Measuring constructs in SLA

Hakyung Sung
University of Hawaii at Manoa

October 14, 2026

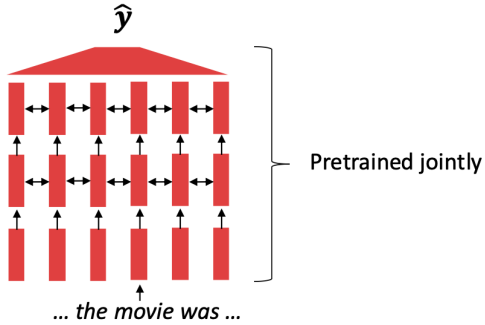
Topics to cover

- Review (20 mins)
- Lab: POS tagging, Dependency parsing (1 hour)
- Measuring constructs in SLA (20 mins)
- Presentations

Review:
Dependency parsing and tagging

Foundational models

Exceptionally effective at building strong probability distributions over language that we can sample from large corpus ▸ See Appendix



[This model has learned how to represent entire sentences through pretraining]

What can we learn from reconstructing the input?

The University of Hawaii at Mānoa is located in _____, Hawaii.
(Location)

What can we learn from reconstructing the input?

I put ___ fork down on the table.
(Grammar)

What can we learn from reconstructing the input?

The woman walked across the street, checking for the traffic over ___ shoulder.
(co-reference)

What can we learn from reconstructing the input?

I went to the ocean to see the fish, turtle, seals, and ___.
(Semantics)

What can we learn from reconstructing the input?

Overall, the value I got from the two hours watching it was the sum total of popcorn and the drink. The movie was ___. (Sentiment)

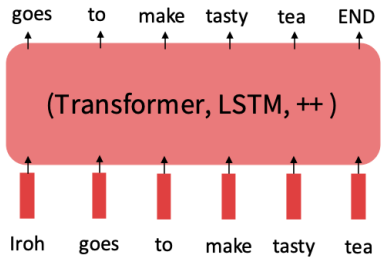
What can we learn from reconstructing the input?

I was thinking about the sequence that goes 1, 1, 2, 3, 5, 8, 13, 21, ___. (Math)

The pretraining paradigm

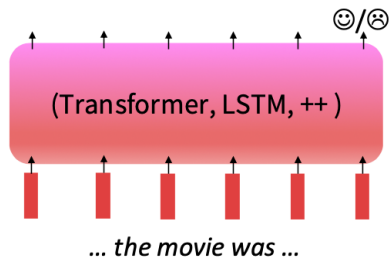
Step 1: Pretrain (on language modeling)

Lots of text; learn general things!



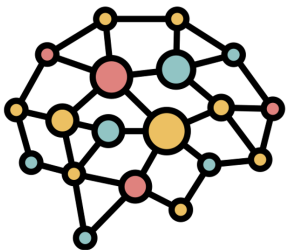
Step 2: Finetune (on your task)

Not many labels; adapt to the task!





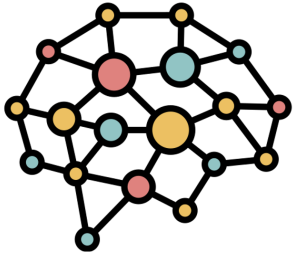
- Leveraging LLMs for different linguistic questions
  - Dependency parsing
  - Semantic role labeling
  - Semantic/discourse analysis



foundational model
(e.g., BERT, GPT2...)

Supervised
learning

Instructed
fine-tuning



foundational model
(e.g., BERT, GPT2...)

