab Type	**Grammar Category / in terms of NLP task**	**Grammar Concept/Question**
Book- L1	family (immediate)	POS + Word Order	types of action verbs, verbs for spatial awareness
	numerals (cardinals)	POS + Word Order	how to arrange verbs?
	plants, fruits, body parts, food	POS + Morphology Inflection + Word Order	how to make Perfect/Simple Present Tense?
	colors	POS + Morphology Inflection	how to make present participle?
	spatial demonstratives (e.g. this/that)	POS + Morphology	What are nouns, pronouns (usage in 1P/2P/3P)
	animals	POS + Morphology	possessive pronoun, locative prepositions, particles
	places (e.g. city, village, garden, jungle)	POS + Phrase Structure + Morphology	how to denote location?
	occupations	POS + Phrase Structure + Morphology	how to make compound nouns?
	day, night, weather, time	POS + Lexical-Semantic	types of adjectives (e.g. countable, immeasurable)

General Info

Agreement

Word Order

Affix Usage

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers

ला (laa) suffix is used when:

current word is in the neighborhood of the word= अटक (atak) current word is in the neighborhood of the word= सुरुवात (sruvaat) current word with Number = Sing current word with Case = Acc current word is the= oblique	Examples

चा (chaa) suffix is used when:

current word with Case = Acc current word is in the neighborhood of the word= मृत्यू (mrityu) current word is the= modifer current word's head is the= compound current word's head with Gender = Masc	Examples

General Info

Agreement

Word Order

Affix Usage

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points

	Vocab Type	Grammar Category / in terms of NLP task	Grammar Concept/Question
Book- L1	family (immediate)	POS + Word Order	types of action verbs, verbs for spatial awareness
	numerals (cardinals)	POS + Word Order	how to arrange verbs?
	plants, fruits, body parts, food	POS + Morphology Inflection + Word Order	how to make Perfect/Simple Present Tense?
	colors	POS + Morphology Inflection	how to make present participle?
	spatial demonstratives (e.g. this/that)	POS + Morphology	What are nouns, pronouns (usage in 1P/2P/3P)
	animals	POS + Morphology	possessive pronoun, locative prepositions, particles
	places (e.g. city, village, garden, jungle)	POS + Phrase Structure + Morphology	how to denote location?
	occupations	POS + Phrase Structure + Morphology	how to make compound nouns?
	day, night, weather, time	POS + Lexical-Semantic	types of adjectives (e.g. countable, immeasurable)

General Info

Agreement

Word Order

Affix Usage

AutoLEX: Selection of Grammar Aspects

- In consultation with the curriculum designers
- Perused existing Kannada textbooks to identify popular grammar points

	Vocab Type	**Grammar Category / in terms of NLP task**	**Grammar Concept/Question**
Book- L1	family (immediate)	POS + Word Order	types of action verbs, verbs for spatial awareness
	numerals (cardinals)	POS + Word Order	how to arrange verbs?
	plants, fruits, body parts, food	POS + Morphology Inflection + Word Order	how to make Perfect/Simple Present Tense?
	colors	POS + Morphology Inflection	how to make present participle?
	spatial demonstratives (e.g. this/that)	POS + Morphology	What are nouns, pronouns (usage in 1P/2P/3P)
	animals	POS + Morphology	possessive pronoun, locative prepositions, particles
	places (e.g. city, village, garden, jungle)	POS + Phrase Structure + Morphology	how to denote location?
	occupations	POS + Phrase Structure + Morphology	how to make compound nouns?
	day, night, weather, time	POS + Lexical-Semantic	types of adjectives (e.g. countable, immeasurable)

General Info

Agreement

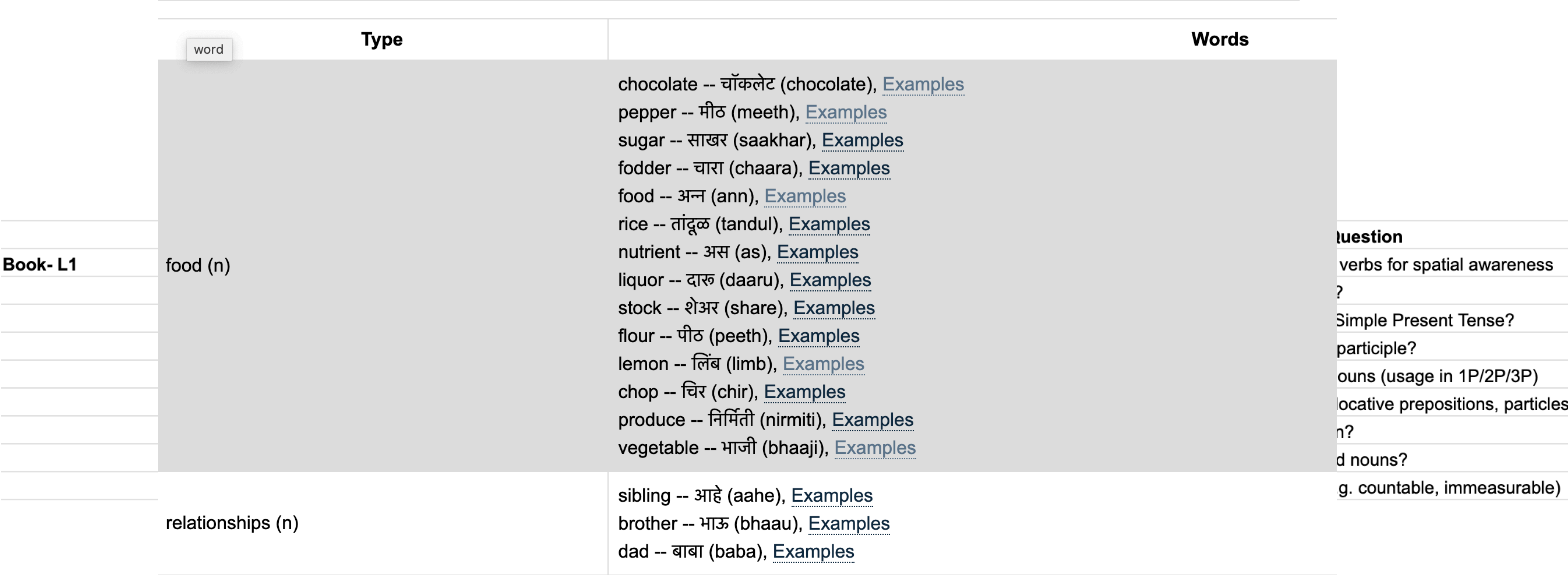
Word Order

Affix Usage

Word Usage

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Selection of Grammar Aspects



Type	Words
food (n)	chocolate -- चॉकलेट (chocolate), Examples pepper -- मीठ (meeth), Examples sugar -- साखर (saakhar), Examples fodder -- चारा (chaara), Examples food -- अन्न (ann), Examples rice -- तांदूळ (tandul), Examples nutrient -- अस (as), Examples liquor -- दारू (daaru), Examples stock -- शेअर (share), Examples flour -- पीठ (peeth), Examples lemon -- लिंब (limb), Examples chop -- चिर (chir), Examples produce -- निर्मिती (nirmiti), Examples vegetable -- भाजी (bhaaji), Examples
relationships (n)	sibling -- आहे (aahe), Examples brother -- भाऊ (bhaau), Examples dad -- बाबा (baba), Examples

Book- L1

uestion
verbs for spatial awareness
?
Simple Present Tense?
participle?
ouns (usage in 1P/2P/3P)
locative prepositions, particles
n?
d nouns?
g. countable, immeasurable)

General Info

Agreement

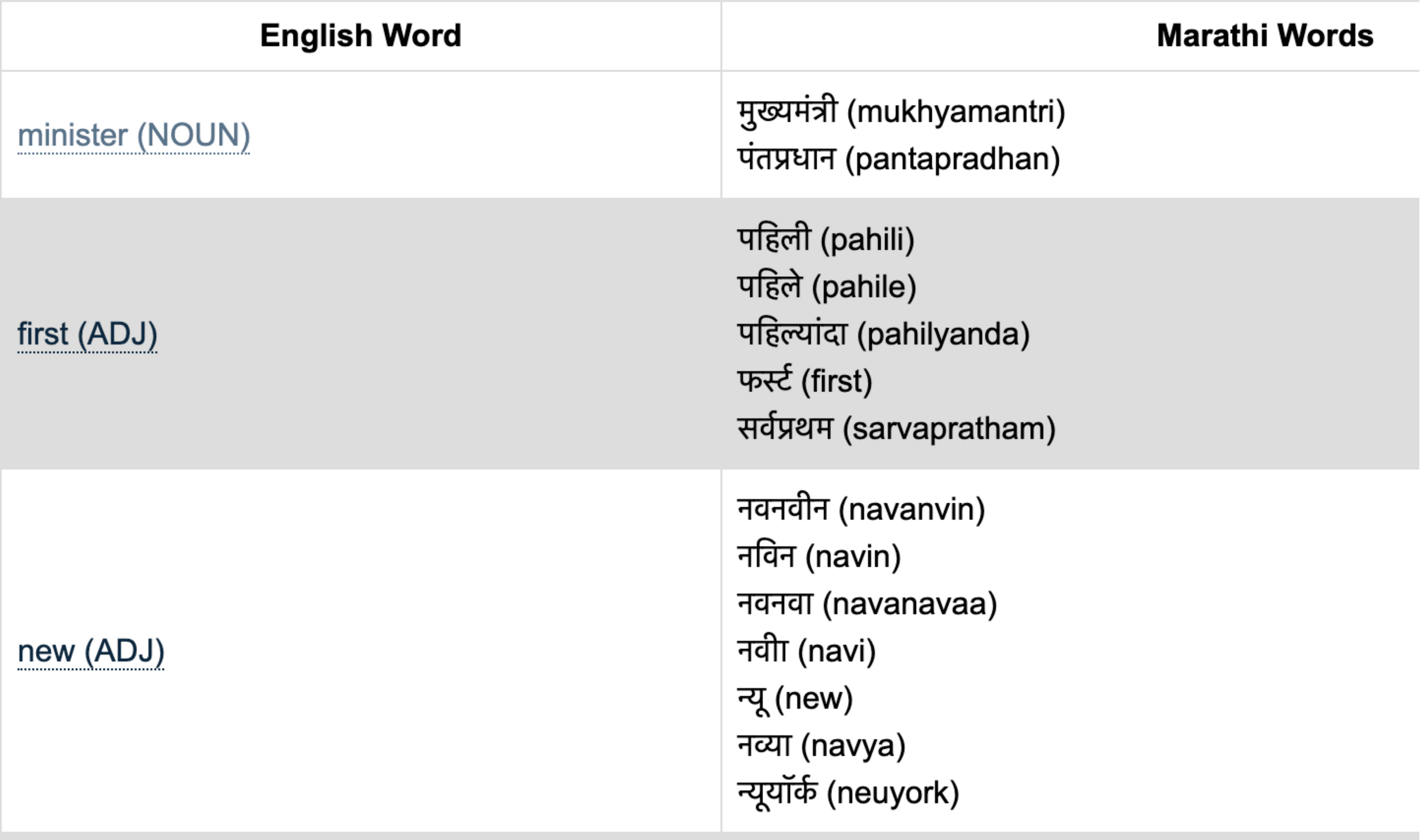
Word Order

Affix Usage

Word Usage

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX



English Word	Marathi Words
minister (NOUN)	मुख्यमंत्री (mukhyamantri) पंतप्रधान (pantapradhan)
first (ADJ)	पहिली (pahili) पहिले (pahile) पहिल्यांदा (pahilyanda) फर्स्ट (first) सर्वप्रथम (sarvapratham)
new (ADJ)	नवनवीन (navanvin) नविन (navin) नवनवा (navanavaa) नवीा (navi) न्यू (new) नव्या (navya) न्यूयॉर्क (neuyork)

General Info

Agreement

Word Order

Affix Usage

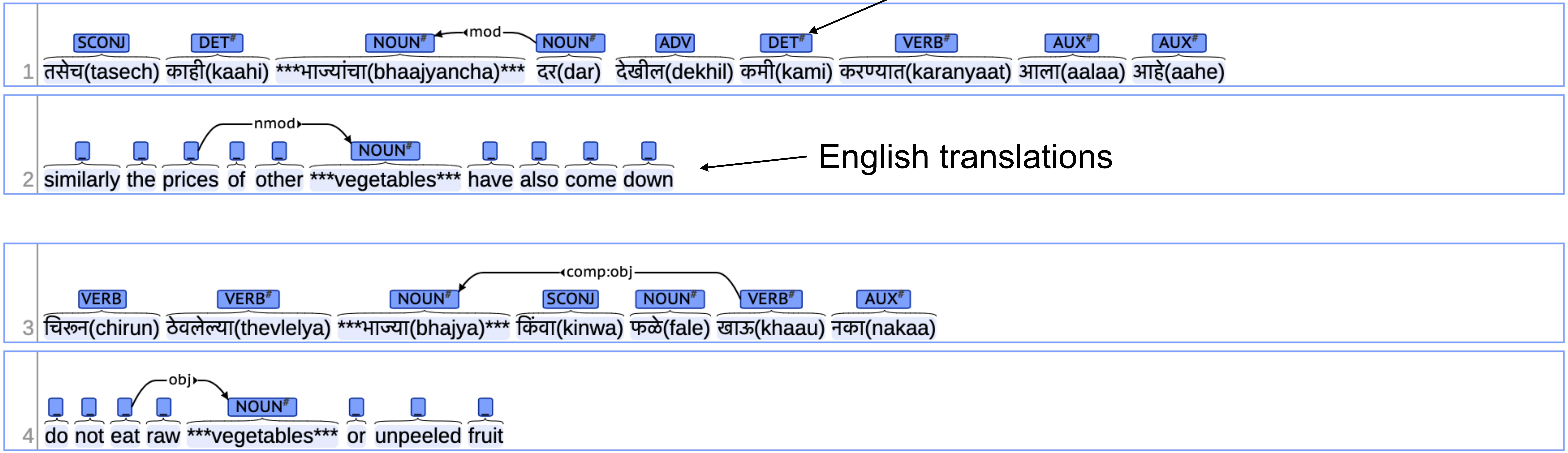
Word Usage

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Selection of Grammar Aspects

Example usages of **भाजी (bhaaji)**

Roman transliteration



1 तसेच(tasech) काही(kaahi) ***भाज्यांचा(bhaajyancha)*** दर(dar) देखील(dekhil) कमी(kami) करण्यात(karanyaat) आला(aalaa) आहे(aahe)

2 similarly the prices of other ***vegetables*** have also come down

3 चिरून(chirun) ठेवलेल्या(thevlelya) ***भाज्या(bhajya)*** किंवा(kinwa) फळे(fale) खाऊ(khaau) नका(nakaa)

4 do not eat raw ***vegetables*** or unpeeled fruit

English translations

General Info

Agreement

Word Order

Affix Usage

Word Usage

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions

kn-en

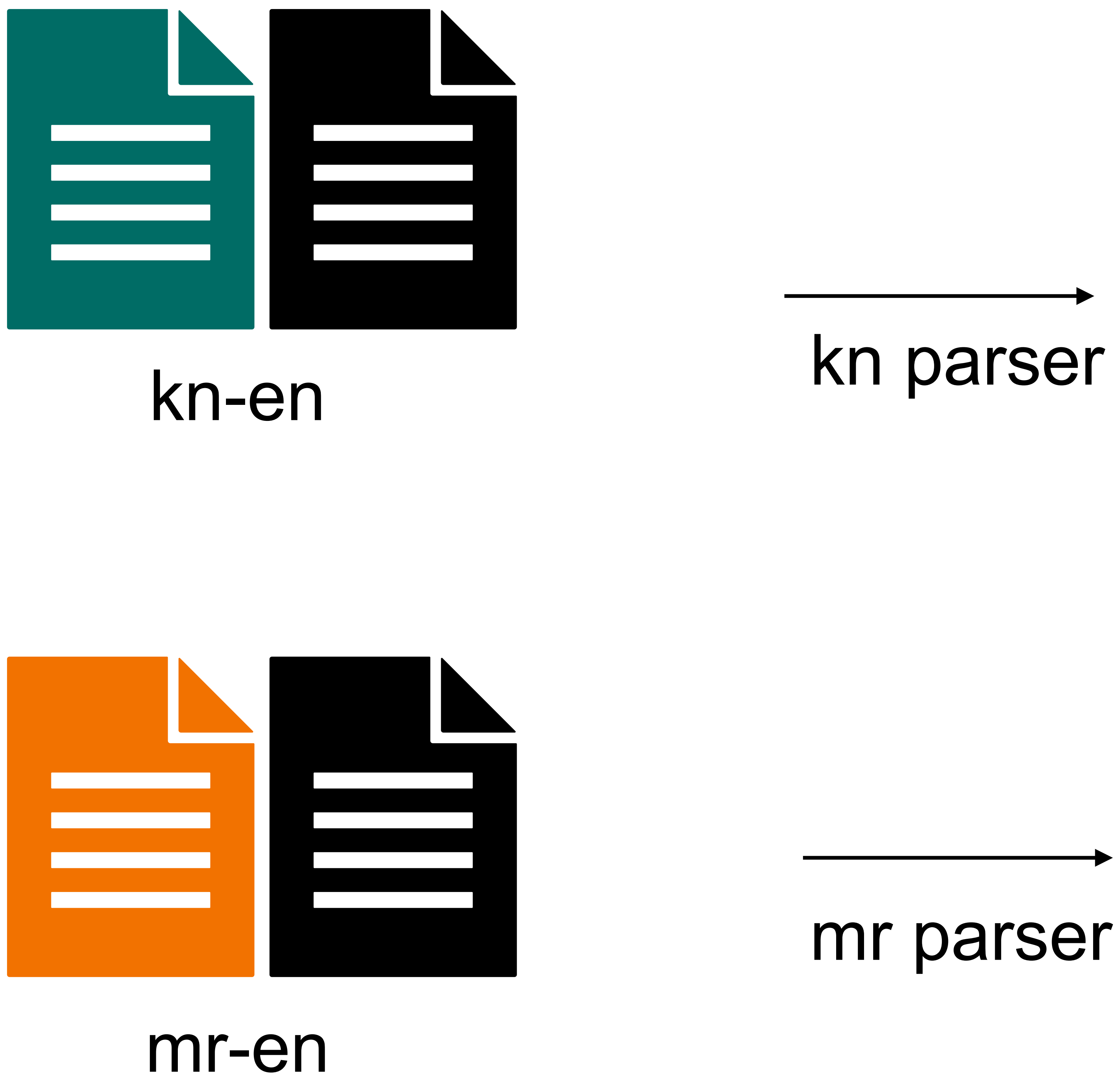
AutoLEX: Extracting Language Descriptions

