M. Elsner 2021: What transfers in morphological inflection?

Experiments with analogical models

Miikka Silfverberg
CMRG
AUG 13, 2021

Outline

General idea

A bit of background

Memory-based learning for morphology

Experiments

Results

Takeaways

General idea

What transfers in morphological inflection?

Several studies have investigated cross-lingual transfer as a method for improving morphological inflection results for low-resource languages

Transfer is useful but poorly understood

In the standard encoder-decoder paradigm, abstract morphological processes cannot easily be decoupled from their concrete exponents (affixes, etc.)

Memory-based learning for morphology offers a way to do this

The paper provides insight into many important questions like what is the role of language relatedness and typological relatedness for transfer learning

Memory-based learning

Training examples:

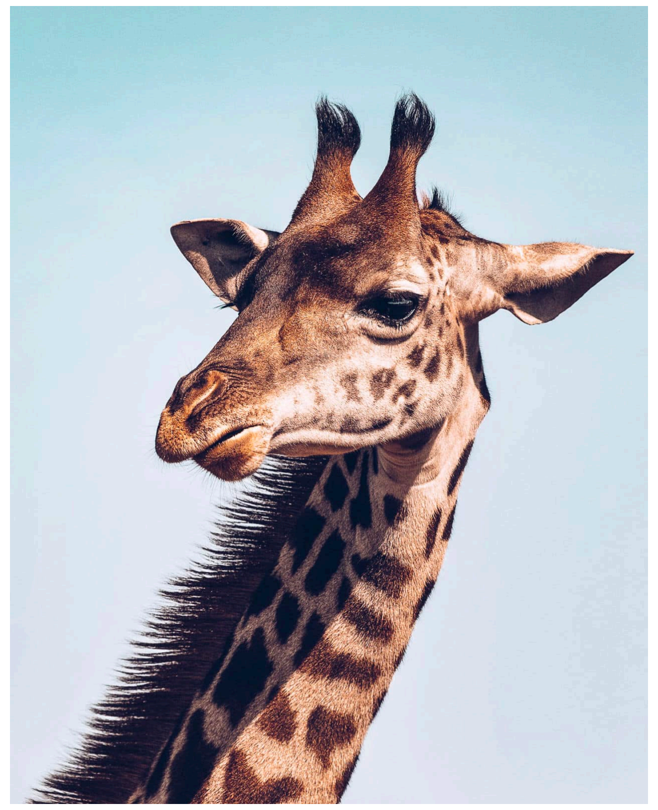
DOG

CAT

CAT

GIRAFFE

Test example:



? -> GIRAFFE

1. Find most similar training example
2. Copy label to test example

Memory-based inflection

Under the standard formulation of the inflection task, a lemma+tag is inflected into a word form

	Lemma	Target specification	$\rightarrow$	Target
Standard inflection generation	waiata	V;PASS		waiatatia

