Remarkable language diversity!
Remarkable language diversity!
Remarkable language diversity!
Remarkable language diversity!
Remarkable language diversity!
Remarkable language diversity!
Different Systems Governing Language

Syntax
Different Systems Governing Language

Syntax

Morphology
Different Systems Governing Language

Syntax

Morphology

eating + ing

Anna
is
eating
an
apple

subj
obj

Anna
apple
eating
is

subj
obj

 Verb

Verb
Different Systems Governing Language

Syntax

Anna is eating an apple

Morphology

eating + eat + ing

खाते + खा + ते + तो
Different Systems Governing Language

**Syntax**

Anna is eating an apple

**Morphology**

eating  eat + ing

**Semantics**

खाते  खा + ते  + तो
/khate/  /kha/  /te/  /to/
Different Systems Governing Language

Syntax

Anna is eating an apple

Morphology

eating, eat + ing

Semantics

खाते /khate/ खा /kha/ + ते /te/

+ तो /to/
Different Systems Governing Language

**Syntax**

Anna, is, eating, an, apple

**Morphology**

eating, eat + ing

खाते, /kha/ + ते, /te/

खाते, /kha/ + ते, /te/

अाना, सफरचंद, खाते, आहे

Anna, apple, eating, is

**Semantics**

खाते /kha/ + ते /te/

तांदुल /tandul/ 'raw rice'

भात /bhaat/ 'cooked rice'
Describing Systems of Language
Describing Systems of Language
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.
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).
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).
What are Language Descriptions?
What are Language Descriptions?

How are subjects ordered wrt verbs in Marathi?
What are Language Descriptions?

How are subjects ordered wrt verbs in Marathi?

Anna
Apple
Eating
Is
What are Language Descriptions?

How are subjects ordered wrt verbs in Marathi?
What are Language Descriptions?

How are subjects ordered wrt verbs in Marathi?

Subject Object Verb

Anna PROP
apple NOUN
eating VERB
is AUX
Why is this challenging?

How are subjects ordered wrt verbs in Marathi?
Why is this challenging?

How are subjects ordered wrt verbs in Marathi?
Why is this challenging?

How are subjects ordered wrt verbs in Marathi?

Subject Object Verb

Anna apple eating is

PROPN NOUN VERB AUX

Where Languages Are Dying

Languages classified as threatened/ endangered in 2022, by region

Europe/Russia** 148
North/Central America* 222
South America 226
Africa 428
Asia 693
Oceania 733

* Including the Caribbean
** Including the Caucasus
Source: Endangered Languages Project
Why is this challenging?

Subject    Object    Verb
Anna    apple    eating    is

How are subjects ordered wrt verbs in Marathi?

Could we help in some way?
NLP already does some of this!

How are subjects ordered wrt verbs in Marathi?

Subject  Object  Verb
NLP already does some of this!

How are subjects ordered wrt verbs in Marathi?

Subject    Object    Verb
Anna    सफरचंद    खाते    आहे
PROPN    NOUN    VERB    AUX
Anna    apple    eating    is
Subject    Object    Verb
Our Proposed Pipeline

Our Proposed Pipeline

(Low-resource) Language Analysis

Our Proposed Pipeline

(Low-resource) Language Analysis

Anna
apple
eating
is

Subject
Object
Verb

AutoLEX: Automatic Language Explorer

Our Proposed Pipeline

(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

Our Proposed Pipeline

(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

Applications

Language Education and Documentation

AutoLEX: Automatic Language Explorer
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?

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?

AutoLEX: Automatic Language Explorer

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

Text corpus
How are subjects ordered in Marathi?

**AutoLEX: Automatic Language Explorer**

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

<table>
<thead>
<tr>
<th>Word Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generally the word order for subject-verb is before i.e. subject before verb</td>
</tr>
<tr>
<td>Some examples are: <strong>Examples</strong></td>
</tr>
<tr>
<td>subject is <strong>after</strong> verb when:</td>
</tr>
<tr>
<td>verb is also governing= काय (kaay)</td>
</tr>
<tr>
<td>subject is nearby= <strong>compound</strong></td>
</tr>
<tr>
<td>subject is governed by a word with <strong>Aspect</strong> = <strong>Simp</strong> ( <strong>Examples</strong> ) OR</td>
</tr>
</tbody>
</table>
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 भूमिकामयक (basy)
  - subject is nearanye compound
  - subject is governed by a word with Aspect = Simp (Examplne)

Carnegie Mellon University
How are subjects ordered in Marathi?