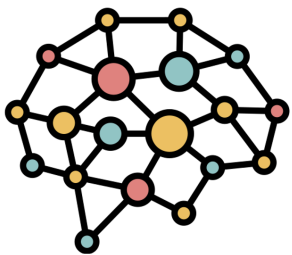
Supervised
learning



Korean
Dependency parser

```
# sent_id = KL-fi-14-1
# text = 내 생각으로 행복은 마음 먹기에 달려있다.
1   내   나+의     PRON    NP+JKG  PronType=Prs    2   nmod:poss   _   _
2   생각으로   생각+으로 ADV NNG+JKB _   6   obl _   _
3   행복은   행복+은   NOUN    NNG+JX  _   6   nsubj   _   _
4   마음 마음 NOUN    NNG _   5   obj _   _
5   먹기에   먹+기+에  VERB    VV+ETN+JKB  _   6   obl _   _
6   달려있다   달리+어+있+다  VERB    VV+EC+VX+EF _   0   root    _   SpaceAfter=No
7   .   .   PUNCT   SF  _   6   punct   _   _
```

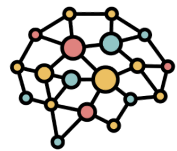
Instructed
fine-tuning



foundational model
(e.g., BERT, GPT2...)

Supervised
learning

Instructed
fine-tuning

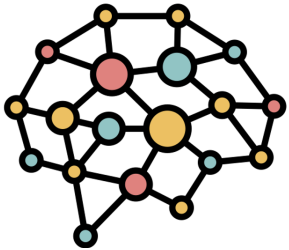


Zero shot

Parse the sentence using Universal Dependencies.
For each token, return its ID, text, head ID, and dependency relation. Use head ID 0 for the root. Output only the table.

Supervised learning

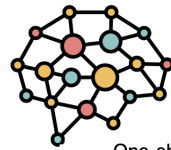
```
# sent_id = KL-fi-14-1
# text = 내 생각으로 행복은 마음 먹기에 달려있다.
1   내   나+의     PRON    NP+JKG  PronType=Prs    2   nmod:poss   _   _
2   생각으로   생각+으로  ADV NNG+JKB _   6   obl _   _
3   행복은   행복+은   NOUN    NNG+JX  _   6   nsubj   _   _
4   마음 마음 NOUN    NNG _   5   obj _   _
5   먹기에   먹+기+에  VERB    VV+ETN+JKB  _   6   obl _   _
6   달려있다   달리+어+있+다  VERB    VV+EC+VX+EF _   0   root    _   SpaceAfter=No
7   .   .   PUNCT   SF  _   6   punct   _   _
```



foundational model
(e.g., BERT, GPT2...)

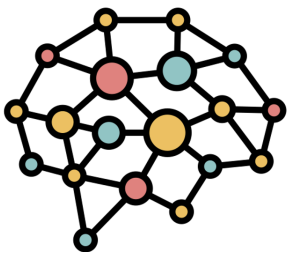
Instructed fine-tuning

evaluation



One-shot prompting

Parse the sentence using Universal Dependencies.
For each token, return its ID, text, head ID, and dependency relation. Use head ID 0 for the root. Output only the table.
e.g., "She runs." → She: 2/nsubj, runs: 0/root



foundational model
(e.g., BERT, GPT2...)

Supervised learning

+

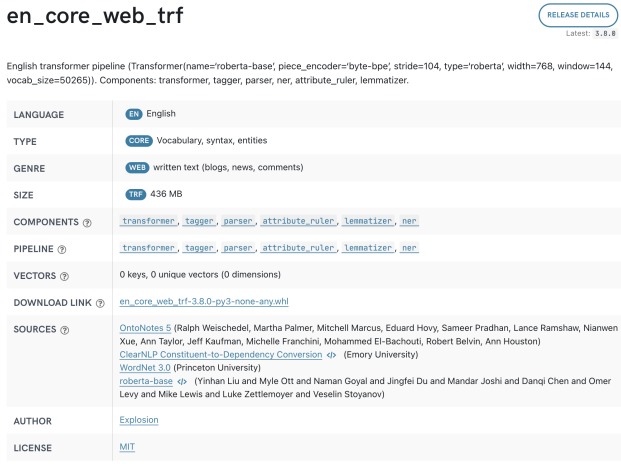
-

Instructed fine-tuning

+

-

cf. spaCy: English



en_core_web_trf

RELEASE DETAILS

Latest: 3.8.0

English transformer pipeline (Transformer(name='roberta-base', piece_encoder='byte-bpe', stride=104, type='roberta', width=768, window=144, vocab_size=50265)). Components: transformer, tagger, parser, ner, attribute_ruler, lemmatizer.

