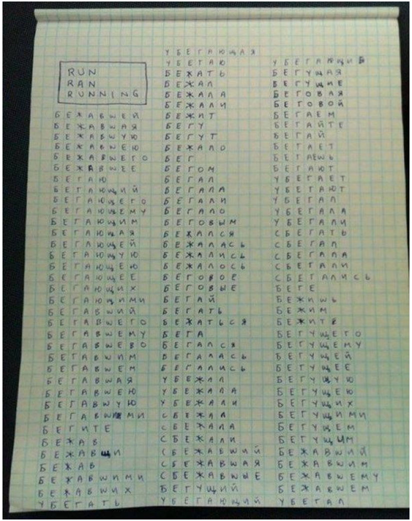
The Secret Life of Words: Exploring Regularity and Systematicity. Part I: Morphological Inflection

Ekaterina Vylomova[ə] ; **Ryan Cotterell**[ʕ,ð]
[ə]University of Melbourne [ʕ]University of Cambridge [ð]ETH Zürich
ekaterina.vylomova@unimelb.edu.au ryan.cotterell@inf.ethz.ch

11 ноября 2020 г.

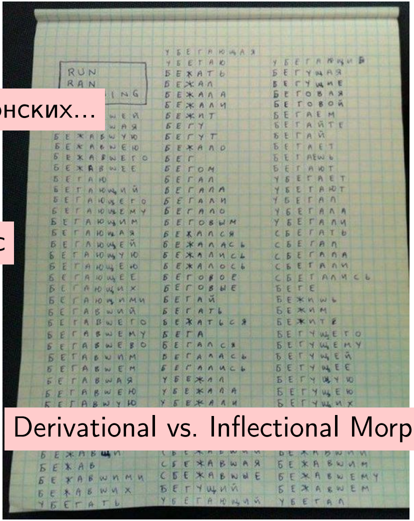
Russian Morphology



Russian Morphology

Всё смешалось в доме Облонских...

Analytic vs. Synthetic



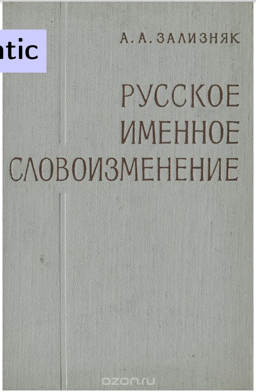
Derivational vs. Inflectional Morphology

Russian Morphology

Morphological Inflection

Can we train a system to learn the regularities from data?

Inflectional Morphology is **Paradigmatic**



Inflectional Morphology: Paradigms (nouns)

Morphological Inflection

беглец + pos=N,case=ACC,num=SG → *беглеца*

Declension of беглéц (anim masc-form ц-stem accent-b) [hide ▲]

	singular	**plural**
nominative	**беглéц** beglée	беглецы́ beglecý
genitive	беглецá beglecá	беглецóв beglecóv
dative	беглецý beglecú	беглецáм beglecám
accusative	беглецá beglecá	беглецóв beglecóv
instrumental	беглецóм beglecóm	беглецáми beglecámi
prepositional	беглецé beglecé	беглецáх beglecáx

ru-noun-table | b | беглец | a=an

Inflectional Morphology: Classes (nouns)

Morphological Inflection

беглец + pos=N,case=ACC,num=SG → *беглеца*

Class	Old class	Nom sg	Nom pl	Gen pl	Declension	Typical gender	Hardness	Examples	Notes
(blank), #	ъ	none	ы/и	ов, [шщчж]ей	2nd	Masculine	Hard	заво́д, язы́к, ча́с	gen pl -ей after sibilants
-а, #-а	ъ-а	none	а	ов, [шщчж]ей	2nd	Masculine	Hard	рука́в/рукава́, по́езд/поезда́	gen pl -ей after sibilants
-ья, #-ья	ъ-ья	none	ья	ьёв/ьев	2nd	Masculine	Hard	дру́г/друзья́	
ин	инъ	ин	е	none	2nd	Masculine	Hard	англича́нин, христиа́ни́н	
ёнок, онок, енок	ёнокъ, онокъ, енокъ	ёнок/о́нок	я́та/а́та	я́т/а́т	2nd	Masculine	Hard	телёнок, внучо́нок	
ёночек, оночек, еночек	ёночекъ, оночекъ, еночекъ	ёночек/о́ночек	я́тки/а́тки	я́ток/а́ток	2nd	Masculine	Hard	цыплёночек, мышо́ночек	
ь-m		ь	и	ей	2nd	Masculine	Soft	до́ждь, ру́бль, ка́мень	
ь-я		ь	я	ей	2nd	Masculine	Soft	кре́ндель/кренделя́	
й		й	и	ёв/ов	2nd	Masculine	Palatal	ча́й, геро́й, ге́ний	includes -ий nouns (old-style -ій)
й-я		й	я	ёв/ов	2nd	Masculine	Palatal	кра́й	includes -ий nouns (old-style -ій)
а		а	ы/и	none, [шщчж]ей	1st	Feminine	Hard	соба́ка, голова́	gen pl -ей after sibilants (stressed only)
я		я	и	ь/й, ей	1st	Feminine	Soft	земля́, револю́ция	stressed gen pl -ей except with vowel stems, otherwise -ь/-й; includes -ия nouns (old-style -ія) but not -ья nouns
ья		ья	ьи	ий, ей	1st	Feminine	Soft	судья́	stressed gen pl -ей (also patterns d and d'), unstressed -ий
о		о	а	none, ей	2nd	Neuter	Hard	стекло́, со́лнце	includes unstressed -е nouns after цшщчж; gen pl -ей after sibilants (stressed only)
о-и, о-ы		о	ы/и	none	2nd	Neuter	Hard	я́блоко/я́блоки, у́хо/у́ши	same as previous
о-ья		о	ья	ьёв/ьев	2nd	Neuter	Hard	перо́/пе́рья, ши́ло/ши́лья, дно́/до́нья	
е, ё		е/ё	я/а	ь/й, ей	2nd	Neuter	Soft	мо́ре, уче́ние	stressed gen pl -ей except with vowel stems, otherwise -ь/-й; includes -ие/-иё nouns (old-style -іе/-іё) but not -ье/-ьё nouns
е́		е́	я/а	ей, [и]й	2nd	Neuter	Soft	бытие́, муде́	nouns with stressed -е́ instead of -ё; includes nouns in -ие́ (old-style -іе́), which have gen pl in -ий
ье, ьё		ье/ьё	ья	ий, ей	2nd	Neuter	Soft	у́стье, копьё	stressed gen pl -ей, unstressed -ий
ь-f		ь	и	ей	3rd	Feminine	Soft	две́рь, гла́сность	
мя		мя	мена	мён	3rd	Neuter	?	и́мя, пле́мя	
$		none	none	none	Indeclinable	—	—	полпути́	

