THE ROLE OF INTERPRETABLE PATTERNS IN DEEP LEARNING FOR MORPHOLOGY

Judit Ács and András Kornai

BME, SZTAKI, University of Washington

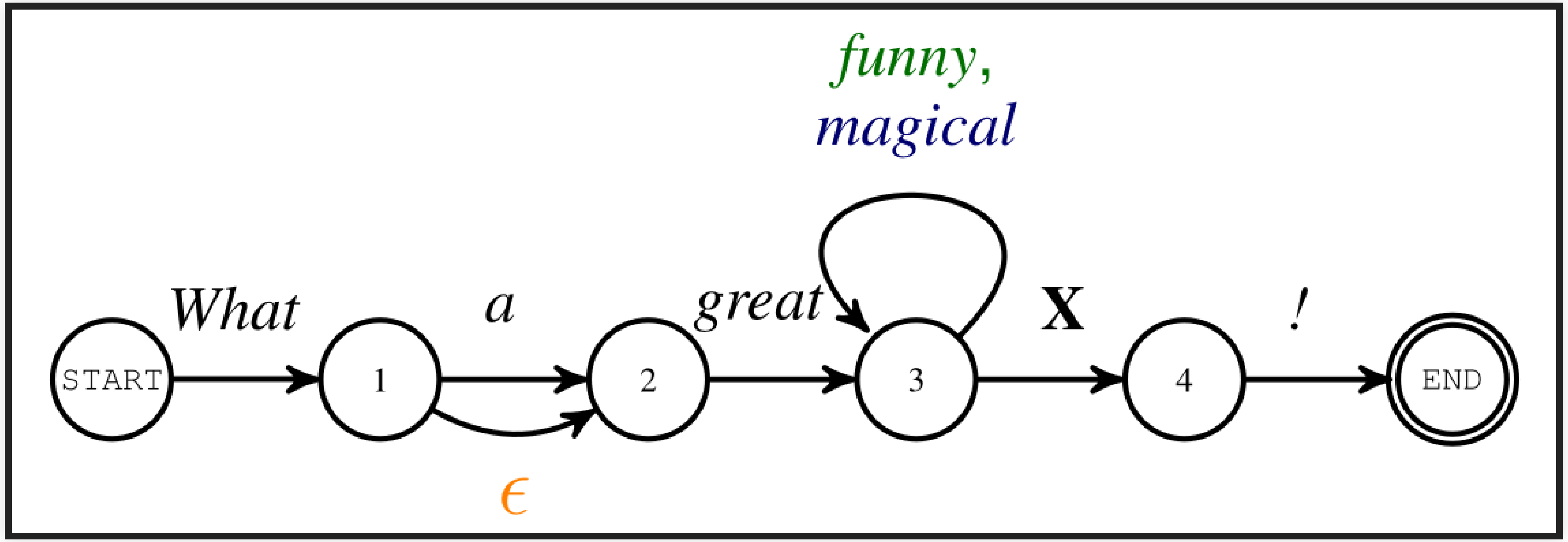
judit@aut.bme.hu

AGENDA

1. Soft Patterns model (Schwartz et al., 2018)
2. Soft Patterns for morphology
3. Model similarity
4. Subword examples

SOFT PATTERNS - SOPA

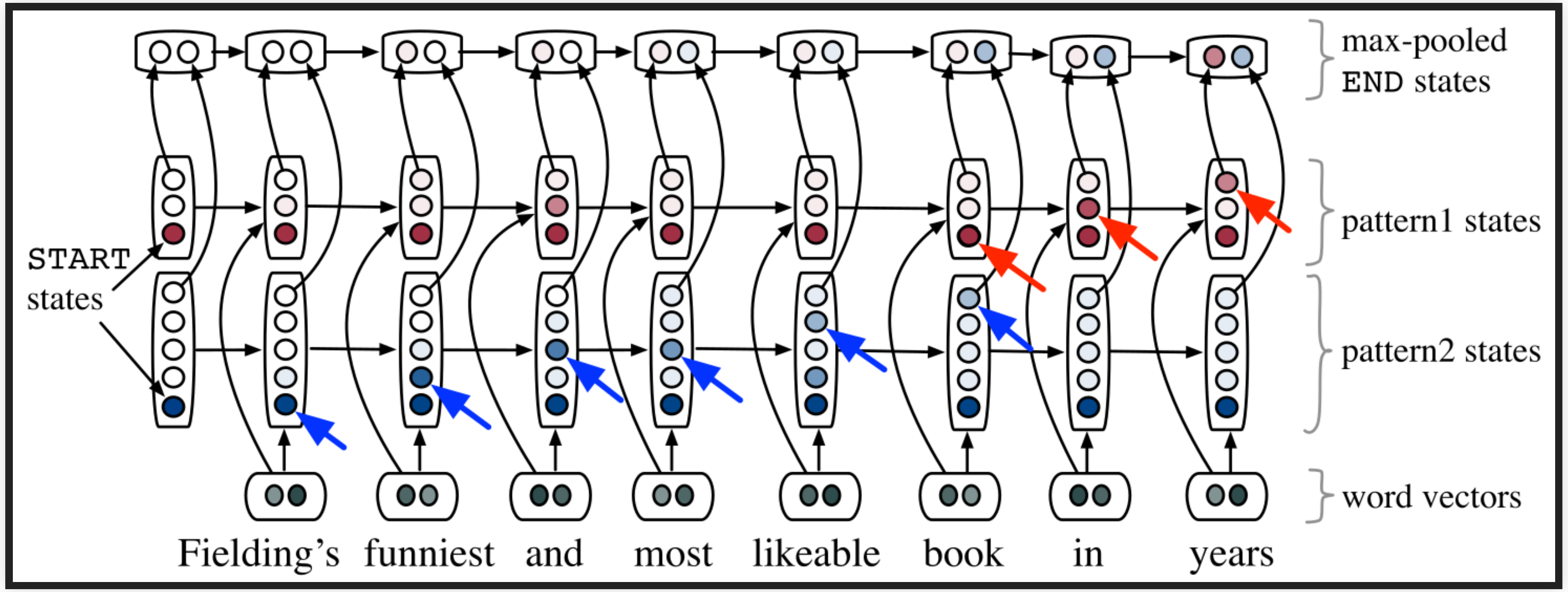
SOFT PATTERNS - SOPA



source

- linear pattern with epsilons and self loops
- fixed pattern length

SOPA UNDER THE HOOD



source

- soft: transition scores are distributions over the entire vocabulary
- each pattern scores every possible subsequence (0-1)
- choose the highest scoring using the Forward algorithm
- fully differentiable, end-to-end training

SOPA FOR SEQ2SEQ

- SoPa was used for (word) sequence classification (sentiment analysis)
- can be used as the encoder of a seq2seq
  - generates a summary of the input
  - generates intermediate outputs that can be used for attention
- LSTM decoder
- the decoder is initialized with the final score of each pattern

SOFT PATTERNS FOR MORPHOLOGY

SOPA SEQ2SEQ FOR MORPHOLOGY

Tasks

1. morphological analysis
2. lemmatization
3. copy

Input-output

- the inputs are inflected word forms in all three tasks
- the output is different

EXAMPLES

	Task	Input	Output
Hungarian	morphological analysis	lepkékben	N IN+ESS PL
French	morphological analysis	désinstalleriez	V COND 2 PL
Hungarian	lemmatization	lepkékben	lepke
French	lemmatization	désinstalleriez	désinstaller
Hungarian	copy	lepkékben	lepkékben
French	copy	désinstalleriez	désinstalleriez

TRAINING DETAILS

SOPA SEQ2SEQ

1. 120 patterns: 40 3-long, 40 4-long and 40 5-long
2. LSTM decoder with 64 hidden cells, 1 layer
3. Luong attention on the intermediate outputs of SoPa

