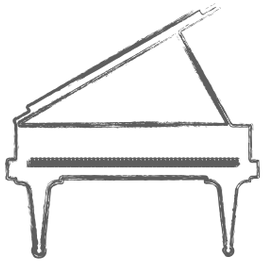
Uncovering Cross-linguistic Structural Transfer in L2 Learning

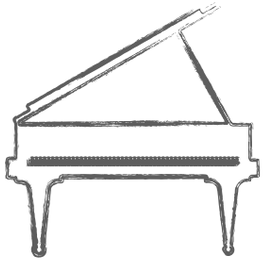
Zoey Liu
University of Florida

Emily Prud'hommeaux
Boston College

Joshua Hartshorne
Boston College

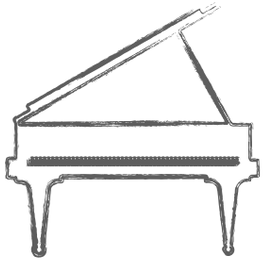


Trailer of This Talk



Trailer of This Talk

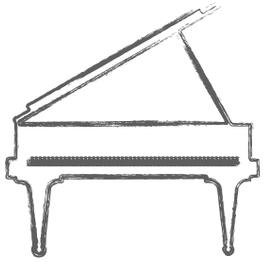
Bigger Picture & Background



Trailer of This Talk

Bigger Picture & Background

Experiments



Trailer of This Talk

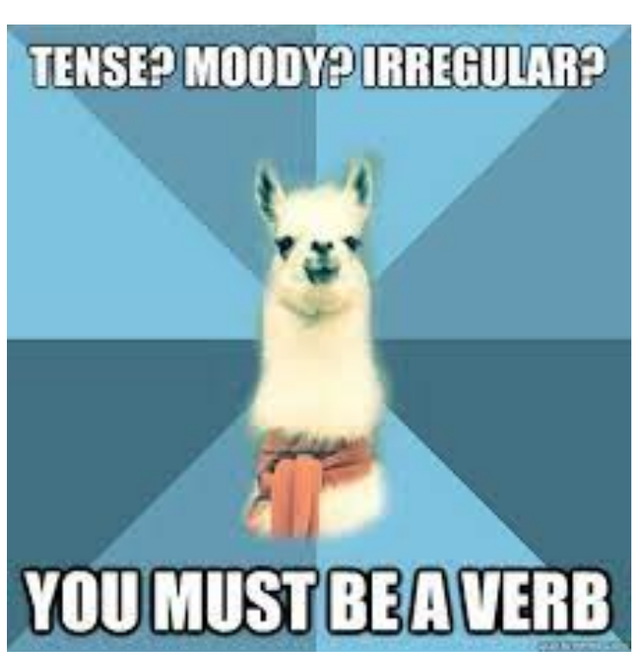
Bigger Picture & Background

Experiments

Keep Looking ahead


CLASSMATE CAN'T GRASP SPANISH CONJUGATION
TRY HUNGARIAN INFLECTION

L2

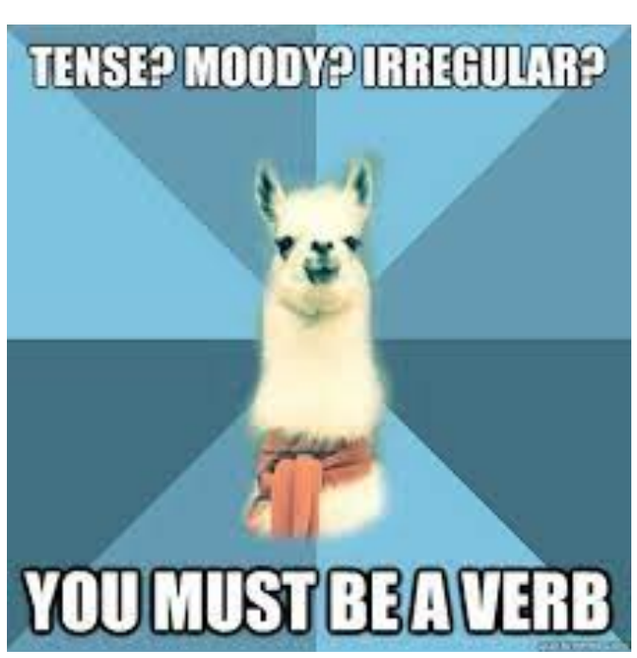

TENSE? MOODY? IRREGULAR?
YOU MUST BE A VERB

L1


CLASSMATE CAN'T GRASP SPANISH CONJUGATION
TRY HUNGARIAN INFLECTION

L2

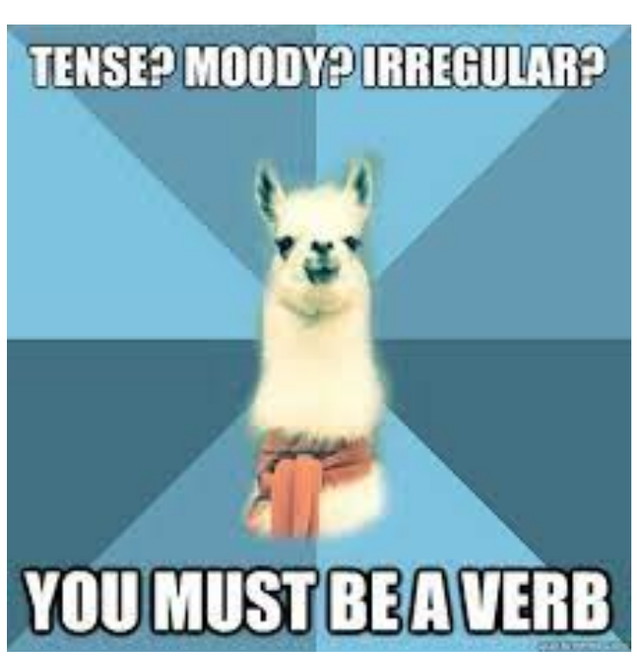
Structural Transfer from L1 —> L2



L1

L2

Crosslinguistic Structural Transfer from L1 —> L2

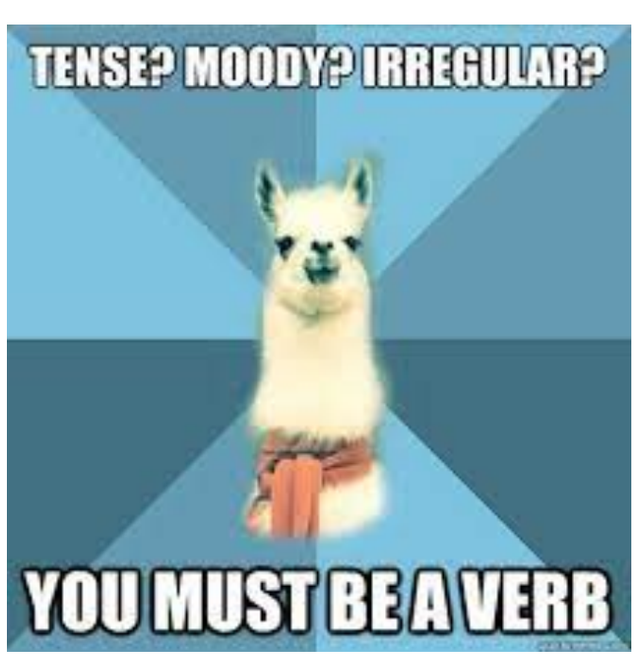


L1

L2

Crosslinguistic Structural Transfer from L1 —> L2

Are there reliable L1 effects *independent of* L2?



L1

L2

Crosslinguistic Structural Transfer from L1 —> L2

Are there reliable L1 effects *independent of* L2?

Spiegelman, 1980

Crosslinguistic Structural Transfer from L1 —> L2

Are L1 effects restricted to specific parts of morphosyntax?

Spiegelman, 1980

Previously…

Are there reliable L1 effects *independent of* L2?