LANGUAGE	EN English
TYPE	CORE Vocabulary, syntax, entities
GENRE	WEB written text (blogs, news, comments)
SIZE	TRF 436 MB
COMPONENTS ?	transformer, tagger, parser, attribute_ruler, lemmatizer, ner
PIPELINE ?	transformer, tagger, parser, attribute_ruler, lemmatizer, ner
VECTORS ?	0 keys, 0 unique vectors (0 dimensions)
DOWNLOAD LINK ?	en_core_web_trf-3.8.0-py3-none-any.whl
SOURCES ?	OntoNotes 5 (Ralph Weischedel, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, Mohammed El-Bachouti, Robert Belvin, Ann Houston) ClearNLP Constituent-to-Dependency Conversion (Emory University) WordNet 3.0 (Princeton University) roberta-base (Yinhan Liu and Myle Ott and Naman Goyal and Jingfei Du and Mandar Joshi and Danqi Chen and Omer Levy and Mike Lewis and Luke Zettlemoyer and Veselin Stoyanov)
AUTHOR	Explosion
LICENSE	MIT

https://spacy.io/models/en

cf. spaCy: Japanese

ja_core_news_trf

RELEASE DETAILS

Latest: `3.8.0`

Japanese transformer pipeline (Transformer(name='cl-tohoku/bert-base-japanese-char-v2', piece_encoder='char', stride=160, type='bert', width=768, window=216, vocab_size=6144)). Components: transformer, morphologizer, parser, ner.

LANGUAGE	JA Japanese
TYPE	CORE Vocabulary, syntax, entities
GENRE	NEWS written text (news, media)
SIZE	TRF 320 MB
COMPONENTS ?	`transformer`, `morphologizer`, `parser`, `attribute_ruler`, `ner`
PIPELINE ?	`transformer`, `morphologizer`, `parser`, `attribute_ruler`, `ner`
VECTORS ?	0 keys, 0 unique vectors (0 dimensions)
DOWNLOAD LINK ?	ja_core_news_trf-3.8.0-py3-none-any.whl
SOURCES ?	UD Japanese GSD v2.8 </> (Omura, Mai; Miyao, Yusuke; Kanayama, Hiroshi; Matsuda, Hiroshi; Wakasa, Aya; Yamashita, Kayo; Asahara, Masayuki; Tanaka, Takaaki; Murawaki, Yugo; Matsumoto, Yuji; Mori, Shinsuke; Uematsu, Sumire; McDonald, Ryan; Nivre, Joakim; Zeman, Daniel) UD Japanese GSD v2.8 NER </> (Megagon Labs Tokyo) cl-tohoku/bert-base-japanese-char-v2 (Inui Laboratory, Tohoku University)
AUTHOR	Explosion
LICENSE	CC BY-SA 3.0

`https://spacy.io/models/ja`

Lab:
POS tagging, Dependency parsing

Assignment 3

- Due: 10/23 (Rubrics will be on Lamakū)
- English POS tagging exercise
- English dependency parsing exercise
- Next lab: Linguistic annotation

One application of tagging/parsing in corpus research

So... why do we tag the corpus (i.e., raw texts)?

Measuring constructs in SLA

Why do we annotate learner corpora?

Our goal

To turn learner language into *empirical* evidence about SLA.

- Learner corpora contain observable language production.
- SLA research asks questions about *theoretical* constructs.
- Annotation connects the text to measurable features.

A construct is not a directly observable feature

Complexity, accuracy, fluency, and proficiency are theoretical constructs.

To investigate them, we need to decide:

- What exactly does the construct mean?
- Which linguistic features represent it?
- How will those features be measured?

Operationalization

Translating a theoretical definition into explicit measurement procedures.
(e.g., how do you measure love? cf. *Is love measurable?*)

Development involves more than accuracy

The complexity–accuracy–fluency (CAF) framework distinguishes several aspects of learner performance (Norris & Oretega, 2009).

Accuracy	Conformity to target-language norms.
Fluency	Efficiency and automatization of language use.
Complexity	Expansion and elaboration of linguistic resources.