In memory-based inflection, we inflect based on an exemplar

Memory-based	waiata	karanga : karangatia		waiatatia

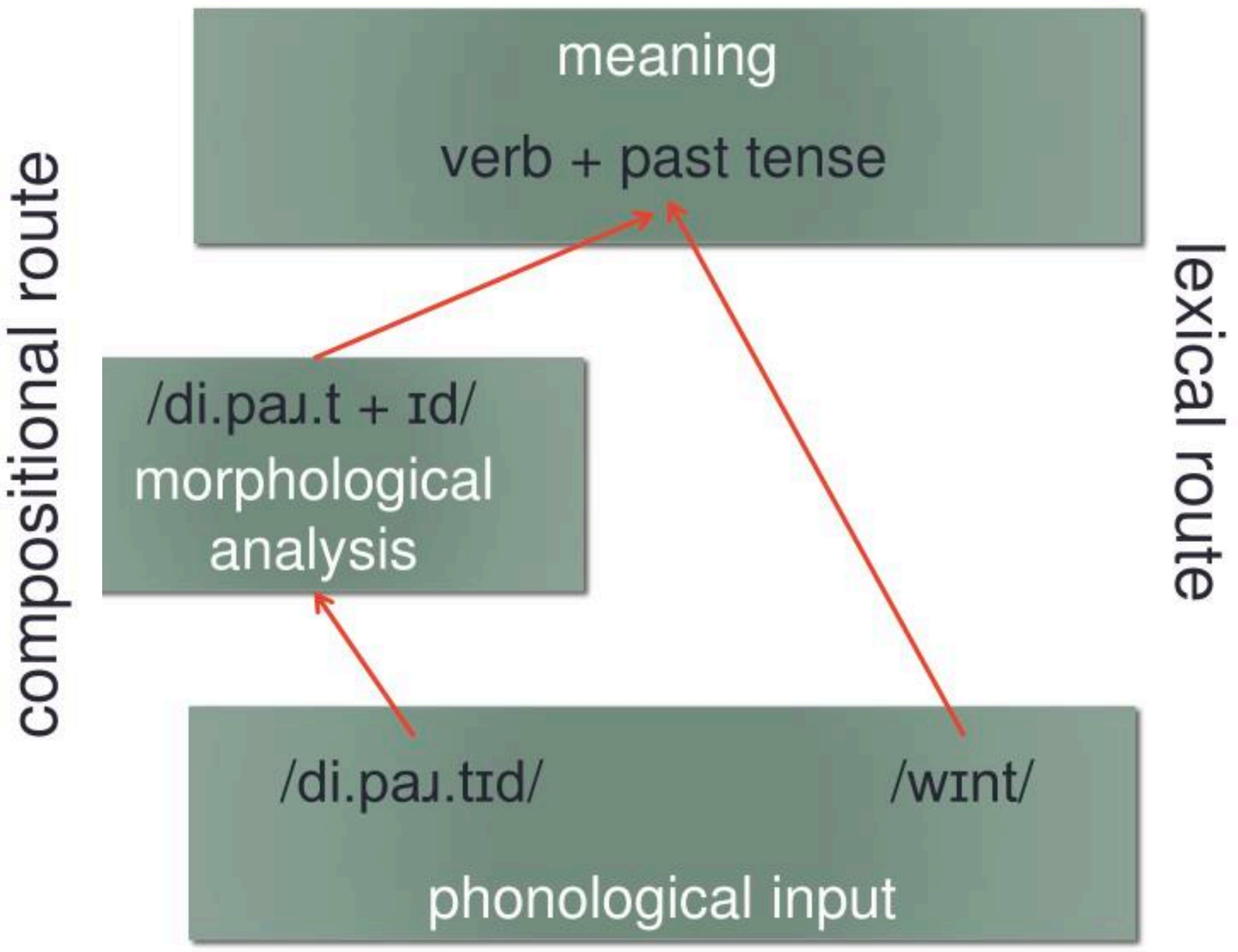
The authors train multilingual models and language+language family tags are added to the input

Dual route model for morphology

The memory-based models for inflection is rooted in cognitive theories of morphological processing

The dual route model for morphological processing uses two different routes:

1. Inflected forms of very frequent lexemes and irregular lexemes are memorized
2. Infrequent regular lexemes are inflected based on analogical memorized lexemes (or possibly some sort of rules)