EN Wiktionary: ru-noun-table | b | беглец | a=an

Inflectional Morphology: Classes (nouns); Differs in En/Ru Wiktionaries

Morphological Inflection

беглец + pos=N,case=ACC,num=SG → *беглеца*

Class	Old class	Nom sg	Nom pl	Gen pl	Declension	Typical gender	Hardness	Examples	Notes
(blank), #	ъ	none	ы/и	ов, [шщчж]ей	2nd	Masculine	Hard	заво́д, язы́к, ча́с	gen pl -ей after sibilants
-а , #-а	ъ-а	none	а	ов, [шщчж]ей	2nd	Masculine	Hard	рука́в/рукава́, по́езд/поезда́	gen pl -ей after sibilants
-ья , #-ья	ъ-ья	none	ья	ьёв/ьев	2nd	Masculine	Hard	дру́г/друзья́	
ин	инъ	ин	е	none	2nd	Masculine	Hard	англича́нин, христиани́н	
ёнок , онок , енок	ёнокъ , онокъ , енокъ	ёнок/о́нок	я́та/а́та	я́т/а́т	2nd	Masculine	Hard	телёнок, внучо́нок	
й		й	и	ёв/ов	2nd	Masculine	Palatal	ча́й, геро́й, ге́ний	includes -ий nouns (old-style -ій)
й-я		й	я	ёв/ов	2nd	Masculine	Palatal	кра́й	includes -ий nouns (old-style -ій)
а		а	ы/и	none, [шщчж]е́й	1st	Feminine	Hard	соба́ка, голова́	gen pl -е́й after sibilants (stressed only)
я		я	и	ь/й, е́й	1st	Feminine	Soft	земля́, револю́ция	stressed gen pl -е́й except with vowel stems, otherwise -ь/-й; includes -ия nouns (old-style -ія) but not -ья nouns
ья		ья	ьи	ий, е́й	1st	Feminine	Soft	судья́	stressed gen pl -е́й (also patterns d and d'), unstressed -ий
о		о	а	none, е́й	2nd	Neuter	Hard	стекло́, со́лнце	includes unstressed -е nouns after цшщчж; gen pl -е́й after sibilants (stressed only)
о-и , о-ы		о	ы/и	none	2nd	Neuter	Hard	я́блоко/я́блоки, у́хо/у́ши	same as previous
о-ья		о	ья	ьёв/ьев	2nd	Neuter	Hard	перо́/пе́рья, ши́ло/ши́лья, дно́/до́нья	
е , ё		е/ё	я/а	ь/й, е́й	2nd	Neuter	Soft	мо́ре, уче́ние	stressed gen pl -е́й except with vowel stems, otherwise -ь/-й; includes -ие/-иё nouns (old-style -іе/-іё) but not -ье/-ьё nouns
е́		е́	я/а	ей, [и]й	2nd	Neuter	Soft	бытие́, муде́	nouns with stressed -е́ instead of -ё; includes nouns in -ие́ (old-style -іе́), which have gen pl in -ий
ье , ьё		ье/ьё	ья	ий, е́й	2nd	Neuter	Soft	у́стье, копьё	stressed gen pl -е́й, unstressed -ий
ь-f		ь	и	ей	3rd	Feminine	Soft	две́рь, гла́сность	
мя		мя	мена	мён	3rd	Neuter	?	и́мя, пле́мя	
$		none	none	none	Indeclinable	—	—	полпути́	

RU Wiktionary:сущ ru m a 5b|основа=беглец|основа1=беглец|слоги=по-слогам|бег|лец

EN Wiktionary: ru-noun-table | b | беглец | a=an

Inflectional Morphology: Wiktionary annotation is Not Cross-linguistically Consistent

Other Languages

Hungarian

Inflection (stem in -e-, front unrounded harmony) [less ▲]

	singular	plural
nominative	**szökevény**	szökevények
accusative	szökevényt	szökevényeket
dative	szökevénynek	szökevényeknek
instrumental	szökevénnyel	szökevényekkel
causal-final	szökevényért	szökevényekért
translative	szökevénnyé	szökevényekké
terminative	szökevényig	szökevényekig
essive-formal	szökevényként	szökevényekként
essive-modal	—	—
inessive	szökevényben	szökevényekben
superessive	szökevényen	szökevényeken
adessive	szökevénynél	szökevényeknél
illative	szökevénybe	szökevényekbe
sublative	szökevényre	szökevényekre
allative	szökevényhez	szökevényekhez
elative	szökevényből	szökevényekből
delative	szökevényről	szökevényekről

[Sylak-Glassman, 2016]

Universal Annotation (by John Sylak-Glassman)

1) 23 dimensions of meaning (TAM, case, number, animacy), 212 features
2) A-morphous morphology (Anderson, 1992)
3) Paradigms extracted from English Wiktionary (Kirov et al., 2016)

UniMorph

Schema and datasets for universal morphological annotation

https://unimorph.github.io/

[Sylak-Glassman, 2016]

Universal Annotation (by John Sylak-Glassman)

1) 23 dimensions of meaning (TAM, case, number, animacy), 212 features
2) A-morphous morphology (Anderson, 1992)
3) Paradigms extracted from English Wiktionary (Kirov et al., 2016)

абажур	абажур	N;ACC;SG
абажур	абажур	N;NOM;SG
абажур	абажура	N;GEN;SG
абажур	абажурам	N;DAT;PL
абажур	абажурами	N;INS;PL
абажур	абажурах	N;ESS;PL
абажур	абажуре	N;ESS;SG
абажур	абажуров	N;GEN;PL
абажур	абажуром	N;INS;SG
абажур	абажуру	N;DAT;SG
абажур	абажуры	N;NOM;PL

UniMorph – Universal Annotation

[Sylak-Glassman, 2016]

Universal Annotation (by John Sylak-Glassman)

1) 23 dimensions of meaning (TAM, case, number, animacy), 212 features
2) A-morphous morphology (Anderson, 1992)
3) Paradigms extracted from English Wiktionary (Kirov et al., 2016)