Learners may use a wider range of structures even while continuing to make errors (accuracy != complexity)

What do we mean by “complex”?

A structure has sometimes been called complex because it is:

1. Structurally elaborate.
2. Cognitively demanding.
3. Acquired relatively late.
4. Associated with more proficient learners.

A problem

These properties may be related, but they do not represent the same construct.

What do we mean by “complex”?

1. **Structurally elaborate $\rightarrow$ Complexity**

Bulté et al. (2025) draw a conceptual distinction here.

2. **Cognitively demanding $\rightarrow$ Difficulty**

Complexity

Structural complexity

The quantity and variety of linguistic components and the relationships among them.

- Quantity: How many components are present?
- Variety: Which component types occur?
- Relationships: How are components connected?
- Hierarchy: How deeply are they embedded?

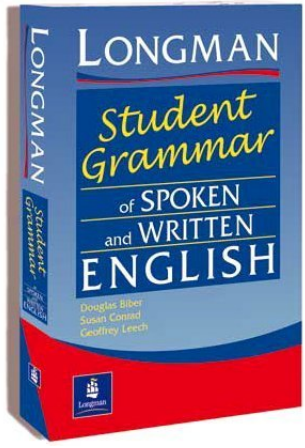
Complexity: More on presentations

- Presentation 1: Kim & Lu (2024) - Syntactic complexity
- Presentation 2: Biber & Larsson (2026) - Grammatical complexity

Complexity dimension	Constitutional	Taxonomical	Hierarchical	Organizational
Lexicon	*Mean word length*	MATTR of lemmas *Entropy-based measures* *Yule's K*		
Morphology	*Mean number of morphological operations*	MCI		
Syntax	Mean words per phrase Mean phrases per clause Mean clauses per T-unit or AS-unit *Number of nodes in tree* *Number of dependents per syntactic unit* *Mean words per (finite) clause, T-unit, AS-unit*		Normalized rate of occurrence of dependent clauses and phrases *Maximum depth of syntactic tree* *Dependency distance* *Hierarchical distance*	MATTR of dependency relations or syntactic structures *Mean sentence similarity score*
Lexico-grammatical		MATTR of lexico-grammatical units		

Table 7: Core and noncore measures of complexity. Adapted from Bulté et al. (2025, p. 559).

Note. Noncore measures are shown in italics, which are not necessarily less appropriate but tend to address highly theory-specific research questions, have relatively limited empirical support, or be difficult to implement in practice.


LONGMAN
Student Grammar
of SPOKEN
and WRITTEN
ENGLISH
Douglas Biber
Susan Conrad
Geoffrey Leech
Longman

Finite Dependent Clause

Syntactic Function	Complexity Feature (`Tag`)	Examples
Adjective Complement	thatcls+jcomp	*We're happy that the hunger strike has **ended**.*
Adjective Complement	xtrapos+thatcl+jcomp	*It's horrible that he **put** up with Claire's nagging.*
Adjective Complement	whcls+jcomp	*I'm unsure why he **likes** it.*
Adjective Complement	xtrapos+whcls+jcomp	*It was incredible what had **happened** to them.*
Adverbial	finitecls+advl	*She finished her work before the deadline **arrived**.*
Noun Phrase Complement	thatcls+ncomp	*The conclusion that global temperatures are **rising** is alarming.*
Noun Phrase Complement	xtrapos+thatcls+ncomp	*It was a surprise that they would **need** more time.*
Noun Phrase Modifier	finitecls+rel	*The idea which **comes** up with great frequency involves several steps.*

https://lcr-ads-lab.github.io/LxGrTagger-Documentation/

Difficulty

Difficulty

The cognitive demands of learning, comprehending, or producing a linguistic structure.

Difficulty depends on:

- Structure: elaboration, salience, transparency.
- Context: frequency and familiarity (psycholinguistic norms).
- Learner: knowledge, experience, cognitive resources.

Structural measures do not establish difficulty

Annotation allows us to describe properties of a text.