AutoLEX: Automatic Language Explorer

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

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
- subject is a near-eye compound
- subject is governed by a word with Aspect = Simp

(Examples)
OR
How are subjects ordered in Marathi?

**AutoLEX: Automatic Language Explorer**

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

1. Formulate the linguistic question into a classification task
2. Extract Features and Construct Training Data
3. Learn an Interpretable Model
4. Extract and Visualize Rules

**Text corpus**

**AutoLEX Pipeline**

- **Word Order**
  - Generally the word order is subject-verb before i.e., subject before verb
  - Some examples are: Examples
  - subject is after verb when:
    - verb is also governing a विरुद्ध (kasya)
    - subject is nearest compound
    - subject is governed by a word with चिन्द्र = Simp
      - चिन्द्र (Examples)
    - OR
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

1. Formulate the linguistic question into a classification task
2. Extract Features and Construct Training Data
3. Learn an Interpretable Model
4. 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
  - subject is nearly compound
  - subject is governed by a word with Aspect = Simple

(Carnegie Mellon University)
AutoLEX: Automatic Language Explorer

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

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

<table>
<thead>
<tr>
<th>Some examples are: Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject is after verb when:</td>
</tr>
<tr>
<td>verb is also governing* विष (kavy)</td>
</tr>
<tr>
<td>subject is nearest compound</td>
</tr>
<tr>
<td>subject is governed by a word with Aspect = Simp (Exemplary)</td>
</tr>
</tbody>
</table>
AutoLEX: Automatic Language Explorer

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

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

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

Expert annotations for feature extraction model

Expert annotations (e.g. UD Treebank)

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?

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
AutoLEX: Automatic Language Explorer

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)

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

- Agreement
- Word Order
- Affix Usage
- Case Marking
- Word Usage

Morpho-Syntax
Types of Grammar Aspects covered in **AutoLEX**

- Agreement
- Word Order
- Affix Usage
- Case Marking
- Word Usage

**Morpho-Syntax**
Types of Grammar Aspects covered in AutoLEX

- Agreement
- Word Order
- Affix Usage
- Case Marking
- Word Usage

Morpho-Syntax

Lexical Semantics
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
- Word Usage

Lexical Semantics
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)

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

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

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

*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

Anna is eating an apple

What is Anna eating?

Verb-Object

Object-Verb
Word Order

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

Anna is eating an apple

Verb-Object

What is Anna eating?

Object-Verb
Word Order: Formulate Linguistic Question

When are objects before or after the verbs?

Anna is eating an apple

What is Anna eating?

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
When are *objects* before or after the *verbs*?
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: $f(\text{eating \ apple})$

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: Formulate Linguistic Question

When are objects before or after the verbs?

Input

\[ f(\text{eating, apple}) \]

Label

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

**Input**

1. $f(\text{eating apple})$
2. $f(\text{What eating})$

**Label**

1. Before
2. After

---

**Formulate the linguistic question into a classification task**

1. **Extract Features and Construct Training Data**
2. **Learn an Interpretable Model**
3. **Extract and Visualize Rules**
Word Order: Formulate Linguistic Question

When are **objects** before or after the **verbs** in English?
Word Order: Formulate Linguistic Question

When are **objects** before or after the **verbs** in English?
Word Order: Formulate Linguistic Question

When are objects before or after the verbs in English?

Anna is eating an apple

What is Anna eating?
When are *objects* before or after the *verbs* in English?

**Word Order: Formulate Linguistic Question**

- Anna
- is
- eating
- an
- apple
  
  **VERB**
  **NOUN**

- What
- is
- Anna
- eating
- ?
  
  **PRON**
  **VERB**

**AUTOLEX: An Automatic Framework for Linguistic Exploration**

Chaudhary, Sheikh, Mortensen, Anastasopoulos, Neubig. *In Submission*
When are **objects** before or after the **verbs** in English?

**Dependency Parsing**

**POS Tagging**

Anna is eating an apple

What is Anna eating?
Word Order: Formulate Linguistic Question

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

- **Dependency Parsing**
- **POS Tagging**
- **Morphological Analysis**

[Diagram showing the POS tagging of words: Anna (NOUN), is (VERB), eating (VERB), an (DET), apple (NOUN), What (PRON), is (VERB), Anna (NOUN), eating (VERB), ? (PUNCT)]
Word Order: Extract Features and Training Data

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

- Anna is eating an apple
  - VERB
  - Tense=Pres
  - NOUN

- What is Anna eating?
  - PRON
  - PronType = Int
  - VERB
  - Tense=Pres

**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?

- Anna is eating an apple
- What is Anna eating?
Word Order: Extract Features and Training Data

When are objects before or after the verbs in English?