kn-en

mr-en

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions



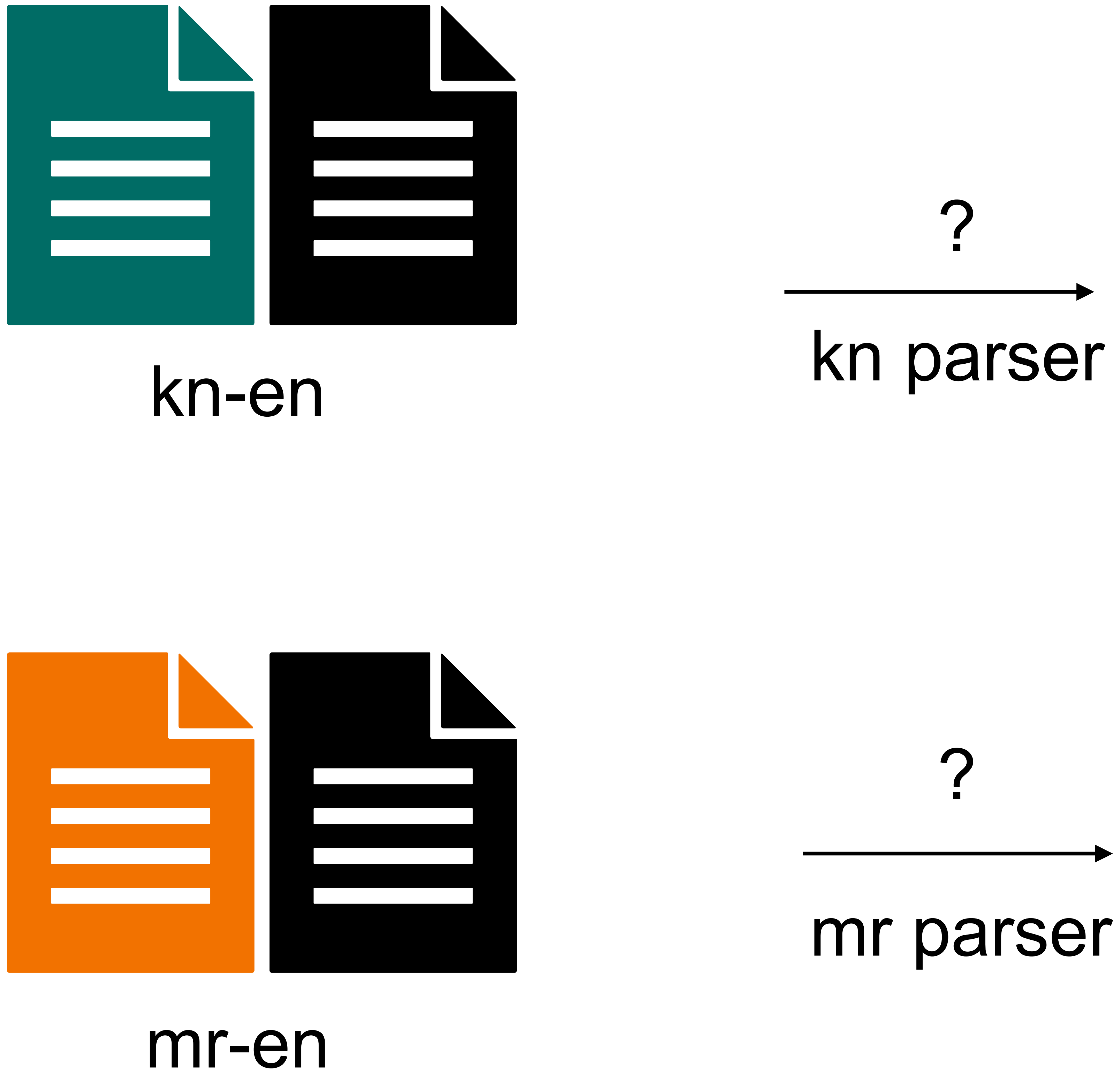
kn-en

kn parser

mr-en

mr parser

AutoLEX: Extracting Language Descriptions



kn-en

?

kn parser

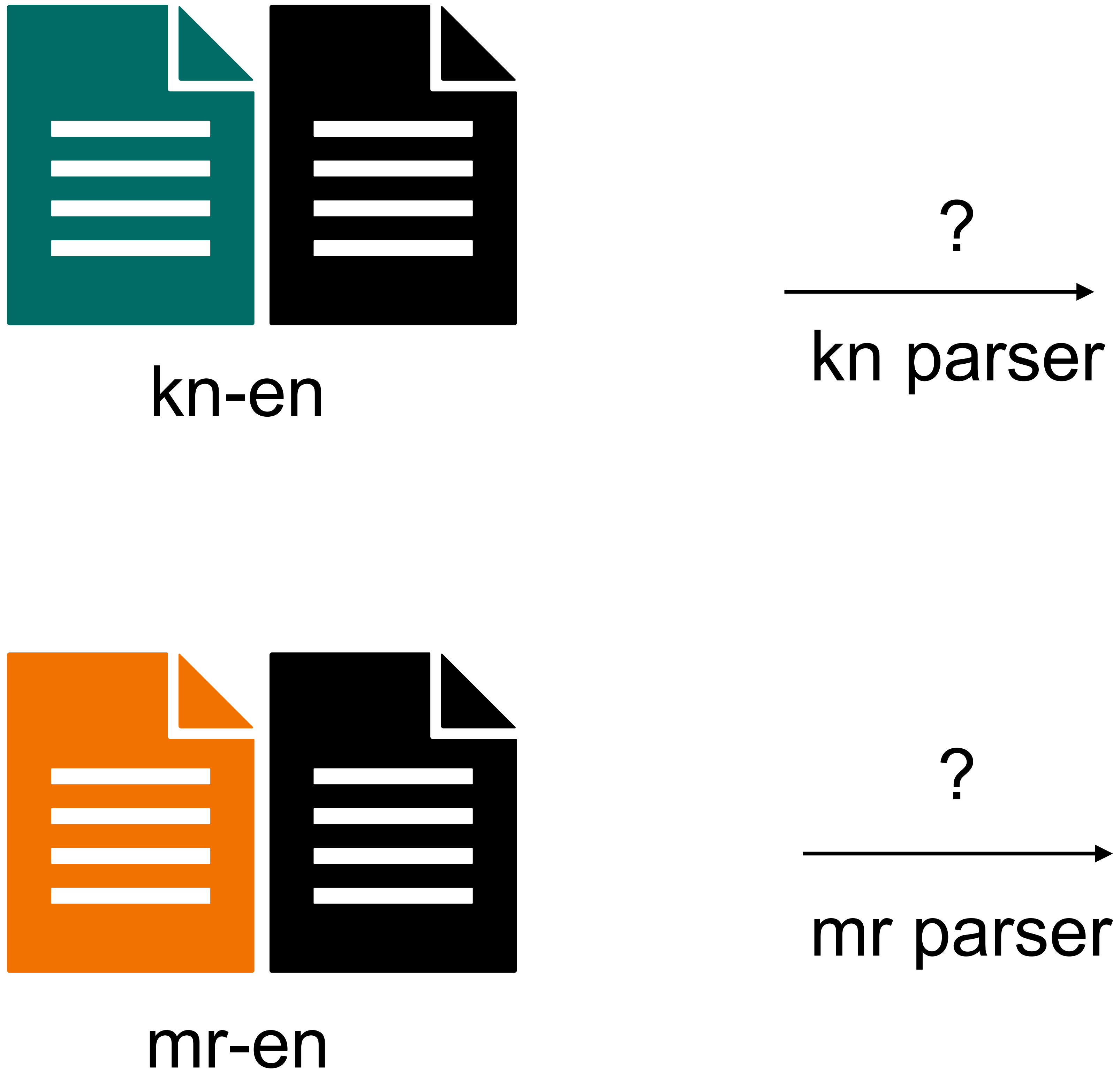
mr-en

?

mr parser

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions



kn has **NO SUD treebank/model**

😕

mr has **VERY SMALL** and **NOISY SUD treebank/model**

AutoLEX: Extracting Language Descriptions

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions

There is **ALWAYS** some data to be found! ☺️

AutoLEX: Extracting Language Descriptions

There is **ALWAYS** some data to be found! ☺️

Paninian Treebanks

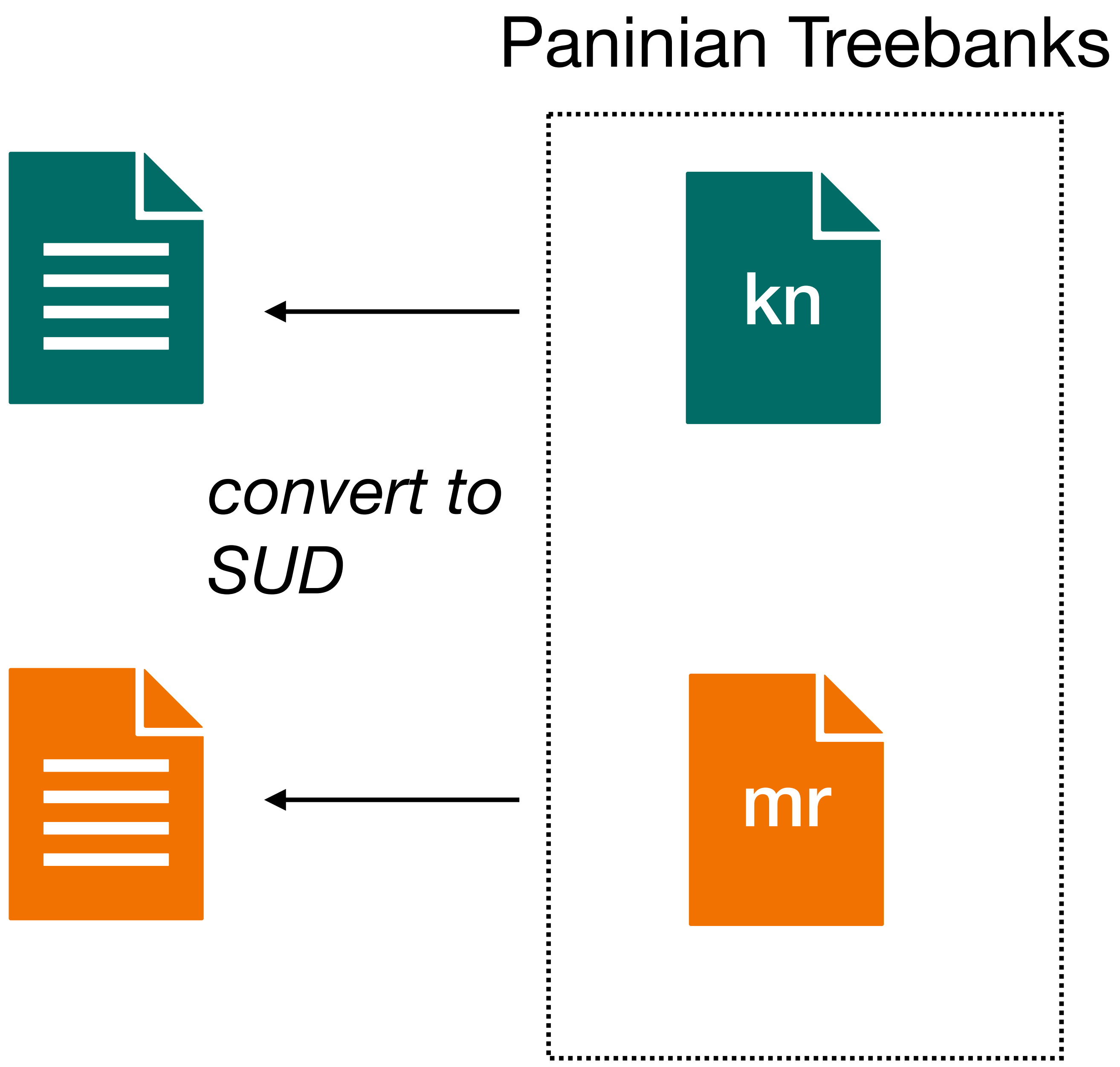
kn

mr

INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY
HYDERABAD

AutoLEX: Extracting Language Descriptions

There is **ALWAYS** some data to be found! ☺️



Paninian Treebanks

kn

mr

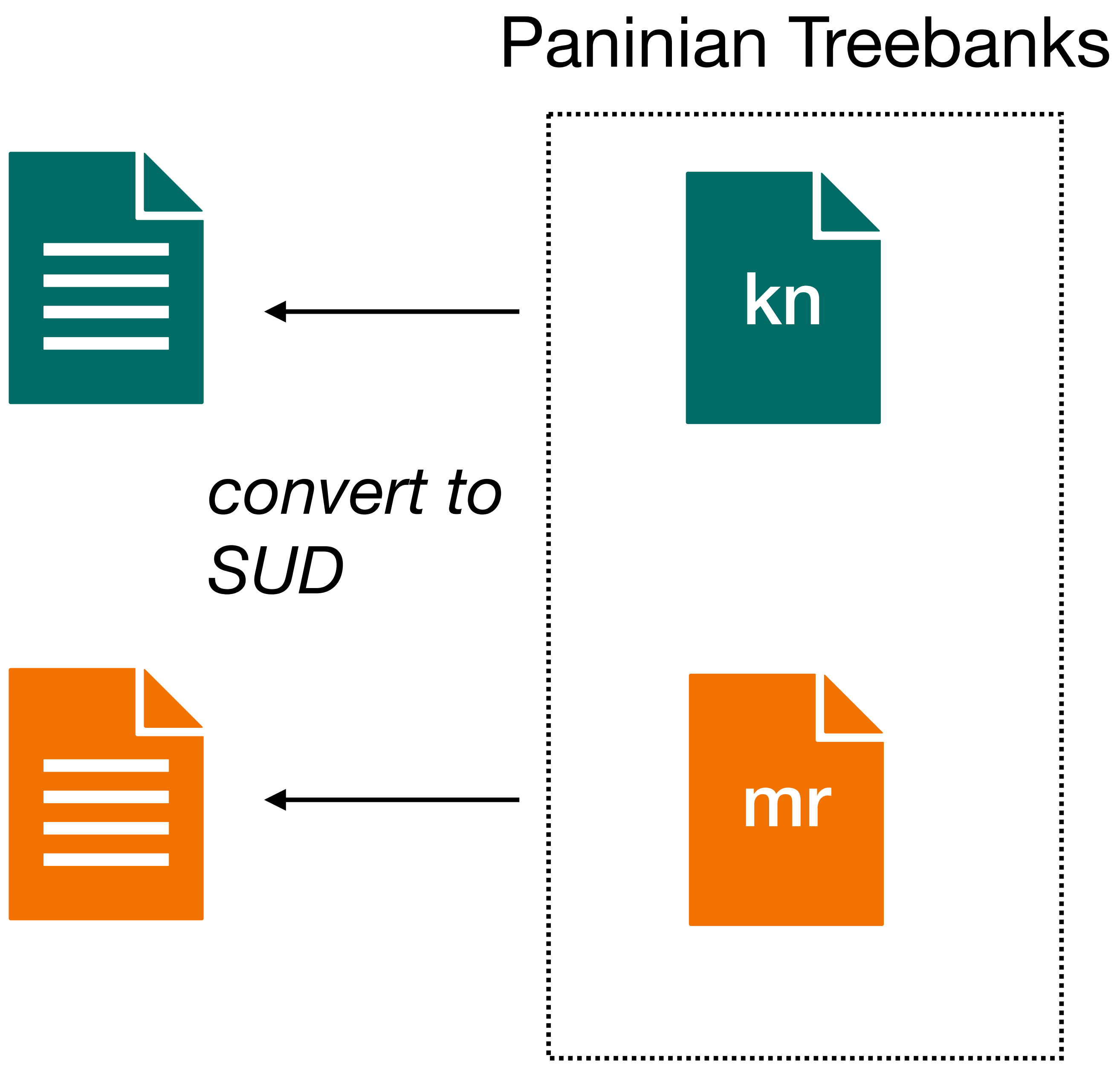
convert to SUD

INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY
HYDERABAD

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions

There is **ALWAYS** some data to be found! ☺️



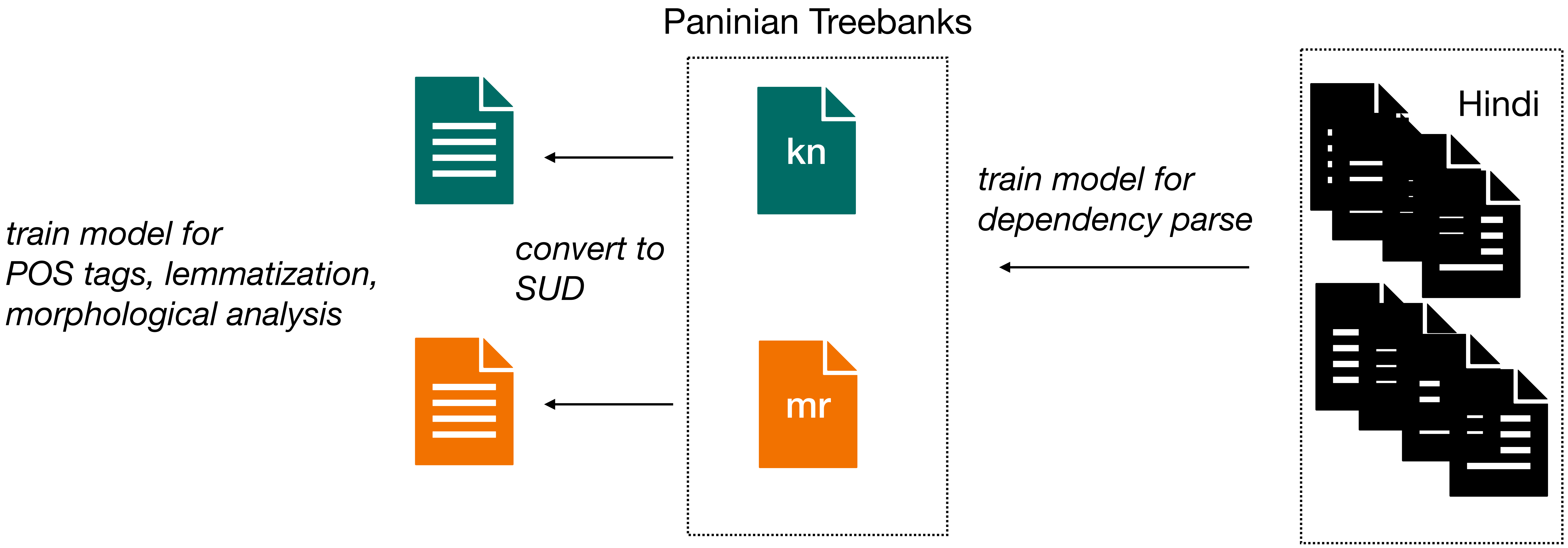
convert to SUD

train model for POS tags, lemmatization, morphological analysis

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions

There is **ALWAYS** some data to be found!