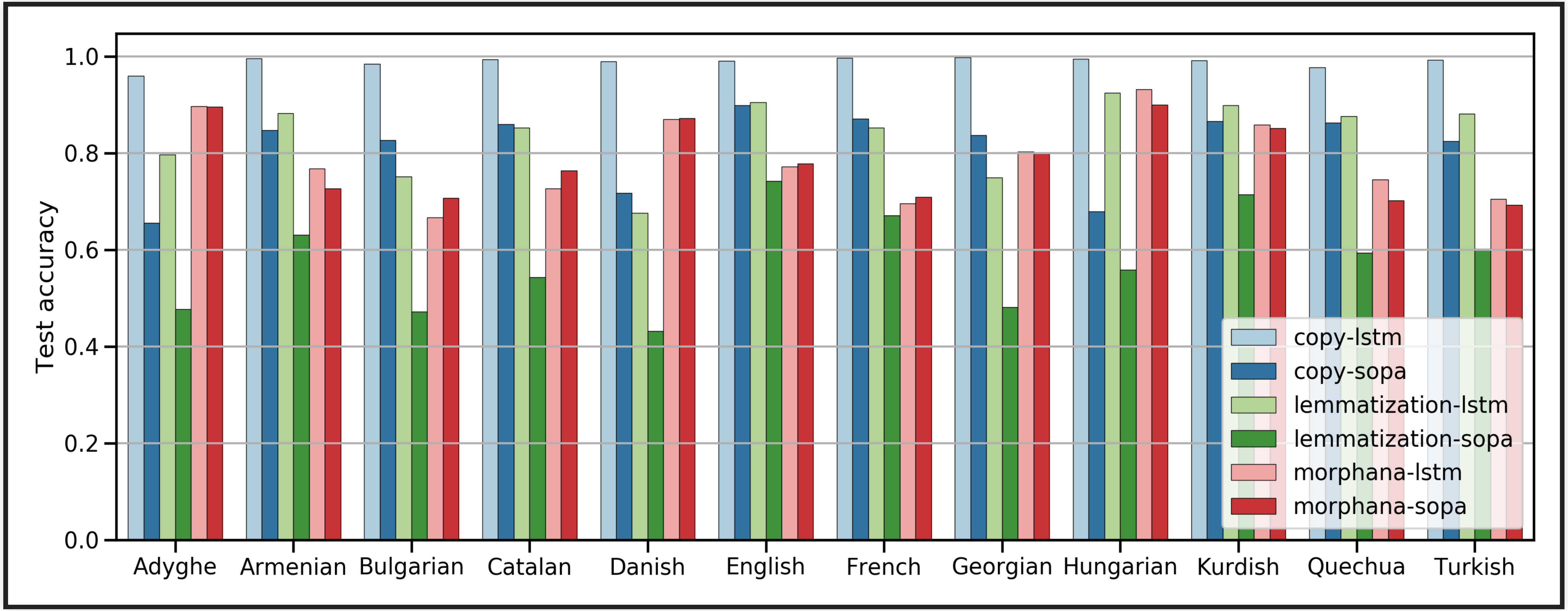
LSTM SEQ2SEQ BASELINE

1. 64 encoder and 64 decoder cells
2. bidirectional encoder, unidirectional decoder
3. 1 layer
4. Luong attention

RESULTS

task	copy		lemmatization		morphana	
	LSTM	SoPa	LSTM	SoPa	LSTM	SoPa
English	99.0	89.9	90.6	74.2	77.2	77.8
Finnish	98.7	65.3	75.6	35.4	78.4	75.1
French	99.7	87.1	85.2	67.1	69.6	70.9
Hungarian	99.4	67.9	92.5	55.9	93.2	90.0

RESULTS



Test accuracy
1.0
0.8
0.6
0.4
0.2
0.0
Adyghe
Armenian
Bulgarian
Catalan
Danish
English
French
Georgian
Hungarian
Kurdish
Quechua
Turkish
copy-lstm
copy-sopa
lemmatization-lstm
lemmatization-sopa
morphana-lstm
morphana-sopa

MODEL SIMILARITY

SIMILARITY BETWEEN TWO SOPA MODELS

- we define a similarity metric between any two models that work on the same input
- morpological analysis, lemmatization and copy take the same input, the inflected form

$$\mathrm{Sim}(M_1, M_2, D) = \frac{1}{|D|} \sum_{d \in D} S(M_1(d), M_2(d)),$$

- D is the dataset,
- and M_1 and M_2 are the models being compared.