Input | Features | Label
--- | --- | ---
Anna is eating an apple | obj | ?

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?

<table>
<thead>
<tr>
<th>Input</th>
<th>Features</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(\text{eating, apple})$</td>
<td></td>
<td>Before</td>
</tr>
</tbody>
</table>
Word Order: Extract Features and Training Data

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

<table>
<thead>
<tr>
<th>Input</th>
<th>Features</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(\text{eating, apple})$</td>
<td></td>
<td>Before</td>
</tr>
<tr>
<td>$f(\text{What, eating})$</td>
<td></td>
<td>After</td>
</tr>
</tbody>
</table>
Word Order: Extract Features and Training Data

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

**Input** | **Features** | **Label**
---|---|---
\[ f(\text{eating, apple}) \] | \text{is-object-NOUN} | Before
\[ f(\text{What, eating}) \] | \text{is-object-PRON} | After
Word Order: Extract Features and Training Data

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

![Diagram showing word order examples: Anna is eating an apple vs. What is Anna eating?]

### Input

<table>
<thead>
<tr>
<th>f( eating apple )</th>
<th>f( What eating )</th>
</tr>
</thead>
</table>

### Features

- **is-object-NOUN**
- **is-verb-tense-present**
- **is-object-PRON**
- **is-object-Interrogative**

### Label

- **Before**
- **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?
Word Order: Learn a Model

When are objects before or after the verbs in English?

is-object-PRON

is-object-Interrogative

root

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?

Leaf 1: Before

is-object-PRON

is-object-Interrogative

root
Word Order: Learn a Model

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

![Diagram showing word order analysis]

- **Leaf 1:** Before
- **Leaf 1:** 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?

**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?
Word Order: Data

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

- Compare the model in a clean setting $\rightarrow$ **Syntactic Universal Dependencies (SUD)**
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**
Word Order: Automated Evaluation

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

Apply model on held-out sentences
Word Order: Automated Evaluation

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

![Diagram showing word order](image)

Apply model on held-out sentences

**Baseline**: most frequent label in test data

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Syntax</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**en-object-verb**
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

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Syntax</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

es-adjective-noun
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!**
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  
great

 libros  
books

 fueron  
were

 comprados  
bought

 por  
by

 la  
the

 niña  
girl

 pequeña  
small
Word Order: Types of Features

grandes

libros

fueron

comprados

por

la

niña

pequeña

great

books

were

bought

by

the

girl

small

ADJ

NOUN

ADJ

NOUN

ADJ

NOUN

ADJ
Word Order: Types of Features

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

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

head-is-noun deprel-is-mod is-adj
Word Order: Types of Features

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

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

dep-lemma head-lemma dep-lemma head-lemma
-is-gran -is-libro -is-niña -is-pequeño
Word Order: Types of Features

**Syntactic**
- is-adj
- head-is-noun
- head-is-noun
- is-adj
- dep-rel-is-mod
- dep-rel-is-mod

**Lexical**
- dep-lemma
- head-lemma
- dep-lemma
- head-lemma
- is-gran
- is-libro
- is-niña
- is-pequeño

**Example Sentence**
- grandes libros fueron comprados por la niña pequeña

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

**Labels**
- ADJ
- NOUN
- mod
Word Order: Types of Features
Word Order: Types of Features
Word Order: Types of Features

Accuracy

- **es-adjective-noun**

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>60</td>
</tr>
<tr>
<td><strong>syntactic</strong></td>
<td>67.5</td>
</tr>
<tr>
<td><strong>syntactic + lexical</strong></td>
<td>75</td>
</tr>
<tr>
<td><strong>90</strong></td>
<td>82.5</td>
</tr>
</tbody>
</table>

**AUTOLEX: An Automatic Framework for Linguistic Exploration**
Chaudhary, Sheikh, Mortensen, Anastasopoulos, Neubig. *In Submission*
Word Order: Types of Features
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
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** → sparse and non-negative transformation over pretrained embeddings

\[\text{dim-0: wrist, shoulder, ligament, ankle, thigh} \quad \text{(Semantic similarity)}\]
\[\text{dim-1: extensively, adequately, traditionally, royally, fully} \quad \text{(Syntactic similarity)}\]
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

  \[
  \begin{align*}
  \text{dim-0:} & \text{ wrist, shoulder, ligament, ankle, thigh} \\
  \text{dim-1:} & \text{ extensively, adequately, traditionally, royally, fully} \\
  \text{dim-2:} & \text{ torque, joystick, grip, wrist, swinging}
  \end{align*}
  \]