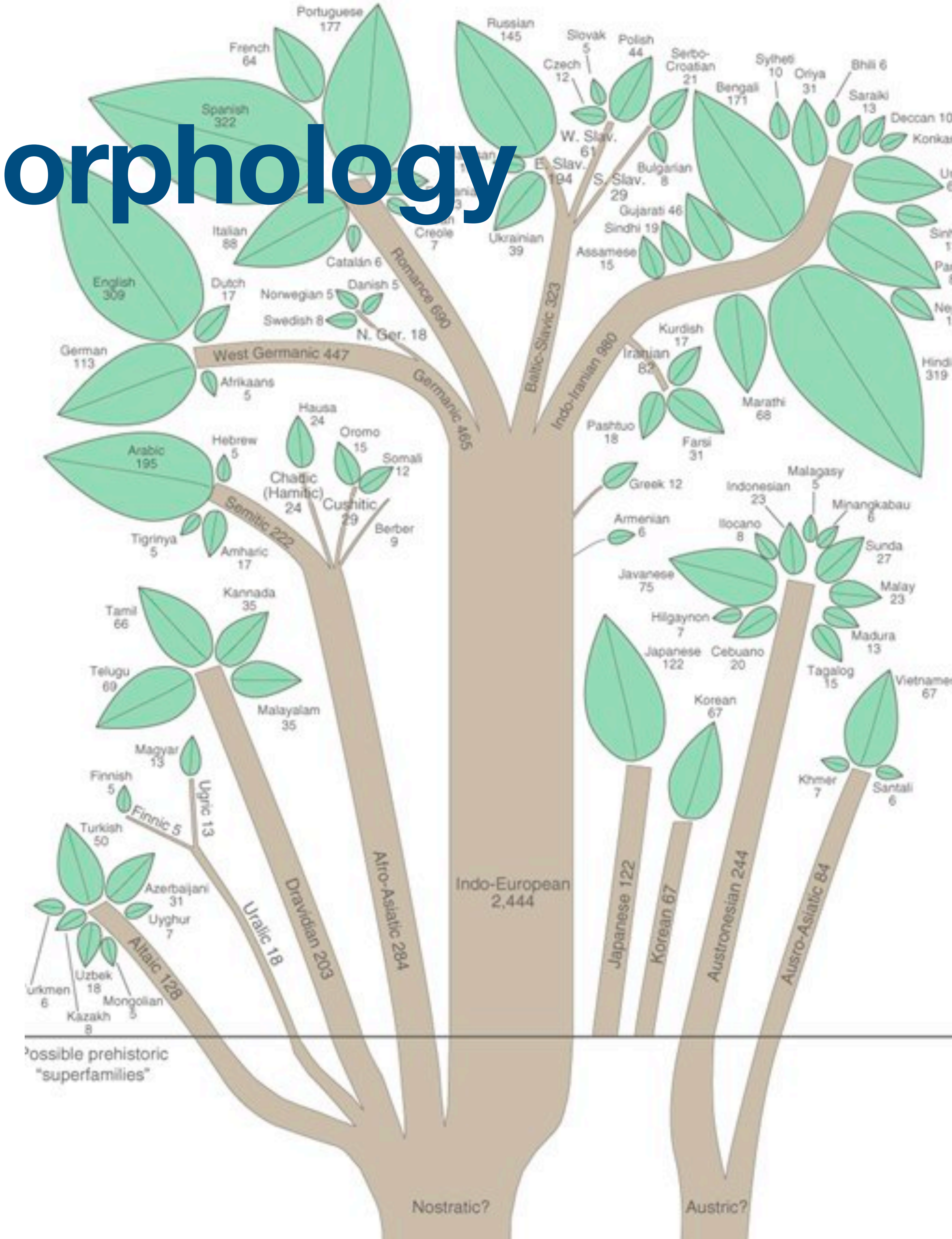
Transfer learning for morphology

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task

In morphology, transfer usually refers to cross-lingual methods where data from several languages are used to train a model

Typically, data from highly resourced languages are used to improve performance for a low-resource language

Consensus: data from related languages helps



When is transfer successful?

It's hard to know in advance if transfer will be successful

In general transfer is more helpful:

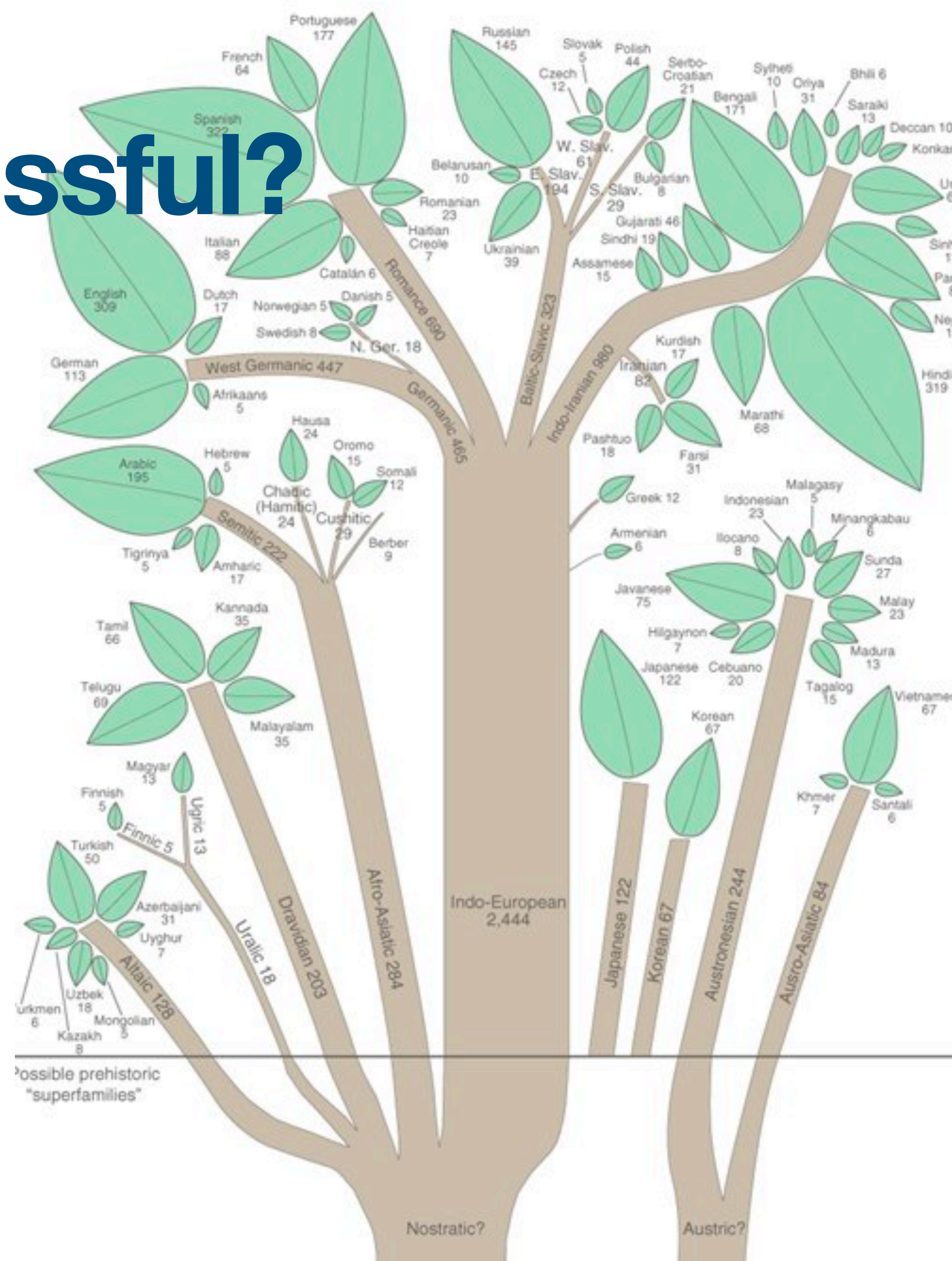
- When the target language data is limited
- When the languages are related or similar
- When both prefer prefixes/suffixes

Kann (2020) notes that Hungarian seems to transfer well to many languages

Hungarian seems to work better for German than English does even though Hungarian and German are unrelated

Lin et al. (2019) make a similar observation concerning Turkish

In general, perhaps agglutinative languages transfer well? Or languages where inflection is challenging?