SAMPLE-BY-SAMPLE SIMILARITY

For each sample:

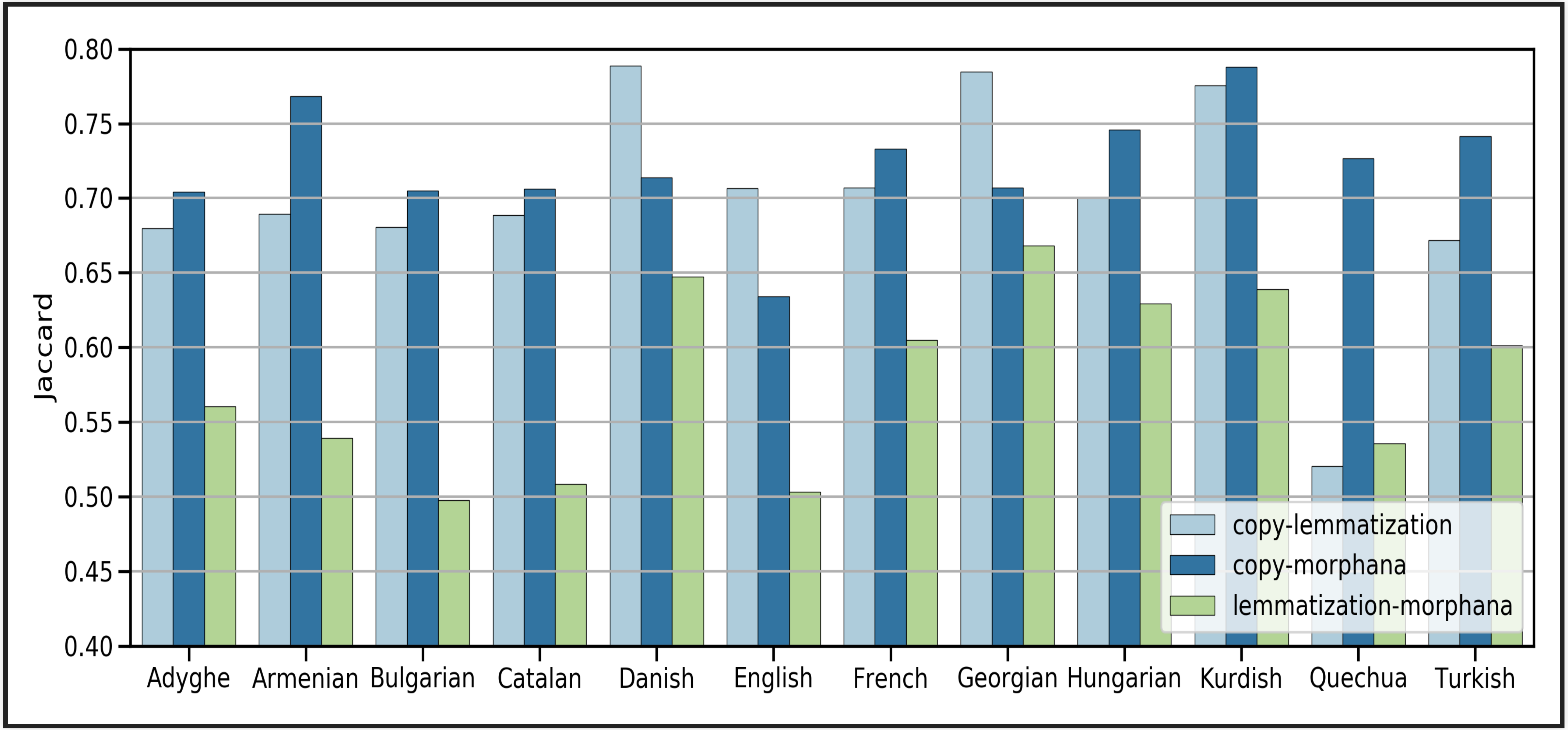
- extract the highest scoring patterns of both models
- find the subword corresponding to these scores
- compute the Jaccard similarity between these subwords

$$\text{Jac}(\text{subword1}, \text{subword2}) = \frac{\text{overlap(subword1,subword2)}}{\text{union(subword1,subword2}}$$