абазинский	абазинская	ADJ;NOM;FEM;SG
абазинский	абазинские	ADJ;ACC;INAN;PL
абазинский	абазинские	ADJ;NOM;PL
абазинский	абазинский	ADJ;INAN;ACC;MASC;SG
абазинский	абазинский	ADJ;NOM;MASC;SG
абазинский	абазинским	ADJ;DAT;PL
абазинский	абазинским	ADJ;INS;MASC;SG
абазинский	абазинским	ADJ;INS;NEUT;SG
абазинский	абазинскими	ADJ;INS;PL
абазинский	абазинских	ADJ;ACC;ANIM;PL
абазинский	абазинских	ADJ;ESS;PL
абазинский	абазинских	ADJ;GEN;PL
абазинский	абазинского	ADJ;ANIM;ACC;MASC;SG
абазинский	абазинского	ADJ;GEN;MASC;SG
абазинский	абазинского	ADJ;GEN;NEUT;SG
абазинский	абазинское	ADJ;ACC;NEUT;SG
абазинский	абазинское	ADJ;NOM;NEUT;SG
абазинский	абазинской	ADJ;DAT;FEM;SG
абазинский	абазинской	ADJ;ESS;FEM;SG
абазинский	абазинской	ADJ;GEN;FEM;SG
абазинский	абазинской	ADJ;INS;FEM;SG
абазинский	абазинском	ADJ;ESS;MASC;SG
абазинский	абазинском	ADJ;ESS;NEUT;SG
абазинский	абазинскому	ADJ;DAT;MASC;SG
абазинский	абазинскому	ADJ;DAT;NEUT;SG
абазинский	абазинскую	ADJ;ACC;FEM;SG

UniMorph – Universal Annotation

[Sylak-Glassman, 2016]

Universal Annotation (by John Sylak-Glassman)

1) 23 dimensions of meaning (TAM, case, number, animacy), 212 features
2) A-morphous morphology (Anderson, 1992)
3) Paradigms extracted from English Wiktionary (Kirov et al., 2016)

вынестись	вынесемся	V;FUT;1;PL
вынестись	вынесетесь	V;FUT;2;PL
вынестись	вынесется	V;FUT;3;SG
вынестись	вынесешься	V;FUT;2;SG
вынестись	вынесись	V;IMP;2;SG
вынестись	вынеситесь	V;IMP;2;PL
вынестись	вынеслась	V;PST;SG;FEM
вынестись	вынеслись	V;PST;PL
вынестись	вынеслось	V;PST;SG;NEUT
вынестись	вынесся	V;PST;SG;MASC
вынестись	вынестись	V;NFIN
вынестись	вынесусь	V;FUT;1;SG
вынестись	вынесутся	V;FUT;3;PL
вынестись	вынесшийся	V.PTCP;ACT;PST
вынестись	вынесясь	V.CVB;PST

SIGMORPHON 2016 Shared Task

[Cotterell et al., 2016]

Morphological (Re-)Inflection (10 Languages)

Task1: *беглец* + pos=N,case=ACC,num=SG → *беглеца*

Task2: *беглецами* + pos=N,case=INS, num=PL + pos=N,case=ACC,num=SG → *беглеца*

Task3: *беглецами* + pos=N,case=ACC,num=SG → *беглеца*

	Standard			Restricted		
System	Task 1	Task 2	Task 3	Task 1	Task 2	Task 3
LMU-1	1.0 (95.56)	1.0 (96.35)	1.0 (95.83)	1.0 (95.56)	1.0 (95.34)	1.0 (90.95)
LMU-2	2.0 (95.56)	2.0 (96.23)	2.0 (95.83)	2.0 (95.56)	2.0 (95.27)	2.0 (90.95)
BIU/MIT-1	—	—	—	4.2 (92.65)	5.2 (77.70)	3.8 (76.39)
BIU/MIT-2	—	—	—	4.2 (93.00)	4.2 (81.29)	—
HEL	—	—	—	3.9 (92.89)	3.5 (86.30)	3.2 (86.48)
MSU	3.8 (84.06)	3.6 (86.06)	3.8 (84.87)	6.2 (84.06)	6.0 (79.68)	6.2 (62.16)
CU	4.6 (81.02)	5.0 (72.98)	5.0 (71.75)	7.3 (81.02)	6.9 (69.89)	5.5 (67.91)
EHU	5.5 (79.24)	—	—	8.0 (79.67)	—	—
COL/NYU	6.5 (67.86)	4.7 (75.59)	4.8 (67.61)	9.2 (67.86)	7.2 (77.34)	6.3 (53.56)
OSU	—	—	—	9.0 (72.71)	—	—
UA	4.6 (81.83)	4.7 (74.06)	4.4 (71.23)	—	—	—
ORACLE.E	97.49	98.15	97.97	98.32	97.84	95.80

SIGMORPHON 2016 Shared Task

[Cotterell et al., 2016]

Morphological (Re-)Inflection (10 Languages)

Task1: *беглец* + pos=N,case=ACC,num=SG → *беглеца*
Task2: *беглецами* + pos=N,case=INS, num=PL + pos=N,case=ACC,num=SG → *беглеца*
Task3: *беглецами* + pos=N,case=ACC,num=SG → *беглеца*

System	Standard Task 1	Standard Task 2	Standard Task 3	Restricted Task 1	Restricted Task 2	Restricted Task 3
LMU-1	1.0 (95.56)	1.0 (96.35)	1.0 (95.83)	1.0 (95.56)	1.0 (95.34)	1.0 (90.95)
LMU-2	2.0 (95.56)	2.0 (96.23)	2.0 (95.83)	2.0 (95.56)	2.0 (95.27)	2.0 (90.95)
BIU/MIT-1	—	—	—	4.2 (92.65)	5.2 (77.70)	3.8 (76.39)
BIU/MIT-2	—	—	—	4.2 (93.00)	4.2 (81.29)	—
HEL	—	—	[illegible]	[illegible]	[illegible]	[illegible]
MSU	3.8 (84.06)	3.6 [illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CU	4.6 (81.02)	5.0 [illegible]	[illegible]	[illegible]	[illegible]	[illegible]
EHU	5.5 (79.24)	—	[illegible]	[illegible]	[illegible]	[illegible]
COL/NYU	6.5 (67.86)	4.7 [illegible]	[illegible]	[illegible]	[illegible]	[illegible]
OSU	—	—	[illegible]	[illegible]	[illegible]	[illegible]
UA	4.6 (81.83)	4.7 [illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ORACLE.E	97.49	98.1[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

LMU+BIU+Helsinki: Neural (seq2seq) +/- aligner

MSU+Col/NYU: rule-based/heuristics

Others: external aligner+WFST/CRF

SIGMORPHON 2016 Shared Task

[Kann and Schuetze, 2016]

Morphological (Re-)Inflection (10 Languages): Neural encoder–decoders

1) character-level input: <s> r u n OUT_POS=V OUT_NUM=SG OUT_TENSE=PRES </s> Output: <s> r u n s </s>
2) Ensembles of seq2seq (GRUs + soft attention (Bahdanau et al., 2015))
3) Enriching the data with combinations of other (non-lemma) forms

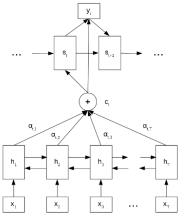


Figure 1: System overview. The input x consists of characters as well as input and output tags. The output y consists of characters only.