Challenges for analysis

Lemma	Target specification	→	Target
waiata	V;PASS		waiatatia

In the standard formulation of the inflection task, it is hard to test if a general suffixation operation has been learned

Morphological processes are inextricably tied to their exponents (e.g. particular affixes)

Memory-based learning

	Lemma	Target specification	→	Target
Memory-based	waiata	karanga : karangatia		waiatatia
	waiata	kaukau : kaukauria		waiatatia

In memory-based learning, the model learns to imitate a given example, for example to identify the suffix *-tia* which should be copied onto the end of the input lemma *waiata*

In principle, the model does not need to memorize individual affixes. Instead it can rely on the exemplar and general knowledge about inflectional processes

One-shot learning

Memory-based learning allows for inflection in languages which were never seen during training

We could train the model to inflect English nouns and verbs:

dog cat:cats -> dogs
wait yell:yelled -> waited

And then apply this model to Swedish inflection:

hund lut:lutar -> hundar

(*hund* ‘dog’, *lut* ‘lye’)

This can only be successful if the model has learned a general suffixation process

Choosing the exemplar

Choosing a good exemplar is important for inflection quality:

dog cat:cats -> dogs ✓
dog ox:oxen -> dogen X

During training, exemplars are chosen either **randomly** or **based on similarity**

The similarity-based strategy requires knowledge about the gold standard output and can only be used during training time

During test time, the random strategy is always applied. 5 random exemplars are sampled and the majority output is chosen

SIGMORPHON 2020 baseline

Character-level transformer model with specialized positional encodings

Vanilla

Token	<s>	V	V.PTCP	PST	s	m	e	a	r	</s>
	+	+	+	+	+	+	+	+	+	+
Position	0	1	2	3	4	5	6	7	8	9
Type										

Feature Invariant

Token	<s>	V	V.PTCP	PST	s	m	e	a	r	</s>
	+	+	+	+	+	+	+	+	+	+
Position	0	0	0	0	1	2	3	4	5	6
		+	+	+	+	+	+	+	+	
Type		F	F	F	C	C	C	C	C	

The model uses fewer layers and attention heads than are typically used in machine translation

The feature invariant positional encodings have a small but positive impact on performance

The model

The base learner in this work is the SIGMORPHON 2020 baseline model using the memory-based formulation of the inflection task

The model is trained on a synthetic inflection task:

mpienjmel rbeaikkea:zlürbeaikkeaüe -> zlümpienjmelüe

The synthetic pre-training set contains prefixation, suffixation and infixation. Some training examples require pure copying

The “language tag” indicates the type of affixation, e.g. LANG_PREFIX_SUFFIX and the family tag is SYNTHETIC

Limitations of memory-based inflection

The major limitation of memory-based inflection as compared to the standard formulation of inflection is that we can't inflect in forms we haven't observed in the training data

Using the standard formulation, we can train the model on examples:

koira N;SG;INE -> koirassa
luuta N;PL;PART -> luutia

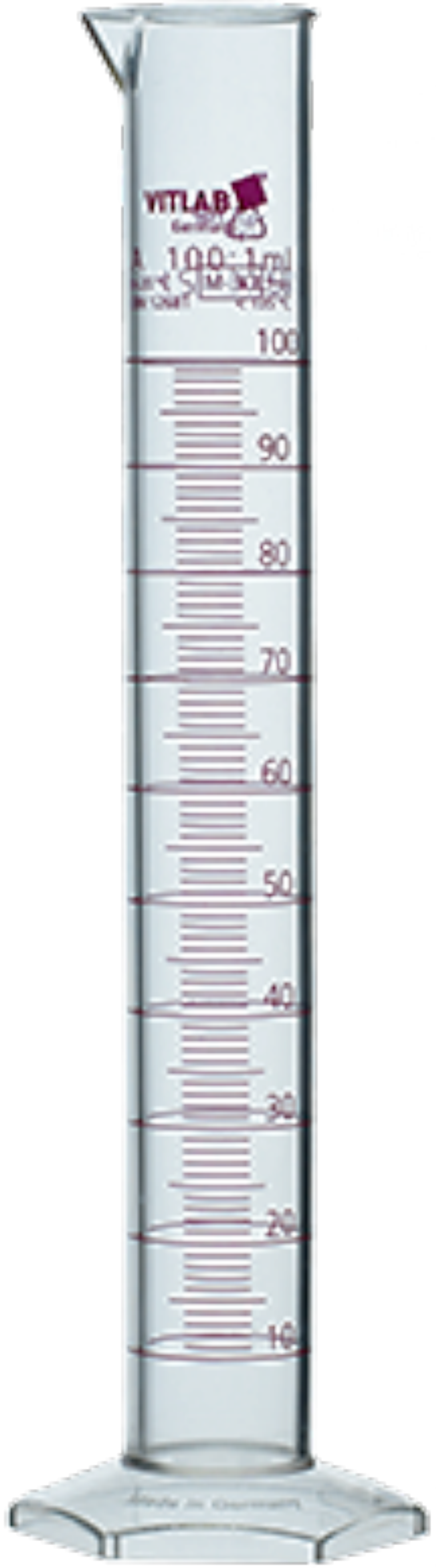
(koira 'dog', luuta 'broom' Finnish)