\(^1\) 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

\[\begin{align*}
\text{dim-0:} & \quad \text{wrist, shoulder, ligament, ankle, thigh} \\
\text{dim-1:} & \quad \text{extensively, adequately, traditionally, royally, fully} \\
\text{dim-2:} & \quad \text{torque, joystick, grip, wrist, swinging}
\end{align*}\]
Word Order: Types of Features

grandes great
 libros books
 fueron were
 Comprados bought
 por by
 la the
 niña girl
 pequeñ small

is-adj
 is-ordinal
 deprel-is-mod

head-is-noun
 deprel-is-mod

is-adj

dep-lemma
 -is-gran
 -is-libro

head-lemma

dep-lemma
 -is-niña
 -is-pequeño

head-lemma
Word Order: Types of Features

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

**syntactic**
- is-adj
- is-ordinal
- dep-lemma
- deprel-is-mod

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

**semantic**
- word-is-like
- -primera
- -página, publicación
- -niño
- -nueva, vuelta
Word Order: Types of Features

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary, Sheikh, Mortensen, Anastasopoulos, Neubig. In Submission
Word Order: Types of Features

- grandes books fueron comprar por la niña pequeña
  - ADJ NOUN were bought by the girl

- grande libros fueron comprados por la niña pequeña
  - ADJ NOUN were bought by the girl

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary, Sheikh, Mortensen, Anastasopoulos, Neubig. In Submission
Word Order: Types of Features
Word Order: Types of Features

AUTOLEX: An Automatic Framework for Linguistic Exploration

Chaudhary, Sheikh, Mortensen, Anastasopoulos, Neubig. In Submission
Word Order: Extract rules
Word Order: Extract rules

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary, Sheikh, Mortensen, Anastasopoulos, Neubig. In Submission
Word Order: Extract rules

Leaf 1:

After: 1000  
Before: 8000

AUTOLEX: An Automatic Framework for Linguistic Exploration
Chaudhary, Sheikh, Mortensen, Anastasopoulos, Neubig.  In Submission
Word Order: Extract rules

Leaf 1:

After: 1000  Before: 1800
Word Order: Extract rules

Leaf 1:

- After: 1000
- Before: 1800

Statistical threshold

Leaf 2: Before
- Is a node
- Dep-lemma is gran

Leaf 3: Before
- Is a node
- Dep-lemma is nuevo

Leaf 1:
- Significance test $\chi^2$
  - Observed distribution is significant
- Effect Size
  - 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
Word Order: Extract rules

Leaf 1: After: 1000 Before: 1800
Label: Before

Statistical threshold

Significance test $\chi^2$
Observed distribution is significant

Effect Size
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
Word Order: Visualize rules

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

The dominant order in the corpus is **after**
Word Order: Visualize rules

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

The dominant order in the corpus is **after**
Word Order: Visualize rules

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

---

**Features that lead to a leaf makes a rule**

- 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

**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

![Diagram showing the word order rule with examples: 'Ahora por ***primera*** ***vez*** [la audiencia ve cara de Lugosi].'](image)

**Exceptions!**

![Diagram showing exceptions: 'Al palacio se une una capilla mediante un pasadizo en ***plata*** ***primera***'].
Agreement
Agreement
Agreement
Agreement

Marathi

तो

/that.M/

मुलगी

/girl/

जेवतो

/eating.M/

आहे

/is/

X
Agreement

Marathi

तो
/that.M/

मुलगी
/girl/

जेवतो
/eating.M/

आहे
/is/

DET

NOUN

VERB

AUX

det
Agreement

Marathi

तो /that.M/ गुलामी /girl/ जेवतो /eating.M/ आहे /is/

ती /that.F/ गुलामी /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/

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

DET NOUN VERB AUX
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
  - cannot 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
   - cannot 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
   - cannot 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
   ○ cannot 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
- cannot 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
   ○ cannot 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
   ○ cannot 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:
Results: Quality Evaluation for Greek
Results: Quality Evaluation for Greek

Has model discovered grammatically valid 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

- Total: 161 rules

**Does model discover grammatically valid rules?**

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

**Are the rules human-readable?**

- 40% rules are too specific/general
- 20% rules are too general
- 10% rules are not a rule
- 30% rules are precise