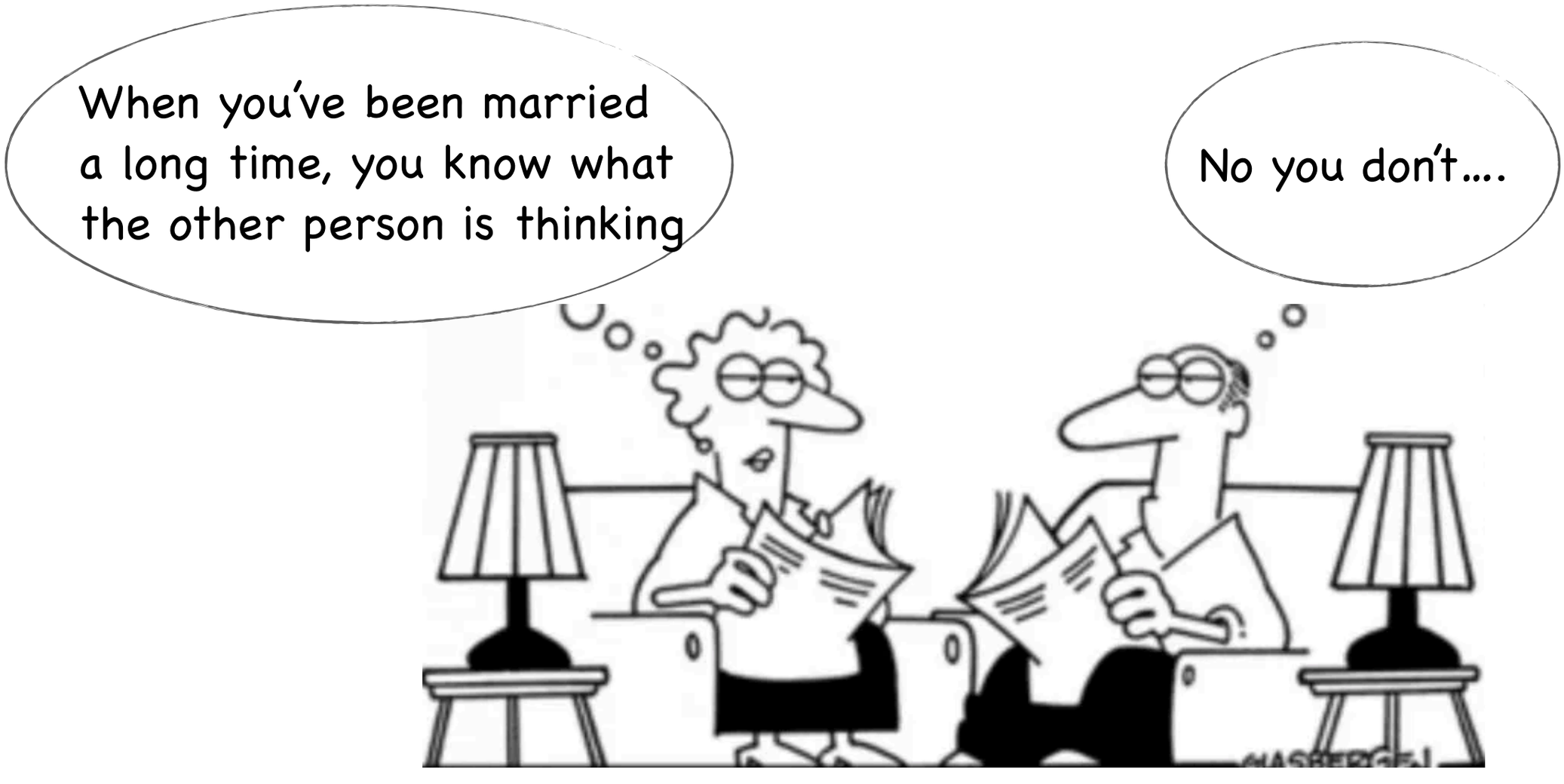
Are L1 effects restricted to specific parts of morphosyntax?

- Focus on narrowly-defined phenomena
- Attend to a handful of language pairs
- N of learners studied is relatively small

Data-driven Approach

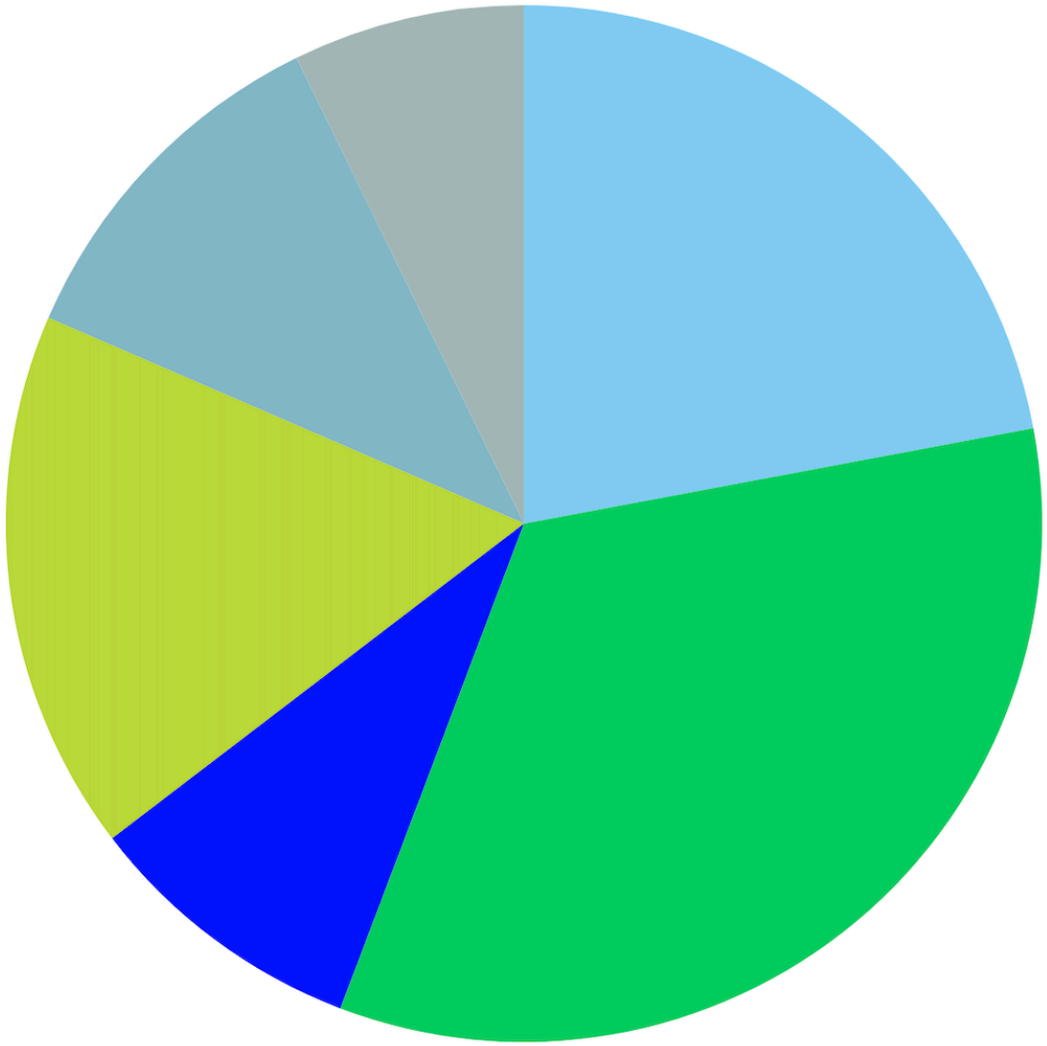
Are there reliable L1 effects *independent of* L2?

Are L1 effects restricted to specific parts of morphosyntax?

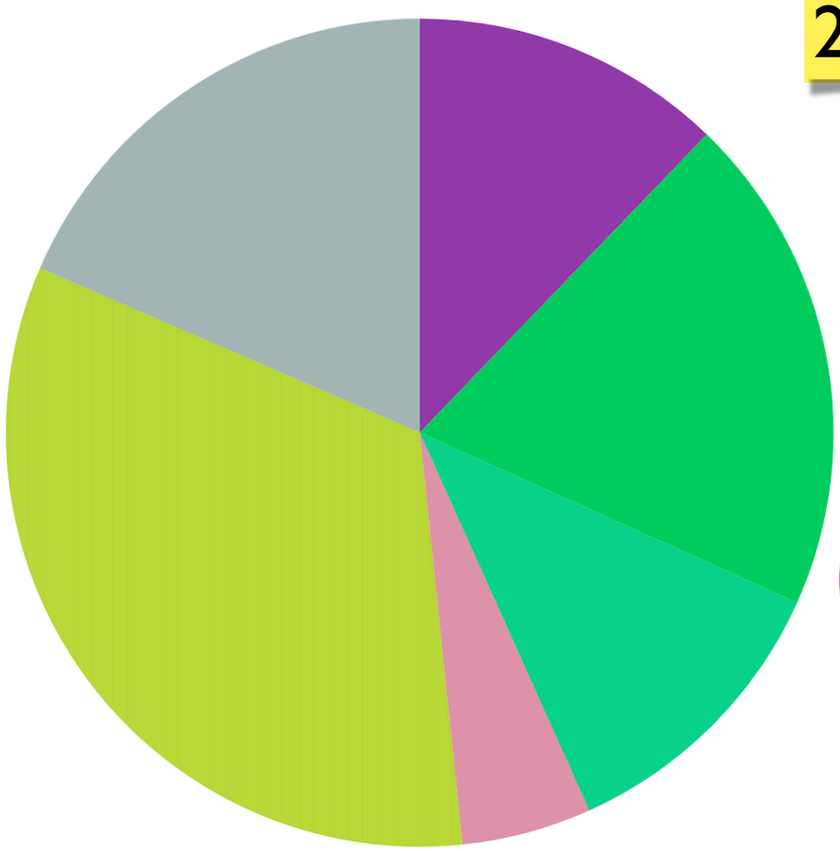


Arabic Czech Dutch English Finnish French German Hungarian Indonesian
Italian Japanese Korean Lithuanian Mandarin Norwegian Polish Portuguese Russian
Serbian Spanish Swedish Turkish Ukranian Vietnamese Other