SIGMORPHON 2016 Shared Task

[Kann and Schuetze, 2016]

Morphological (Re-)Inflection (10 Languages): Neural encoder–decoders

1) character-level input: <s> r u n OUT_POS=V OUT_NUM=SG OUT_TENSE=PRES </s> Output: <s> r u n s </s>
2) Ensembles of seq2seq (GRUs + soft attention (Bahdanau et al., 2015))
3) Enriching the data with combinations of other (non-lemma) forms

	T2, given	T2, restricted		T2, standard	
Dataset	**no. samples**	**no. samples**	**factor**	**no. samples**	**factor**
Arabic	14,400	28,800	2	458,814	32
Finnish	14,400	28,800	2	116,206	8
Georgian	14,400	28,800	2	196,396	14
German	14,400	28,800	2	166,148	12
Hungarian	21,600	43,200	2	643,630	30
Maltese	21,600	43,200	2	1,629,446	75
Navajo	14,385	28,770	2	160,332	11
Russian	14,400	28,800	2	129,302	9
Spanish	14,400	28,800	2	211,030	15
Turkish	14,400	28,800	2	392,136	27

SIGMORPHON 2016 Shared Task

[Kann and Schuetze, 2016]

Morphological (Re-)Inflection (10 Languages): Neural encoder–decoders

1) character-level input: <s> r u n OUT_POS=V OUT_NUM=SG OUT_TENSE=PRES </s> Output: <s> r u n s </s>
2) Ensembles of seq2seq (GRUs + soft attention (Bahdanau et al., 2015))
3) Enriching the data with combinations of other (non-lemma) forms

Language	Task 1	Task 2	Task 3
Arabic	95.47%	97.38%	96.52%
Finnish	96.80%	97.40%	96.56%
Georgian	98.50%	99.14%	98.87%
German	95.80%	97.45%	95.60%
Hungarian	99.30%	99.67%	99.50%
Maltese	88.99%	88.17%	87.83%
Navajo	91.48%	96.64%	96.20%
Russian	91.46%	91.00%	89.91%
Spanish	98.84%	98.74%	97.96%
Turkish	98.93%	97.94%	99.31%

Table 2: Exact-match accuracy per language for the standard track of the SIGMORPHON 2016 Shared Task.

SIGMORPHON 2016 Shared Task

[Kann and Schuetze, 2016]

Morphological (Re-)Inflection (10 Languages): Neural encoder–decoders

1) character-level input: <s> r u n OUT_POS=V OUT_NUM=SG OUT_TENSE=PRES </s> Output: <s> r u n s </s>
2) Ensembles of seq2seq (GRUs + soft attention (Bahdanau et al., 2015))
3) Enriching the data with combinations of other (non-lemma) forms

Language	Task 1	Task 2	Task 3
Arabic	95.47%	91.09%	82.80%
Finnish	96.80%	96.81%	93.18%
Georgian	98.50%	98.50%	96.21%
German	95.80%	96.22%	92.41%
Hungarian	99.30%	99.42%	98.37%
Maltese	88.99%	86.88%	84.25%
Navajo	91.48%	97.81%	83.50%
Russian	91.46%	90.11%	87.13%
Spanish	98.84%	98.45%	96.69%
Turkish	98.93%	98.38%	95.00%

Table 3: Exact-match accuracy per language for the restricted track of the SIGMORPHON 2016 Shared Task.

SIGMORPHON 2016 Shared Task

[Aharoni and Goldberg, 2017]

Morphological (Re-)Inflection (10 Languages): Neural encoder–decoders

1) Extract input–output string alignments; 2) Train seq2seq (LSTM-based) models to learn a sequence of operations (hard monotonic attention)

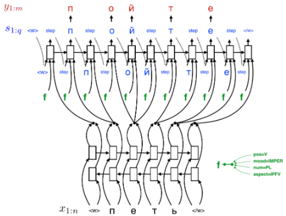


Figure 1: The hard attention network architecture. A round tip expresses concatenation of the inputs it receives. The attention is promoted to the next input element once a step action is predicted.

SIGMORPHON 2016 Shared Task

[Aharoni and Goldberg, 2017]

Morphological (Re-)Inflection (10 Languages): Neural encoder–decoders

1) Extract input–output string alignments; 2) Train seq2seq (LSTM-based) models to learn a sequence of operations (hard monotonic attention)

	suffixing+stem changes			circ.	suffixing+agg.+v.h.			c.h.	templatic		
	RU	DE	ES	GE	FI	TU	HU	NA	AR	MA	Avg.
MED	91.46	95.8	98.84	98.5	95.47	98.93	96.8	91.48	**99.3**	**88.99**	95.56
Soft	92.18	96.51	98.88	**98.88**	**96.99**	**99.37**	**97.01**	**95.41**	**99.3**	88.86	**96.34**
Hard	**92.21**	**96.58**	**98.92**	98.12	95.91	97.99	96.25	93.01	98.77	88.32	95.61

Table 3: Results on the SIGMORPHON 2016 morphological inflection dataset. The text above each language lists the morphological phenomena it includes: circ.=circumfixing, agg.=agglutinative, v.h.=vowel harmony, c.h.=consonant harmony

SIGMORPHON 2016 Shared Task

[Aharoni and Goldberg, 2017]

Morphological (Re-)Inflection (10 Languages): Neural encoder–decoders

1) Extract input–output string alignments; 2) Train seq2seq (LSTM-based) models to learn a sequence of operations (hard monotonic attention)

Errors

глядеть pos=V,tense=PRS,per=1,num=SG,aspect=IPFV gold: гляжу predicted: глядею

увлекаться pos=V,tense=PRS,per=1,num=SG,aspect=IPFV gold: увлекаюсь predicted: увлеклюсь

звать pos=V,tense=PRS,per=3,num=SG,aspect=IPFV gold: зовёт predicted: звает

SIGMORPHON 2016 Shared Task

[Aharoni and Goldberg, 2017]

Morphological (Re-)Inflection (10 Languages): Neural encoder–decoders

1) Extract input–output string alignments; 2) Train seq2seq (LSTM-based) models to learn a sequence of operations (hard monotonic attention)

Errors

зять pos=N,case=GEN,num=PL gold: зятьёв predicted: зятей

перстень pos=N,case=GEN,num=PL gold: перстней predicted: перстеее

телекамера pos=N,case=GEN,num=PL gold: телекамер predicted: телекаморо

SIGMORPHON 2016 Shared Task

[Aharoni and Goldberg, 2017]