Percentage of rules:
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

Total: 161 rules

- **CaseMarking**
  - Precise: 0.5
  - Too-specific: 0.25
  - Too-general: 0.5
  - Not-a-rule: 0.25

- **WordOrder**
  - Yes: 0.25
  - Partially-correct: 0.25
  - No: 0.5

- **Agreement**
  - Yes: 0.5
  - Partially-correct: 0.25
  - No: 0.25
Results: Quality Evaluation

Does the model discover *new rules?
Results: Quality Evaluation

**Does the model discover *new rules?**

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

<table>
<thead>
<tr>
<th>Features that make up this rule</th>
<th>Active Features</th>
<th>Inactive Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>adjective's head is a= PRON</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

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!
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
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 UDIFY¹ 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 UDIFY¹ parser with all available data

---

¹75 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 UDIFY¹ 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 UDIFY¹ parser with all available data
  - For Subj-V, Obj-V word orders **nearly 50%** showed valid patterns

¹75 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!
Revisiting our research goal
Revisiting our research goal

(Low-resource) Language Analysis
Revisiting our research goal

(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

Subject  Object  Verb

Anna  apple  eating  is

安娜  苹果  吃  是

Subject  Object  Verb

तांदूळ  भात
/tandul/  ’raw rice’  /bhaat/  ’cooked rice’
Revisiting our research goal

(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

Subject  Object  Verb

Anna  apple  eating  is

तांदूळ  (‘raw rice’)  भात  (‘cooked rice’)
Revisiting our research goal

(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

Applications

Subject

Object

Verb

Language Education and Documentation
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!
• 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!

Limit in language coverage!
Language Education

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

- Creating a curriculum is a challenging process → grammar coverage, examples, exercises …
Computer-assisted language learning (CALL) systems are in **high demand**!

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
Word Usage: Semantic Subdivisions

- Different languages carve up the semantic space differently

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

**Outside**

- muro
- pared

**Inside**

- wall
Results

Learner accuracy

- **Spanish**
  - 7 learners, 9 words
  - 9 learners, 10 words

- **Greek**
  - 9 learners, 10 words

- **Number of Examples**
  - 5
  - 10
  - 20
  - 30
  - 40
  - 50
  - 80
  - 120

- **Learner accuracy**
  - 0
  - 0.225
  - 0.45
  - 0.675
  - 0.9

- **with rules**
- **without rules**
Results

Learner accuracy

Number of Examples

Spanish
7 learners, 9 words

Greek
9 learners, 10 words
When shown rules, learners learn better and faster!

Spanish
7 learners, 9 words

Greek
9 learners, 10 words
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
Language Education

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

![Marathi map](image-url)

Indo-Aryan language family
Language Education

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

Indo-Aryan language family

Dravidian language family

“Classical language status”
Language Education

• Apply AutoLEX to extract grammar aspects for teaching two Indian languages outside of India.

Marathi - मराठी  
Kannada - ಕನ್ನಡ

Indo-Aryan language family
Dravidian language family
“Classical language status”
Language Education

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

![Marathi - मराठी](image1)
Indo-Aryan language family

![Kannada - ಕನ್ನಡ](image2)
Dravidian language family
“Classical language status”

-------

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

Carnegie Mellon University
Language Education

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

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

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

![Marathi - मराठी](image1.png)  ![Kannada - ಕನ್ನಡ](image2.png)

- 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
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

- **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

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

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

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

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

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

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

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

General Info
## AutoLEX: Selection of Grammar Aspects

<table>
<thead>
<tr>
<th>Lemma</th>
<th>Morphosyntactic Attributes</th>
<th>Fem</th>
<th>NA</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>कर (kar)</td>
<td>-</td>
<td>करण (karun)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>2;Plur</td>
<td>-</td>
<td>करा (kara)</td>
<td>-</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>Acc;Sing</td>
<td>-</td>
<td>-</td>
<td>करण्याचा (karanya)</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>Nom;Sing</td>
<td>केली (keli)</td>
<td>-</td>
<td>कराने (karane)</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>3;Past;Sing</td>
<td>केली (keli)</td>
<td>-</td>
<td>केले (kele)</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>1;Plur</td>
<td>-</td>
<td>-</td>
<td>करावे (karaave)</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>3;Sing</td>
<td>करावी (karavi)</td>
<td>-</td>
<td>करावे (karaave)</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>3;Nom;Past;Sing</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>3;Past;Plur</td>
<td>केळा (kela)</td>
<td>-</td>
<td>केली (keli)</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>Acc</td>
<td>-</td>
<td>केळामुळे (kelyamule)</td>
<td>-</td>
</tr>
<tr>
<td>कर (kar)</td>
<td>3;Plur;Pres</td>
<td>-</td>
<td>-</td>
<td>करतात (kartaat)</td>
</tr>
</tbody>
</table>

