A Word-and-Paradigm workflow for fieldwork annotation

Maria Copot
Sara Court
Noah Diewald
Stephanie Antetomaso
Micha Elsner

The proposal

A workflow for morphosyntactic annotation of underdescribed languages.

- modular
  - State-of-the-art technology can be immediately integrated
  - Can interface with existing annotation software
- consultant-friendly
  - Relies on a same vs different task
  - No linguistic training necessary
- emergent categories
  - data labeling and segmenting can be done post-annotation and won't constrain the process

We present a proof of concept.

Current standard annotation practices

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

Overview

- Motivation
  - More inclusive fieldwork practices
  - Theoretical hygiene
- Our contribution
  - A detailed proposal
  - Experiments
    - Different levels of language knowledge
    - Natural fieldwork setting
- Discussion and future directions

Inclusive fieldwork methods

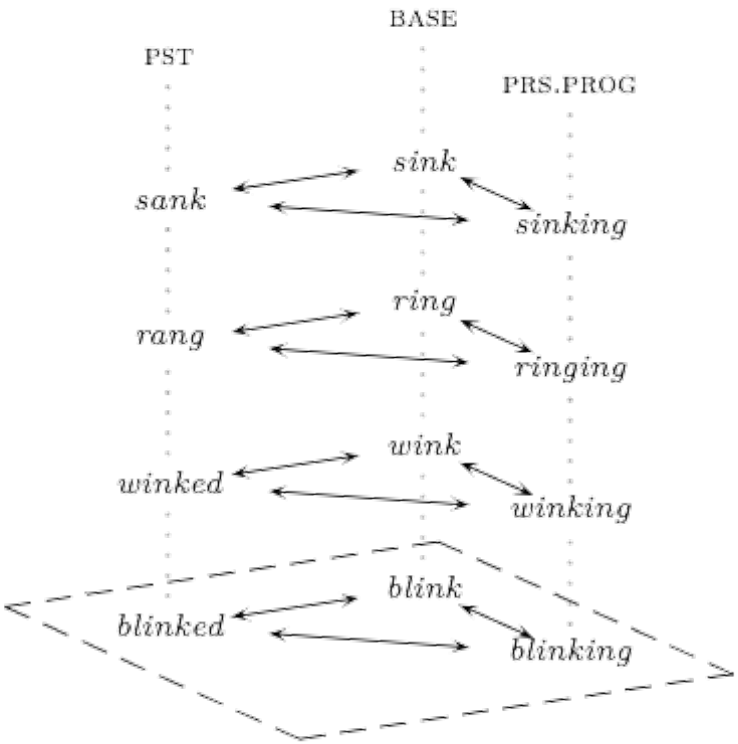
Community Engagement

- Navigate researcher-community collaboration ethically
  - Give community members maximum agency
- There is no single “best” strategy for increasing community engagement
- Software tools for community-led annotation
  - Lower technical barrier for entry for broader participation
  - Increase community involvement and agency in fieldwork

Theoretical underpinnings

Word and Paradigm morphology

- Establishing parallel relationships of form and meaning between words



- Covariation, not segmentation – looking outwards
  - The word is the smallest unit.
- Concepts like paradigm cell or lexeme are emergent
  - The result of establishing contrasts and similarities along different dimensions

Current model and workflow

Step 1: Unsupervised paradigm induction

- Obtain initial unlabeled paradigms using a machine learning method
  - In our work: Jin et al. 2020 (part of a SIGMORPHON shared task)
  - unsupervised model outputting forms related by edit trees
- System uses both form and content to group surface forms into paradigms

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

- A good starting point, but automatic methods cannot solve the task independently.

Step 2: Extract concordances

- Extract examples of each form in context from the corpus:

LEXEME

annotator	form	model output	
...you're still going to	hear	True	them.
She thought she could	hear	True	Gomez laughing.
...signalling of problems of	hearing	True	and understanding.
...gray marble mausoleum at the	heart	True	of the city.