Morphological (Re-)Inflection (10 Languages): Neural encoder–decoders

1) Extract input–output string alignments; 2) Train seq2seq (LSTM-based) models to learn a sequence of operations (hard monotonic attention)

Errors

лоботряс pos=N,case=ACC,num=PL gold: лоботрясов predicted: лоботрясы

львица pos=N,case=ACC,num=PL gold: львиц predicted: львица

милиционер pos=N,case=ACC,num=PL gold: милиционеров predicted: милиционеры

светлячок pos=N,case=ACC,num=PL gold: светлячков predicted: светлячки

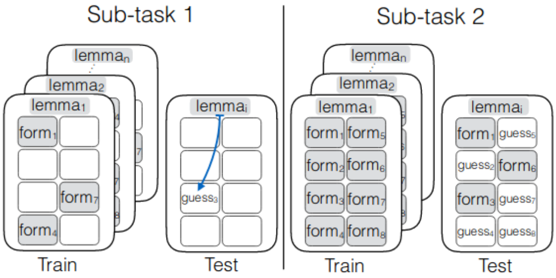
скот pos=N,case=ACC,num=PL gold: скотов predicted: скоты

счёт pos=N,case=ACC,num=PL gold: счета predicted: счеты

SIGMORPHON 2017 Shared Task

Universal Morphological Reinflection (52 Languages)

Task1: Morphological Reinflection
Task2: Paradigm Completion

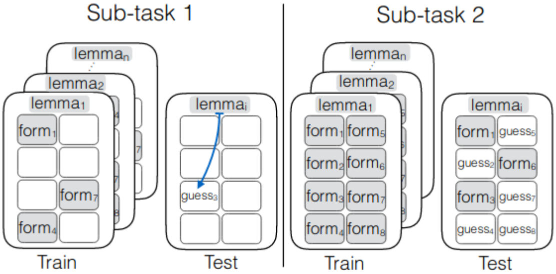


SIGMORPHON 2017 Shared Task

[Cotterell et al., 2017]

Universal Morphological Reinflection (52 Languages)

3 Settings: Low (100 samples), Medium (1000), High (10,000)
Sampled based on their token frequency in Wikipedia corpus (with resampling for syncretic slots)



SIGMORPHON 2017 Shared Task

[Cotterell et al., 2017]

Universal Morphological Reinflection (52 Languages)

3 Settings: Low (100 samples), Medium (1000), High (10,000)
Sampled based on their token frequency in Wikipedia corpus (with resampling for syncretic slots)

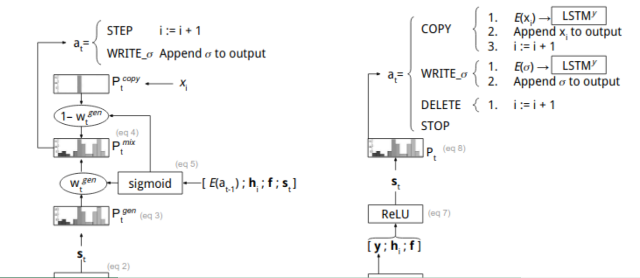
	Sub-task 1			Sub-task 2		
	High	Medium	Low	High	Medium	Low
Albanian	99.00(UE-LMU)	89.40(CU-1)	31.00(CU-1)	98.35(LMU-2)	88.81(LMU-1)	66.63(LMU-2)
Arabic	94.50(CLUZH-7)	79.70(LMU-2)	37.00(CLUZH-7)	95.48(LMU-2)	90.21(LMU-2)	80.43(LMU-2)
Armenian	97.50(UE-LMU)	91.50(LMU-2)	58.70(CLUZH-7)	98.78(LMU-2)	97.77(LMU-2)	93.92(LMU-2)
Basque	100.00(UTNII-1)	89.00(UE-LMU)	20.00(LMU-2)	—	94.14(LMU-2)	93.02(LMU-2)
Bengali	100.00(UE-LMU)	99.00(CLUZH-1)	68.00(CLUZH-3)	92.61(LMU-1)	91.72(LMU-2)	90.19(LMU-2)
Bulgarian	98.10(UE-LMU)	82.50(LMU-2)	57.10(CU-1)	85.93(LMU-2)	55.95(LMU-2)	49.58(LMU-2)
Catalan	98.40(CLUZH-1)	92.60(CLUZH-7)	66.40(CU-1)	99.35(LMU-2)	97.06(LMU-2)	94.16(baseline)
Czech	94.10(UE-LMU)	86.30(CU-1)	44.00(CLUZH-7)	86.00(LMU-1)	58.61(LMU-2)	34.96(LMU-2)
Danish	94.50(UE-LMU)	83.60(LMU-2)	75.50(CLUZH-7)	75.74(LMU-2)	71.15(baseline)	53.11(CU-1)
Dutch	96.90(UE-LMU)	86.50(LMU-2)	53.60(baseline)	89.30(LMU-2)	86.53(LMU-2)	56.64(LMU-2)
English	97.20(UE-LMU)	94.70(LMU-2)	90.60(UA-1)	91.60(baseline)	84.00(baseline)	84.40(CU-1)
Estonian	98.90(UE-LMU)	82.40(UE-LMU)	32.90(CLUZH-7)	97.90(LMU-2)	92.43(LMU-2)	77.42(LMU-2)
Faroese	87.80(CLUZH-7)	68.10(CLUZH-7)	42.40(CLUZH-7)	71.90(LMU-2)	68.31(LMU-2)	57.55(LMU-2)
Finnish	95.10(UE-LMU)	78.40(UE-LMU)	19.70(CLUZH-7)	93.67(LMU-2)	89.48(LMU-2)	76.30(LMU-2)
French	89.50(UE-LMU)	80.30(CLUZH-7)	66.00(CLUZH-7)	98.83(LMU-2)	95.38(LMU-2)	87.45(LMU-2)
Georgian	99.40(LMU-2)	93.40(CLUZH-7)	85.60(LMU-2)	96.20(LMU-2)	89.67(LMU-2)	86.82(LMU-2)
German	93.00(UE-LMU)	80.00(CLUZH-4)	68.10(CLUZH-4)	85.88(LMU-2)	77.56(LMU-2)	74.66(LMU-2)

SIGMORPHON 2017 Shared Task

[Makarov et al., 2017]

Universal Morphological Reinflection (52 Languages): Neural encoder–decoders

1) (Align & Copy): Based on Aharoni and Goldberg, 2017
2) Extract input–output string alignments (add COPY/edit operations) 2) Train seq2seq (LSTM-based) models to learn a sequence of operations (hard monotonic attention)



SIGMORPHON 2017 Shared Task

[Makarov et al., 2017]

Universal Morphological Reinflection (52 Languages)