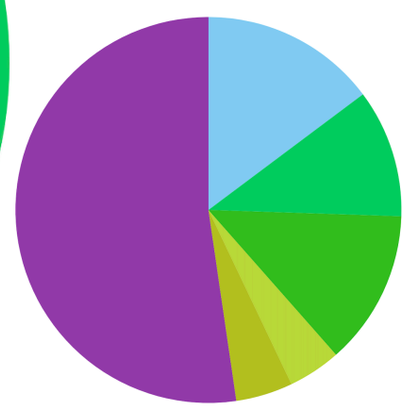
275 L1-L2 pairs



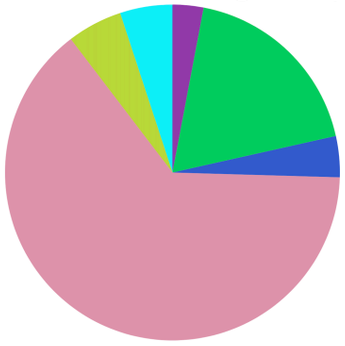
English (N=61,634)



Korean (N=30,028)



Spanish (N=8,935)



Czech (N=5,390)

Chinese (N=2,632)

Portuguese (N=2,216)

Croatian (N=2,099)

Norwegian (N=1,335)

Italian (N=812)

Latvian (N=807)

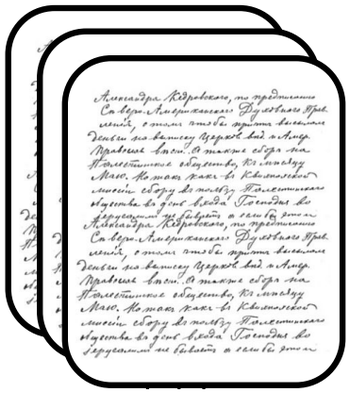
German (N=647)

Finnish (N=419)

Icelandic (N=48)

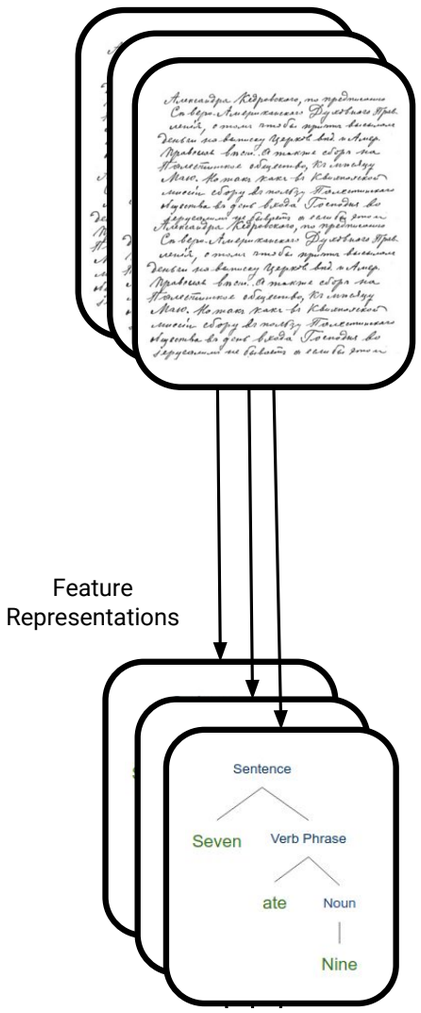
Native Language Identification as a Tool

Raw Corpora

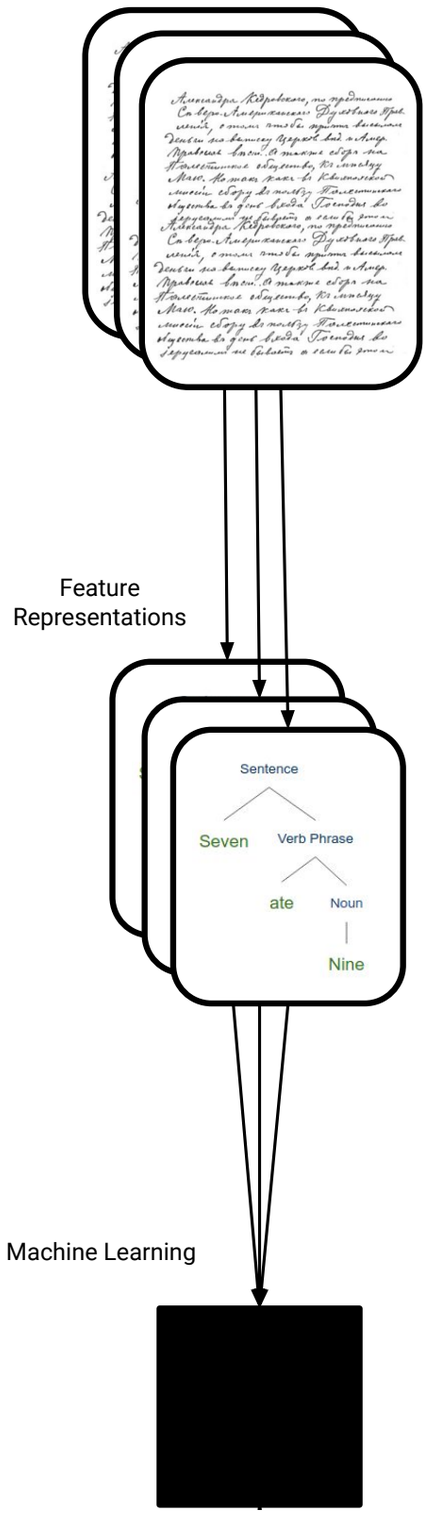


Native Language Identification as a Tool

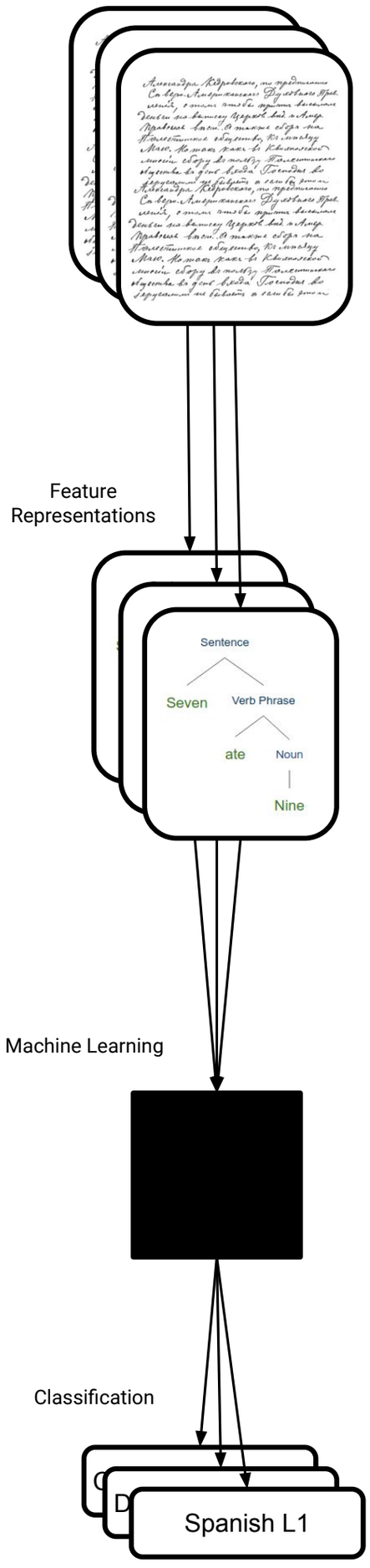
Native Language Identification as a Tool

