A Community-Centered Workflow for Morphological Annotation

- Sara Court¹ Maria Copot² Noah Diewald¹ Stephanie Antetomaso¹ Micha Elsner¹
- ¹The Ohio State University
- ²Université de Paris Cité, LLF

Background

- Navigate researcher-community collaboration ethically
 - Community members as collaborators, not research subjects
 - Give community members maximum agency
- Engaging the community is complicated by current **software and practices**
 - Technical barrier to use existing software
 - Need for linguistic training for e.g. applying morphological labels

cat-s would='ve chase-d mice cat-PL COND=PERF chase-PST PL\mouse

Diminished usefulness for understudied languages

- Theoretical issues
 - Early commitment to an analysis
 - Assumption of segmental patterns
- Practical issues
 - Suboptimal use of human time
 - Requires linguistic training

- Requires non-trivial technological ability
- Low integration with other software prevents linguists from making use of rapid advances in NLP and CL

- Word-and-Paradigm Morphology
 - A more intuitive annotation process and software interface
- Computational methods and machine learning
 - Automating the initial steps of analysis
 - Suggesting most informative data points to analyse next
 - Automatically extending the annotation and analysis to new data

Word and Paradigm morphology

 Establishing parallel relationships of form and meaning between words



Word and Paradigm morphology

• Establishing parallel relationships of form and meaning between words



• The word is the smallest unit

- Defined by its place in a system of contrasts, not by its component parts
- Concepts like **paradigm cell** or **lexeme** are emergent
 - The result of establishing contrasts and similarities between words along different dimensions

The annotator's task: decide whether two items in context share a property or not

- 1. Do two words belong to the same **lexeme**?
- **2.** Are pairs of words belonging to different lexemes in an analogous relationship?

sink \sim sunk ring \sim rung wink \sim winked

This allows us to identify the **cells** that define the system

Goal: paradigmatic morphological analysis from documentary corpora

- Computational automation of the initial steps of the analysis
- The annotator corrects the initial analysis
 - Simple task: same or different?
- Active learning
 - Updates the analysis after each annotator correction
 - Directs the annotator's attention to the most informative data points

The workflow

- List of lemmas
 - + corpus
 - + machine-learning method (Jin et al. 2020)
 - = initial unlabeled paradigms
- System searches a documentary corpus to identify related forms for each lexeme and group surface forms into paradigms

	Cell						
Lexeme	1	2	3	4	5	6	
HEAR	hear	heard	-	hearing	heart	-	
HELP	help	-	helped	helping	-	helps	

Step 2: Same or different? (Lexemes)

- Automatically extract examples of each **form in context** from the corpus
- The annotator marks items that don't belong with the others

File							
Lexicon Analogies Paradigms Texts							
Lexemes	Senses	Concordances					
DANCE DRIVE LIVE HEAR WORK	HEAR-V HEARING-N	 you're still going to hear them. She thought she could hear Gomez laughing. signalling of problems of hearing and understanding. gray marble mausoleum at the heart of the city. gray marble mausoleum at the heart of the city. 					

Step 2: Same or different? (Lexemes)

• Over time, the system uses annotator corrections to improve its suggestions via active learning



- Items are reordered in real time to make efficient use of annotator time
 - The software gradually learns which items belong together and which to exclude with minimal user input
 11

Step 3: Same or different? (Analogies)

• Pairwise analogy to group forms instantiating the same paradigm cell.

File		
Lexicon Analogies Parac	ligms Texts	
Analogies	Conce	ordances
	Х	Xing
X ~ Xment X ~ Xer	Ø We publish these	Time for publishing
X ~ Xing X ~ Xed	If we learn how	8 Second language learning is
X ~ X	We go regularly to	She's not going to like
		· · ·
	-	

- The annotator's task is the same: mark words that don't belong, and confirm those that do
- Annotation order is determined using active learning, as before 12

The result: Unlabeled paradigms



Case Study: Wao Terero

Wao Terero provides a demonstration of this workflow in the field.