3 Settings: Low (100 samples), Medium (1000), High (10,000)
Sampled based on their token frequency in Wikipedia corpus (with resampling for syncretic slots)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	t
⟨s⟩		f		l		o			g				⟨/s⟩	y
⟨s⟩	STEP	f	STEP	l	STEP	o	STEP	STEP	g	STEP	STEP	STEP	⟨/s⟩	a_t
⟨s⟩	⟨s⟩	f	f	l	l	i	i	e	g	g	e	n	⟨/s⟩	x_i
0	0	1	1	2	2	3	3	4	5	5	6	7	8	i

1	2	3	4	5	6	7	8	9	t
f	l			o	g				y
COPY	COPY	DELETE	DELETE	o	COPY	DELETE	DELETE	STOP	a_t
f	l	i	e	g	g	e	n	–	x_i
1	2	3	4	5	5	6	7	7	i

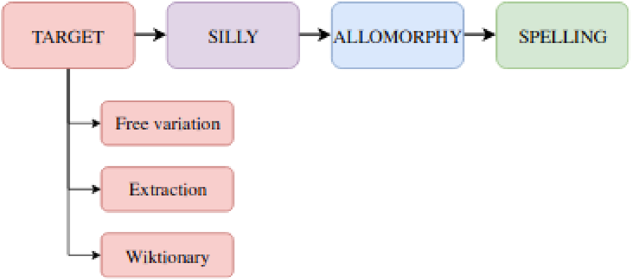
Table 1: Examples of generating German “flog” from “fliegen”: HACM (left), HAEM (right). i is the attention pointer, x_i the currently attended lemma character, a the sequence of actions, y the output, t the index over actions.

SIGMORPHON 2017 Shared Task

[Gorman et al., 2019]

Error taxonomy

What are common errors that neural systems make?



SIGMORPHON 2017 Shared Task

[Gorman et al., 2019]

Error taxonomy

What are common errors that neural systems make?

Types of Errors

✓Free variation error: more than one acceptable form exists
✓Extraction errors: flaws in UniMorph's parsing of Wiktionary
✓Wiktionary errors: errors in the Wiktionary data itself
✓Silly errors: "bizarre" errors which defy any purely linguistic characterization ("*membled" instead of "mailed" or enters a loop such as "ynawemaylmyylmyylmyylmyylmyylmyym..." instead of "ysnewem")
✓Allomorphy errors: misapplication of existing allomorphic patterns
✓Spelling errors: forms that do not follow language-specific orthographic conventions

SIGMORPHON 2017 Shared Task

[Gorman et al., 2019]

Error taxonomy

What are common errors that neural systems make?

	Target	Silly		Allomorphy		Spelling	
Language		UE-LMU-1	CLUZH-7	UE-LMU-1	CLUZH-7	UE-LMU-1	CLUZH-7
Dutch	8	1	1	19	16	5	7
English	3	0	0	18	18	7	11
Finnish	11	7	7	33	48	0	0
German	3	4	10	54	67	9	9
Hungarian	83	21	9	37	44	1	0
Italian	5	5	1	11	16	0	2
Latin	119	2	0	76	93	0	0
Polish	5	6	3	60	67	2	4
Portuguese	1	1	0	6	7	1	2
Romanian	54	3	5	61	69	1	2
Russian	7	7	0	48	45	23	28
Spanish	7	2	1	12	12	6	6
Total	306	59	37	435	502	55	71

[Gorman et al., 2019]

Error taxonomy

What are common errors that neural systems make?

Allomorphy Errors

✓Stem-final vowels in Finnish (*pohjanpystykorvojen); Consonant gradation in Finnish (*ei kiemurda)
✓Ablaut in Dutch and German (*pront; *saufte)
✓Umlaut (*Einwohnerzähle, *Förmer), plural suffixes, Verbal prefixes in German (*umkehre)
✓Linking vowels in Hungarian (*masszázsakból instead of *masszázsokból)
✓Yers (*klęsek instead of klęsk), Genitive singular suffixes in Polish (*izotopa)
✓Animacy in Polish and Russian (грузин vs. магазин in ACC.SG)
✓Aspect in Russian (*будешь сорвать)
✓Internal inflection in Russian compounds (*государствах-донорах, *лёгких промышленности (ACC.PL))

CoNLL–SIGMORPHON 2018 Shared Task

[Cotterell et al., 2018]

Universal Morphological Reinflection (103 Languages)

Task1: Morphological Reinflection (Low, Medium, High)
Task2: Inflection in Context

CoNLL–SIGMORPHON 2018 Shared Task

[Cotterell et al., 2018]

Universal Morphological Reinflection (103 Languages)

Task1: Morphological Reinflection (Low, Medium, High)
Task2: Inflection in Context

	High	Medium	Low
uzh-01	**96.00 / 0.08**	**86.64 / 0.26**	57.18 / 1.00
uzh-02	95.97 / 0.08	86.38 / 0.27	**57.21 / 1.02**
bme-02	94.66 / 0.11	67.26 / 0.88	2.43 / 6.91
iitbhu-iiith-01	94.43 / 0.11	82.90 / 0.34	49.79 / 1.18
iitbhu-iiith-02	94.43 / 0.11	84.19 / 0.32	52.60 / 1.10
bme-03	93.97 / 0.12	67.36 / 0.75	3.63 / 6.75
bme-01	93.88 / 0.12	67.43 / 0.75	3.74 / 6.72
msu-04	91.87 / 0.23	76.40 / 0.55	31.40 / 2.16
iit-varanasi-01	91.73 / 0.16	70.17 / 0.66	23.33 / 2.40
waseda-01	91.12 / 0.19	77.38 / 0.67	44.09 / 1.68
msu-03	90.52 / 0.25	75.74 / 0.55	25.86 / 2.38
axsemantics-01	84.19 / 0.40	*58.00 / 1.10*	*72.00 / 0.96*
msu-02	82.68 / 0.41	69.45 / 0.79	41.61 / 1.86
racai-01	*79.93 / 0.43*	— / —	— / —
hamburg-01	77.53 / 0.44	74.03 / 0.54	40.28 / 1.45
axsemantics-02	74.77 / 0.68	60.00 / 1.03	*14.89 / 3.89*
msu-01	74.33 / 0.78	*64.57 / 0.93*	— / —
tuebingen-oslo-03	63.05 / 1.15	30.98 / 2.25	1.39 / 5.70
tuebingen-oslo-02	56.60 / 1.34	29.72 / 2.36	4.43 / 5.06
kucst-01	*54.37 / 1.57*	*32.28 / 2.23*	*2.79 / 5.28*
tuebingen-oslo-01	49.52 / 1.67	20.97 / 2.81	0.00 / 7.94
ua-08	— / —	— / —	53.22 / 1.35
ua-05	— / —	— / —	50.53 / 1.34
ua-06	— / —	— / —	49.73 / 1.46
ua-03	— / —	— / —	44.82 / 1.45
ua-02	— / —	— / —	*41.61 / 2.47*
ua-07	— / —	— / —	*39.52 / 1.76*
ua-01	— / —	— / —	38.22 / 2.02
ua-04	— / —	— / —	21.25 / 3.43
baseline	77.42 / 0.51	63.53 / 0.90	38.89 / 1.88
oracle-fc	99.87 / –	98.27 / –	77.23 / –
oracle-e	98.90 / –	93.74 / –	74.88 / –