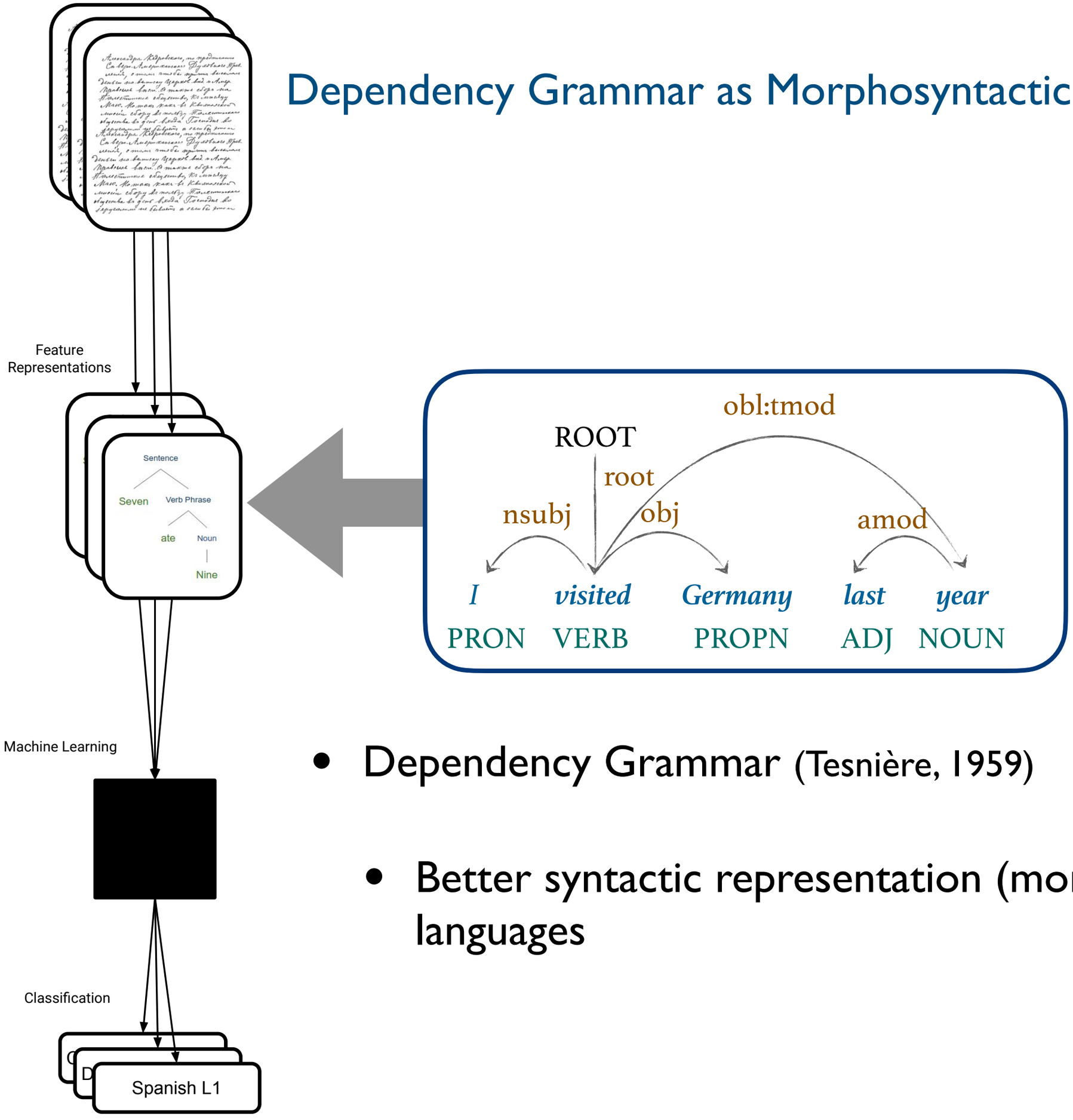


Native Language Identification as a Tool



Native Language Identification as a Tool

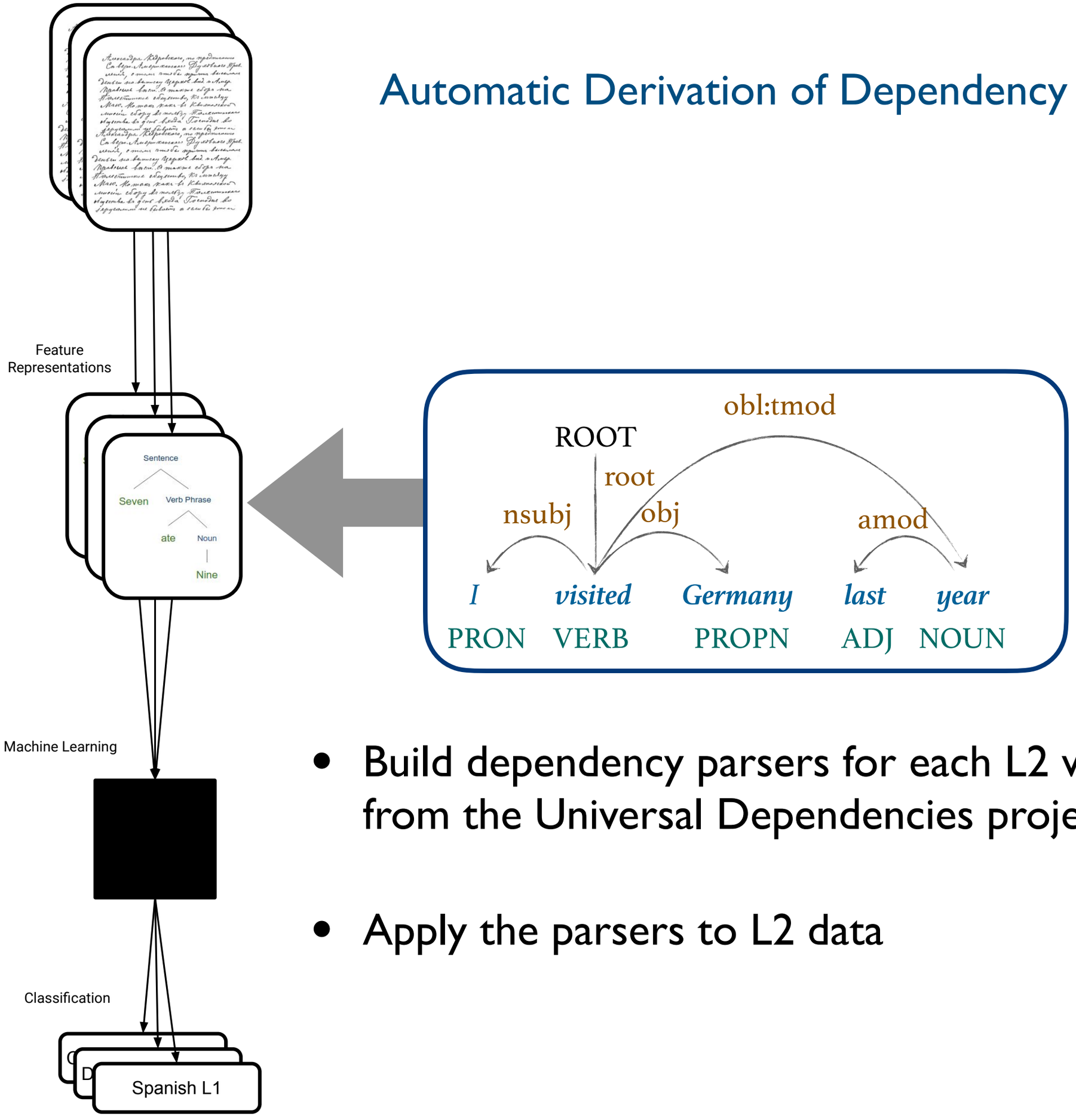




Dependency Grammar as Morphosyntactic Representation

- Dependency Grammar (Tesnière, 1959)
  - Better syntactic representation (more flexible) across languages