The model can learn to inflect:

koira N;PL;INE -> koirissa

Without ever seeing examples of N;PL;INE forms

Experiments



The authors train a multilingual inflection model on 45 development languages from five language families (in addition to the synthetic data) and fine-tuned on target language family

The model is tested on 45 test languages

Two settings:

1. The model is fine-tuned on the **test language**
2. The multilingual model is applied **without fine-tuning**

In setting 2, the only training signal from the target language comes from the exemplar

Fine-tuned results

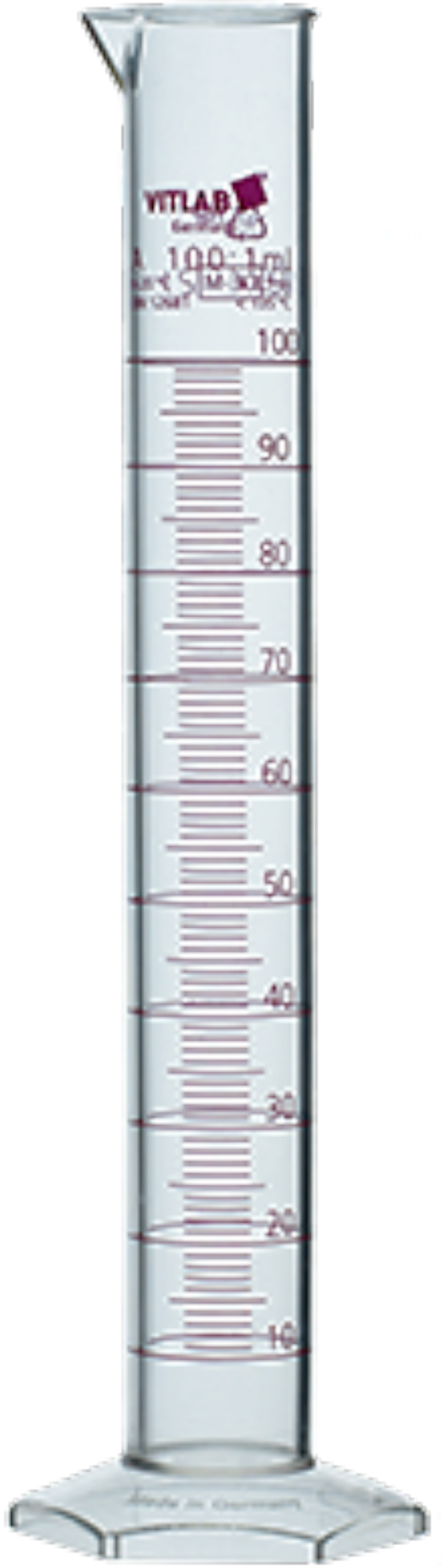
Family	Random	Similarity	Base
Austronesian (4)	83 (13)	67 (21)	81 (18)
Germanic (10)	87 (10)	51 (16)	90 (9)
Niger-Congo (9)	98 (4)	94 (9)	97 (3)
Oto-Manguean (10)	82 (16)	39 (23)	86 (12)3
Uralic (11)	92 (6)	46 (14)	93 (0.05)
Overall	89 (12)	57 (26)	90 (11)