CoNLL–SIGMORPHON 2018 Shared Task

[Cotterell et al., 2018]

Universal Morphological Reinflection (103 Languages)

Task1: Morphological Reinflection (Low, Medium, High)
Task2: Inflection in Context

Pashto	100.00(waseda-1)	85.00(uzh-1)	48.00(uzh-2)
Persian	99.90(bme-2)	93.40(uzh-2)	67.60(uzh-2)
Polish	93.40(uzh-2)	82.40(uzh-2)	49.40(ua-6)
Portuguese	98.60(uzh-2)	94.80(uzh-2)	75.80(uzh-2)
Quechua	99.90(uzh-2)	98.90(uzh-1)	70.20(uzh-2)
Romanian	89.00(uzh-2)	77.60(uzh-1)	46.20(uzh-1)
Russian	94.40(uzh-2)	86.90(uzh-1)	53.50(uzh-1)
Sanskrit	96.50(uzh-1)	85.90(uzh-2)	58.00(uzh-1)
Scottish-gaelic	—	94.00(iitbhu-iiith-1)	74.00(iitbhu-iiith-2)
Serbo-croatian	92.40(uzh-2)	86.10(uzh-1)	44.80(ua-3)
Slovak	97.10(uzh-1)	78.60(uzh-1)	51.80(uzh-2)
Slovene	97.40(uzh-1)	86.20(uzh-1)	58.00(uzh-2)

SIGMORPHON 2019 Shared Task

[McCarthy et al., 2019]

Morphological Analysis in Context and Cross-Lingual Transfer for Inflection (100 Language Pairs)

Task1: Cross-lingual Transfer for Morphological Inflection (10k HR +100 LR)
Task2: Inflection in Context

Low-resource target training data (Asturian)

`facer`	"fechu"	V;V.PTCP;PST
`aguar`	"aguà"	V;PRS;2;PL;IND
⋮	⋮	⋮

High-resource source language training data (Spanish)

`tocar`	"tocando"	V;V.PTCP;PRS
`bailar`	"bailaba"	V;PST;IPFV;3;SG;IND
`mentir`	"mintió"	V;PST;PFV;3;SG;IND
⋮	⋮	⋮

Test input (Asturian)

`baxar`		V;V.PTCP;PRS

Test output (Asturian)

"baxando"

SIGMORPHON 2019 Shared Task

[McCarthy et al., 2019]

Morphological Analysis in Context and Cross-Lingual Transfer for Inflection (100 Language Pairs)

Task1: Cross-lingual Transfer for Morphological Inflection (10k HR +100 LR)
Task2: Inflection in Context

Team	Avg. Accuracy	Avg. Levenshtein
AX-01	18.54	3.62
AX-02	24.99	2.72
CMU-03	**58.79**	1.52
IT-IST-01	49.00	**1.29**
IT-IST-02	50.18	1.32
Tuebingen-01†	34.49	1.88
Tuebingen-02†	20.86	2.36
UAlberta-01*	48.33	1.23
UAlberta-02*†	54.75	1.03
UAlberta-03*†	8.45	4.06
UAlberta-04*†	11.00	3.86
UAlberta-05*	4.10	3.08
UAlberta-06*†	26.85	2.65
Baseline	48.55	1.33

SIGMORPHON 2019 Shared Task

[McCarthy et al., 2019]

Morphological Analysis in Context and Cross-Lingual Transfer for Inflection (100 Language Pairs)

Task1: Cross-lingual Transfer for Morphological Inflection (10k HR +100 LR)
Task2: Inflection in Context

czech–kashubian	52.0	78.0	CMU-03	polish–kashubian	74.0	78.0	CMU-03
czech–latin	8.4	42.0	CMU-03	polish–old-church-slavonic	40.0	58.0	CMU-03
danish–middle-high-german	72.0	82.0	it-ist-02	portuguese–russian	27.5	76.3	CMU-03
danish–middle-low-german	36.0	44.0	it-ist-01	romanian–latin	6.7	41.3	CMU-03
danish–north-frisian	28.0	46.0	CMU-03	russian–old-church-slavonic	34.0	64.0	CMU-03
danish–west-frisian	42.0	43.0	CMU-03	russian–portuguese	50.5	88.4	CMU-03
danish–yiddish	76.0	67.0	it-ist-01	sanskrit–bengali	33.0	65.0	CMU-03
dutch–middle-high-german	76.0	78.0	it-ist-01 / it-ist-02	sanskrit–pashto	34.0	43.0	CMU-03
dutch–middle-low-german	42.0	52.0	it-ist-02	slovak–kashubian	54.0	76.0	CMU-03
dutch–north-frisian	32.0	46.0	CMU-03	slovene–old-saxon	10.6	53.2	CMU-03
dutch–west-frisian	38.0	51.0	it-ist-02	sorani–irish	27.6	66.3	CMU-03
dutch–yiddish	78.0	64.0	it-ist-01	spanish–friulian	53.0	81.0	CMU-03
english–murrinhpatha	22.0	42.0	it-ist-02	spanish–occitan	57.0	78.0	CMU-03
english–north-frisian	31.0	42.0	CMU-03	swahili–quechua	13.9	92.1	CMU-03

SIGMORPHON 2019 Shared Task

[Anastasopoulos and Neubig, 2019]

Morphological Analysis in Context and Cross-Lingual Transfer for Inflection (100 Language Pairs)

Task1: Cross-lingual Transfer for Morphological Inflection (10k HR +100 LR)
Task2: Inflection in Context

Model	Accuracy	Median
Wu and Cotterell (2019)	48.5	45.5
this work	48.8	67.0
$+\mathcal{H}$	60.1	66.6
$+\mathcal{H} + \mathcal{L}_l$	60.8	66.0
+multi-language transfer	**63.8**	64.0
oracle	68.2	74.0

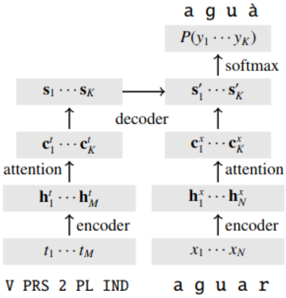
Table 1: Macro-averaged accuracy over 100 language test pairs. Our best model outperforms the baseline by 15 percentage points.