**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

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

General Info  Agreement
AutoLEX: Selection of Grammar Aspects

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

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

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

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

- In consultation with the curriculum designers

<table>
<thead>
<tr>
<th>ला (laa) suffix is used when:</th>
</tr>
</thead>
<tbody>
<tr>
<td>current word is in the neighborhood of the word= अटक (atak)</td>
</tr>
<tr>
<td>current word is in the neighborhood of the word= त्रुवत (sruvaat)</td>
</tr>
<tr>
<td>current word with Number = Sing</td>
</tr>
<tr>
<td>current word with Case = Acc</td>
</tr>
<tr>
<td>current word is the= oblique</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>चा (cha) suffix is used when:</th>
</tr>
</thead>
<tbody>
<tr>
<td>current word with Case = Acc</td>
</tr>
<tr>
<td>current word is in the neighborhood of the word= मृत्यु (mrityu)</td>
</tr>
<tr>
<td>current word is the= modifier</td>
</tr>
<tr>
<td>current word's head is the= compound</td>
</tr>
<tr>
<td>current word's head with Gender = Masc</td>
</tr>
</tbody>
</table>

**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

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

### 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

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

**General Info**  
**Agreement**  
**Word Order**  
**Affix Usage**  
**Word Usage**
# AutoLEX: Selection of Grammar Aspects

<table>
<thead>
<tr>
<th>word</th>
<th>Type</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>chocolate</td>
<td>性格 (chocolate)</td>
<td>Examples</td>
</tr>
<tr>
<td>pepper</td>
<td>情形 (meeth)</td>
<td>Examples</td>
</tr>
<tr>
<td>sugar</td>
<td>糖 (saakhar)</td>
<td>Examples</td>
</tr>
<tr>
<td>fodder</td>
<td>饲料 (chaara)</td>
<td>Examples</td>
</tr>
<tr>
<td>food</td>
<td>食物 (ann)</td>
<td>Examples</td>
</tr>
<tr>
<td>rice</td>
<td>米 (tandul)</td>
<td>Examples</td>
</tr>
<tr>
<td>nutrient</td>
<td>营养 (as)</td>
<td>Examples</td>
</tr>
<tr>
<td>liquor</td>
<td>酒 (daaru)</td>
<td>Examples</td>
</tr>
<tr>
<td>stock</td>
<td>货物 (share)</td>
<td>Examples</td>
</tr>
<tr>
<td>flour</td>
<td>面粉 (peeth)</td>
<td>Examples</td>
</tr>
<tr>
<td>lemon</td>
<td>柠檬 (limb)</td>
<td>Examples</td>
</tr>
<tr>
<td>chop</td>
<td>切 (chir)</td>
<td>Examples</td>
</tr>
<tr>
<td>produce</td>
<td>生产 (nirmiti)</td>
<td>Examples</td>
</tr>
<tr>
<td>vegetable</td>
<td>蔬菜 (bhaaji)</td>
<td>Examples</td>
</tr>
<tr>
<td>sibling</td>
<td>兄弟 (aahe)</td>
<td>Examples</td>
</tr>
<tr>
<td>brother</td>
<td>兄弟 (bhaau)</td>
<td>Examples</td>
</tr>
<tr>
<td>dad</td>
<td>爸爸 (baba)</td>
<td>Examples</td>
</tr>
</tbody>
</table>

## General Info

Teacher Perception of Automatically Extracted Grammar Concepts for L2 Language Learning

Chaudhary, Sampath, Sheshadri, Anastasopoulos, Neubig. In Submission.
<table>
<thead>
<tr>
<th>English Word</th>
<th>Marathi Words</th>
</tr>
</thead>
</table>
| minister (NOUN) | मुख्यमंत्री (mukhyamantri)  
पंतप्रधान (pantapradhan) |
| first (ADJ) | पहिली (pahili) 
पहिले (pahile) 
पहिल्यांदा (pahilyanda) 
फ़र्स्ट (first) 
सर्वप्रथम (sarvapratham) |
| new (ADJ) | नवनवीन (navanvin) 
नविन (navin) 
नवनवया (navanavaa) 
नवी (navi) 
न्यू (new) 
न्याय (navya) 
न्यूयॉर्क (neuyork) |
**AutoLEX**: Selection of Grammar Aspects