- Linguistic isolate spoken in Ecuadorian Amazon
 - Estimated 1,200-3,000 speakers
 - No standard orthography
- Collaboration with native speakers (Spanish-Wao bilinguals)

- Two **native speaker consultants** from the Wao community of Geyepade served as annotators.
 - Neither consultant had taken a course in linguistics
- 10 minutes of training, with Spanish verbal paradigms
 - annotate as many items (lexemes and paradigm cells) as possible within **1 hour**
- Annotators found the task **understandable** and **interesting**, with high inter-annotator agreement across annotated examples



Copot et. al (2021)

Proposed workflow has high potential to **increase community engagement** Future Work

Future work: Annotator input



- Currently, annotators can only confirm or reject examples
- Future versions of the software will allow **direct user input** to correct incomplete paradigms

Conclusion

Word-and-Paradigm annotation makes direct comparisons in context

- Intuitive for untrained consultants
 - Increases community participation
- Defers difficult decisions about segmentation and labeling
 - Paradigmatic analysis of morphological system as a whole
- Modular architecture:
 - Future improvements in state of the art machine learning can immediately benefit annotator
- Annotation output may be used for linguistic analysis as well as community resource development

Many thanks to our consultants, Flora and Alberto Boyotai!

Appendix

Concordance Workflow

				on how we organize thoughts and information gathered from our environments into
	They also sometimes	focus	True	meaningful categories of thought , which will be discussed later .
Х	Personal experiences of discrimination and bias have been the	focus	True	of much social science research . [1 - 3]
			_	completely to languages , and in 1887 earned his master's master 's degree in French ,
X	In 1881 he shifted his	tocus	True	with English and Latin as his secondary languages .
	The events in the Arab world notwithstanding , we must continue to	focus	True	on the Middle East peace process .
	Which elements of specific artworks do they	focus	True	on ?
x	This	focus	True	stemmed from his close patron relationships with several prominent female ascetics who were members of affluent senatorial families . [7]
	First , specialization in a particular small job allows workers to	focus	True	on the parts of the production process where they have an advantage .
	The ultimate result of workers who can	focus	True	on their preferences and talents , learn to do their specialized jobs better , and work in larger organizations is that society as a whole can produce and consume far more than if each person tried to produce all of his or her own goods and services.
				primarily on language teaching reform and on phonetics , but he is best known for his later
	His early work	focused	True	work on syntax and on language development .
				no work has been completed on the aesthetic appreciation of collections or of devotional
	While studies of the psychology of art have	focused	True	themes .
	Because assembling such a full census is difficult, past studies have tended to avoid this task and have instead used samples of researchers [8-11], usually specific to a			
	particular field [12 - 16] , and often	focused	True	on the scientific elite [17 , 18] .

x	In the present	study	True	, we examine the outcomes of such a period of no exposure on the <u>repurpcopping</u> of L2 grammar: that is, whether a substantial period of no exposure leads to decreased proficiency and / or leas native-like neural processes (* use it or lose it " [20]), no such changes , or perhaps whether even higher proficiency and / or more native-like processing may bodiserved.
×	These language measures were compared in most studies to the same measures in a different set of subjects who had not experienced a period of limited exposure [17], [18], [22], [23], or to retrospective ratings of the same subjects [21], with only one longitudinal	study	True	testing the same subjects before and after a period of limited exposure [24].
x	Moreover , one	study	True	found no changes at all in performance, across proficiency levels, after either 2 or 4 years of limited exposure [22].
x	Finally, in some cases a gain in performance has been observed : after 1.5 years of limited exposure in one	study	True	, particularly for L2 learners with immersion as well as classroom training [24], and in another study after 2 years, though only for some abilities, such as listening and reading comprehension [18].
х	And one of our friends , our common friends , he introduced us during	study	True	hall , and we just kind of hit it off from there .
x	A new	study	True	documenting iodine nutritional status in Australian school children has revealed many are not getting enough iodine - which can lead to mental and growth retardation .
x	Tasmania was excluded from the	study	True	 where an voluntary iodine fortification program using iodised salt in bread, is ongoing.
x	Professor Cres Eastman , Director of the National Iodine Nutrition	study	True	, and Chairman of the Australian Centre for Control of Iodine Deficiency Disorders , says it is crucial that children and pregnant women in particular have an adequate intake of iodine .
	It attempts to explain how and why we think the way we do by	studying	True	the interactions among human thinking , emotion , creativity , language , and problem solving , in addition to other cognitive processes .
	He entered the University of Copenhagen in 1877 when he was 17, initially	studying	True	law but not forgetting his language studies .
	So I'm I 'm gonna gon na be	studying	True	on Thank sgiving Day, Black Friday, and the whole weekend even though we 're going getting our Christmas tree I think on Saturday, and I I'm I'm not decorating I'm I 'm not doing anything fun, Ike I have to study.
	So I'm I 'm going to go in with like a different color and add in the chapters I'm I 'm gonna gon na be	studying	True	for that , but I'm I 'm super stressed .
	Importantly ,	studying	True	eye movements offers an insight that does not depend on the participants' participants ' beliefs , memories or subjective impressions of the artwork .