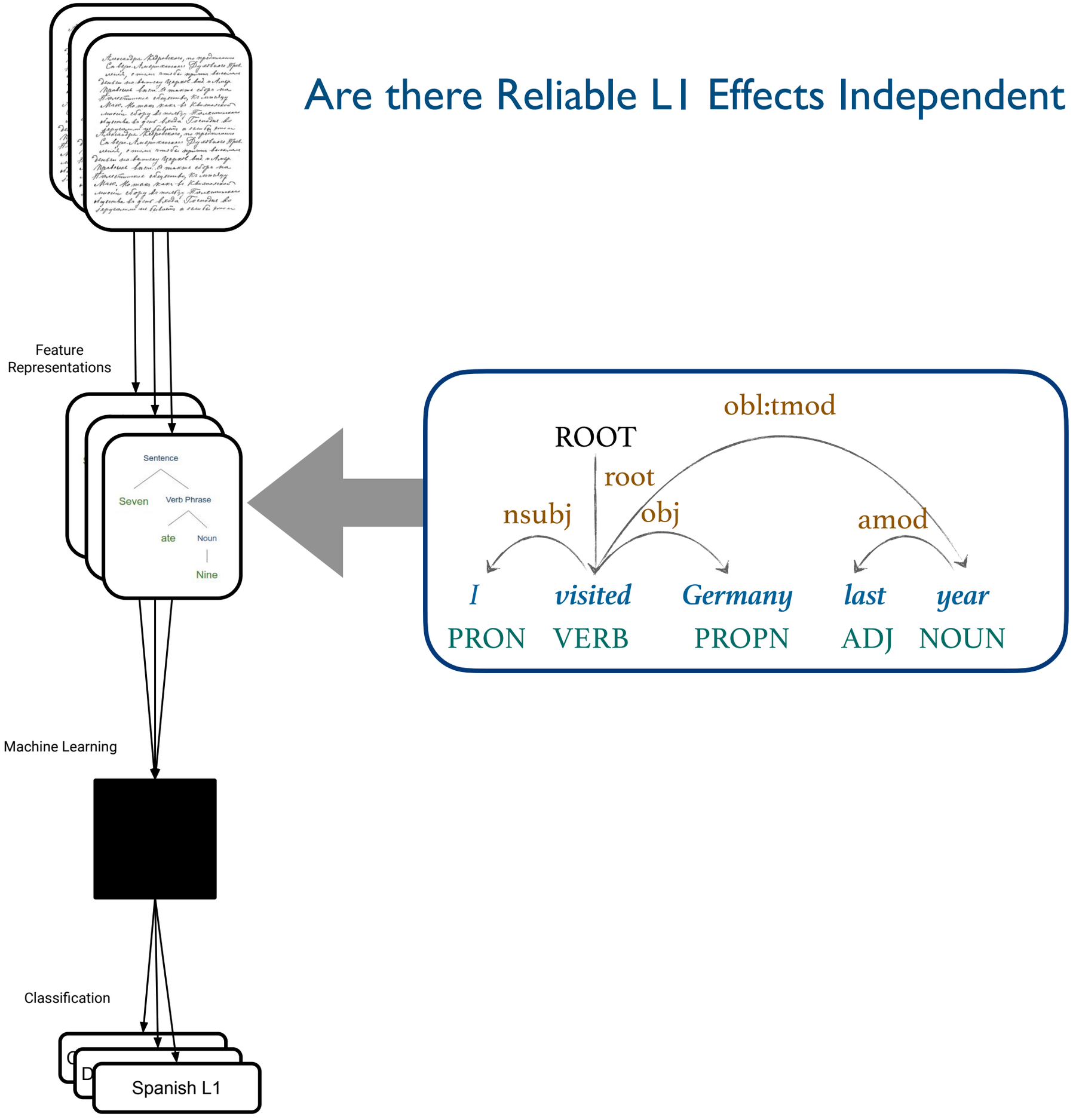
Automatic Derivation of Dependency Structures



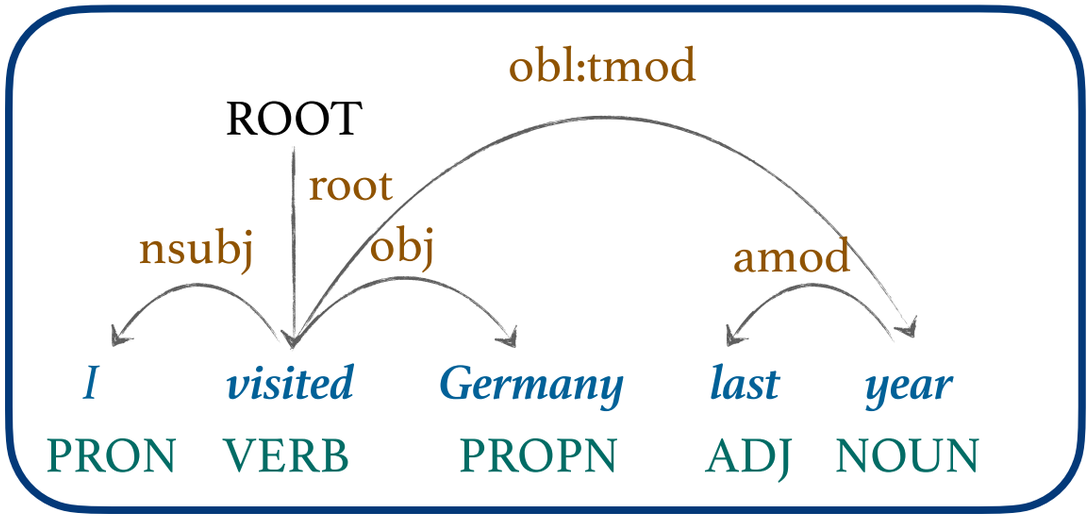
- Build dependency parsers for each L2 with training data from the Universal Dependencies project

- Apply the parsers to L2 data

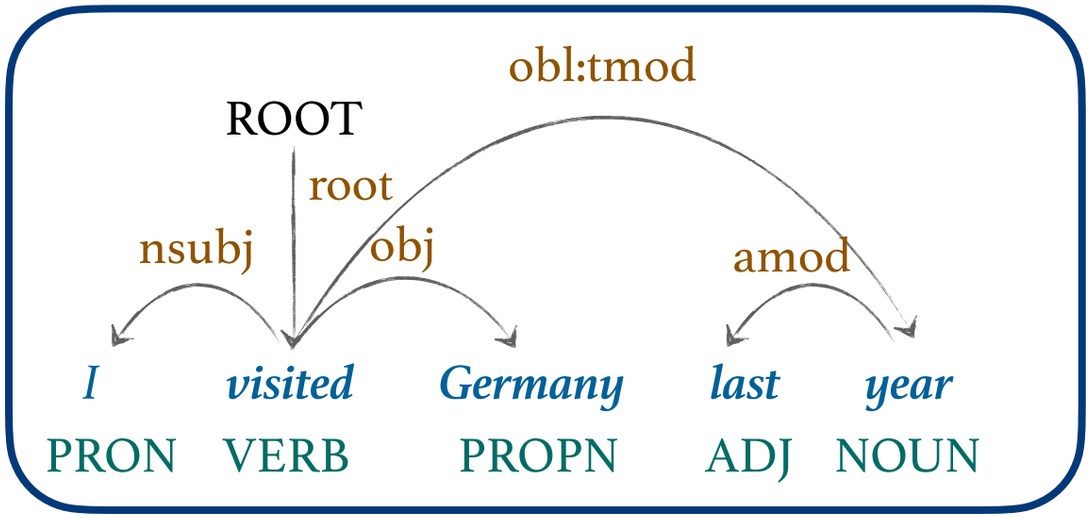
Are there Reliable L1 Effects Independent of L2?



Classifying L1 based on Trigram Features



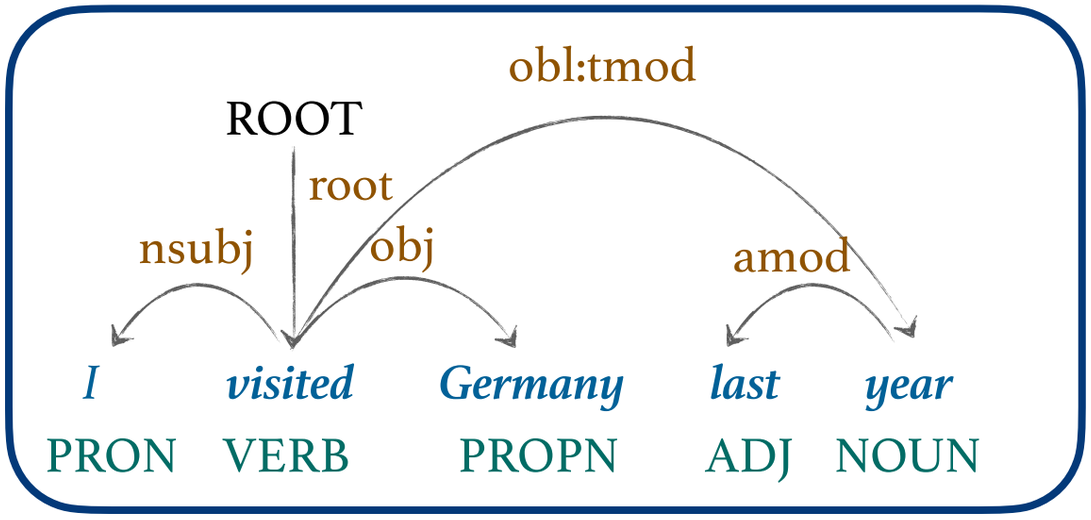
Classifying L1 based on Trigram Features



Dependency trigrams

nsubj + root + obj + root + obj + amod + obj + amod + obl:tmod + ...