In general, results are lower than for the traditional baseline

The similarity-based strategy has a data mismatch between training and test time leading to poor performance compared to the random strategy

Performance on Niger-Congo languages is strong (but the data might be biased)

Oto-Manguean and Germanic languages are hard due to inflectional classes



One-shot results

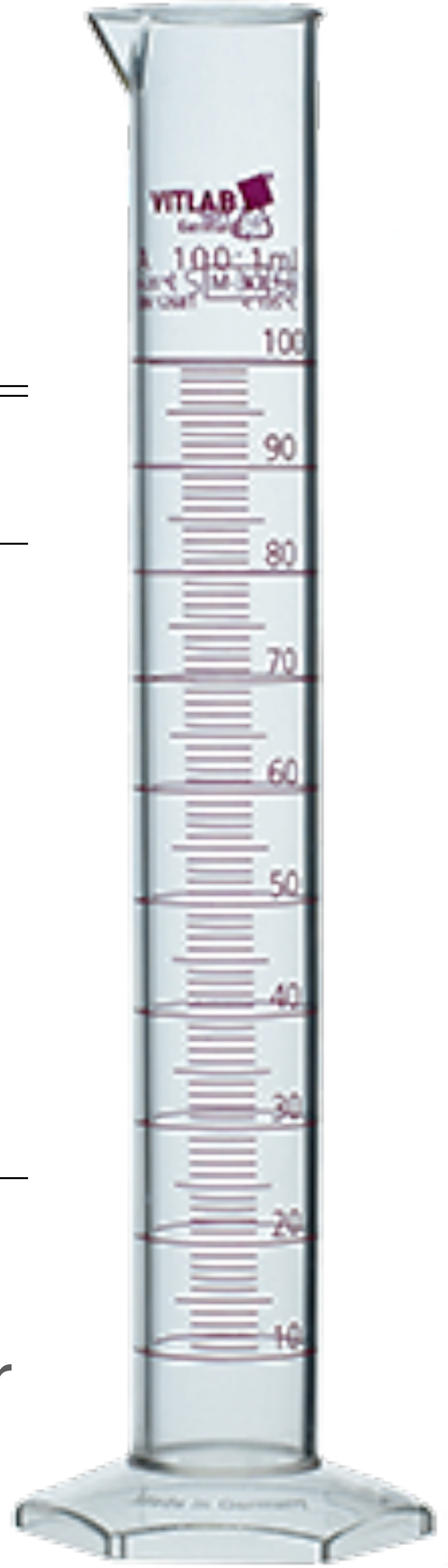
One-shot learning is not competitive compared to fine-tuning

Related languages in the training data help

For some languages, performance is better than for the baseline system (perhaps data sparsity)

In contrast to fine-tuning, the similarity-based strategy works much better here

Family	Random	Similarity	Base
Germanic (3)	29 (13)	38 (22)	80 (13)
Niger-Congo (1)	75 (0)	88 (0)	100 (0)
Uralic (5)	21 (9)	28 (12)	76 (26)
Afro-Asiatic (3)	7 (3)	26 (18)	96 (3)
Algic (1)	2 (0)	14 (0)	68 (0)
Dravidian (2)	7 (7)	13 (3)	85 (9)
Indic (4)	4 (5)	4 (2)	98 (3)
Iranian (3)	35 (39)	34 (32)	82 (19)
Romance (8)	6 (4)	53 (19)	99 (1)
Sino-Tibetan (1)	21 (0)	9 (0)	84 (0)
Siouan (1)	13 (0)	13 (0)	96 (0)
Songhay (1)	21 (0)	82 (0)	88 (0)
Southern Daly	4 (0)	6 (0)	90 (0)
Tungusic (1)	28 (0)	27 (0)	57 (0)
Turkic (9)	7 (8)	19 (11)	96 (7)
Uto-Aztecan (1)	33 (0)	30 (0)	81 (0)
Overall	14 (18)	30 (25)	90 (15)



It is possible that the random strategy learns to modify the affix when the exemplar and input lemma are divergent. This might hurt performance in one-shot learning where we don't know anything about the target language.