train model for POS tags, lemmatization, morphological analysis

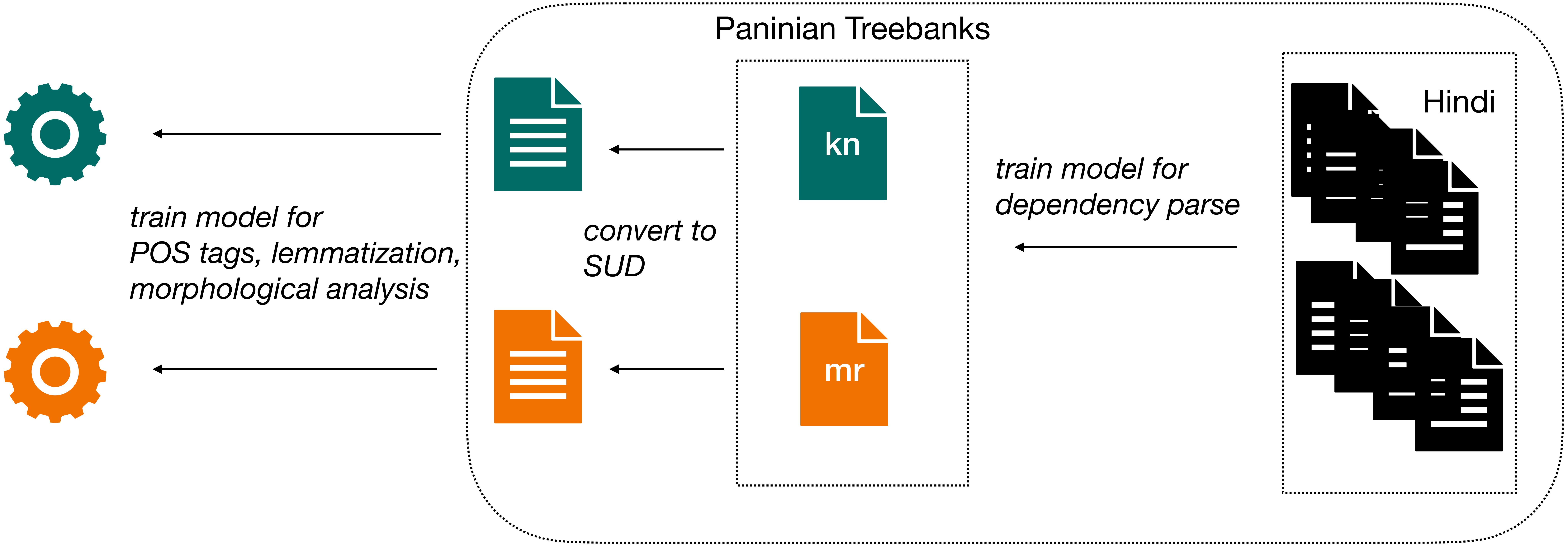
train model for dependency parse

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions

There is **ALWAYS** some data to be found! ☺️



Paninian Treebanks

Hindi

train model for dependency parse

kn

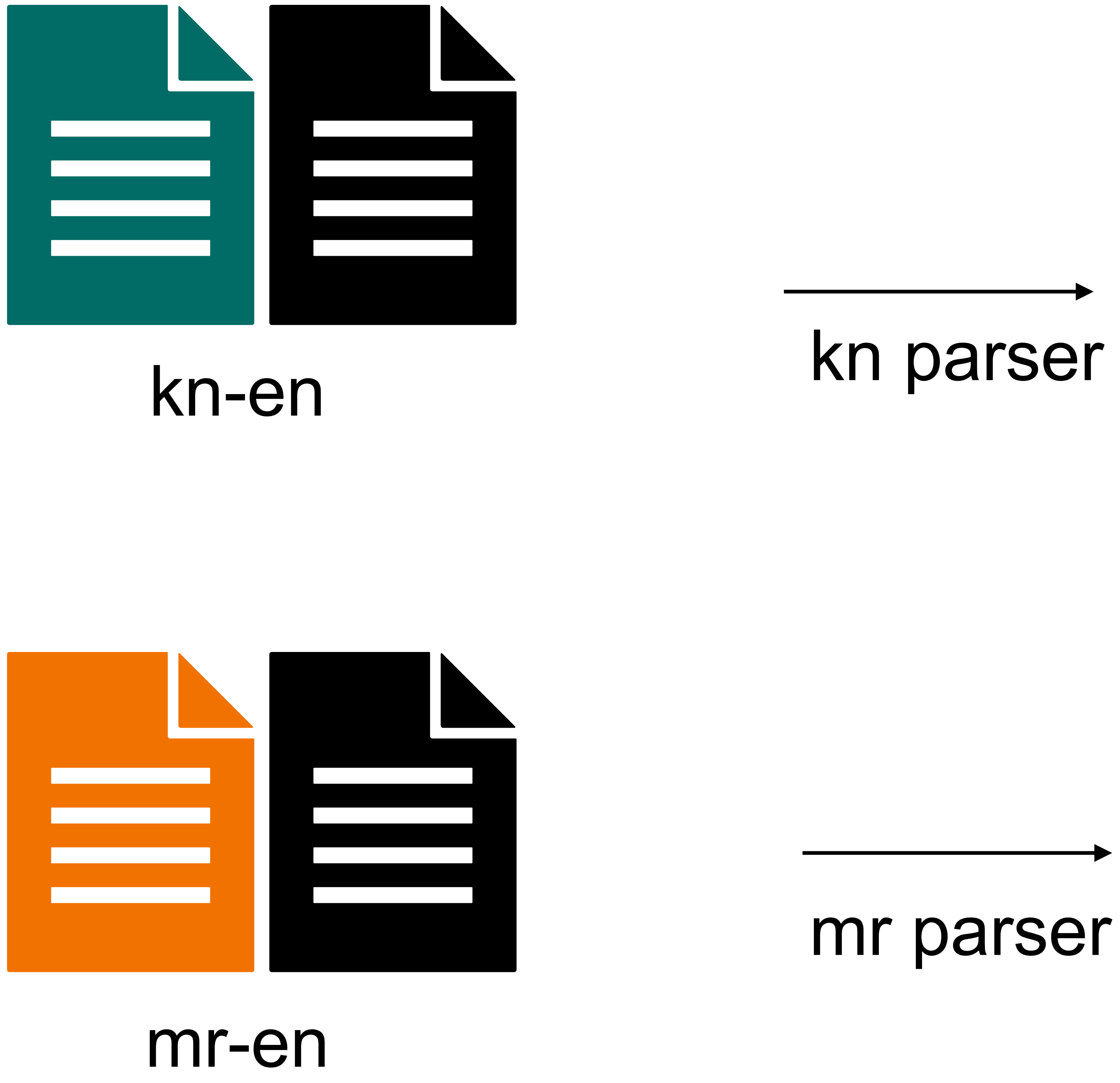
mr

convert to SUD

train model for POS tags, lemmatization, morphological analysis

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions



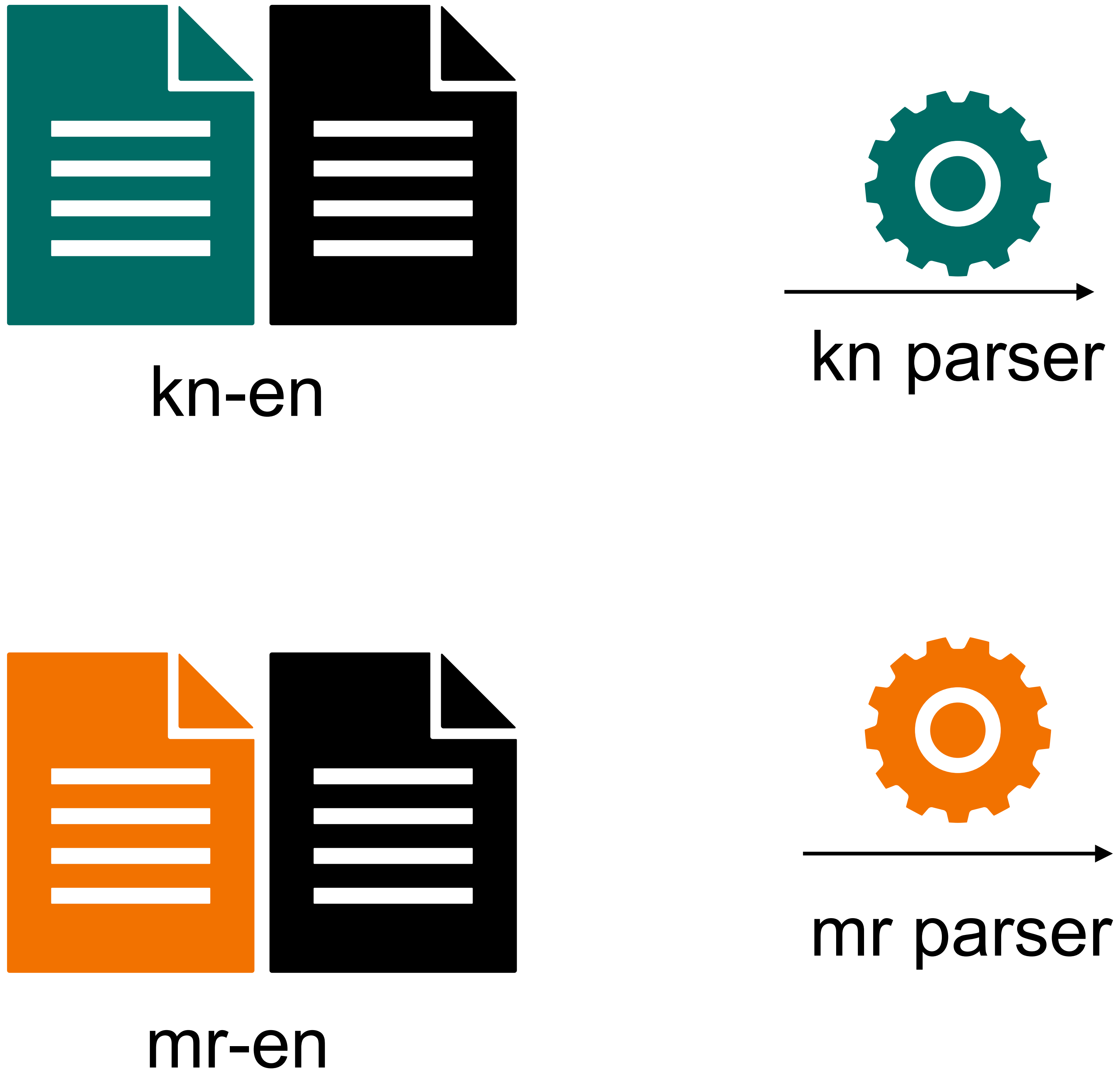
kn-en

kn parser

mr-en

mr parser

AutoLEX: Extracting Language Descriptions



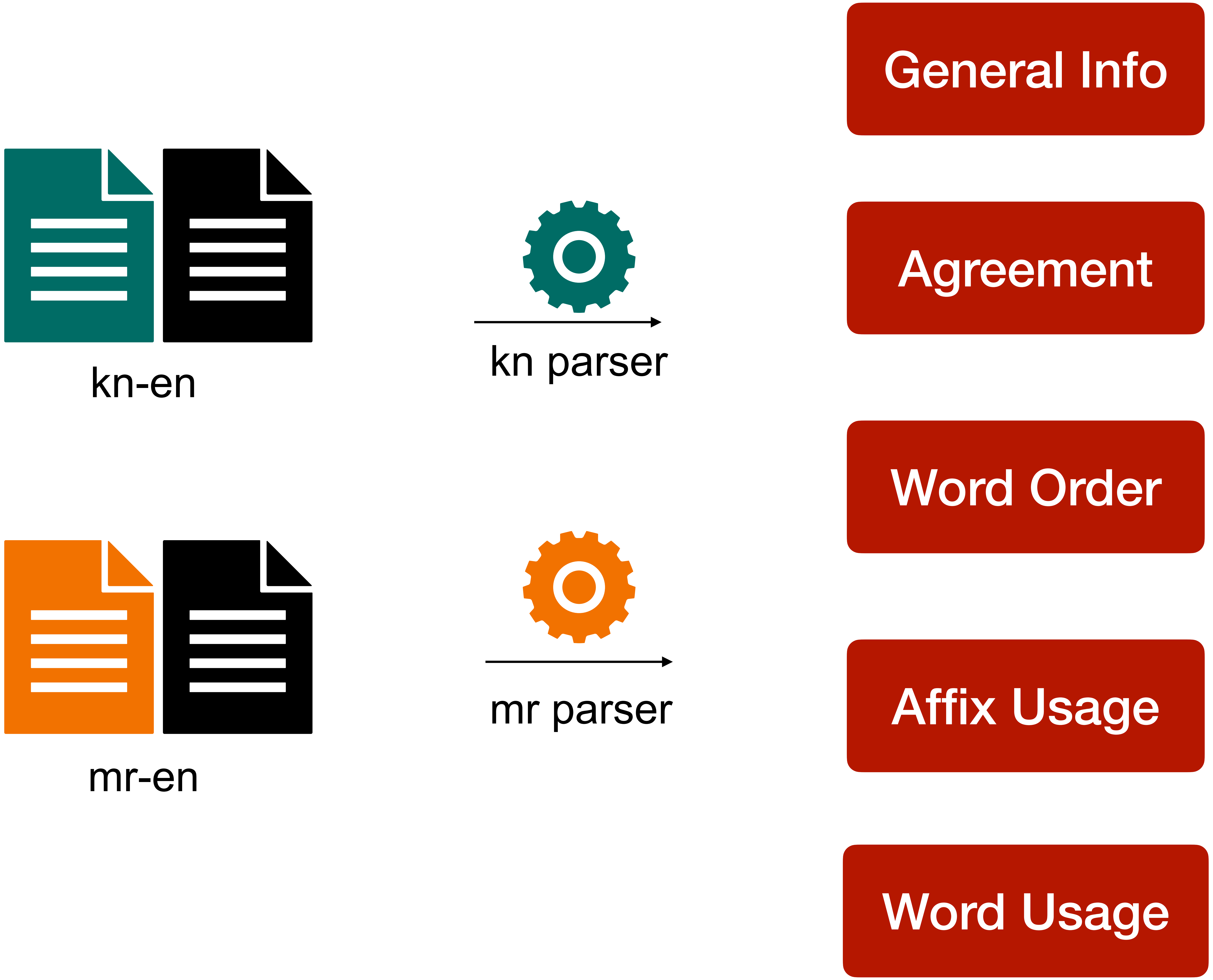
kn-en

kn parser

mr-en

mr parser

AutoLEX: Extracting Language Descriptions



kn-en

kn parser

mr-en

mr parser

General Info

Agreement

Word Order

Affix Usage

Word Usage

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Extracting Language Descriptions

kn-en

kn parser

mr-en

mr parser

General Info

Agreement

Word Order

Affix Usage

Word Usage

Linguistic Question

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Quality Study

General Info

Agreement

Word Order

Affix Usage

Word Usage

AutoLEX: Quality Study

Affix Usage

AutoLEX: Quality Study

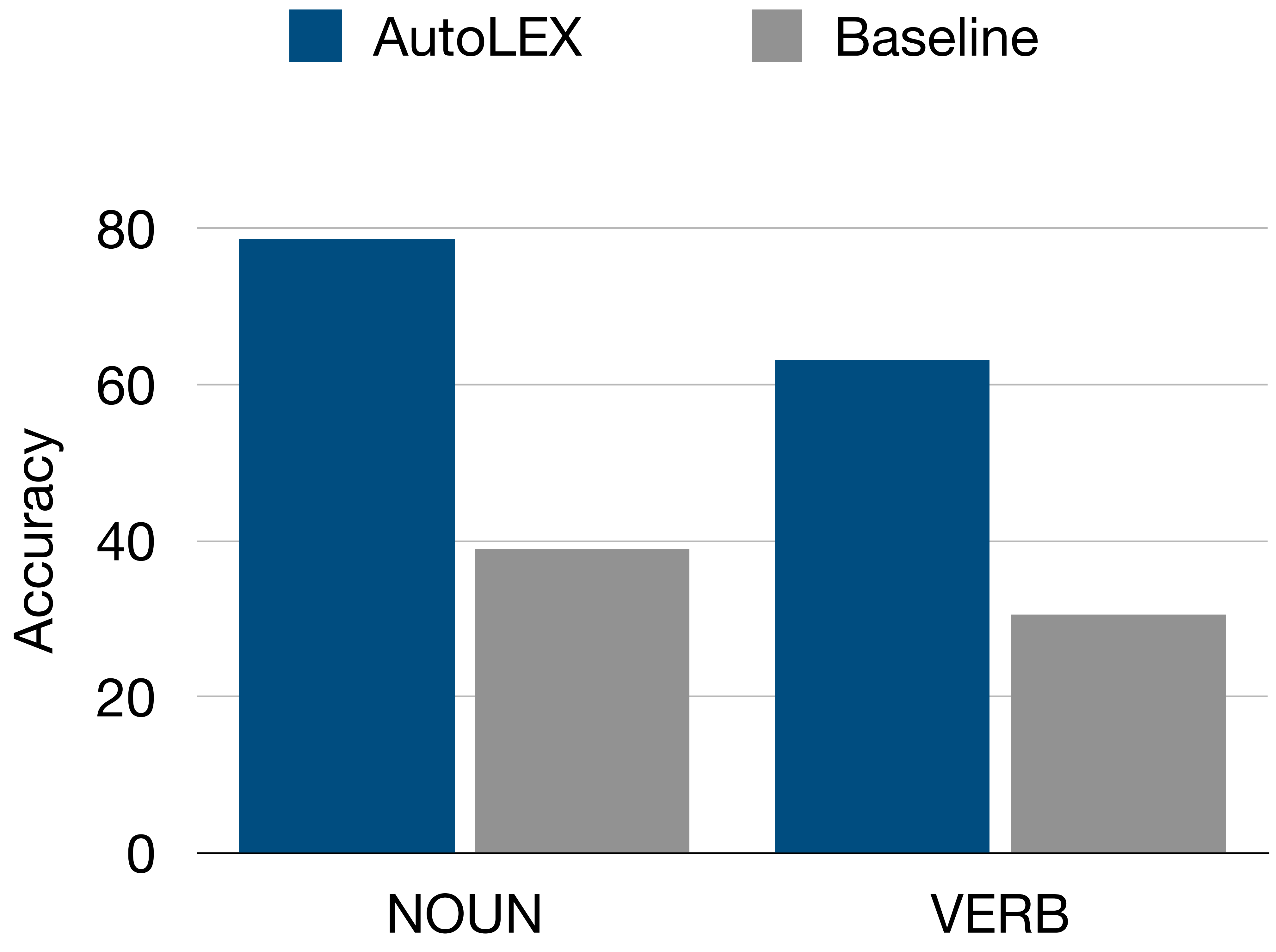
What are the common suffixes for **Kannada nouns**
and when is each used?

Affix Usage

AutoLEX: Quality Study

What are the common suffixes for **Kannada nouns** and when is each used?

Affix Usage



Automated Evaluation

AutoLEX: Quality Study

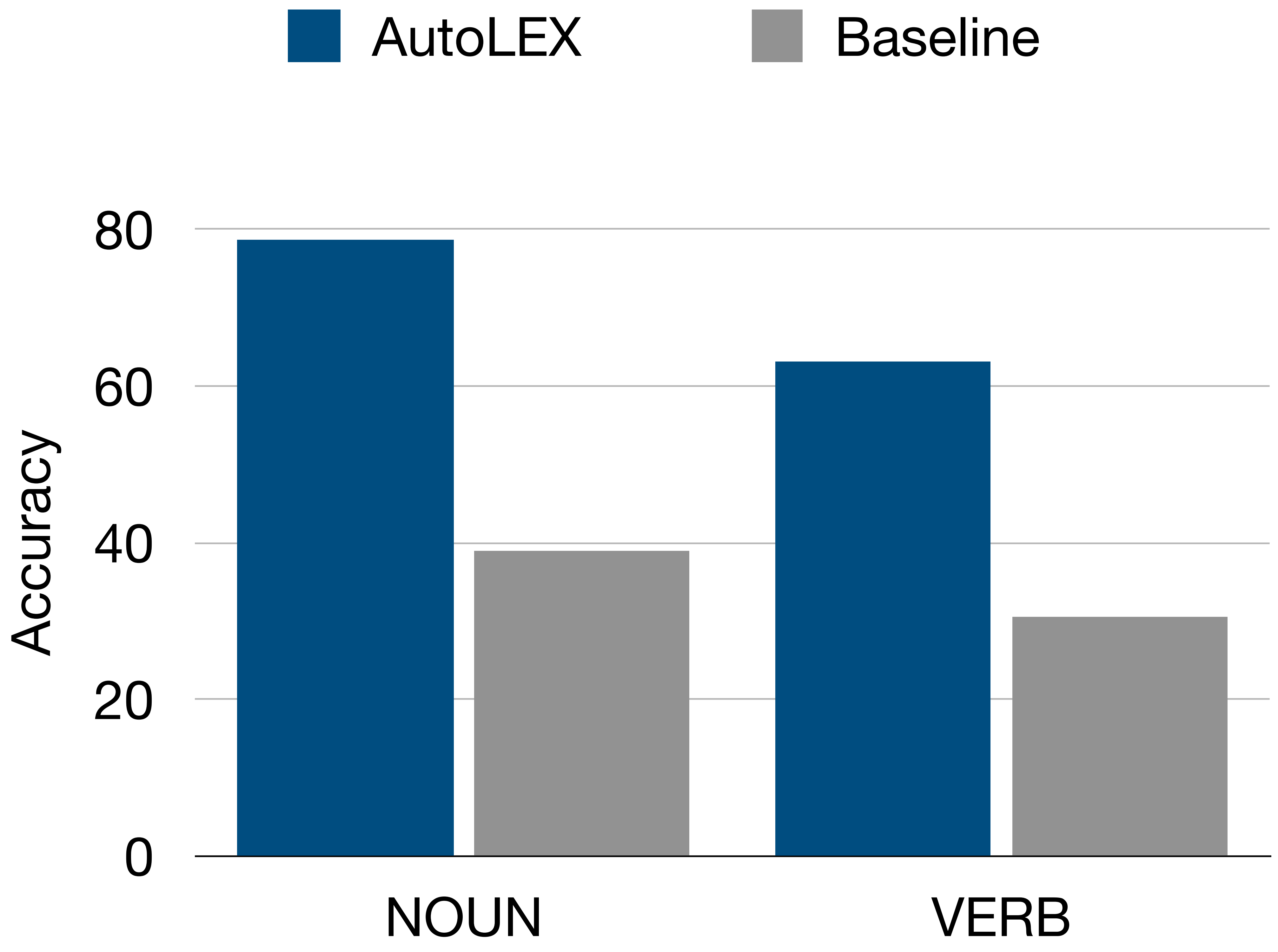
What are the common suffixes for Kannada nouns
and when is each used?