Classifying L1 based on Trigram Features



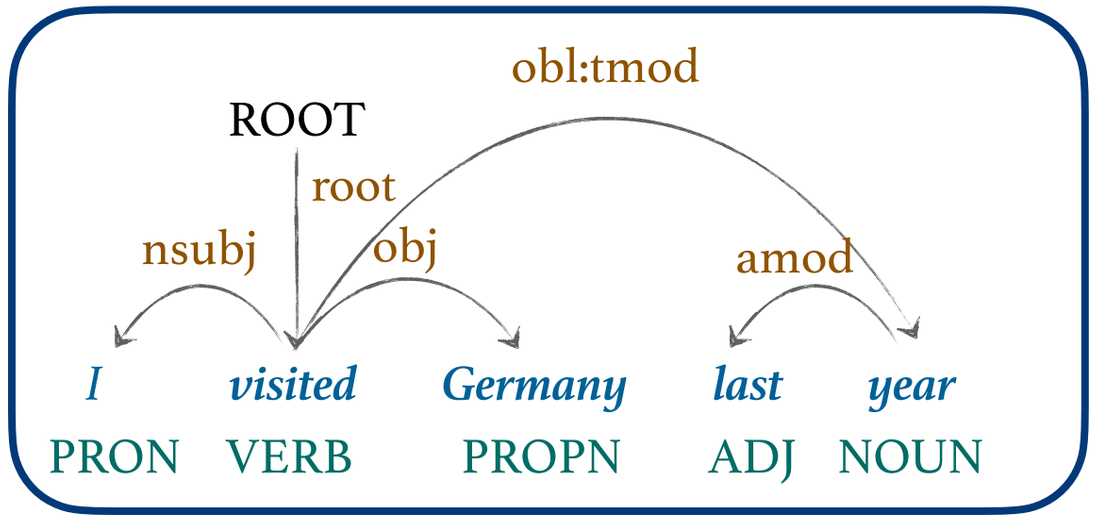
Dependency trigrams

nsubj + root + obj + root + obj + amod + obj + amod + obl:tmod + ...

POS trigrams

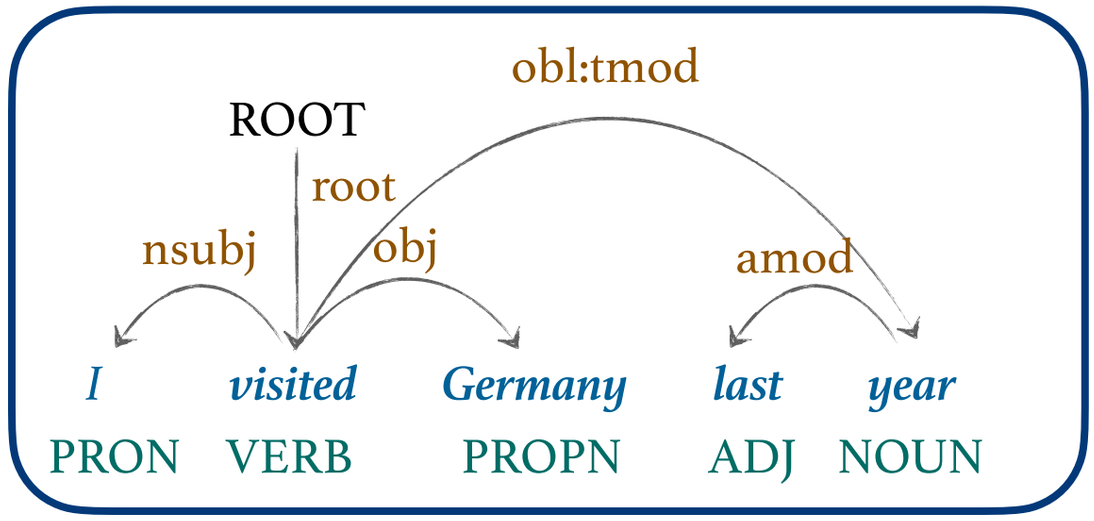
PRON + VERB + PROPN + VERB + PROPN + ADJ + VERB + PROPN + NOUN + ...

Classifying L1 based on Trigram Features



nsubj + root + obj + … + obj + amod + obl:tmod + PRON + VERB + PROPN + … + VERB + PROPN + NOUN

Classifying L1 based on Trigram Features



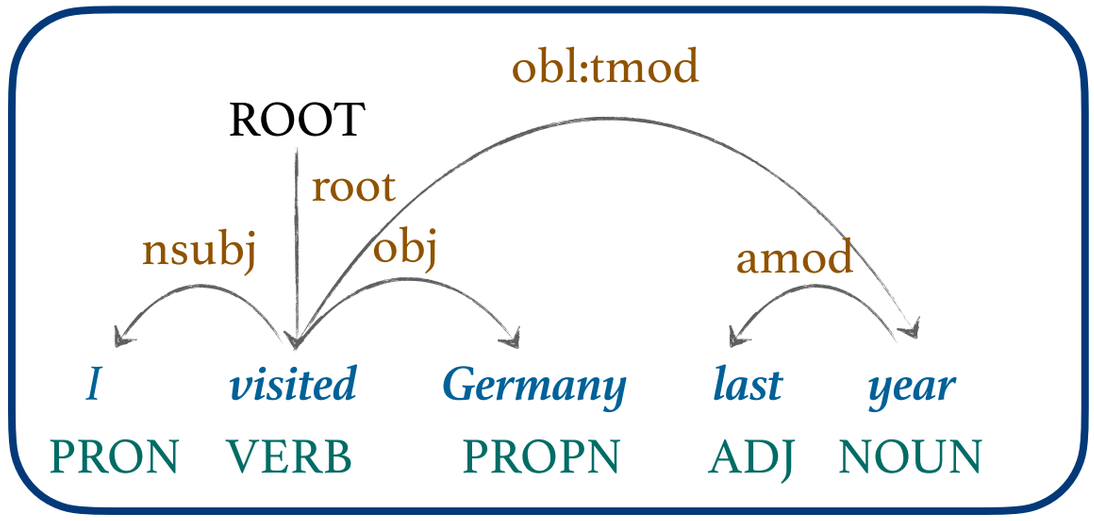
nsubj + root + obj + ... + obj + amod + obl:tmod + PRON + VERB + PROPN + ... + VERB + PROPN + NOUN

tf-idf

Vectors

(.3, .7, 0, .3, .9, . . .)

Classifying L1 based on Trigram Features



nsubj + root + obj + ... + obj + amod + obl:tmod + PRON + VERB + PROPN + ... + VERB + PROPN + NOUN

.....

.....

tf-idf

Vectors

(.3, .7, 0, .3, .9, ...)

(.1, .9, .5, .6, .8, ...)

(.5, .5, .3, .2, 1, ...)

Classifying L1 based on Trigram Features

Vectors

(.3, .7, 0, .3, .9, …)

(.1, .9, .5, .6, .8, …)

(.5, .5, .3, .2, 1, …)

…

Classifying L1 based on Trigram Features

Vectors		L1s
(.3, .7, 0, .3, .9, …)		Mandarin
(.1, .9, .5, .6, .8, …)	→	German
(.5, .5, .3, .2, 1, …)		Japanese
…		…

- Ridge classifier
  - A linear classifier able to perform multinomial classification
  - Does not assume that errors are normally distributed
  - Fast computation (why we chose this classifier)

Classifying L1 based on Trigram Features

Vectors		L1s
(.3, .7, 0, .3, .9, …)		Mandarin
(.1, .9, .5, .6, .8, …)	→	German
(.5, .5, .3, .2, 1, …)		Japanese
…		…

- Three baselines
  - Majority: predicting the most frequent L1
  - Random: randomly predicting L1s
  - Stratified: predicting L1s based on their distribution on the learner corpora

Classifying L1 based on Trigram Features

Vectors		L1s
(.3, .7, 0, .3, .9, …)		Mandarin
(.1, .9, .5, .6, .8, …)	→	German
(.5, .5, .3, .2, 1, …)		Japanese
…		…

Model	*Precision*	*Recall*	*F1*
Majority	0.01	0.04	0.01
Random	0.08	0.01	0.02
Stratified	0.10	0.04	0.04
Ridge	0.41	0.41	0.41

There is consistent transfer effect across L1-L2 pairs

Vectors		L1s
(.3, .7, 0, .3, .9, …)		Mandarin
(.1, .9, .5, .6, .8, …)	→	German
(.5, .5, .3, .2, 1, …)		Japanese
…		…

Model	*Precision*	*Recall*	*F1*
Majority	0.01	0.04	0.01
Random	0.08	0.01	0.02
Stratified	0.10	0.04	0.04
Ridge	0.41	0.41	0.41

But what is Transferred?