Comparison between similarity-based and random strategies in one-shot learning

Lemma	Exemplar	Rand. Sel.	Sim. Sel	Target
arrissar	posar : posarien	arrissaren	arrissarien	arrissarien
disputar	descriure : descriuria	disputarta	disputaria	disputaria
repetir	cremar : cremo	repetirer	repetio	repeteixo
engolir	forjar : forjava	engolire	engoliva	engolia
llevar-se	terminar : termino	llevar-se	llevor-se	llevo

The similarity-based model is more likely to copy the affix directly from the exemplar

Training purely on synthetic data

It can be hard to tell whether transfer fails because the model didn't learn a generalizable affixation operation or because there was a mismatch between lexical/phonological triggers

Experiments on synthetic data can help to more clearly understand the type of abstract processes learned by the model

The authors generate fully synthetic probing sets including:

- Affixation
- Reduplication
- Regular phonological alternations (gemination)
- Two different scripts (mix training scripts and Cyrillic script)

Trained models are prompted to inflect examples in the synthetic languages

Results for probing on synthetic languages

Especially affixation has been learned

The random-based model learns language-specific rules for affixation which do not transfer

Both models can reduplicate but mostly for Austronesian languages

Gemination is hard but occurs for Uralic models

Reduplication and gemination show that typology of the training languages matters

Results worse across the board for Cyrillic script

Model	Fam/Lg.	Pref	Pref (Cyrl)	Suff	Suff (Cyrl)	Redup.	Redup. (Cyrl)	Gem.
Rand.	austro	62	36	26	38	0 (10)	0	0
	austro/tgl	0	1	0	0	28 (90)	3 (7)	0
	ger	1	0	25	36	0 (3)	0	0
	ger/deu	0	0	8	10	0 (3)	0	0
	n-congo	92	55	40	41	0 (3)	0	0
	n-congo/swa	100	76	36	25	0 (3)	0	0
	oto	20	15	21	33	0 (3)	0	0
	oto/ote	35	30	1	9	0 (3)	0	0
	uralic	3	0	23	34	0 (3)	0	0
	uralic/fin	0	0	7	22	0 (3)	0	0
	synth	84	62	97	91	0 (3)	0	0
	synth/suff	28	1	100	97	0 (3)	0	0
Sim.	austro	86	75	94	85	30 (30)	0	0
	austro/tgl	30	35	75	63	88 (88)	8 (8)	0
	ger	85	55	99	96	3 (3)	0	8
	ger/deu	86	55	99	98	0	0	5
	n-congo	99	96	98	93	0 (3)	0	3
	n-congo/swa	99	98	88	57	0	0	0
	oto	88	76	95	87	18 (18)	0	0
	oto/ote	96	84	59	17	5 (5)	0	0
	uralic	59	10	97	95	0	0	17
	uralic/fin	52	4	98	98	0	0	12
	synth	94	84	99	95	8 (10)	0	2
	synth/suff	86	42	100	99	0	0	2

Discussion

The results offer some support for the fact that agglutinative languages support transfer because it is easy to segment forms into stems and affixes which biases to model to learn general affixation

Transfer sometimes seems to fail because the model learns overly language-specific behaviour. The authors state that this is evident for Germanic and Uralic languages in the synthetic experiments but I didn’t quite understand this point

The synthetic experiments show a clear link between typology and learning of morphological processes which means that transfer can be successful without cognates

Future work: replace language family tags with WALS features

Takeaways

Memory-based analogy provides a relatively strong basis for inflection prediction

In one-shot transfer exemplars should closely match the elicited inflections

One-shot transfer using this mechanism can achieve higher accuracy than previously thought, even when no genetically related languages are available in training

The model learns to reduplicate arbitrary CV sequences, but applies this process only when targeting a language with reduplication

Learning of morphological processes in general appears to be driven by the input typology