Example usages of **भाजी (bhaaji)**

1. तेलेग (tasegh) काळी (kaahi) ***भाज्यांचा (bhaajyancha)*** दर (dar) देखील (dekhi) कमी (kami) करण्यात (karanyaat) आला (aalaa) आहे (aahe)

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

3. विसलन (chirun) ठेवलेल्या (thevlelya) ***भाज्या (bhaiya)*** किंवा (kinwa) फक्त (fale) खाऊ (khaau) नका (nakaa)

4. do not eat raw ***vegetables*** or unpeeled fruit
AutoLEX: Extracting Language Descriptions
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

mr-en
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-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-en
  - kn parser
  - kn has NO SUD treebank/model

- mr-en
  - mr parser
  - mr has VERY SMALL and NOISY SUD treebank/model
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
AutoLEX: Extracting Language Descriptions

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

Paninian Treebanks

convert to SUD

kn

mr
**AutoLEX: Extracting Language Descriptions**

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

Paninian Treebanks

- train model for POS tags, lemmatization, morphological analysis
- convert to SUD

**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**

- train model for POS tags, lemmatization, morphological analysis
- convert to SUD

**Hindi**

- train model for dependency parse

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

- train model for POS tags, lemmatization, morphological analysis
- convert to SUD

Hindi

- train model for dependency parse

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
**AutoLEX: Extracting Language Descriptions**

- **kn-en**
  - kn parser

- **mr-en**
  - mr parser

**Steps:**
- **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

1. General Info
2. Agreement
3. Word Order
4. Affix Usage
5. Word Usage
AutoLEX: Quality Study

Affix Usage
What are the common suffixes for Kannada nouns and when is each used?
AutoLEX: Quality Study

Affix Usage

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

Accuracy

AutoLEX Baseline

[Bar chart showing accuracy for NOUN and VERB categories]

Automated Evaluation
AutoLEX: Quality Study

Affix Usage

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

NOUN: 9*/18 valid
VERB: 7/13 valid

Expert Evaluation

Automated Evaluation

Accuracy

<table>
<thead>
<tr>
<th></th>
<th>AutoLEX</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUN</td>
<td>80</td>
<td>40</td>
</tr>
<tr>
<td>VERB</td>
<td>60</td>
<td>20</td>
</tr>
</tbody>
</table>
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
Affix Usage

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

‘she’s doing it for herself’

→ ಮಾಡಲ + ಅ + ಸಹಜ + (ಅಂ+). ಸಹಲ + ವಲಸ + ವತ + ಎ

‘mAD’ ‘i’ ‘koL’
sandhi
reflexive

‘ut.Ta’ ‘id’ ‘A.L’ ‘e’
gender/number
present tense
AutoLEX: Teacher Perception Study
AutoLEX: Teacher Perception Study

- Evaluate the **relevance**, **utility**, and **presentation** of the materials
AutoLEX: Teacher Perception Study

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

- Recruited **12 Kannada teachers** and **5 Marathi 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
**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
**AutoLEX: Teacher Perception Study**

- How **relevant** are the extracted materials to the teaching needs?
**AutoLEX: Teacher Perception Study**

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

<table>
<thead>
<tr>
<th>Percentage of Materials</th>
<th>Kannada</th>
<th>Marathi</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- General Info  | Word Usage  | Suffix Usage  | Word Order  | Agreement  

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?

![Bar chart showing percentage of materials for General Info, Word Usage, Suffix Usage, Word Order, and Agreement for Kannada and Marathi languages.](chart.png)
AutoLEX: Teacher Perception Study

• How **relevant** are the extracted materials to the teaching needs?
AutoLEX: Teacher Perception Study

- How **relevant** are the extracted materials to the teaching needs?
AutoLEX: Teacher Perception Study

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

![Bar chart showing the percentage of materials in Kannada and Marathi for different categories: General Info, Word Usage, Suffix Usage, Word Order, Agreement. The chart indicates that the school's primary focus is on beginner-levels.](chart.png)

*Chaudhary, Sampath, Sheshadri, Anastasopoulos, Neubig. In Submission.*
**AutoLEX: Teacher Perception Study**

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage of Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Info</td>
<td>52.5%</td>
</tr>
<tr>
<td>Word Usage</td>
<td>70%</td>
</tr>
<tr>
<td>Suffix Usage</td>
<td>70%</td>
</tr>
<tr>
<td>Word Order</td>
<td>70%</td>
</tr>
<tr>
<td>Agreement</td>
<td>70%</td>
</tr>
</tbody>
</table>

- School’s primary focus on **beginner-levels**
- More beneficial for **advanced learners**

---

**Kannada**

- More relevant for **beginner-levels**
- More beneficial for **advanced learners**

**Marathi**

- Less relevant for **beginner-levels**
- Less beneficial for **advanced learners**
**AutoLEX**: Teacher Perception Study
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?