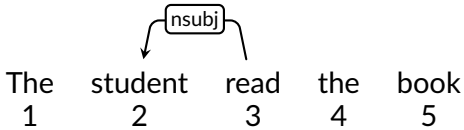
Claims about difficulty need evidence about learners or language users, such as:

- Processing time.
- Learning outcomes or subjective ratings ($\approx$ proficiency).

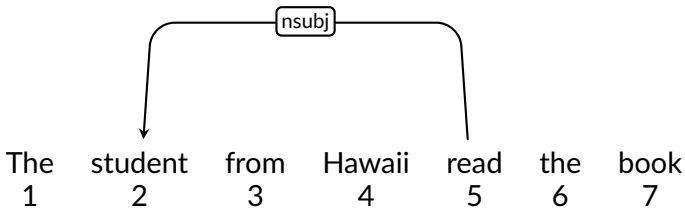
Structural measures do NOT establish difficulty

Example

Dependency distance measures a structural relationship.



Dependency distance: $|3 - 2| = 1$



Dependency distance: $|5 - 2| = 3$

Why POS tagging?

Identifying grammatical categories

POS tagging assigns a category to each token (e.g., noun, verb, derivative morphemes).

The students wrote detailed reports.

`The/DET students/NOUN wrote/VERB detailed/ADJ reports/NOUN`

We can now retrieve grammatical categories across many learner texts.

What can POS tags help us investigate?

- Frequencies and distributions of word classes.
- Category sequences (e.g., `ADJ + NOUN`)
- Changes in category use across learners, tasks, or time points.

Categories alone are not enough

POS tags identify what kind of word a token is. They do not explicitly identify which words are syntactically connected.

Why dependency parsing?

Identifying syntactic relationships

Dependency parsing identifies heads and dependents and labels their grammatical relationships.

The students wrote detailed reports.

- *students* $\rightarrow$ *wrote*: subject (`nsubj`).
- *reports* $\rightarrow$ *wrote*: object (`obj`).
- *detailed* $\rightarrow$ *reports*: adjectival modifier (`amod`).

From syntactic relationships to measures

Dependency annotation supports the investigation of:

- Elaboration: number and types of modifiers.
- Subordination: patterns of dependent clauses.
- Hierarchy: depth in a dependency tree.
- Distance: separation between heads and dependents.

Annotation connects theory to evidence

Research question

Do more proficient English learners use more *modifiers per noun* in writing?

Annotation connects theory to evidence

Research question

Do more proficient English learners use more *modifiers per noun* in writing?

1. Define
2. Annotate
3. Calculate
4. Validate

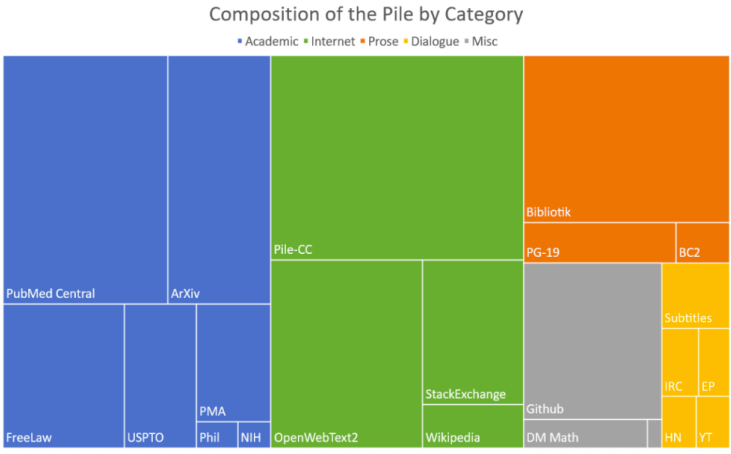
Midway presentation

Midway presentation and report

- Mark the date 10/28
- Any questions?
- Attendance?

Appendix

Where does this data come from?



Model	Training Data
BERT	BookCorpus, English Wikipedia
GPT-1	BookCorpus
GPT-3	CommonCrawl, WebText, English Wikipedia, and 2 book databases (“Books 1” and “Books 2”)
GPT-3.5+	Undisclosed

*Acknowledgment: Some figures/slides based on materials by CS224N @ Stanford University.