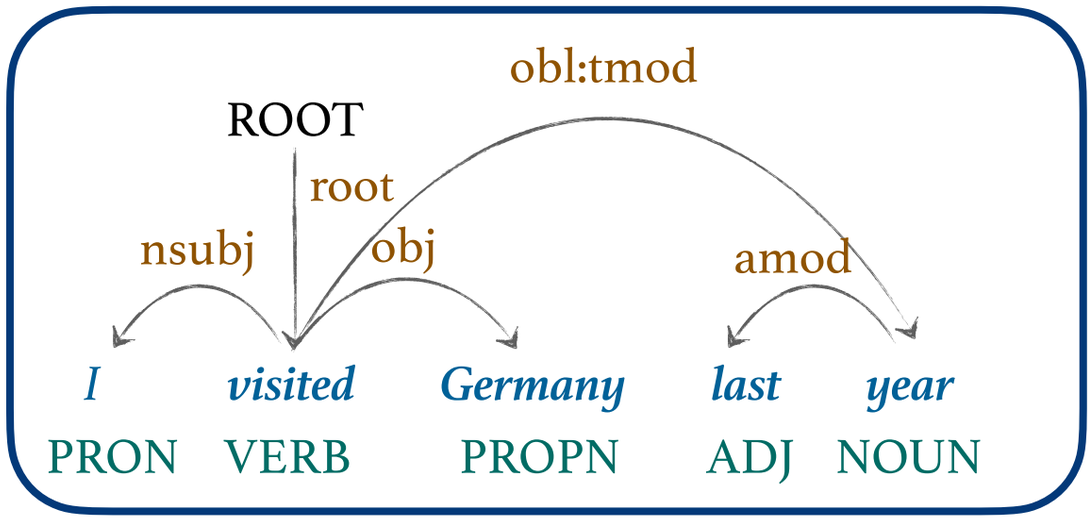
Model	*Precision*	*Recall*	*F1*
Majority	0.01	0.04	0.01
Random	0.08	0.01	0.02
Stratified	0.10	0.04	0.04
Ridge	0.41	0.41	0.41

Are L1 effects restricted to specific parts of morphosyntax?

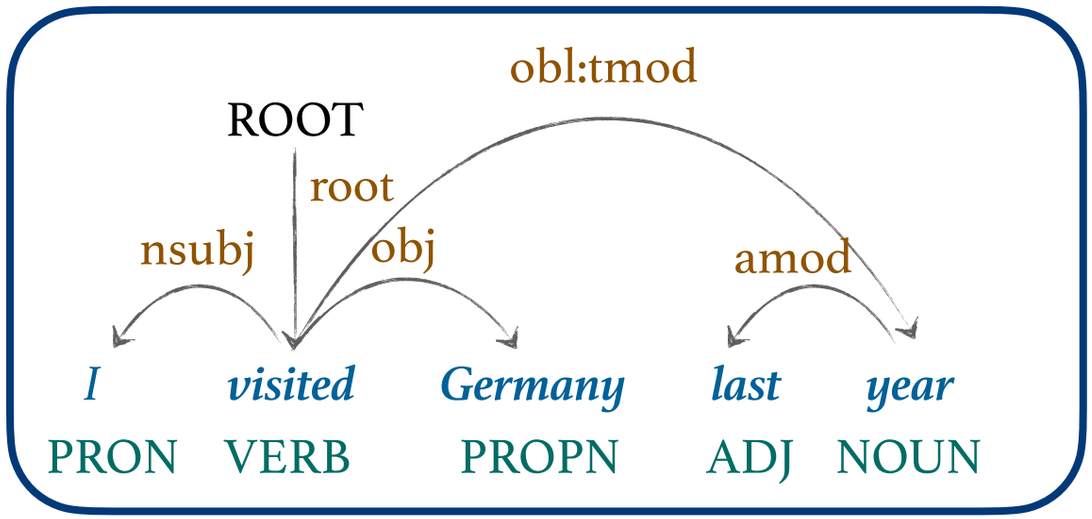
Vectors		L1s
(.3, .7, 0, .3, .9, …)		Mandarin
(.1, .9, .5, .6, .8, …)	→	German
(.5, .5, .3, .2, 1, …)		Japanese
…		…

Model	*Precision*	*Recall*	*F1*
Majority	0.01	0.04	0.01
Random	0.08	0.01	0.02
Stratified	0.10	0.04	0.04
Ridge	0.41	0.41	0.41

Are L1 effects restricted to specific parts of morphosyntax?

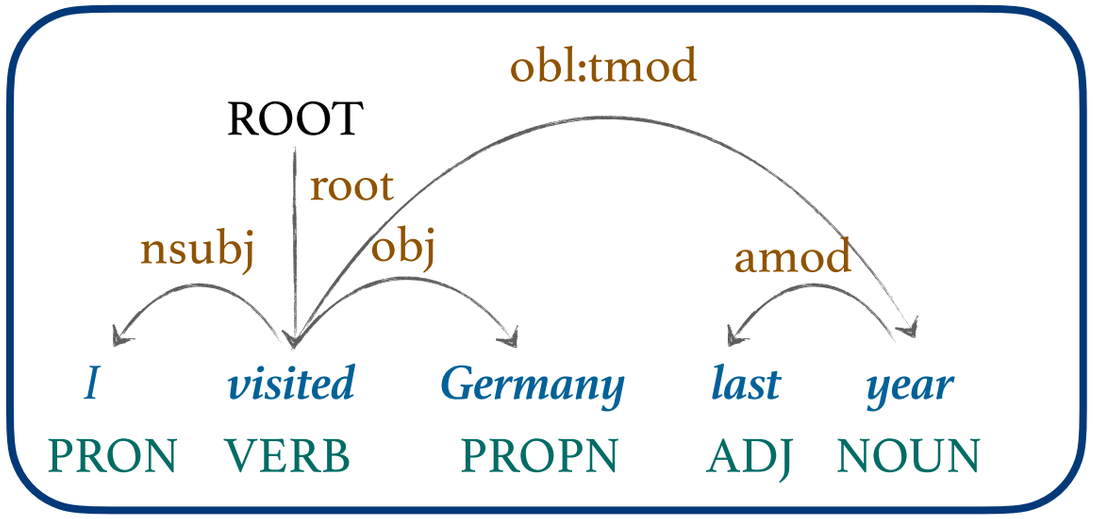


Are L1 effects restricted to specific parts of morphosyntax?



Trigram features

Are L1 effects restricted to specific parts of morphosyntax?



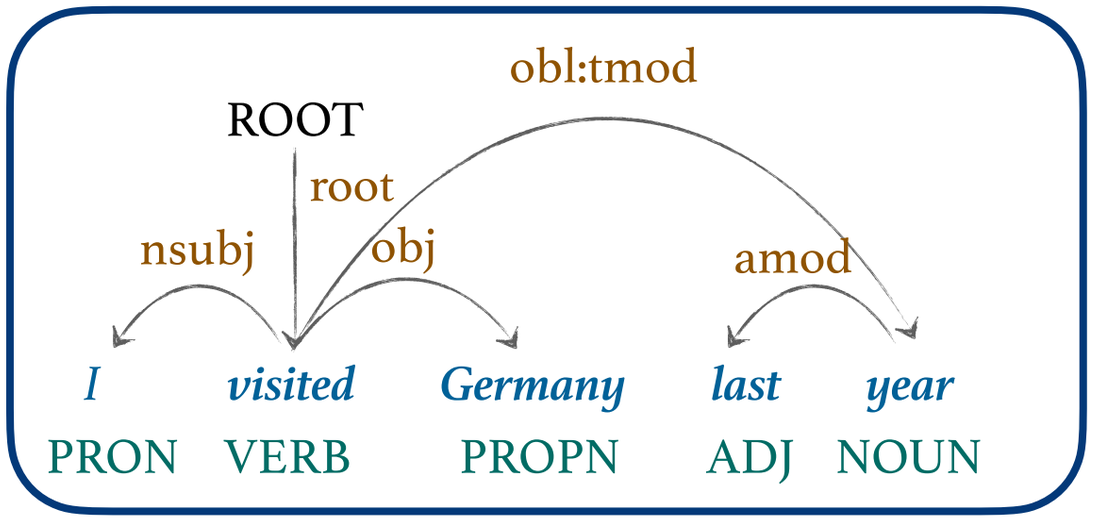
Trigram features

hard to interpret

Hand-curated features

- Raw texts features:
  - Number of sentences and words
- Morphological features
  - *Distribution* of verbs and auxiliaries
  - *Distribution* of aspect, number, mood, etc
  - Etc …
- Dependency parse features
  - Average depth of parse tree
  - Proportion of head-final dependencies
  - *Distribution* of dependency relations
  - *Distribution* of main constituent orders
  - …

Are L1 effects restricted to specific parts of morphosyntax?



Trigram features

hard to interpret

Hand-curated features

- Raw texts features:
  - Number of sentences and words
- Morphological features
  - *Distribution* of verbs and auxiliaries
  - *Distribution* of aspect, number, mood, etc
  - Etc …
- Dependency parse features
  - Average depth of parse tree
  - Proportion of head-final dependencies
  - *Distribution* of dependency relations
  - *Distribution* of main constituent orders
  - …

$$H(X) = -\sum_{i=1}^{n} P(x_i) log P(x_i)$$

Are L1 effects restricted to specific parts of morphosyntax?

Model	Precision	Recall	F1
Trigrams	0.41	0.41	0.41
Hand-curated feature set			

much less info than trigrams

Are L1 effects restricted to specific parts of morphosyntax?