Affix Usage

NOUN: **9*/18 valid**

VERB: **7/13 valid**

Expert Evaluation



Automated Evaluation

AutoLEX: Quality Study

What are the common suffixes for Kannada nouns and when is each used?

Affix Usage

NOUN: **9*/18 valid**

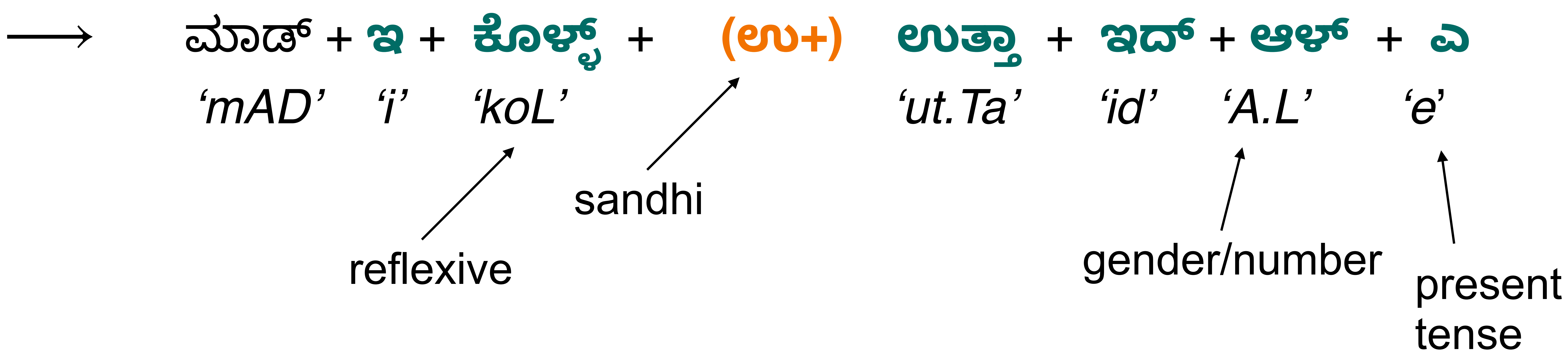
VERB: **7/13 valid**

Expert Evaluation

AutoLEX: Quality Study

What are the common suffixes for **Kannada nouns** and when is each used?

Affix Usage



ಮಾಡಿಕೊಳ್ಳುತ್ತಿದ್ದಾಳೆ ⟶ ಮಾಡ್ + ಇ + ಕೊಳ್ಳ್ + (ಉ+) ಉತ್ತಾ + ಇದ್ + ಆಳ್ + ಎ

‘she’s doing it for herself’ *‘mAD’ ‘i’ ‘koL’ ‘ut.Ta’ ‘id’ ‘A.L’ ‘e’*

AutoLEX: Teacher Perception Study

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

- Evaluate the **relevance**, **utility**, and **presentation** of the materials

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

- Evaluate the **relevance**, **utility**, and **presentation** of the materials
- Recruited **12 Kannada teachers** and **5 Marathi teachers**

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

- Evaluate the **relevance**, **utility**, and **presentation** of the materials
- Recruited **12 Kannada teachers** and **5 Marathi teachers**

Introduce **AutoLEX**
to teachers

AutoLEX: Teacher Perception Study

- Evaluate the **relevance**, **utility**, and **presentation** of the materials
- Recruited **12 Kannada teachers** and **5 Marathi teachers**

☎ →

Introduce **AutoLEX**
to teachers

🔍 →

1-2 weeks for
exploration of materials

AutoLEX: Teacher Perception Study

- Evaluate the **relevance**, **utility**, and **presentation** of the materials
- Recruited **12 Kannada teachers** and **5 Marathi teachers**

Introduce **AutoLEX** to teachers → 1-2 weeks for exploration of materials → Fill the questionnaire

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

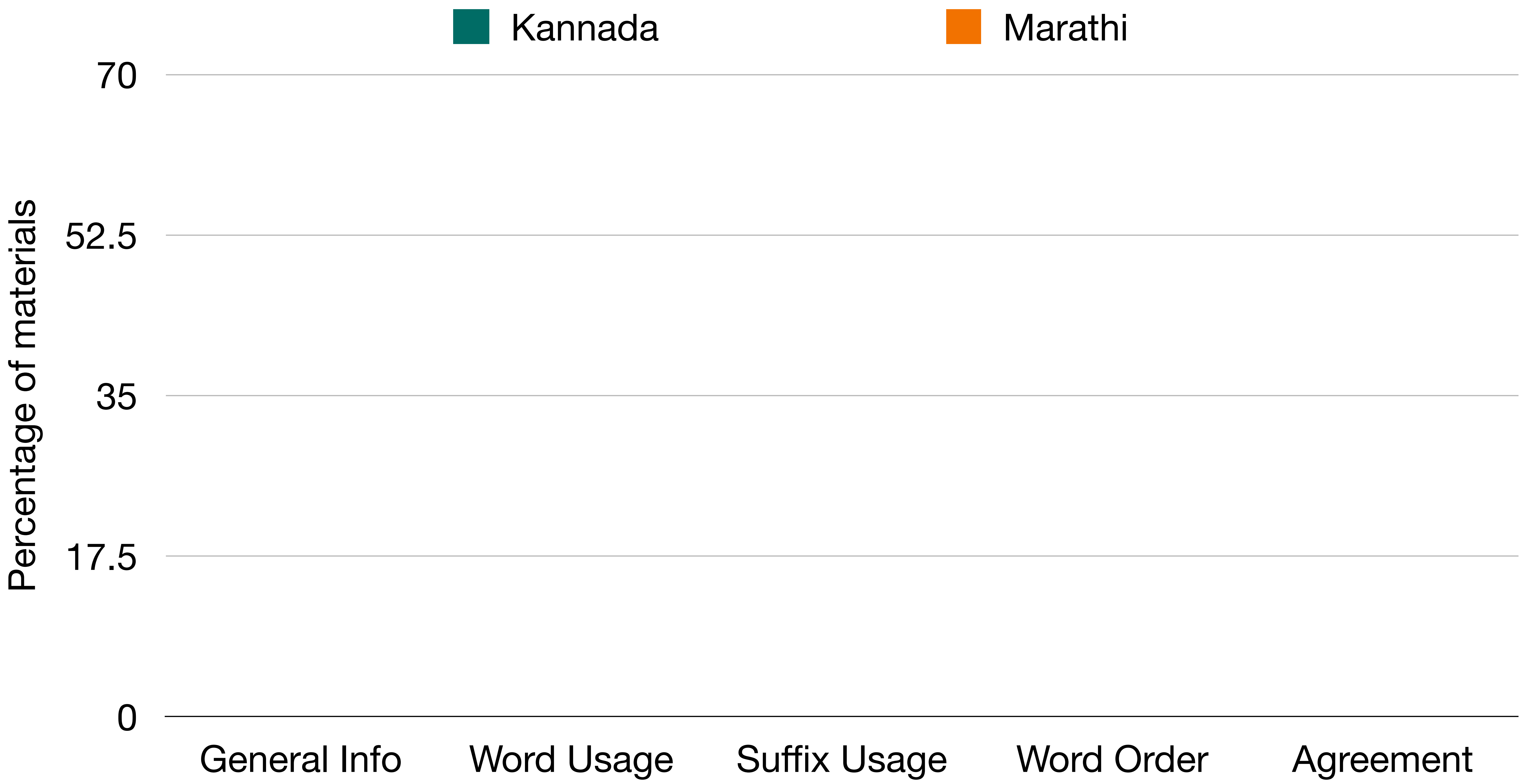
AutoLEX: Teacher Perception Study

- How **relevant** are the extracted materials to the teaching needs?

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
***Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* **In Submission.**

AutoLEX: Teacher Perception Study

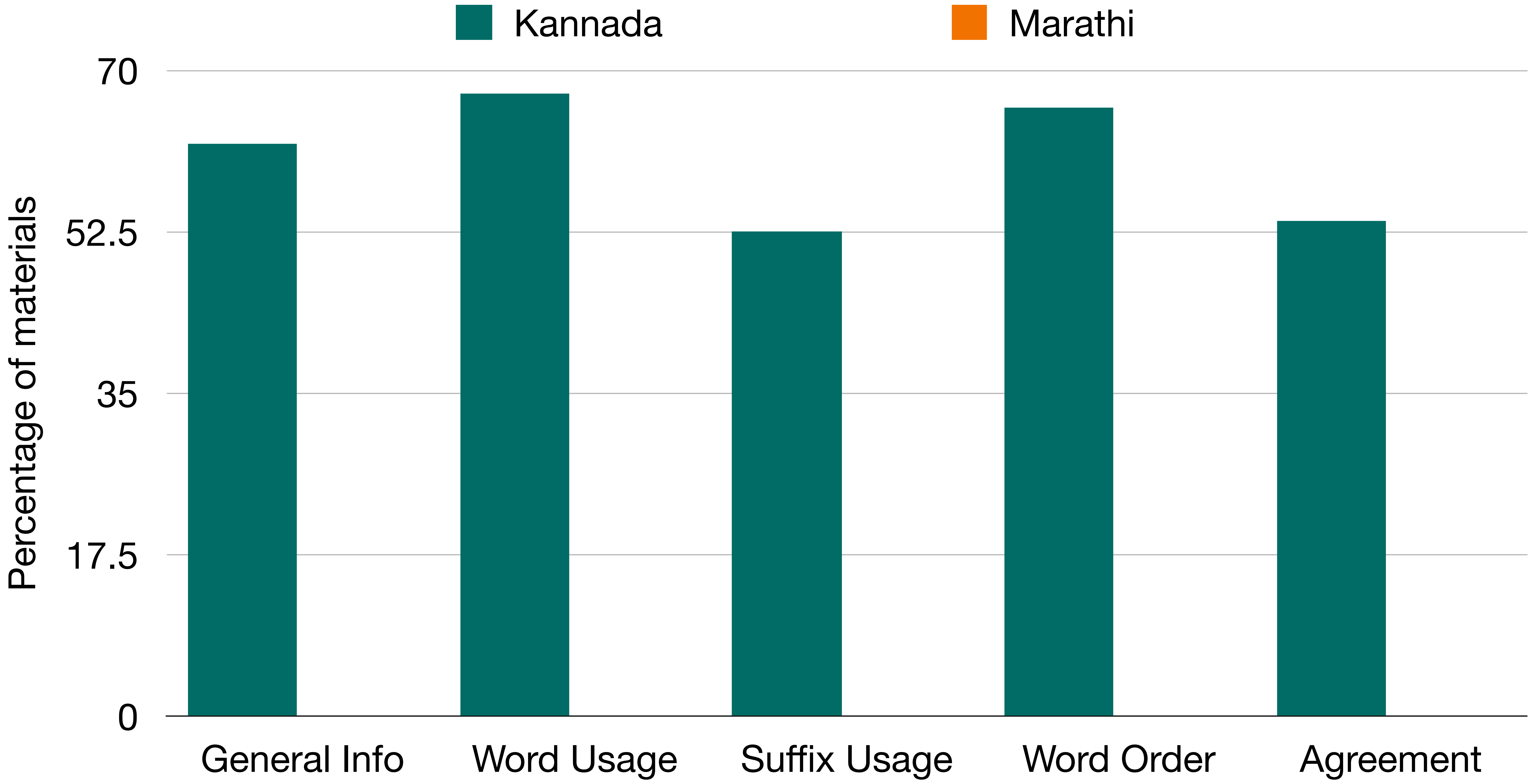
- How **relevant** are the extracted materials to the teaching needs?



Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

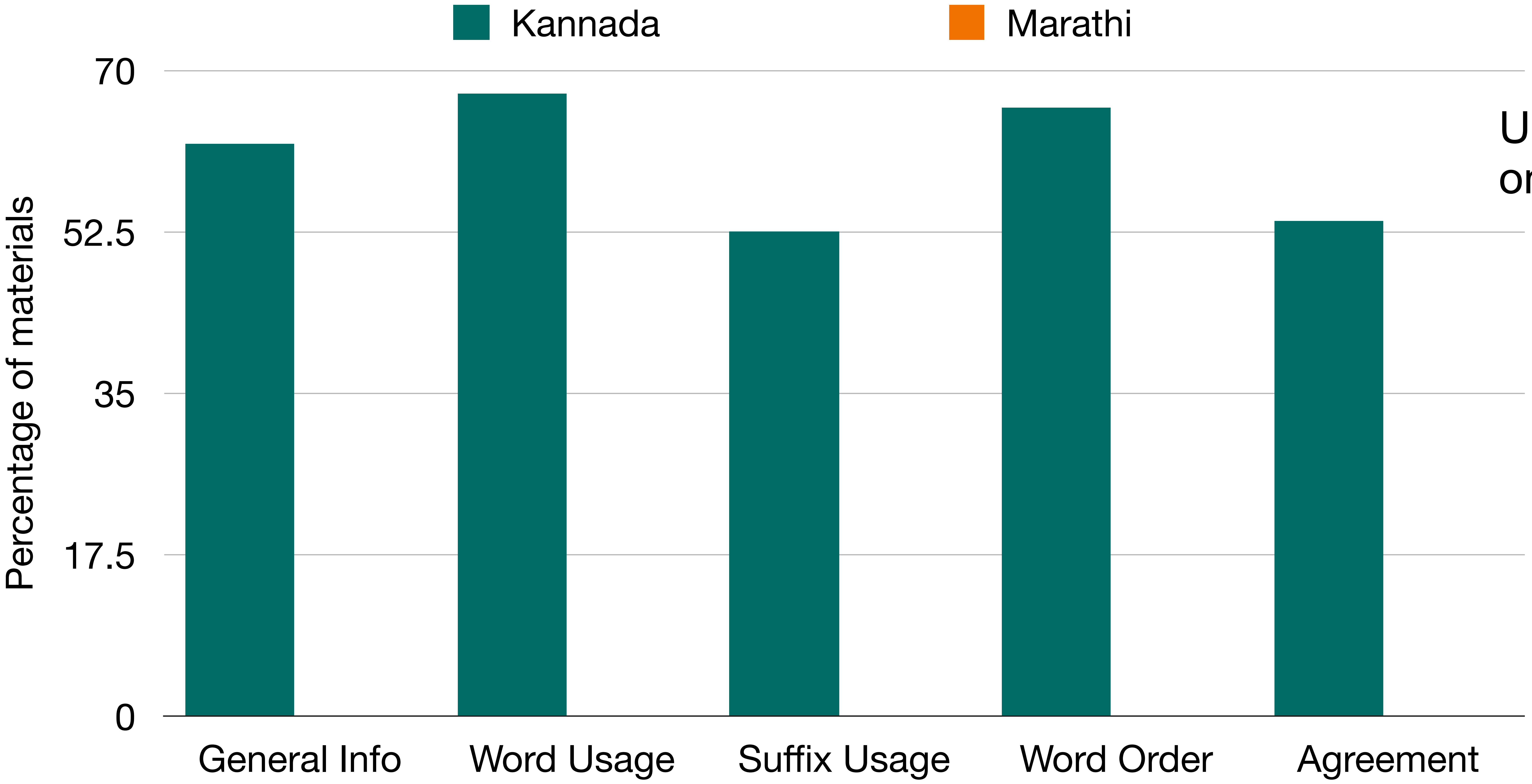
- How **relevant** are the extracted materials to the teaching needs?



Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

- How **relevant** are the extracted materials to the teaching needs?