Concordance Workflow

×	These language	measures	True	were compared in most studies to the same measures in a different set of subjects who had not experienced a period of limited exposure [17], [18], [22], [23], or to retrospective ratings of the same subjects [21], with only one longitudinal study testing the same subjects before and after a period of limited exposure [24].
X	Overall , the	results	True	of the six studies have been taken to suggest the following .
Г	A period of limited exposure generally	leads	True	to attrition (loss) of L2 performance or knowledge [17], [18], [21], [23].
×	Such	loss	True	has been observed after as little as a few months of limited exposure , e.g., after a 1 - 7 month [23] or 6 month delay [21], as well as after 2 years [18], though in one case it was observed only by 3 - 5 years , and not earlier [37].
E	Although attrition may take place within the first few years , some studies suggest that it then	appears	True	to level off , with no further losses occurring [17], [18].
X	Moreover, one study found no	changes	True	at all in performance, across proficiency levels, after either 2 or 4 years of limited exposure [22].
×	Finally, in some cases a gain in performance has been observed : after 1.5 years of limited exposure in one study, particularly for L2 learners with immersion	as	True	well as classroom training [24], and in another study after 2 years, though only for some abilities, such as listening and reading comprehension [18].
	n	remains	True	unclear what might explain such gains, which have been attributed to motivation and to L2 experience during the period of ostensibly initied exposure [24], or to factors related to general maturation, cognitive development, or continued academic training [18].
×	We see this	as	True	the ideal opportunity to provide these educative sessions not only to our own librarians , but also to the academic librarians of other Dutch research libraries .
Т	Each day	starts	True	with one or more lectures by researchers , that address the conceptual knowledge needed .
X	The afternoon sessions will be devoted to the hands-on training of skills , following the Library Carpentry model	as	True	closely as possible .
×	Byron received	his	True	early formal education at Aberdeen Grammar School , and in August 1799 entered the school of Dr. William Glennie , in Dulwich , [17]
×	Placed under the care of a Dr. Bailey , he was encouraged to exercise in moderation but not restrain himself from " violent " bouts in an attempt to overcompensate for	his	True	deformed foot
×	His mother interfered with	his	True	studies, often withdrawing him from school, with the result that he lacked discipline and his classical studies were neolected.
X	In Byron's Byron 's later memoirs, " Mary Chaworth is portrayed	as	True	the first object of his adult sexual feelings . " [20]
×	His nostalgic poems about	his	True	Harrow friendships, Childish Recollections (1806), express a prescient " consciousness of sexual differences that may in the end make England untenable to him." [23]
×	About	his	True	"protegie" he wrote, "He has been my almost constant associate since October, 1805, when I entered Trinity College. His voice first attracted my attention, his countenance fixed it, and his manners attached me to him for ever."
X	In	his	True	memory Byron composed Thyrza, a series of elegies. [25]
Γ	01	interest	False	here, studies suggest that, despite the difficuties in acquiring L2 grammar, adult learners can approximate native-like levels of use and neurocognitive processing [12]-[15].
T	However, it is	not	False	enough to have attained such native-like levels
T	Crucially, it is also desirable to retain them, even in the absence of continued	practice	False	or exposure to the L2.

Edit Trees (Jin et al., 2020)



Figure 3: Visualization of the EDIT TREES representing (a) work \mapsto worked and (b) continue \mapsto continued.