Model	Precision	Recall	F1
Trigrams	0.41	0.41	0.41
Hand-curated feature set	0.26	0.31	0.23

much less info than trigrams

Majority	0.01	0.04	0.01
Random	0.08	0.01	0.02
Stratified	0.10	0.04	0.04

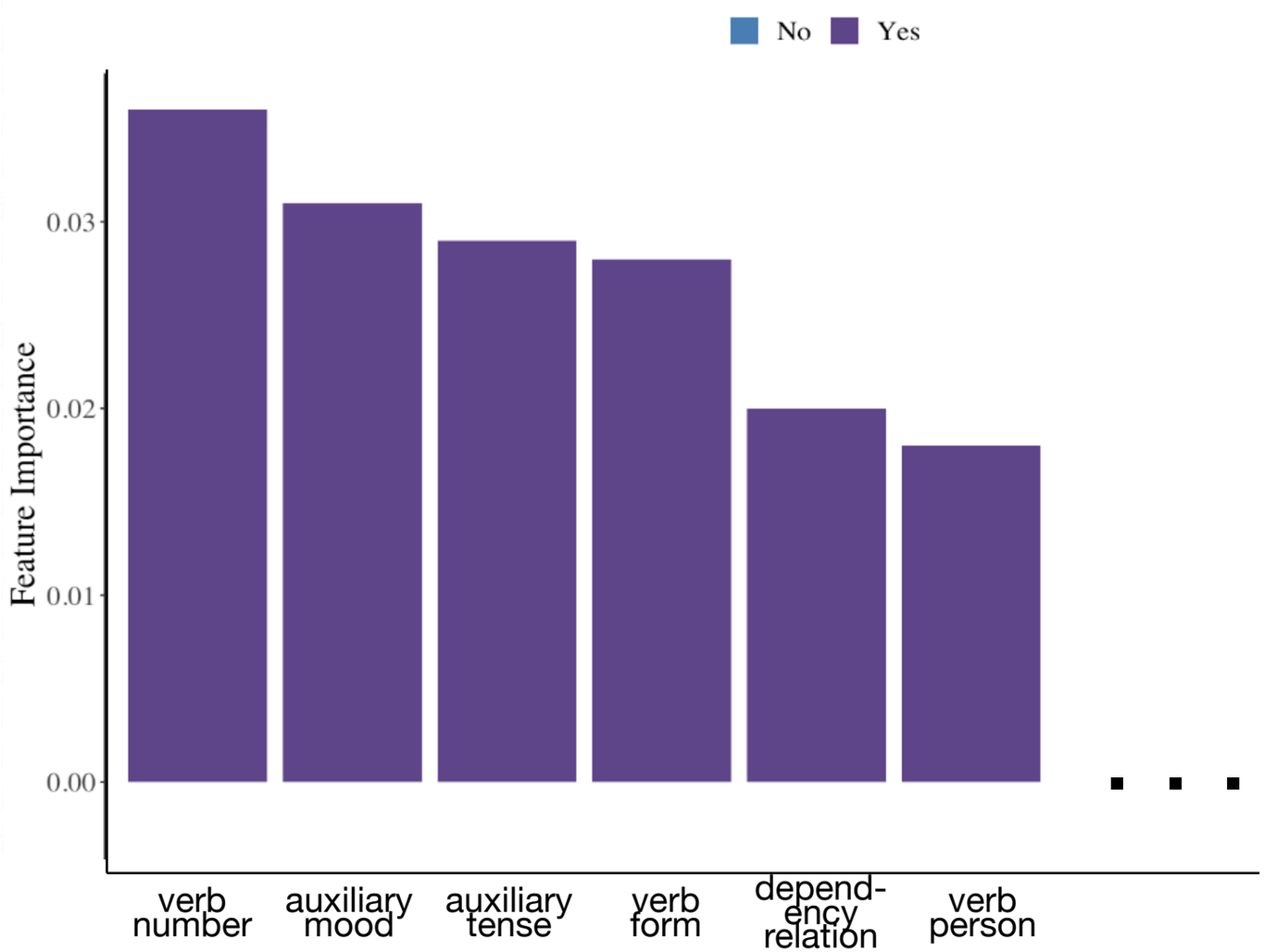
Which Features are Predictive?

Which Features are Predictive?

F*eature importance* of each feature x

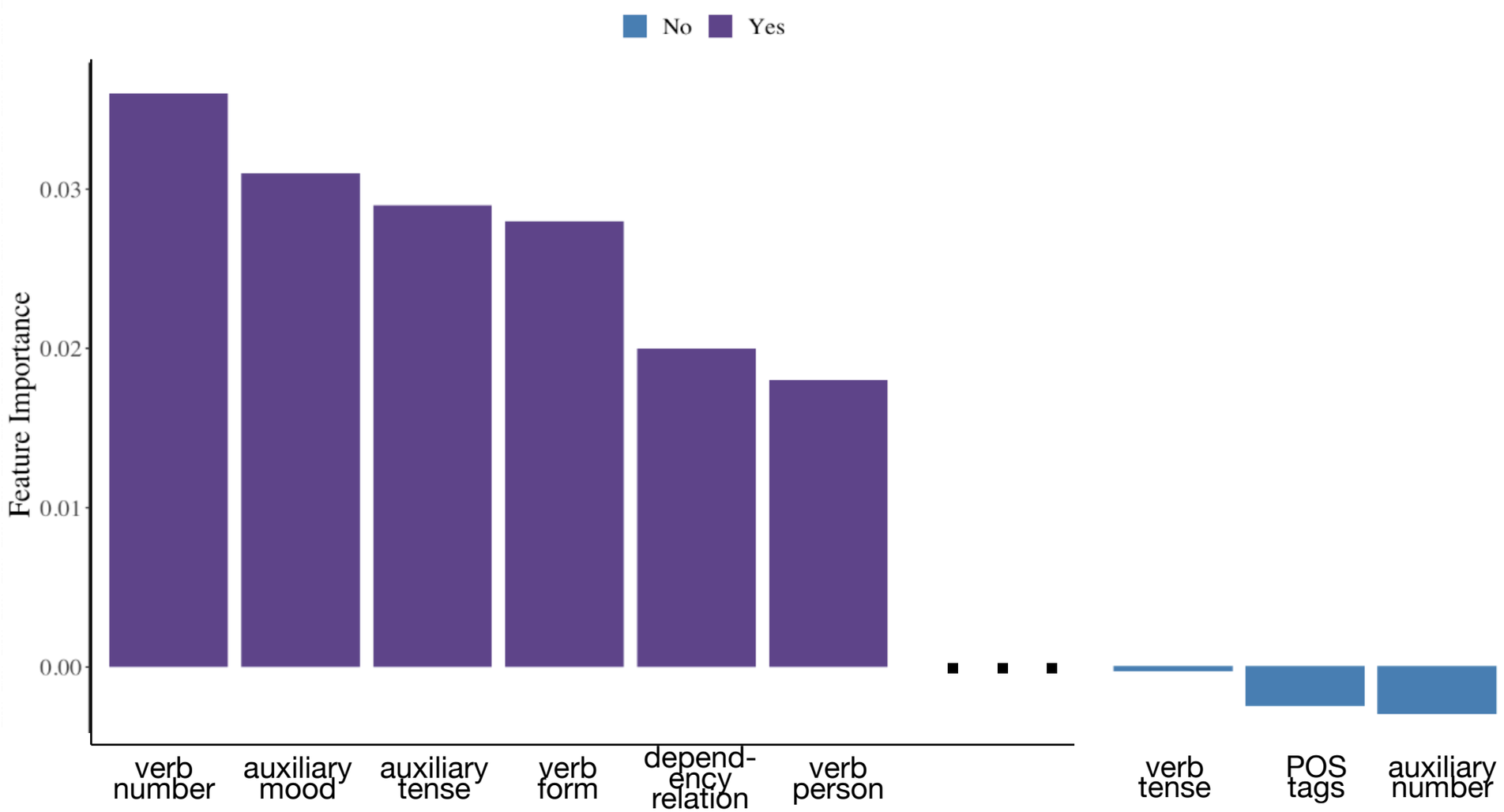
Which Features are Predictive?

F*eature importance* of each feature x = (F1 score including *x*) - (F1 score excluding *x*)



Which Features are Predictive?

F*eature importance* of each feature x = (F1 score including *x*) - (F1 score excluding *x*)



Are there reliable L1 effects *independent of* L2?

Are L1 effects restricted to specific parts of morphosyntax?

Limitations & Ongoing Work

Are there reliable L1 effects *independent of* L2?

Are L1 effects restricted to specific parts of morphosyntax?

- Feature sets are too large (need dimensionality reduction)
- Features aren't always *that* interpretable
- Feature sets are probably incomplete
- Single feature set for all L2 is tricky

Thank you!

Questions?