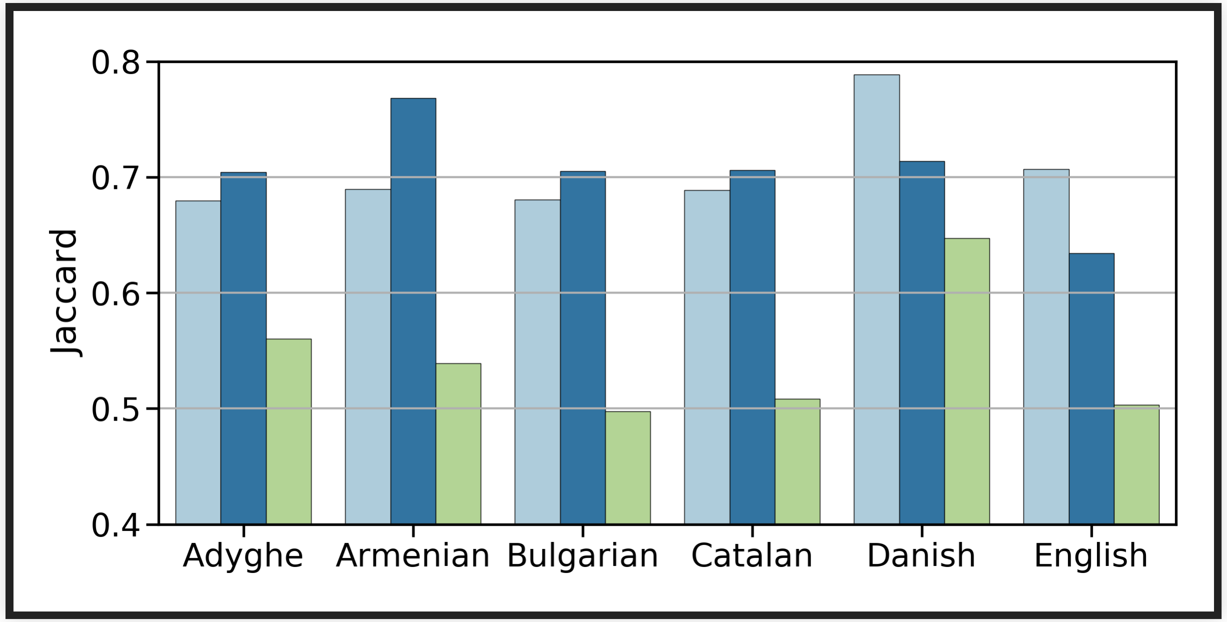
SIMILARITY EXAMPLE

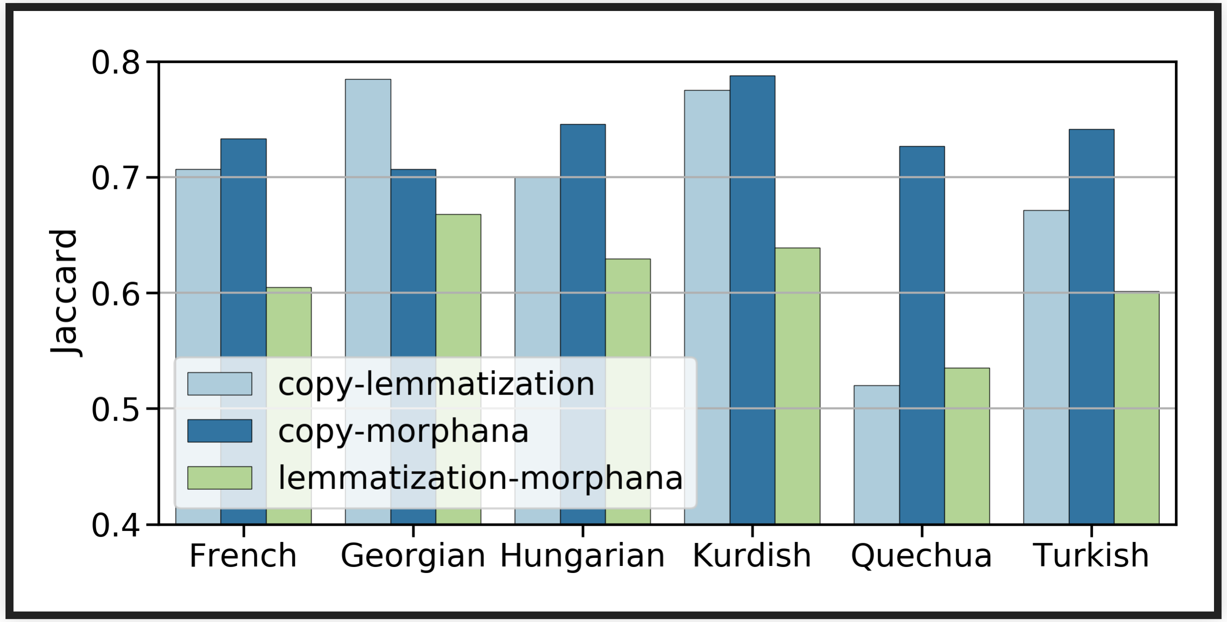
	^ablakban$	^ablakban$	^ablakban$	^ablakkban$	Max
^ablakban$	0	0.2	1	0.75	1
^ablakban$	0	0.5	0.5	0.75	0.75
^ablakban$	0	0.5	0	0.167	0.5
^ablakban$	0	0.75	0.167	0.33	0.75
Max	0	0.75	1	0.75	**0.685**