sent id = GUM academic art-22 # s_type = decl # text = Importantly, studying eve movements offers an insight that does not depend on the participants' beliefs, memories or subjective impressions of the artwork. ADV RB Degree=Pos 6 advmod 6:advmod Discourse=evaluation: 41->39:1 SpaceAfter=No Importantly importantly PUNCT punct l:punct studying study VERB VBG VerbForm=Ger 6 csubj 6:csubj _ eve NOUN NN Number=Sing 5 compound 5:compound Entity=(event-92-giv;act-2-coref(object-93-giv;inact-1-coref) eve NOUN NNS Number=Plur obi 3:obi Entity=92) movements VBZ Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin offers offer VERB root 0:root an a DET DT Definite=Ind PronType=Art det 8:det Entity=(abstract-102-new-2-sgl insight insight NOUN NN Number=Sing 6 obj 6:obj 12:nsubj that that DRON WDT PronType=Rel 12 nsubi 8:ref Discourse=elaboration:42->41:0 10 does do AUX VBZ Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin 12 aux 12:aux _ not not PART RB Polarity=Neg 12 adymod 12:adymod 12 depend depend VERB VB VerbForm=Inf 8 acl:relcl 8:acl:relcl 13 on on ADP IN 17 case 17:case 14 the DPT DT Definite=Def PronType=Art 15 det 15:det Bridge=92<104 Entity=(abstract-103-new-4-sgl(person-104-acc;inf-2-sgl 15-16 participants NOUN NNS Number=Plur participants participant nmod:poss 17:nmod:poss 16 ' a PART POS 15 case 15:case Entity=104) beliefs belief NOUN NNG Number=Plur 12 obl 12:obl:on Entity=103) SpaceAfter=No 18 PUNCT 19 19:punct punct 12:obl:on|17:conj:or 19 memories memory NOUN NNS Number=Plur coni Entity=(abstract-105-new-1-sgl) 20 or or CCONJ CC 22 00 22100 21 subjective subjective AD.T JJDegree=Pos 22 amod 22:amod Entity=(abstract-106-new-2-sql 22 impressions impression NOUN NNS Number=Plur 17 coni 12:obl:on 17:conj:or 23 of of ADP IN 25 case 25:case 24 DT Definite=Def | PronType=Art the the DET det 25:det Entity=(object-87-giv:inact-2-coref artwork artwork NOUN NN Number=Sing nmod 22:nmod:of Entity=87)106)102) SpaceAfter=No 26 PUNCT 6 punct 6:punct

Analogy-based Annotation Workflow

	sting analog	ies		,
~	INDEX	NAME	PROVENANCE	MEMBE
~	0	accord_1~according_1	jin_etal	
\checkmark	1	add_1~add_1	jin_etal	
\checkmark	2	add_1~added_1	jin_etal	
\checkmark	3	add_1~adds_1	jin_etal	
~	4	base_1~bas_1	jin_etal	
8 records	selected.			1-8 o
Create n	ew analogy.			,
(select a	word and s	ense) ~ (select a word and	i sense)	
			Search	
Search.			Joeanon	

Lexicon browser

Search...

Show all words ...

how focus

look

	Name	(#positive)
1	look_1	62
		1-1 of 1

 \sim

Viewing lexical split: look_1

Relat	Dist	Relative	
accord_1~according	1	looking_1	0
add_1~adde	1	looked_1	1
add_1~add	1	looks_1	2
1-3 o			

Experiments and results

- Universal Dependencies datasets for English and Croatian provide a gold standard for evaluation
- Annotators: 4 linguists (2 per language), fluent English speakers
 - English: upper estimate of model + annotator performance
 - Croatian: unfamiliar language
- Formalized annotation guidelines provide instructions and guidance for dealing with ambiguity
- Annotators had 30 minutes to annotate lexeme data and 30 minutes for cell data

	I	Lexeme				Cell	
	Acc.	Marked	Corr.		Acc.	Marked	Corr.
English			Englis	h			
Base	81%	-	-	Base	67%	-	-
A1	84%	58	50	A1	97%	129	120
A2	83%	43	33	A2	94%	119	108
Croatian			Croatia	an			
Base	66%	-	-	Base	90%	-	-
A3	67%	19	19	A3	90%	8	-1
A4	66%	12	12	A4	90%	28	16

Coll

Wao Terero provides a demonstration of this workflow in the field.

- Linguistic isolate spoken in Ecuadorian Amazon
 - Estimated 1,200-3,000 speakers
 - No standard orthography
- Part of ongoing fieldwork and language documentation project
 - Collaboration with native speakers (Spanish-Wao bilinguals)

Experiments: Wao Terero

• Model input:

- Wao Terero New Testament
- Multi-syllabic target lemmas
- Two **native speaker consultants** from the Wao community of Geyepade serve as annotators.
 - Neither consultant has taken a course in linguistics
 - Annotators given 10 minutes of training on task using Spanish verbal paradigms
- A non-native linguistics Ph.D. student also completed the annotation experiment.

• No gold annotations. We instead measure **annotation speed** and collect **qualitative feedback**.

67 tokens/h Wao consultants (each) 776 tokens/h Fieldworker

- Differences in speed reflect different annotation strategies:
 - Meaning in context vs. orthographic similarity
- Annotators found the task **understandable** and **valuable**, but the data was challenging
 - More natural texts and better heuristics for dealing with ambiguous lexeme categories may improve future performance