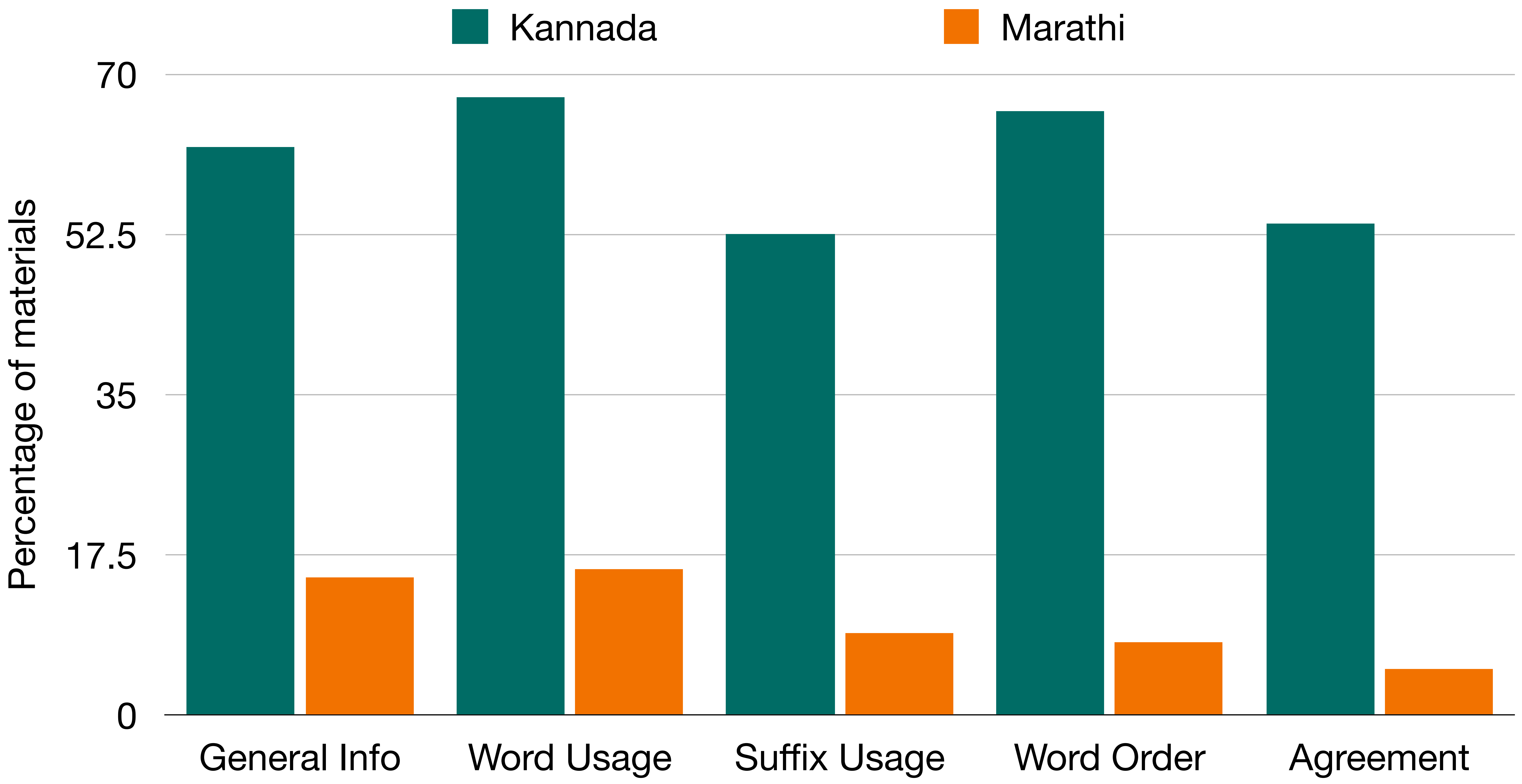


Underlying corpora based
on **formal and traditional usage**

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

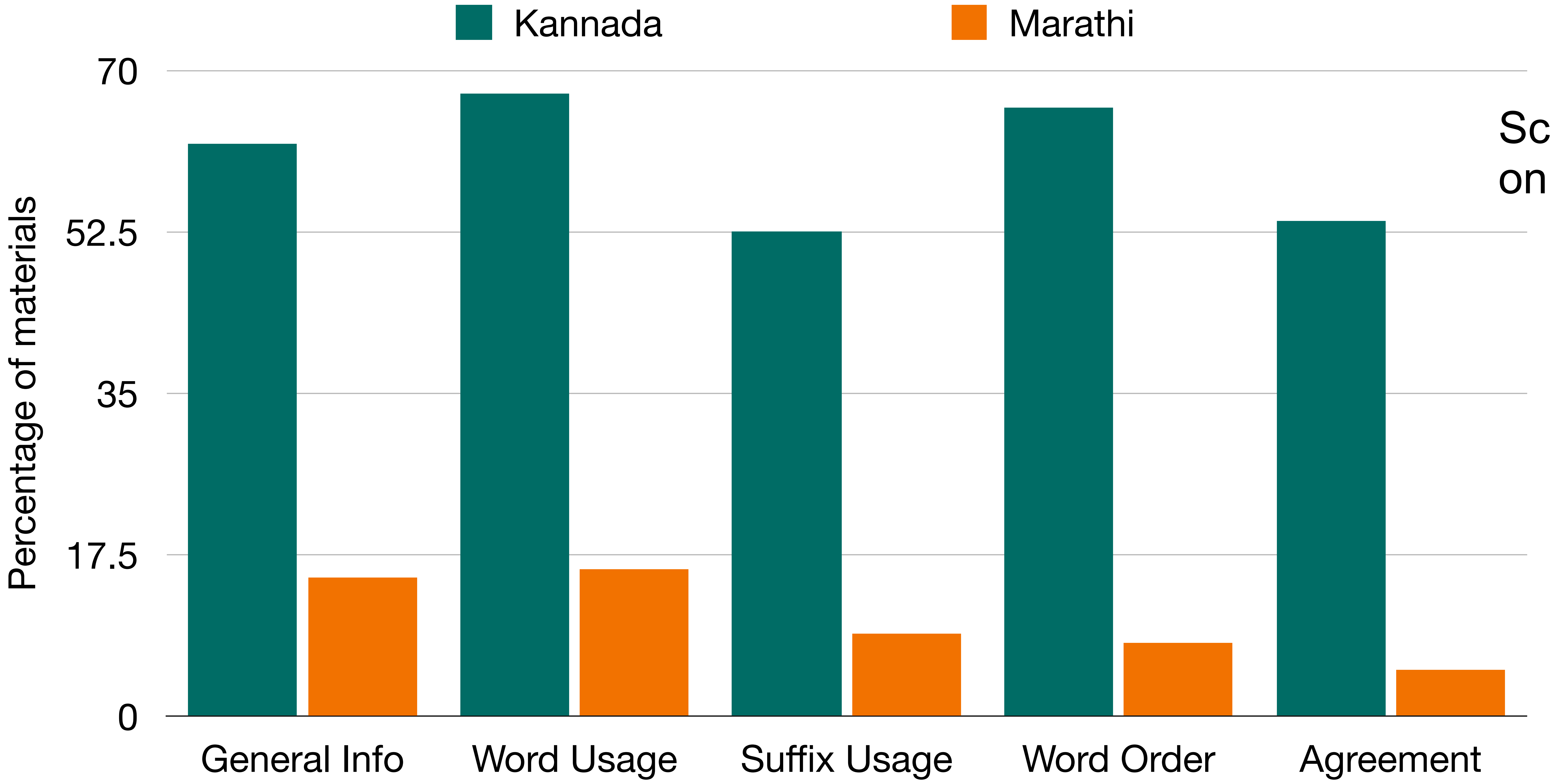
- How **relevant** are the extracted materials to the teaching needs?



Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

- How **relevant** are the extracted materials to the teaching needs?

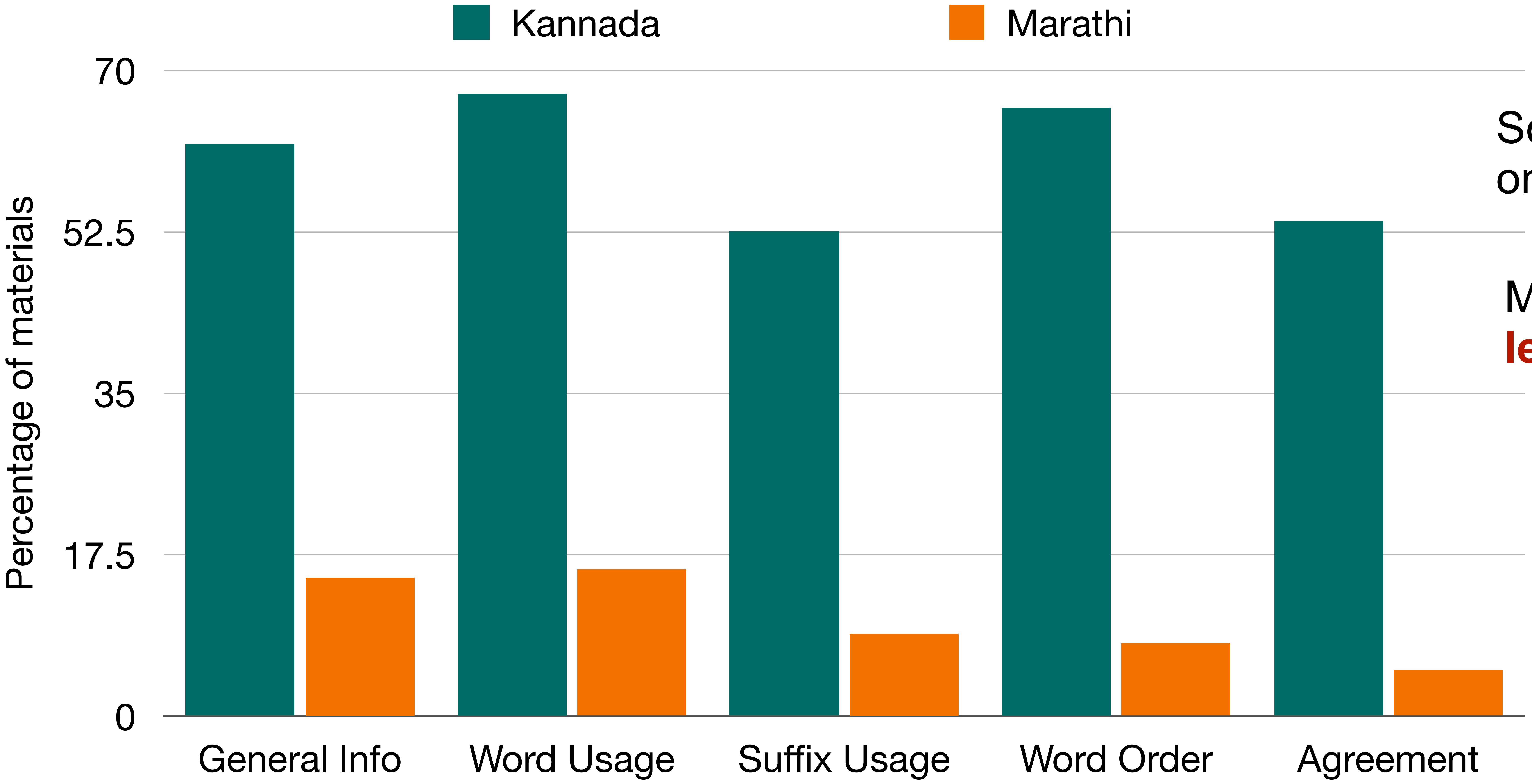


School's primary focus on **beginner-levels**

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary, Sampath, Sheshadri, Anastasopoulos, Neubig. **In Submission.**

AutoLEX: Teacher Perception Study

- How **relevant** are the extracted materials to the teaching needs?



School's primary focus on **beginner-levels**

More beneficial for **advanced learners**

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

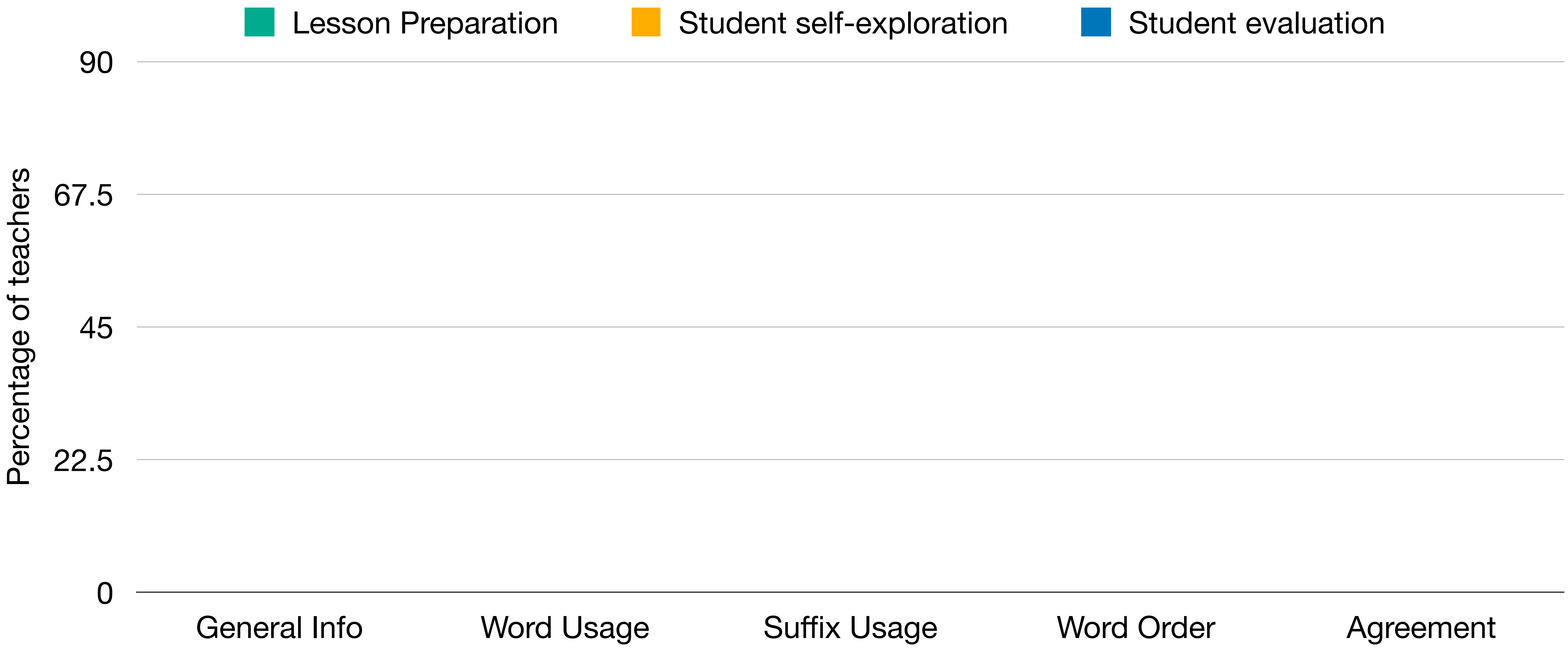
Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

- For what **purposes would the teachers use** the extracted materials for their teaching needs?

AutoLEX: Teacher Perception Study

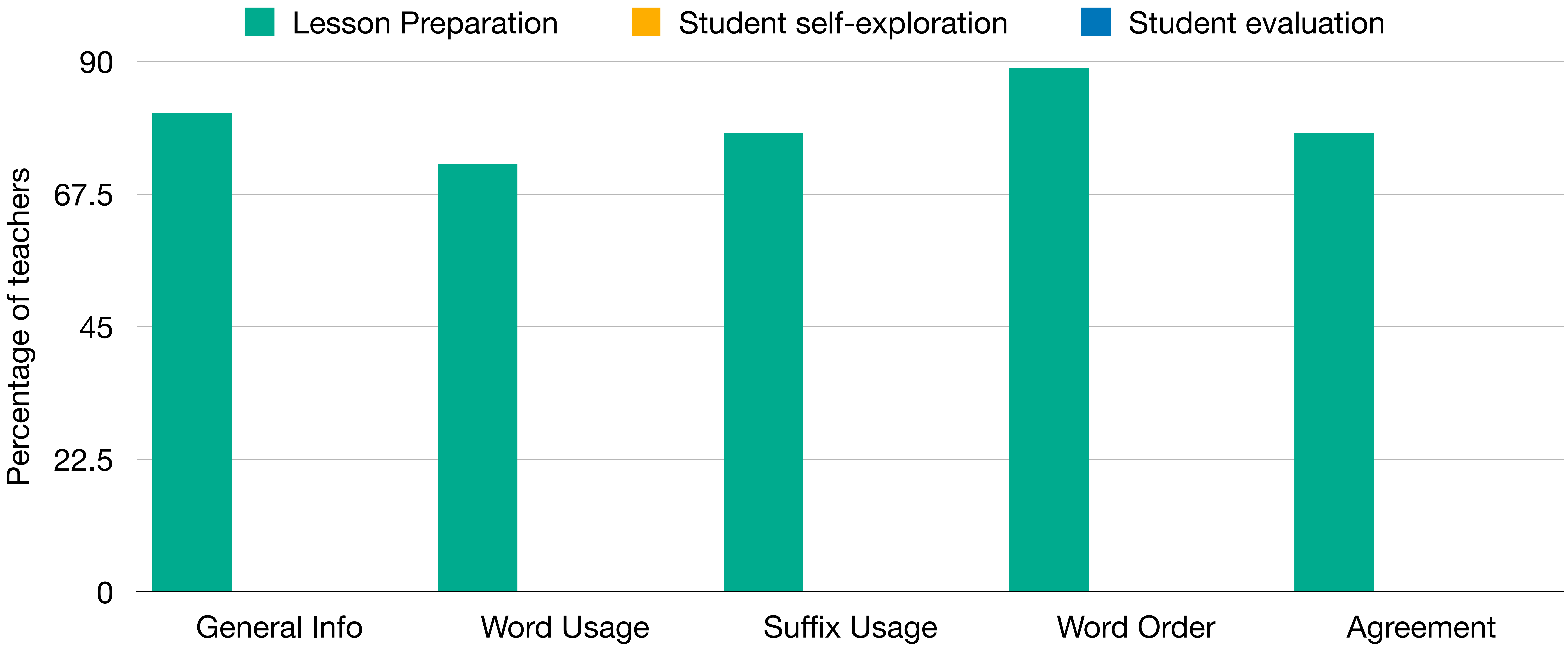
- For what **purposes would the teachers use** the extracted materials for their teaching needs?



Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

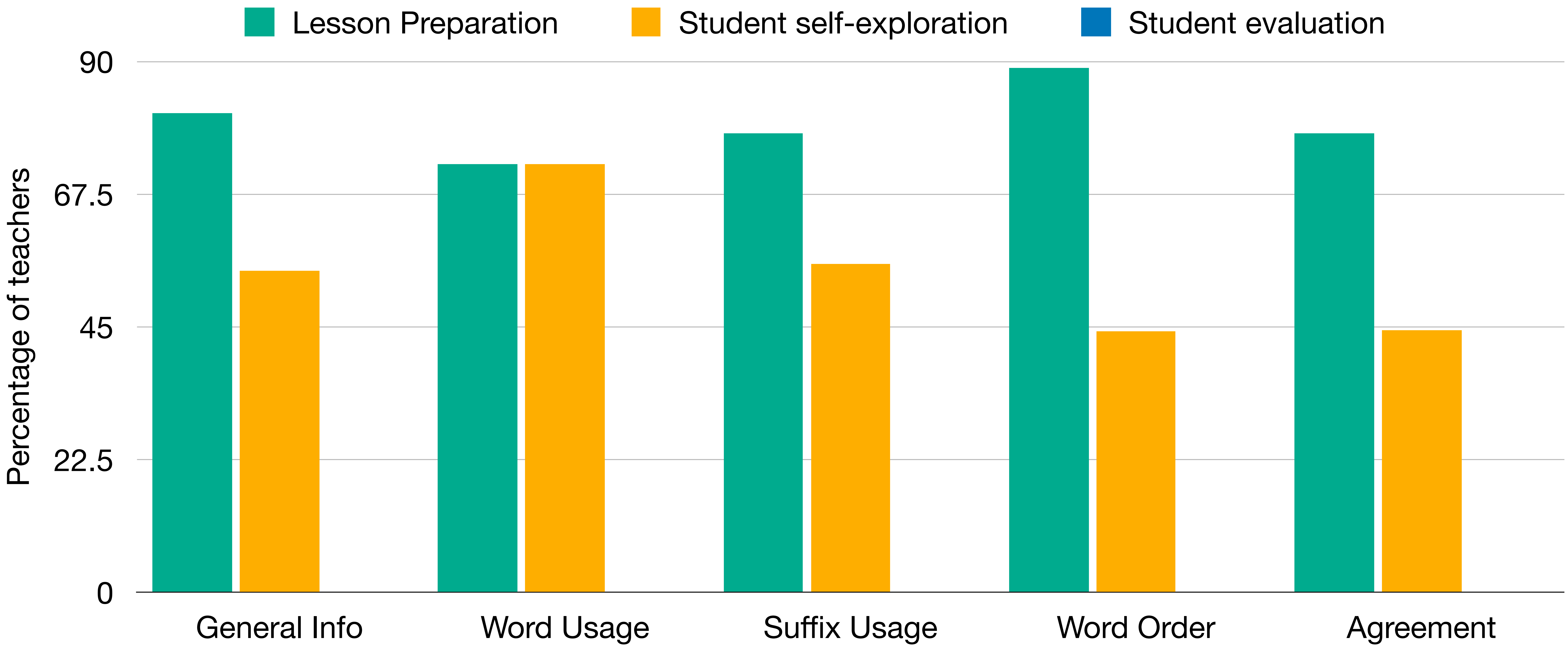
- For what **purposes would the teachers use** the extracted materials for their teaching needs?



Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

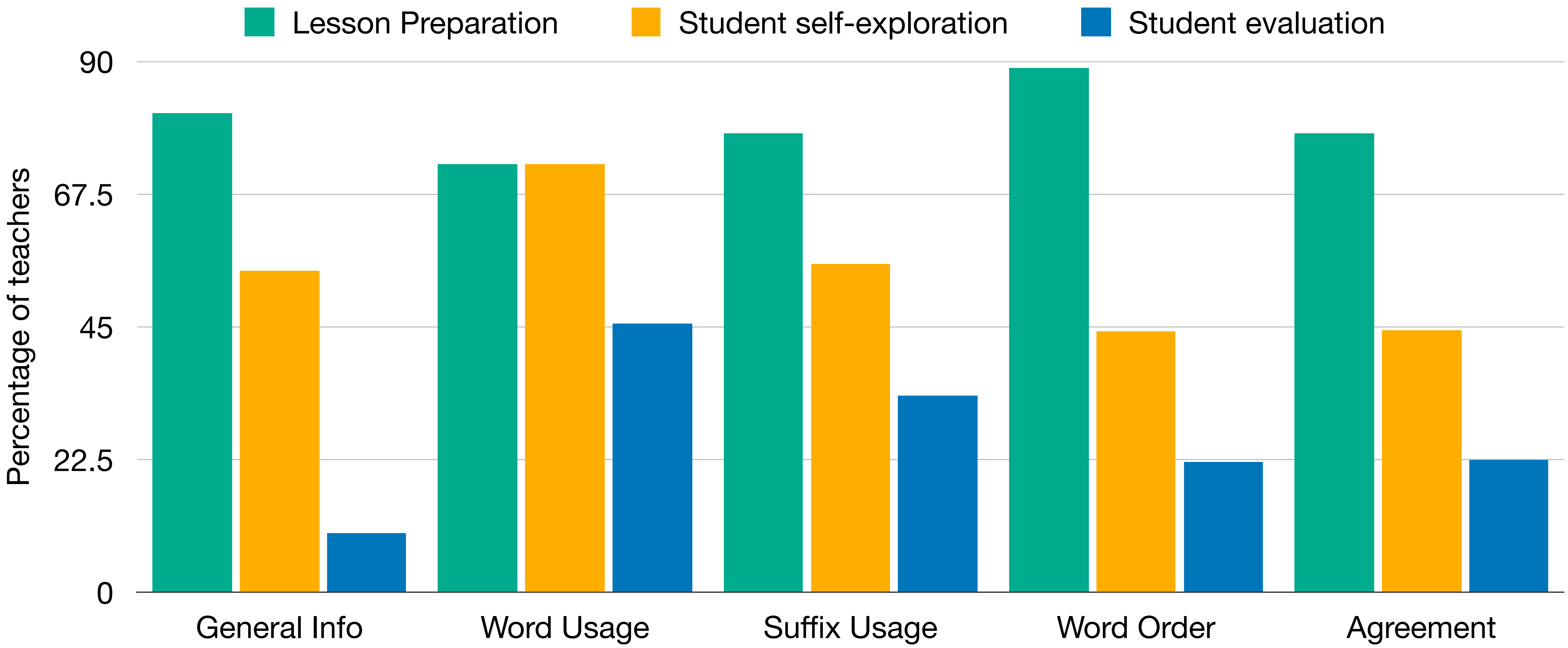
- For what **purposes would the teachers use** the extracted materials for their teaching needs?



Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Perception Study

- For what **purposes would the teachers use** the extracted materials for their teaching needs?



Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning
Chaudhary**, Sampath, Sheshadri, Anastasopoulos, Neubig.* ***In Submission.

AutoLEX: Teacher Testimonials

The illustrative examples, and the grouped synonyms. It does need some work on accuracy in some places, but this is a great start!

I used this tool to teach an American adult who takes private lessons, found this tool helpful in addressing her grammar questions.

given that these word pairs have been extracted from natural text, its interesting to see that there are certain word senses which are so frequently used in the real world which currently we haven't covered in our lesson but are we are now thinking of adding them.

If this tool could be used to target the older kids it would be very helpful. However, the past present and future tenses of the verbs are interesting and this tool managed to impress me with the vast database, Unfortunately, the words used are very technical, and make excellent tool to improve writing skills.

Providing teachers the ability to input a curated set of data (stories written in good and correct language) to prepare relevant examples from may be helpful. Working more collaboratively will help us a lot

Other Applications

Other Applications

Language
Descriptions

Other Applications

Automatic Multilingual Grammar Checker

Language
Descriptions

Other Applications

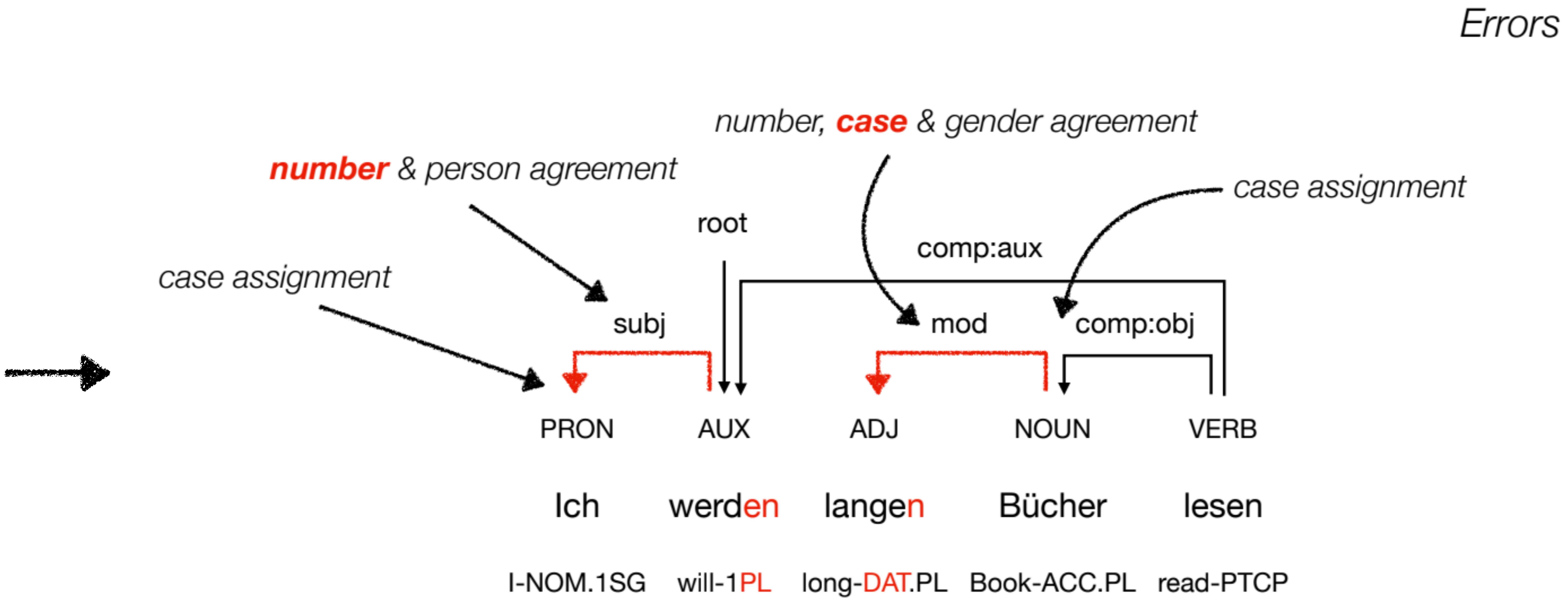
Automatic Multilingual Grammar Checker

Language
Descriptions →

Other Applications

Automatic Multilingual Grammar Checker

Language
Descriptions



Errors

*number, **case** & gender agreement*

***number** & person agreement*

case assignment

root

comp:aux

case assignment

subj

mod

comp:obj

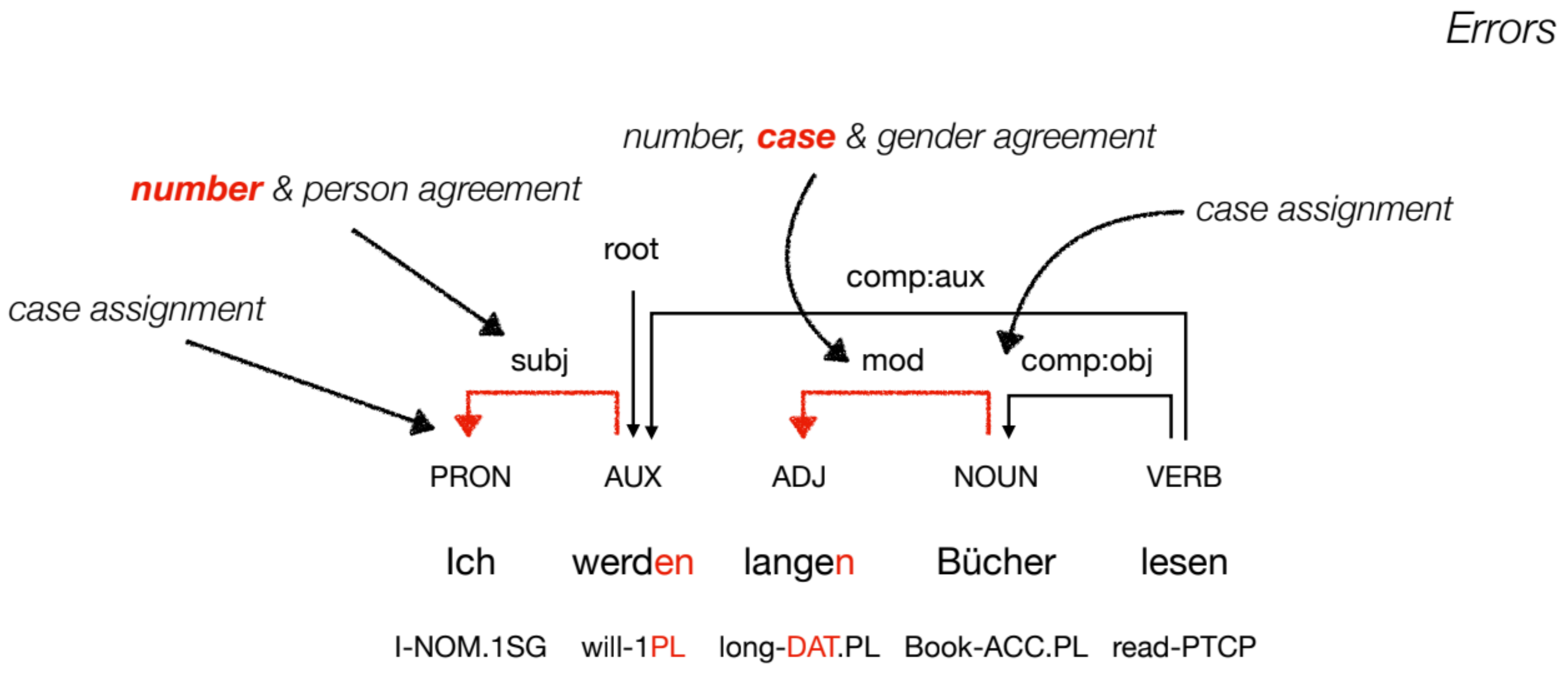
PRON AUX ADJ NOUN VERB

Ich werden langen Bücher lesen

I-NOM.1SG will-1PL long-DAT.PL Book-ACC.PL read-PTCP

Evaluating the Morphosyntactic Well-formedness of Generated Texts
*Pratapa, Anastasopoulos, RIjhwani, **Chaudhary,** Mortensen, Sheikh, Neubig, Tsvetkov.* ***EMNLP 2020***

Other Applications



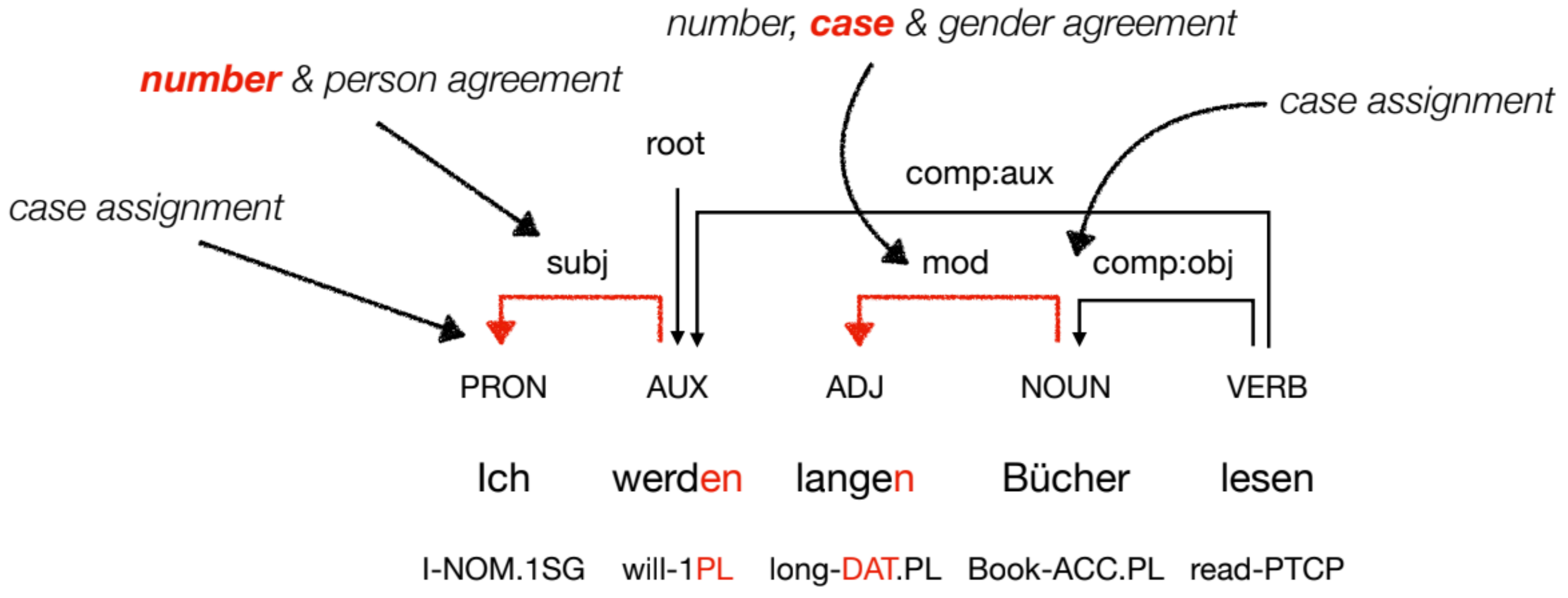
Evaluating the Morphosyntactic Well-formedness of Generated Texts
*Pratapa, Anastasopoulos, RIjhwani, **Chaudhary,** Mortensen, Sheikh, Neubig, Tsvetkov.* ***EMNLP 2020***

Language
Descriptions →

Other Applications

Evaluating Context-Usage in MT models

Errors

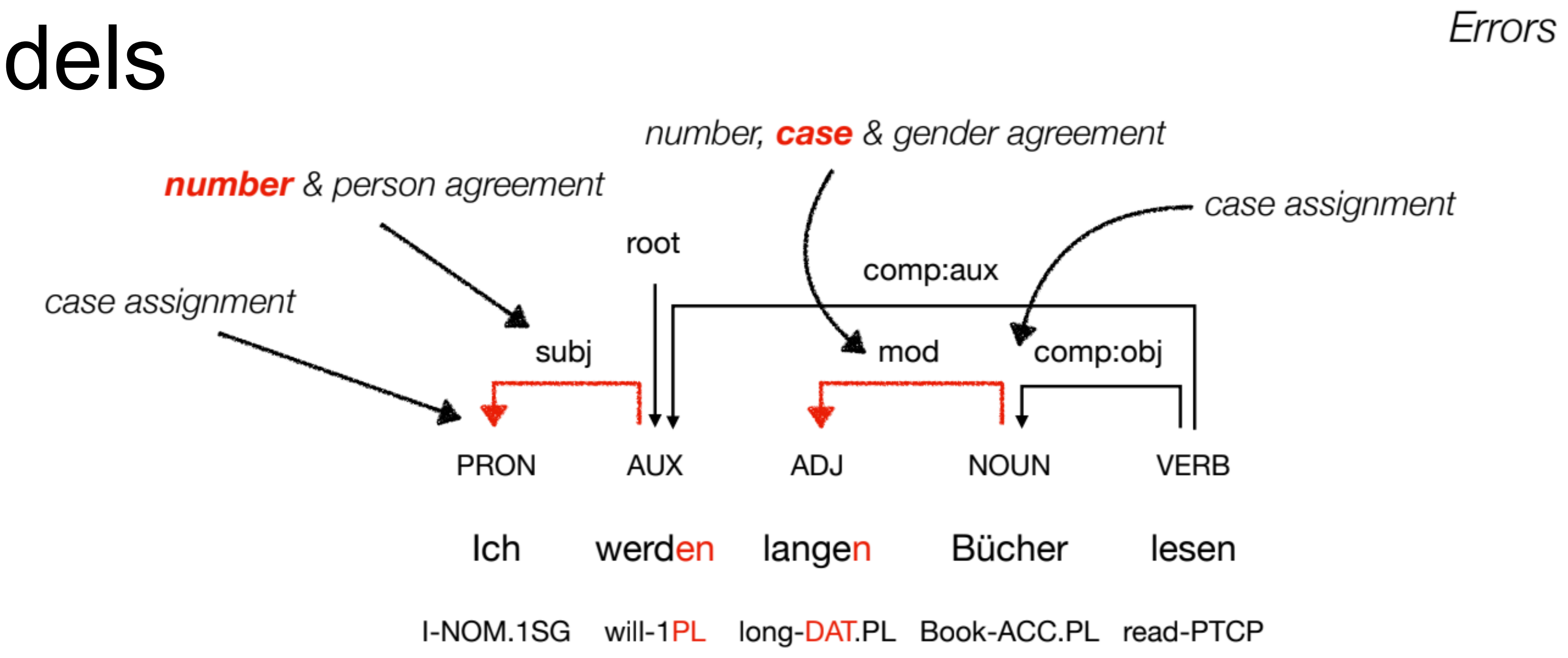


Evaluating the Morphosyntactic Well-formedness of Generated Texts
*Pratapa, Anastasopoulos, Rljhwani, **Chaudhary,** Mortensen, Sheikh, Neubig, Tsvetkov.* ***EMNLP 2020***