- Concordances for cells also contain some random negative examples (presumed not to belong to the cell)

CELL

annotator	form	model output	
...mechanisms underlying the	learning	True	and processing of L2 grammar ...
...periods of limited ...exposure	following	True	L2 training are not uncommon ...
...may be found in different situations	including	True	when one studies a language ...
...such as listening and	reading	True	comprehension ...
The training	lasted	False	varying lengths of time...

Step 3: Mark same or different

- The annotator marks items that don't belong with the others:

		Lexeme		
annotator		form	model output	
	...you're still going to	hear	True	them.
	She thought she could	hear	True	Gomez laughing.
X	...signalling of problems of	hearing	True	and understanding.
X	...gray marble mausoleum at the	heart	True	of the city.

- In our pilot study, we asked annotators to exclude derived forms like 'hearing', but this is a design decision.

The output of the method

- Grouped unlabeled cells and lexemes
  - Corrected by annotators
- The groupings can be used for all purposes of linguistic description and analysis, and are convertible into IGTs if desired.

Experiments and results

Experiments: English & Croatian

- 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

English & Croatian Results

Lexeme

	Acc.	Marked	Corr.
English			
Base	81%	-	-
A1	84%	58	50
A2	83%	43	33
Croatian			
Base	66%	-	-
A3	67%	19	19
A4	66%	12	12

Cell

	Acc.	Marked	Corr.
English			
Base	67%	-	-
A1	97%	129	120
A2	94%	119	108
Croatian			
Base	90%	-	-
A3	90%	8	-1
A4	90%	28	16

English & Croatian Results

Lexeme

	Acc.	Marked	Corr.
English			
Base	81%	-	-
A1	84%	58	50
A2	83%	43	33
Croatian			
Base	66%	-	-
A3	67%	19	19
A4	66%	12	12

Cell

	Acc.	Marked	Corr.
English			
Base	67%	-	-
A1	97%	129	120
A2	94%	119	108
Croatian			
Base	90%	-	-
A3	90%	8	-1
A4	90%	28	16

English & Croatian Results

Lexeme

	Acc.	Marked	Corr.
English			
Base	81%	-	-
A1	84%	58	50
A2	83%	43	33
Croatian			
Base	66%	-	-
A3	67%	19	19
A4	66%	12	12

Cell

	Acc.	Marked	Corr.
English			
Base	67%	-	-
A1	97%	129	120
A2	94%	119	108
Croatian			
Base	90%	-	-
A3	90%	8	-1
A4	90%	28	16

English & Croatian Results

Lexeme

	Acc.	Marked	Corr.
English			
Base	81%	-	-
A1	84%	58	50
A2	83%	43	33
Croatian			
Base	66%	-	-
A3	67%	19	19
A4	66%	12	12

Cell

	Acc.	Marked	Corr.
English			
Base	67%	-	-
A1	97%	129	120
A2	94%	119	108
Croatian			
Base	90%	-	-
A3	90%	8	-1
A4	90%	28	16

Experiments: 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
- 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.

Results: Wao Terero

- 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

Discussion and future directions

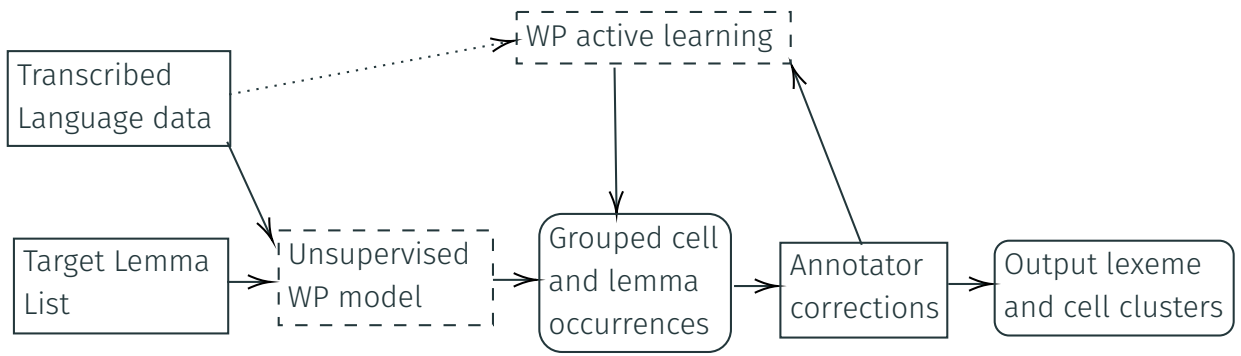
Benefits of the Workflow for Linguistic Fieldwork