SIMILARITY RESULTS



SIMILARITY RESULTS





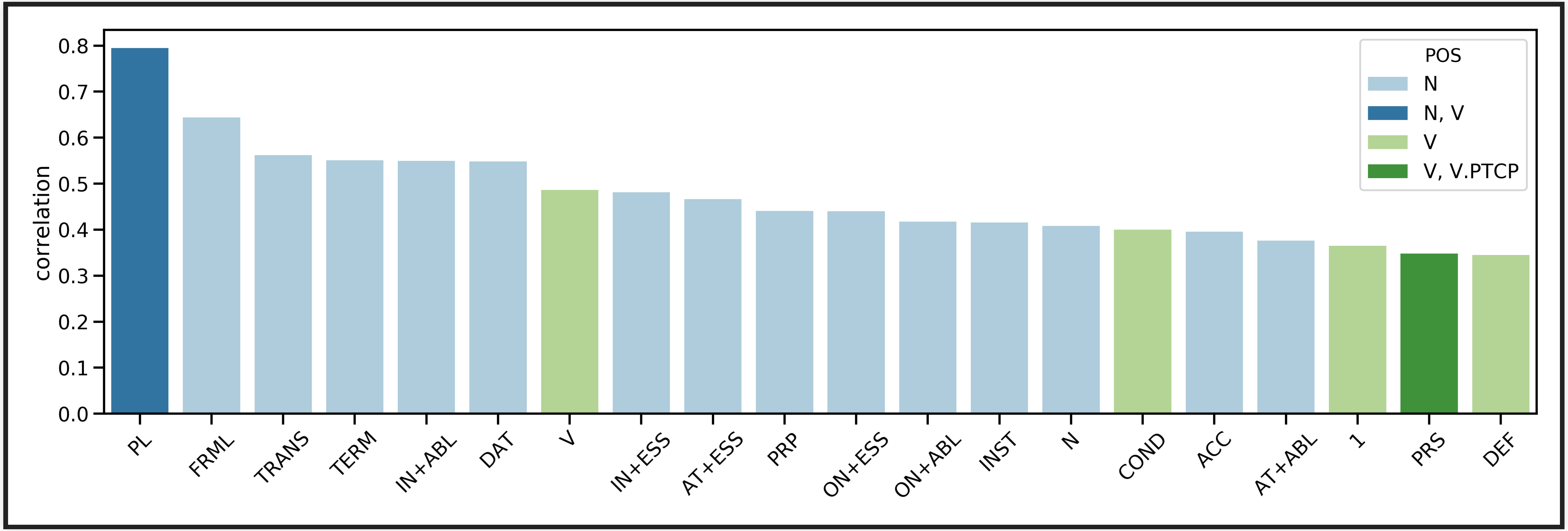
- lemmatization-morphana are the least similar, except in Quechua
  - Quechuan lemmas are much shorter than the inflected forms (13.6 vs. 5.7)
  - Danish is the opposite (9.6 vs. 7.5)
  - 95% of Danish samples are nouns
- copy-morphana are similar in highly inflecting languages

SUBWORD EXAMPLES

EXTRACTING TOP SUBWORDS

- each pattern scores each subword from zero to one
- the highest score is kept as the final score
- we compute the correlation of this score with the presence of every tag
- we extract the subwords corresponding to this pattern

CORRELATIONS



TOP SUBWORDS 1.