![Chart showing the percentage of teachers using extracted materials for different purposes: Lesson Preparation, Student self-exploration, and Student evaluation.

- Lesson Preparation: 90%
- Student self-exploration: 67.5%
- Student evaluation: 22.5%]
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?
• For what purposes would the teachers use the extracted materials for their teaching needs?

![Bar chart showing the percentage of teachers using extracted materials for lesson preparation, student self-exploration, and student evaluation across different grammar concepts.]

**AutoLEX: Teacher Perception Study**

- General Info
- Word Usage
- Suffix Usage
- Word Order
- Agreement

Percentage of teachers
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.

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.

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.
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

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
Other Applications

Evaluating the Morphosyntactic Well-formedness of Generated Texts

Pratapa, Anastasopoulos, Rikhwani, Chaudhary, Mortensen, Sheikh, Neubig, Tsvetkov. EMNLP 2020

Evaluating Context-Usage in MT models
Other Applications

Evaluating Context-Usage in MT models

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**

<table>
<thead>
<tr>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>En</strong> Look after her a lot. Okay. Any questions? Have we got her report? Yes, it's in the infirmary already.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context-aware baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>En</strong> Look after her a lot. Okay. Any questions? Have we got her report? Yes, it’s in the <strong>infirmary</strong> already.</td>
</tr>
<tr>
<td><strong>Fr</strong> Dorlotez-la. D’accord. Vous avez des questions ? On dispose de son rapport ? Oui, <strong>elle</strong> est déjà à l’infirmerie.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model w/ attention regularization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>En</strong> Look after her a lot. <strong>Okay</strong>. Any questions? Have we got her report? Yes, it’s in the infirmary already.</td>
</tr>
<tr>
<td><strong>Fr</strong> Dorlotez-la. D’accord. Vous avez des questions ? On dispose de son report ? <strong>Oui</strong>, il est déjà à l’hôpital</td>
</tr>
</tbody>
</table>

---

**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

Language Descriptions

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

<table>
<thead>
<tr>
<th>Human</th>
<th>Fr</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Context-aware baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>En</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model w/ attention regularization</th>
</tr>
</thead>
<tbody>
<tr>
<td>En</td>
</tr>
</tbody>
</table>

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

---

**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:

<table>
<thead>
<tr>
<th>ISO</th>
<th>Language</th>
<th>Treebank</th>
<th>Linguistic Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>en</td>
<td>English</td>
<td>EWT</td>
<td>General Information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agreement</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WordOrder</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CaseMarking</td>
</tr>
<tr>
<td>el</td>
<td>Greek</td>
<td>GDT</td>
<td>General Information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agreement</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WordOrder</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CaseMarking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Learn.Vocab</td>
</tr>
<tr>
<td>es</td>
<td>Spanish</td>
<td>GSD</td>
<td>General Information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agreement</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WordOrder</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CaseMarking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Learn.Vocab</td>
</tr>
<tr>
<td>mr</td>
<td>Marathi</td>
<td>SAM-EN</td>
<td>General Information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Learn.Vocab</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>WordOrder</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Suffix Usage</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Agreement</td>
</tr>
</tbody>
</table>
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/](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
--- | --- | ---
Anna | apple | eating
| is |

['tanduḷḷu' /'tanduḷḷu/ 'raw rice' | 'bhaat' /'bhaat/ 'cooked rice' |

Applications

Language Education and Documentation

https://www.autolex.co/
(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

Applications

What's Next?

https://www.autolex.co/
What’s Next?

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

AutoLEX: Automatic Language Explorer

- Subject
- Object
- Verb

https://www.autolex.co/
(Low-resource) Language Analysis

AutoLEX: Automatic Language Explorer

Subject | Object | Verb
--- | --- | ---
Anna | apple | eating
pera | is | Verb

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
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**
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)

https://www.autolex.co/