Word-and-Paradigm annotation makes direct comparisons in context

- Intuitive for untrained consultants
  - Increases community participation
- Defers difficult decisions about segmentation and labeling
  - Output can still be used to create Interlinear Glossed Texts
- Modular architecture: future improvements in state of the art can immediately benefit annotator

Future Directions

- Interactive environment
  - Allow annotators to view proposed paradigm tables alongside the text to identify missing forms
  - Integrate more efficient search tools, e.g., word clouds of similar forms
- Active learning
  - Introduce a supervised learner using active learning heuristics to minimize annotation effort
  - Direct the annotator's attention to the least certain distinctions
  - Eliminate repetitive annotation of "easy" instances



Many thanks to our consultants,
Flora and Alberto Boyotai!

Appendix

Concordance Workflow

	They also sometimes	focus	True	on how we organize thoughts and information gathered from our environments into meaningful categories of thought , which will be discussed later .
X	Personal experiences of discrimination and bias have been the	focus	True	of much social science research . [1 - 3]
X	In 1881 he shifted his	focus	True	completely to languages , and in 1887 earned his master's master 's degree in French , 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 .
	His early work	focused	True	primarily on language teaching reform and on phonetics , but he is best known for his later work on syntax and on language development .
	While studies of the psychology of art have	focused	True	no work has been completed on the aesthetic appreciation of collections or of devotional 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 neurocognition of L2 grammar : that is , whether a substantial period of no exposure leads to decreased proficiency and / or less 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 be observed .
X	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] .
X	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 Thanksgiving 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 , like 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

X	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 .
	A period of limited exposure generally	leads	True	to attrition (loss) of L2 performance or knowledge [17] , [18] , [21] , [23] .
X	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 [17] .
	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] .
X	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] .
	It	remains	True	unclear what might explain such gains , which have been attributed to motivation and to L2 experience during the period of ostensibly limited exposure [24] , or to factors related to general maturation , cognitive development , or continued academic training [18] .
X	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 .
	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 .
X	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]
X	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 .
X	His mother interfered with	his	True	studies , often withdrawing him from school , with the result that he lacked discipline and his classical studies were neglected .
X	In Byron's Byron 's later memoirs , " Mary Chaworth is portrayed	as	True	the first object of his adult sexual feelings . " [20]
X	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]
X	About	his	True	" protégé " 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]
	Of	interest	False	here , studies suggest that , despite the difficulties in acquiring L2 grammar , adult learners can approximate native-like levels of use and neurocognitive processing [12] – [15] .
	However , it is	not	False	enough to have attained such native-like levels .
	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)

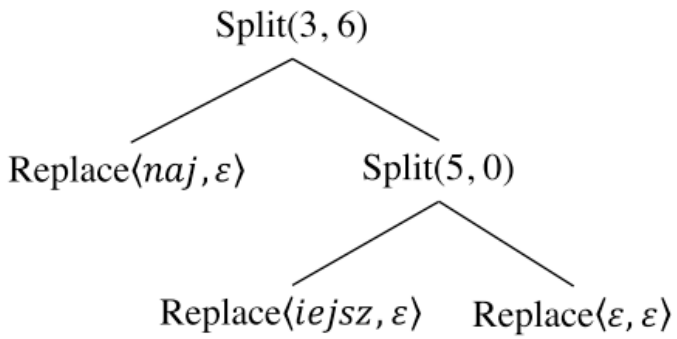


Figure 2: Visualization of the EDIT TREE constructed from *najtrudniejszy* to *trudny* (Chrupała, 2008).