Language Descriptions →

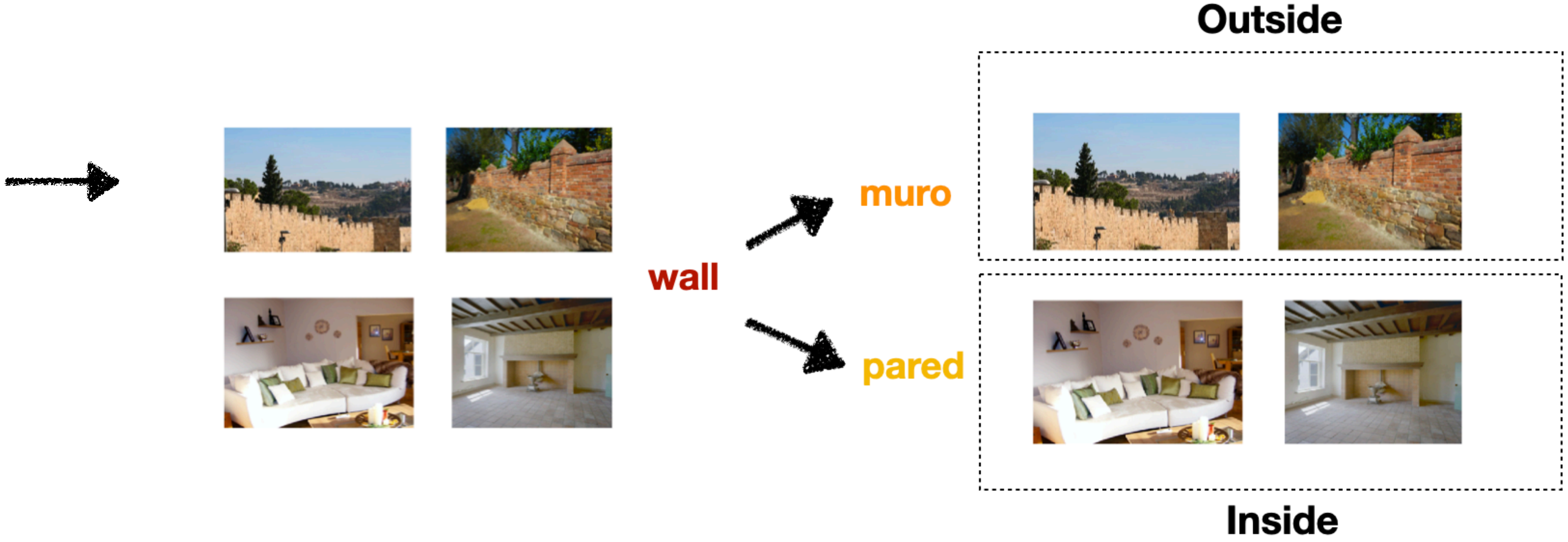
Other Applications

Evaluating Context-Usage in MT models



Evaluating the Morphosyntactic Well-formedness of Generated Texts

Language Descriptions

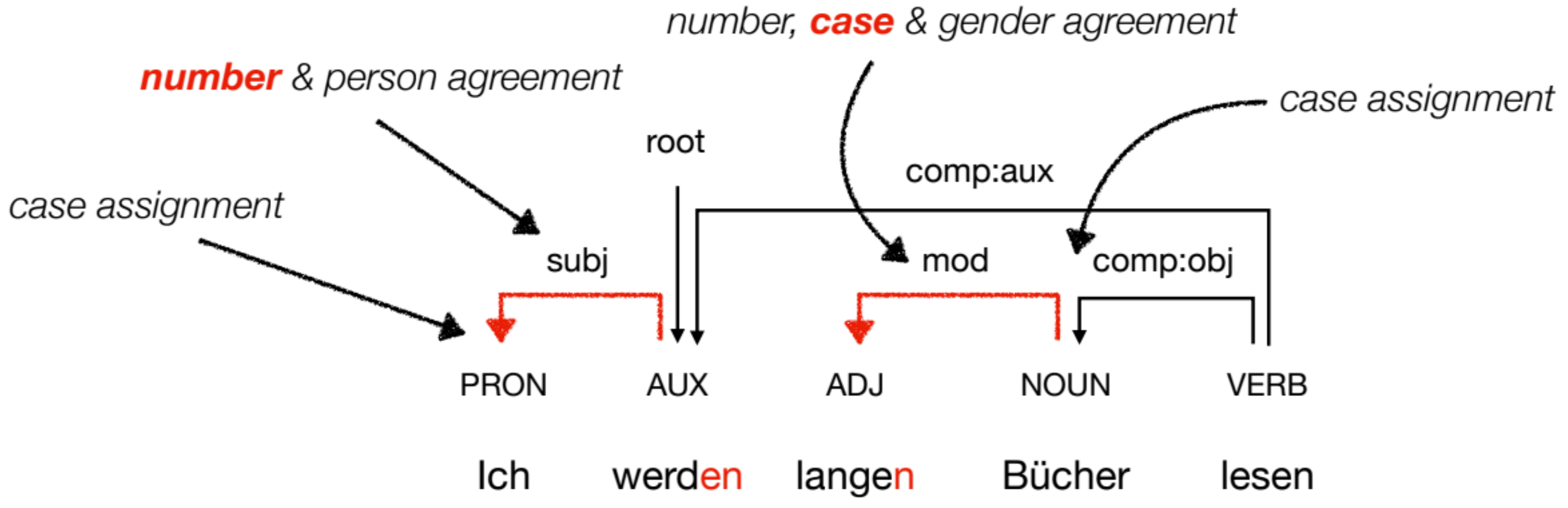


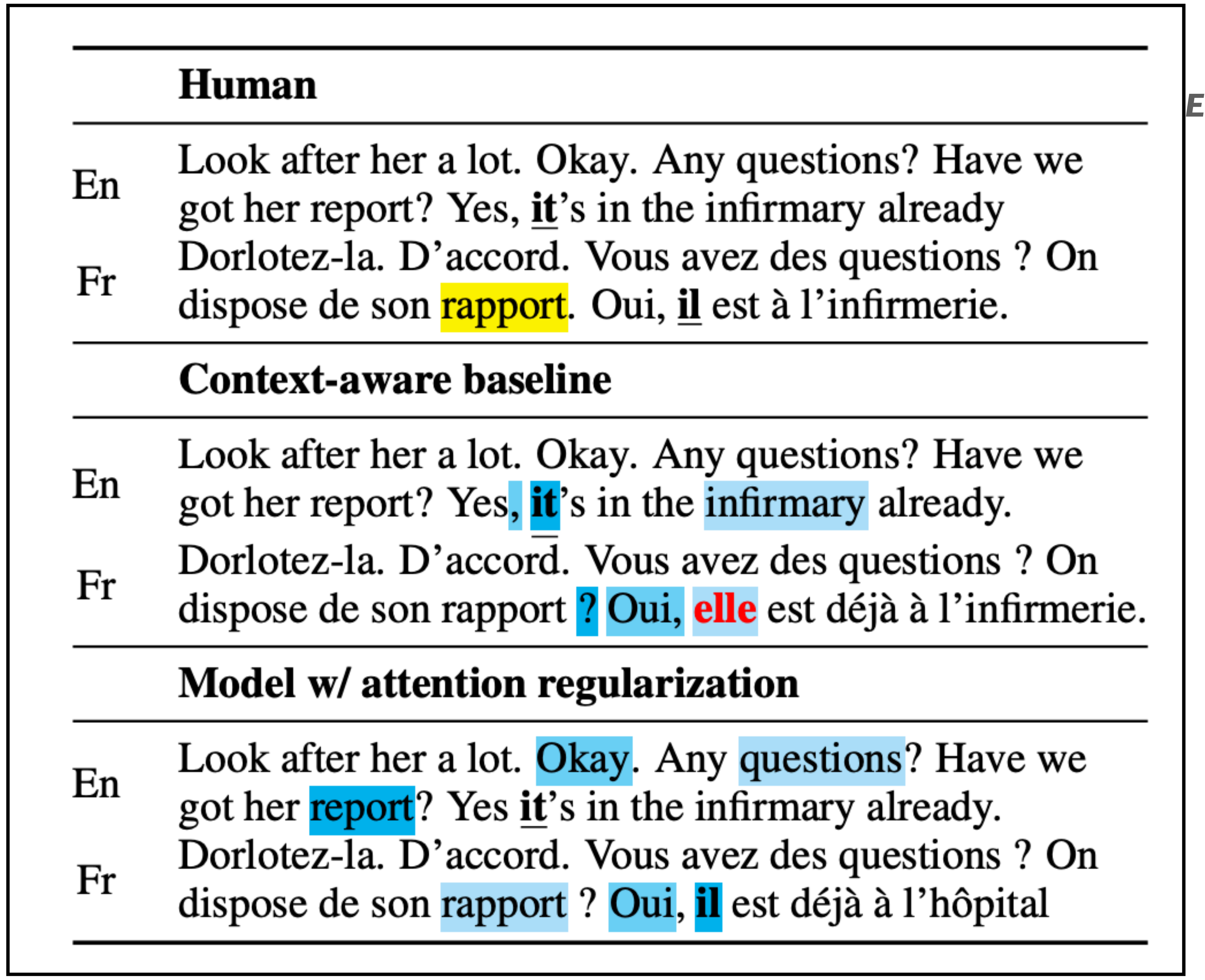
When is Wall a Pared and when a Muro? Extracting Rules Governing Lexical Selection
Chaudhary**, Yin, Anastasopoulos, Neubig.* ***EMNLP 2021

Other Applications

Evaluating Context-Usage in MT models

Errors





	Human
En	Look after her a lot. Okay. Any questions? Have we got her report? Yes, **it**'s in the infirmary already
Fr	Dorlotez-la. D'accord. Vous avez des questions ? On dispose de son rapport. Oui, **il** est à l'infirmerie.
	Context-aware baseline
En	Look after her a lot. Okay. Any questions? Have we got her report? Yes, **it**'s in the infirmary already.
Fr	Dorlotez-la. D'accord. Vous avez des questions ? On dispose de son rapport ? Oui, **elle** est déjà à l'infirmerie.
	Model w/ attention regularization
En	Look after her a lot. Okay. Any questions? Have we got her report? Yes **it**'s in the infirmary already.
Fr	Dorlotez-la. D'accord. Vous avez des questions ? On dispose de son rapport ? Oui, **il** est déjà à l'hôpital

EMNLP 2020

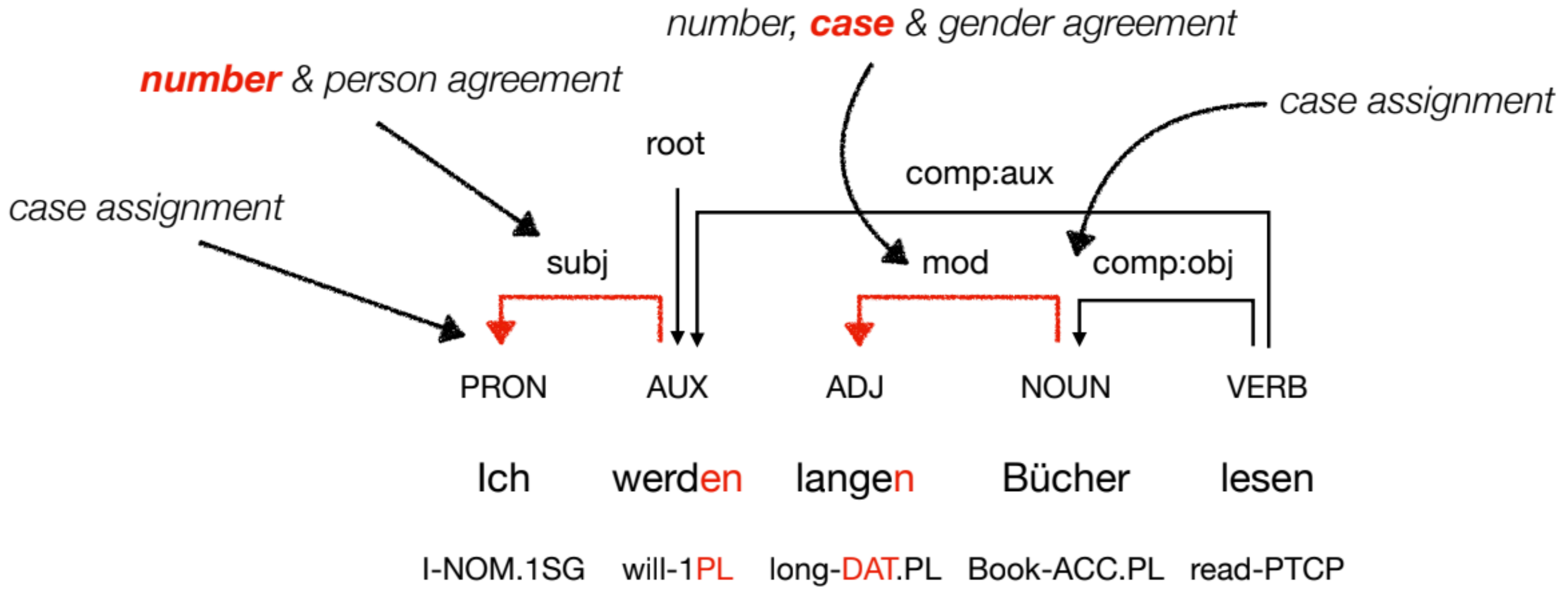
Language Descriptions →

Do Context-Aware Translation Models Pay the Right Attention?
*Yin, Fernandes, Pruthi, **Chaudhary**, Martins, Neubig.* ***ACL 2021***

Other Applications

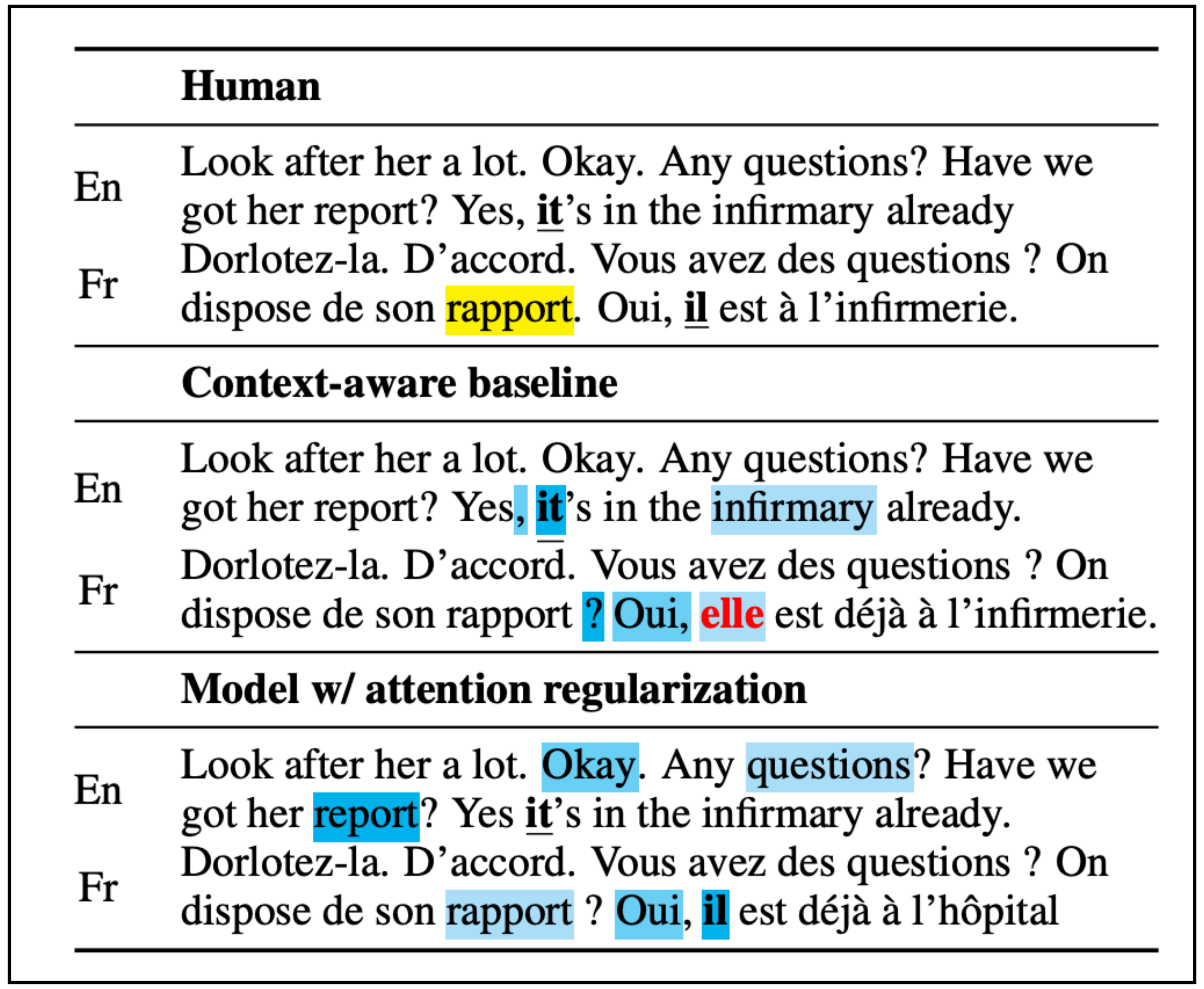
Evaluating Context-Usage in MT models

Errors



Evaluating the Morphosyntactic Well-formedness of Generated Texts
*Pratapa, Anastasopoulos, RIjhwani, **Chaudhary,** Mortensen, Sheikh, Neubig, Tsvetkov.* ***EMNLP 2020***

Language Descriptions →



	Human
En	Look after her a lot. Okay. Any questions? Have we got her report? Yes, it's in the infirmary already
Fr	Dorlotez-la. D'accord. Vous avez des questions ? On dispose de son rapport. Oui, il est à l'infirmerie.
	Context-aware baseline
En	Look after her a lot. Okay. Any questions? Have we got her report? Yes, it's in the infirmary already.
Fr	Dorlotez-la. D'accord. Vous avez des questions ? On dispose de son rapport ? Oui, elle est déjà à l'infirmerie.
	Model w/ attention regularization
En	Look after her a lot. Okay. Any questions? Have we got her report? Yes it's in the infirmary already.
Fr	Dorlotez-la. D'accord. Vous avez des questions ? On dispose de son rapport ? Oui, il est déjà à l'hôpital

Do Context-Aware Translation Models Pay the Right Attention?
*Yin, Fernandes, Pruthi, **Chaudhary**, Martins, Neubig.* ***ACL 2021***