tag	morphemes	subwords	subwords when correct	correlation
PL		k_, kig_, t_, kért_, ká_	k_, kig_, kért_, ká_, kal_	0.794
FRML	-ként	kon, ásokb, áln, tokb, nokkén	nokkén, ásokkén, tekkén, ákkén, akkén	0.644
TRANS	-vá/-vé	_, é_, én_	_, é_	0.562
TERM	-ig	kig_, kén, ól_, ba_, ból_	kig_, ig_, mig_, big_, iig_	0.551
IN+ABL	-ból/-ből	ból, ig, ké, ért, ből	ból, ből	0.549
DAT	-nak/-nek	nak_, nek_, tek_, nk_, ok_	nak_, nek_	0.549
V		zt, n, ^, rt, ít	ln, ít, zt, zn, lj	0.486
IN+ESS	-ban/-ben	an_, en_, ént_, hoz_, sza	an_, en_	0.481
AT+ESS	-nál/-nél	nál_, ért_, nek_, nél_, ről_	nál_, nél_	0.467
PRP	-ért	ént_, ért_, kig_, be_, ban_	ért_	0.441
ON+ESS	-on/-en/-ön	n_, nt_, rt_, on_, á_	on_, n_, án_, in_, űn_	0.440
ON+ABL	-ról/-ről	ól_, ől_, ak_, t_, ál_	ól_, ől_	0.417

TOP SUBWORDS 2.

tag	morphemes	subwords	subwords when correct	correlation
INST	-val/-vel	ól_, ál_, ig_, al_, ra_	al_, kel_, vel_, el_, bel_	0.416
N		ól_, ént_, ől_, ál_, en_	ól_, ént_, ől_, ál_, al_	0.408
COND		nek, nak, nok, nátok, nunk	nénk, nátok, nétek, nának, nék	0.400
ACC		rt_, ig_, ak_, ek_, hoz_	at_, et_, át_, őt_, ást_	0.396
AT+ABL	-tól/-től	ól_, ől_, ak_, t_, ál_	ól_, ől_	0.376
1		nak_, nk_, ban_, nek_, be_	nk_, m|_, uk_, nék_, tük_	0.365
PRS		ól_, ént_, ért_, ok_, ál_	ák_, ok_, tek_, unk_, ék_	0.348
DEF		ól_, ént_, ért_, ok_, ál_	ák_, tek_, tuk_, ok_, ta_	0.345
SG		ól_, ént_, ől_, ál_, en_	ól_, ől_, ént_, ál_, áért_	0.337
IND		ól_, ént_, ért_, ok_, ál_	ák_, ok_, tek_, tak_, tuk_	0.326
INDF		l_, ak_, ek_, a_, nt_	nk_, ak_, ek_, l_, sz_	0.324
IN+ALL	-ba/-be	ól_, ként_, ban_, be_, ba_	be_, ba_, ba|_	0.309

TOP SUBWORDS 3.

tag	morphemes	subwords	subwords when correct	correlation
AT+ALL	-hoz/-hez/-höz	rt_, ig_, ak_, ek_, hoz_	hoz_, hez_, eth, höz_, műh	0.299
ON+ALL	-ra/-re	l_, rt_, t_, ig_, al_	ra_, re_	0.284
3		zt, n, ^, rt, ít	ln, zt, zn, ít, lt	0.283
2		ól_, ént_, ért_, ok_, ál_	ok_, tek_, ál_, od_, ád_	0.270
PST		ól_, ént_, ért_, ok_, ál_	tak_, tuk_, tam, tek_, ta_	0.266
NFIN		zt, n, ^, rt, ít	ln, zn, ít, dn, gn	0.232
SBJV		_, k_, kb, kk, kr	k_, _, ke	0.220
V.PTCP		t_, te, tá, ta, á_	ó_, t_, ő_	0.113
NOM		_, k_, kb, kk, kr	k_, _, kr, pc	0.110
V.CVB		l_, rt_, t_, ig_, al_	va_, ve_	0.089
FUT		an, én, ny, en, in	ndó, zan	0.088

CONCLUSION

- we applied SoPa as the encoder of seq2seq models for morphology
- we showed that it's often competitive with the LSTM baseline
- we defined a similarity metric between SoPa models
- we showed that task similarities vary among languages

BIBLIOGRAPHY

Schwartz, Roy, Sam Thomson, and Noah A. Smith. 2018. "SoPa: Bridging CNNs, RNNs, and Weighted Finite-State Machines." In *Proc. 56th Acl Annual Meeting*, 295–305. Melbourne, Australia.