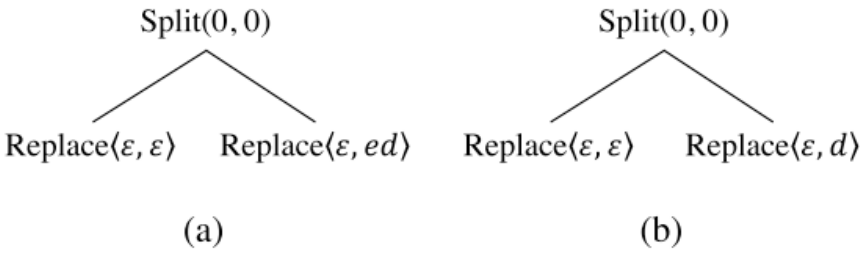


Figure 3: Visualization of the EDIT TREES representing (a) *work* ↦ *worked* and (b) *continue* ↦ *continued*.

UD Treebanks

```
# sent_id = GUM_academic_art-22
# s_type = decl
# text = Importantly, studying eye movements offers an insight that does not depend on the participants' beliefs, memories or subjective impressions of the artwork.
1       Importantly     importantly     ADV     RB      Degree=Pos      6       advmod  6:advmod        Discourse=evaluation:41->39:1|SpaceAfter=No
2       ,       ,       PUNCT   ,       _       1       punct   1:punct _
3       studying        study   VERB    VBG     VerbForm=Ger    6       csubj   6:csubj _
4       eye     eye     NOUN    NN      Number=Sing     5       compound        5:compound      Entity=(event-92-giv:act-2-coref(object-93-giv:inact-1-coref)
5       movements       movement        NOUN    NNS     Number=Plur     3       obj     3:obj   Entity=92)
6       offers  offer   VERB    VBZ     Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin   0       root    0:root  _
7       an      a       DET     DT      Definite=Ind|PronType=Art       8       det     8:det   Entity=(abstract-102-new-2-sgl
8       insight insight NOUN    NN      Number=Sing     6       obj     6:obj|12:nsubj  _
9       that    that    PRON    WDT     PronType=Rel    12      nsubj   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  _
11      not     not     PART    RB      Polarity=Neg    12      advmod  12:advmod       _
12      depend  depend  VERB    VB      VerbForm=Inf    8       acl:relcl       8:acl:relcl     _
13      on      on      ADP     IN      _       17      case    17:case _
14      the     the     DET     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'    _       _       _       _       _       _       _
15      participants    participant     NOUN    NNS     Number=Plur     17      nmod:poss       17:nmod:poss    _
16      '       's      PART    POS     _       15      case    15:case Entity=104)
17      beliefs belief  NOUN    NNS     Number=Plur     12      obl     12:obl:on       Entity=103)|SpaceAfter=No
18      ,       ,       PUNCT   ,       _       19      punct   19:punct        _
19      memories        memory  NOUN    NNS     Number=Plur     17      conj    12:obl:on|17:conj:or    Entity=(abstract-105-new-1-sgl)
20      or      or      CCONJ   CC      _       22      cc      22:cc   _
21      subjective      subjective      ADJ     JJ      Degree=Pos      22      amod    22:amod Entity=(abstract-106-new-2-sgl
22      impressions     impression      NOUN    NNS     Number=Plur     17      conj    12:obl:on|17:conj:or    _
23      of      of      ADP     IN      _       25      case    25:case _
24      the     the     DET     DT      Definite=Def|PronType=Art       25      det     25:det  Entity=(object-87-giv:inact-2-coref
25      artwork artwork NOUN    NN      Number=Sing     22      nmod    22:nmod:of      Entity=87)106)102)|SpaceAfter=No
26      .       .       PUNCT   .       _       6       punct   6:punct _
```

Analogy-based Annotation Workflow