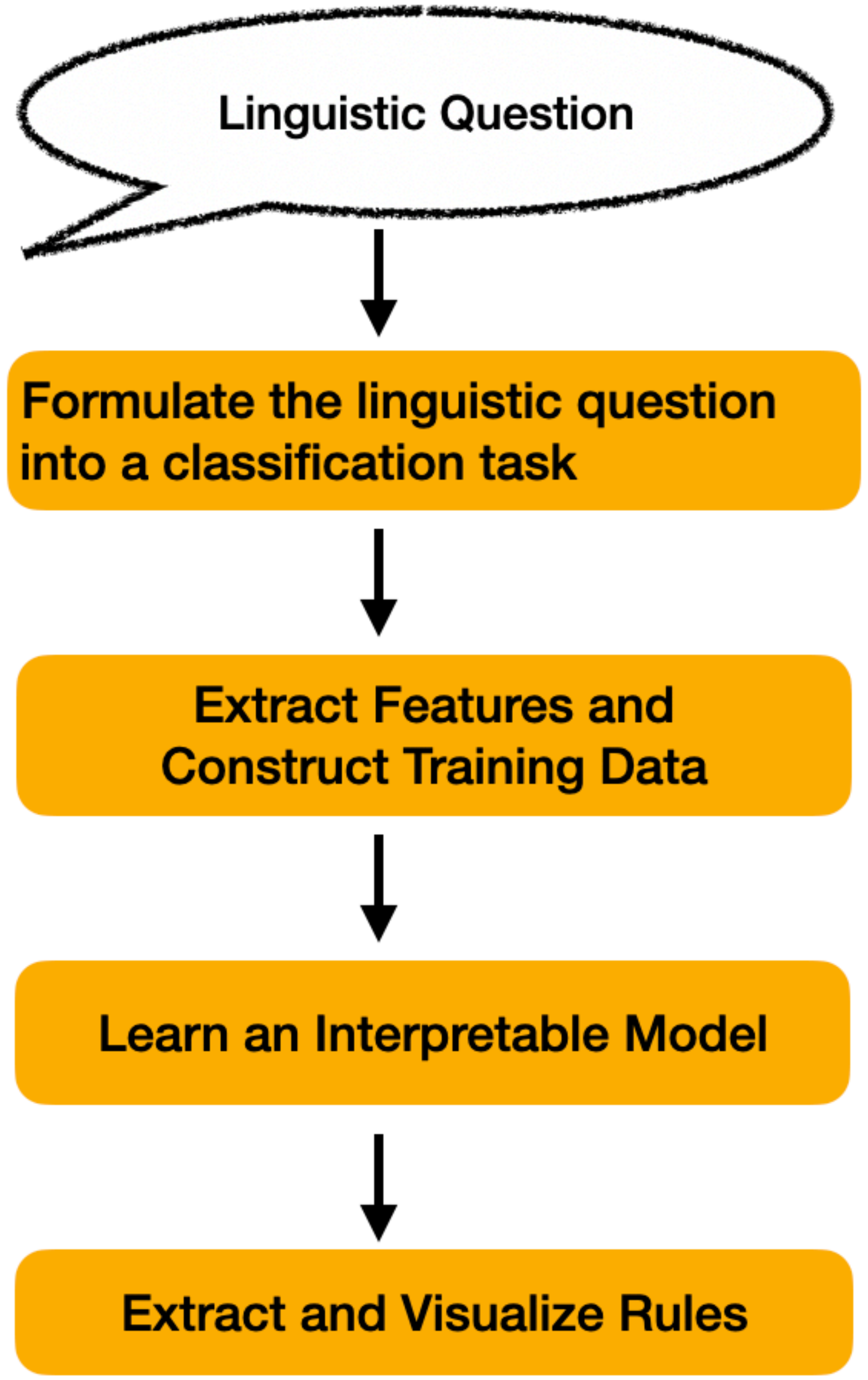
Contributions

Contributions

- **AutoLEX: framework to extract and visualize language descriptions**

Contributions

- **AutoLEX**: framework to extract and visualize language descriptions



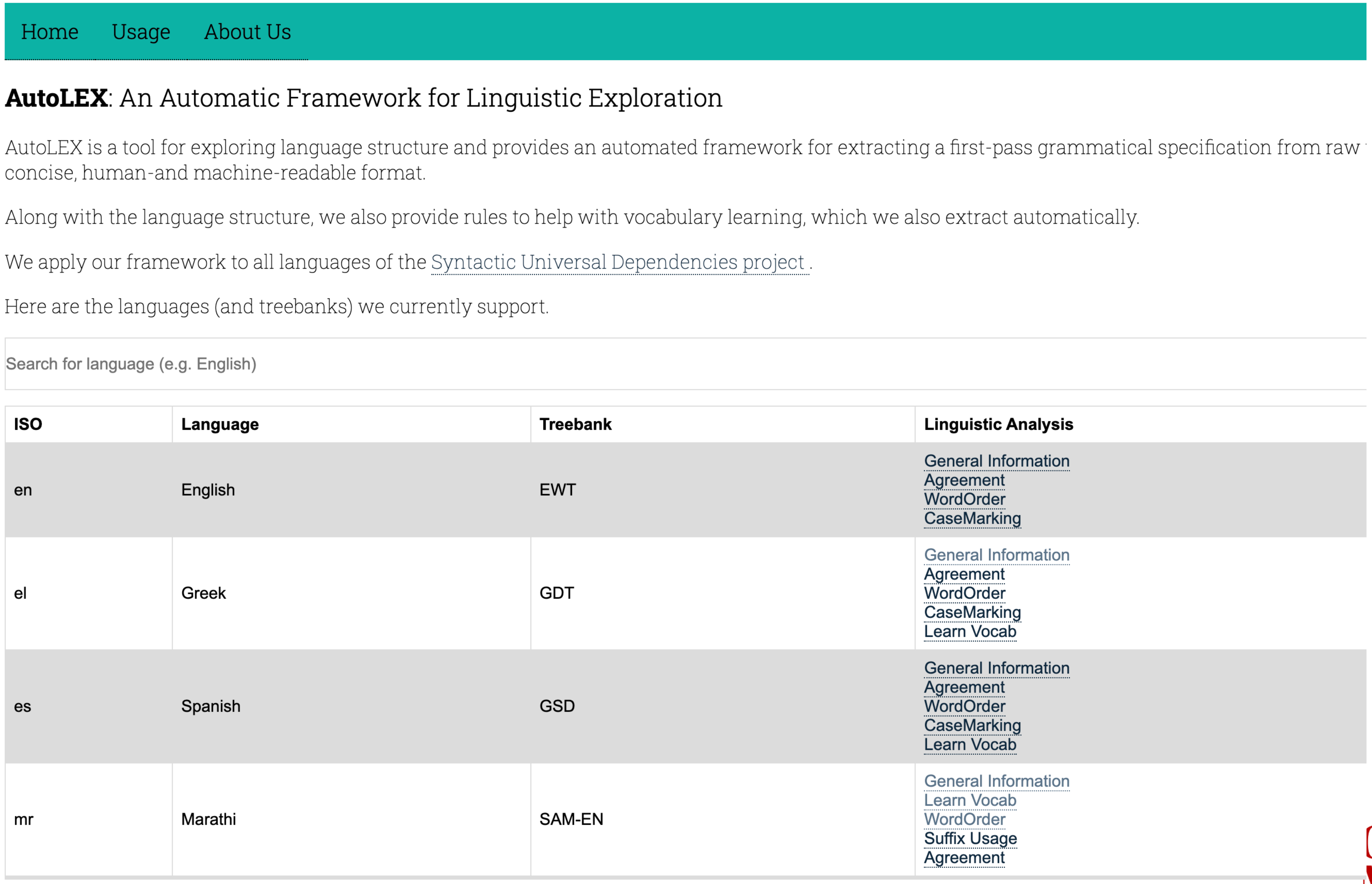
Linguistic Question

Formulate the linguistic question into a classification task

Extract Features and Construct Training Data

Learn an Interpretable Model

Extract and Visualize Rules



AutoLEX: An Automatic Framework for Linguistic Exploration

AutoLEX is a tool for exploring language structure and provides an automated framework for extracting a first-pass grammatical specification from raw concise, human-and machine-readable format.

Along with the language structure, we also provide rules to help with vocabulary learning, which we also extract automatically.

We apply our framework to all languages of the Syntactic Universal Dependencies project .

Here are the languages (and treebanks) we currently support.

ISO	Language	Treebank	Linguistic Analysis
en	English	EWT	General Information Agreement WordOrder CaseMarking
el	Greek	GDT	General Information Agreement WordOrder CaseMarking Learn Vocab
es	Spanish	GSD	General Information Agreement WordOrder CaseMarking Learn Vocab
mr	Marathi	SAM-EN	General Information Learn Vocab WordOrder Suffix Usage Agreement

Contributions

- **AutoLEX**: framework to extract and visualize language descriptions

http://www.autolex.co/interface/

Contributions

- **AutoLEX**: framework to extract and visualize language descriptions

http://www.autolex.co/interface/

- **Real-World Utility**: established collaborations with teacher communities

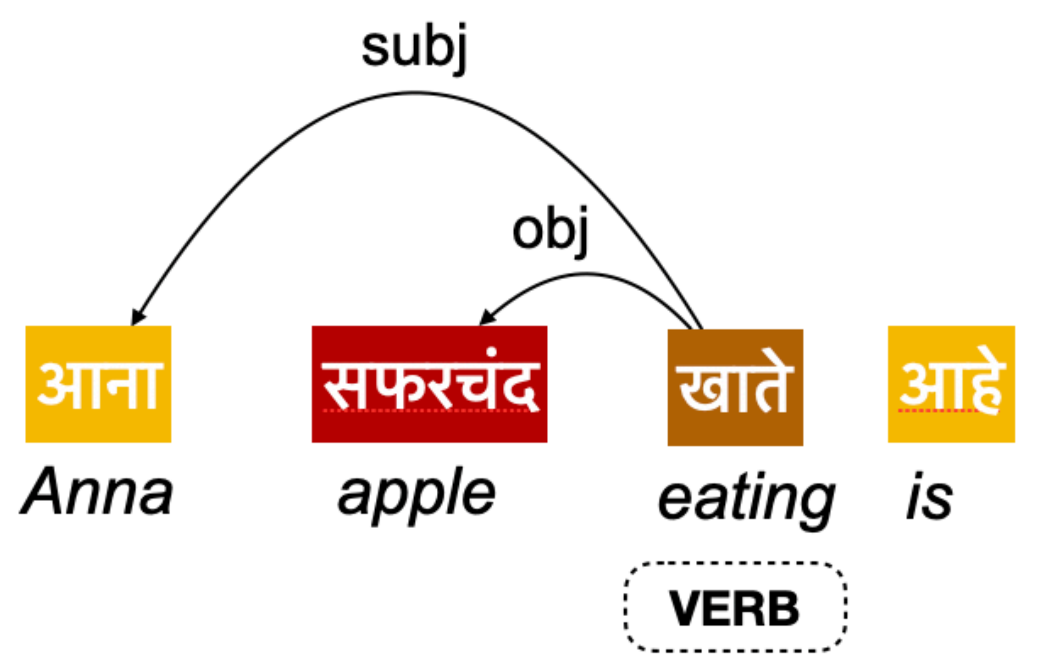
Contributions

- **AutoLEX**: framework to extract and visualize language descriptions

http://www.autolex.co/interface/

- **Real-World Utility**: established collaborations with teacher communities
- **Under-Resourced NLP**: effectively utilize existing data and collect new data

(Low-resource) Language Analysis



AutoLEX: Automatic Language Explorer

Subject **Object** **Verb**

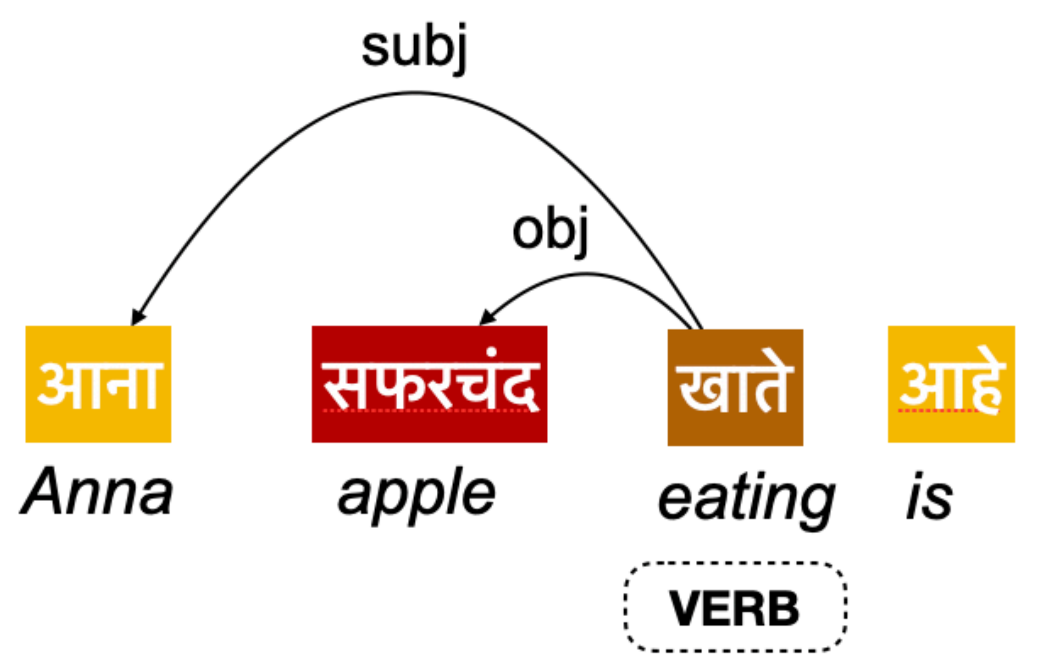
तांदूळ	भात
/tandul/	/bhaat/
‘raw rice’	‘cooked rice’

Applications

Language Education and Documentation

https://www.autolex.co/

(Low-resource) Language Analysis



AutoLEX: Automatic Language Explorer

Subject **Object** **Verb**

तांदूळ
/tandul/
'raw rice'

भात
/bhaat/
'cooked rice'

Applications

Language Education and Documentation

https://www.autolex.co/

What's Next?

(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

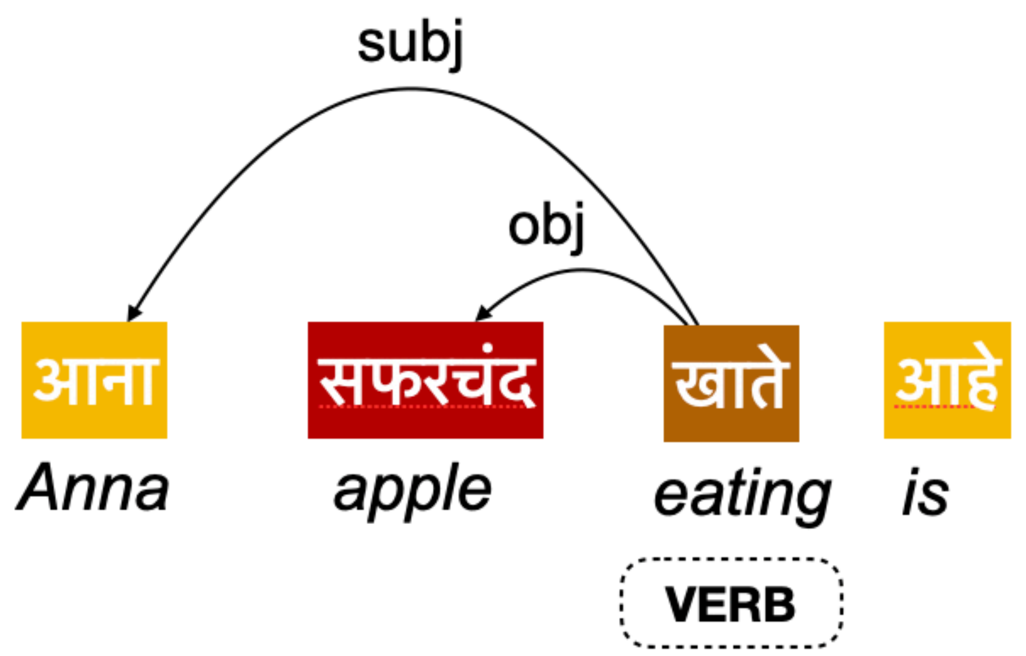
Applications

https://www.autolex.co/

What's Next?

- We **demonstrated utility** on 4 languages, about 7,000 more to go

(Low-resource) Language Analysis



AutoLEX: Automatic Language Explorer

Subject **Object** **Verb**

तांदूळ
/tandul/
'raw rice'

भात
/bhaat/
'cooked rice'

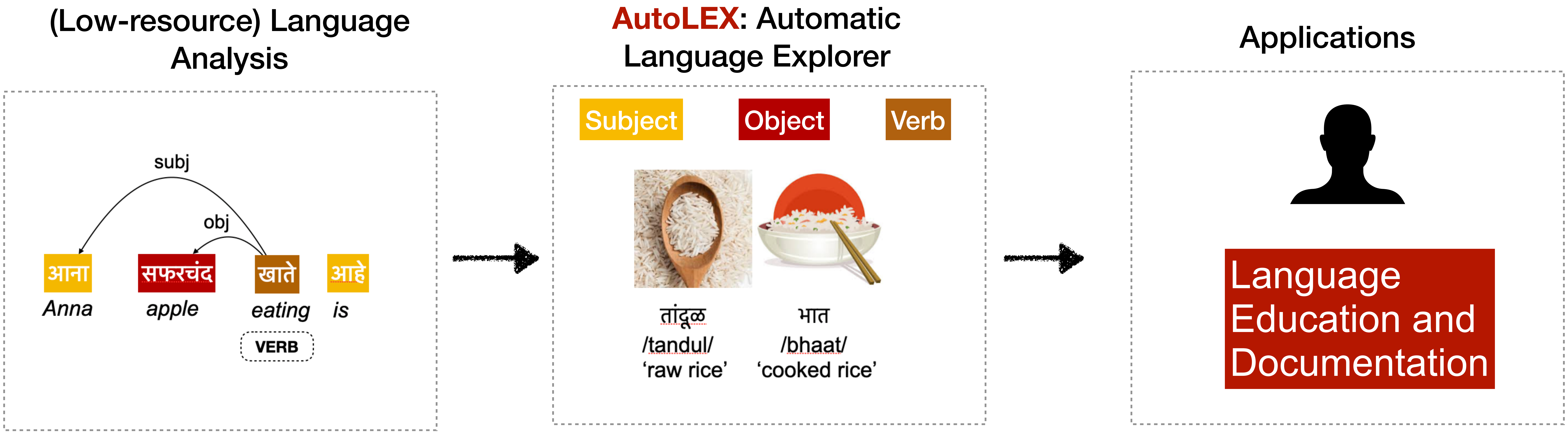
Applications

Language Education and Documentation

https://www.autolex.co/

What's Next?

- We **demonstrated utility** on 4 languages, about 7,000 more to go
- **Low-resource language analysis** still doesn't work well enough



(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

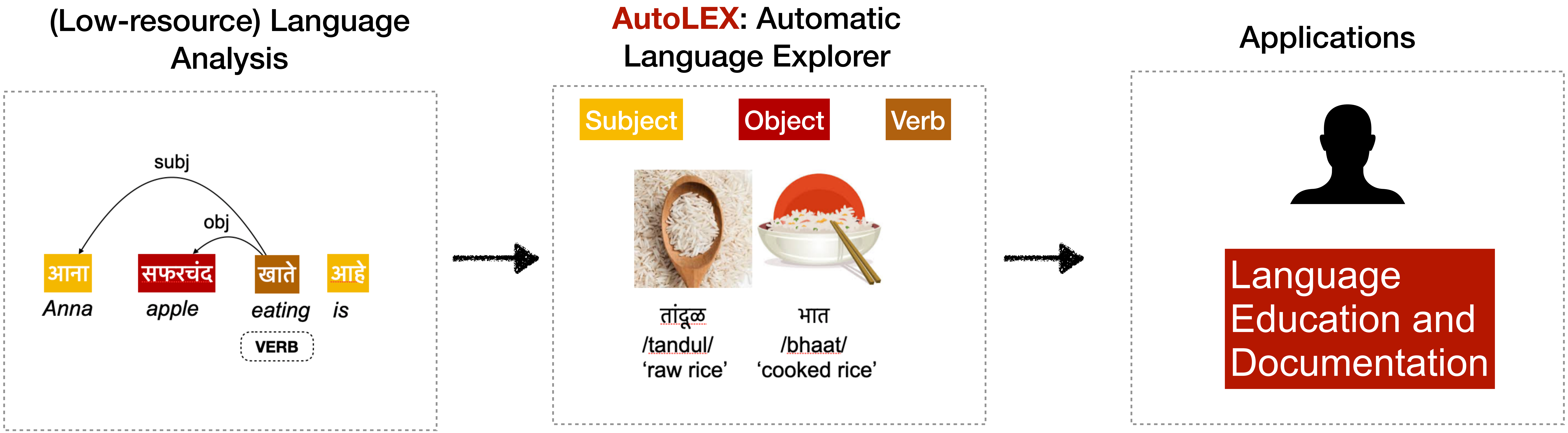
Applications

Language Education and Documentation

https://www.autolex.co/

What's Next?

- We **demonstrated utility** on 4 languages, about 7,000 more to go
- **Low-resource language analysis** still doesn't work well enough
- Better **rule extraction methods**



(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

Applications

https://www.autolex.co/

What’s Next?

- We **demonstrated utility** on 4 languages, about 7,000 more to go
- **Low-resource language analysis** still doesn’t work well enough
- Better **rule extraction methods**
- Close link w/ **data provenance** (conversational text >> legal text)