SIGMORPHON 2019 Shared Task

[Anastasopoulos and Neubig, 2019]

Morphological Analysis in Context and Cross-Lingual Transfer for Inflection (100 Language Pairs)

Task1: Cross-lingual Transfer for Morphological Inflection (10k HR +100 LR)
Task2: Inflection in Context



[Anastasopoulos and Neubig, 2019]

Morphological Analysis in Context and Cross-Lingual Transfer for Inflection (100 Language Pairs)

Task1: Cross-lingual Transfer for Morphological Inflection (10k HR +100 LR)
Task2: Inflection in Context

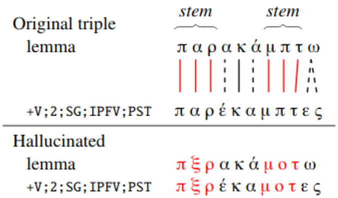


Figure 2: Example of our hallucination process (Greek). The lemma and inflected forms are aligned at the character level. The inside of *stem*-considered parts (highlighted) are substituted with random characters, creating hallucinated triples (bottom).

SIGMORPHON 2020 Shared Task

SIGMORPHON Shared Task 2016–2019

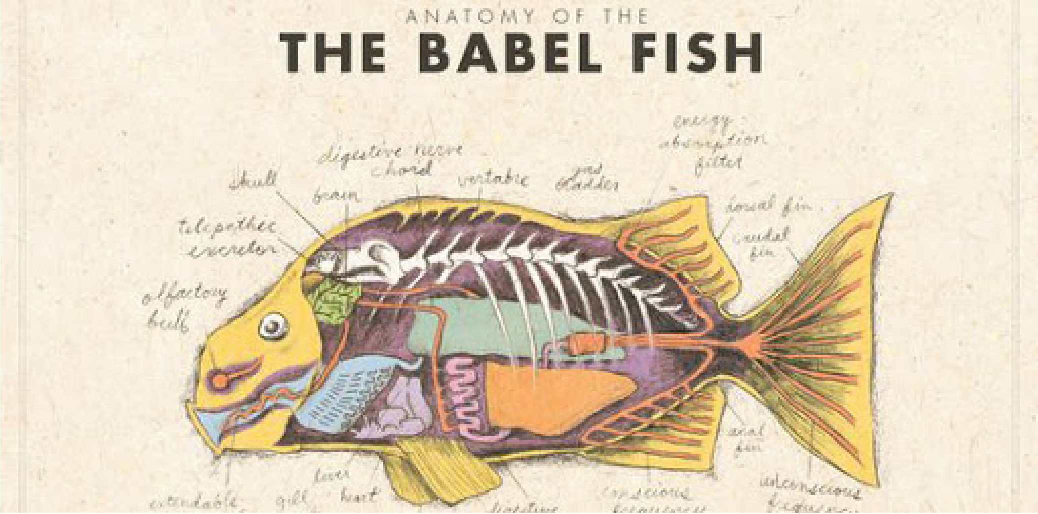
PLAY + PRESENT PARTICIPLE → *playing*
played + PRESENT PARTICIPLE → *playing*

Lemma	Tag	Form
RUN	PAST	*ran*
RUN	PRES;1SG	*run*
RUN	PRES;2SG	*run*
RUN	PRES;3SG	*runs*
RUN	PRES;PL	*run*
RUN	PART	*running*

But much less well
in low-resource setting

2018 :∼ 96% accuracy on avg.
in high-resource setting

Increasing Multilinguality



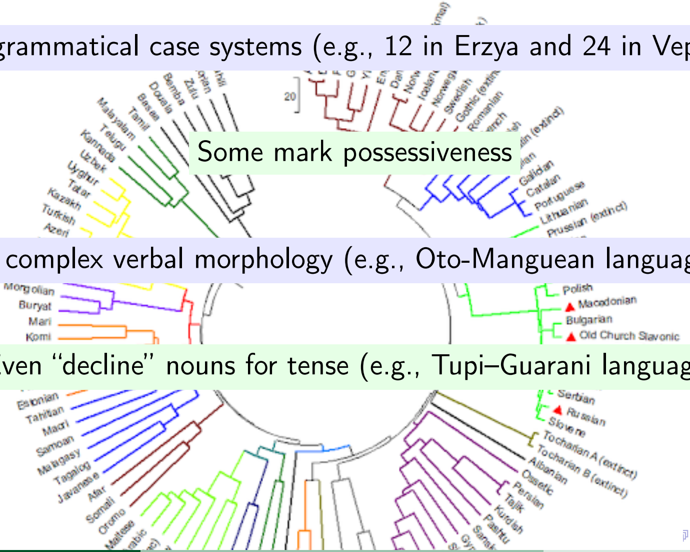
Languages differ in many ways!

Some exhibit rich grammatical case systems (e.g., 12 in Erzya and 24 in Veps)

Some mark possessiveness

Others might have complex verbal morphology (e.g., Oto-Manguean languages)

Even "decline" nouns for tense (e.g., Tupi–Guarani languages)



Languages differ in many ways!

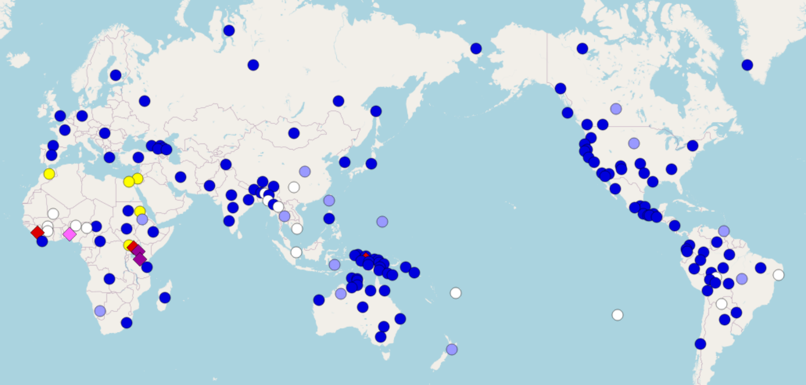
Let's Discuss The Following Dimensions:

Fusion

Inflectional Synthesis

Position of Case Affixes

Fusion (WALS 20A)



●	Exclusively concatenative	125
○	Exclusively isolating	16
◆	Exclusively tonal	3
◆	Tonal/isolating	1
◆	Tonal/concatenative	2
●	Ablaut/concatenative	5
●	Isolating/concatenative	13

Fusion (WALS 20A)

From isolating to concatenative

Concatenative morphology is the most common system

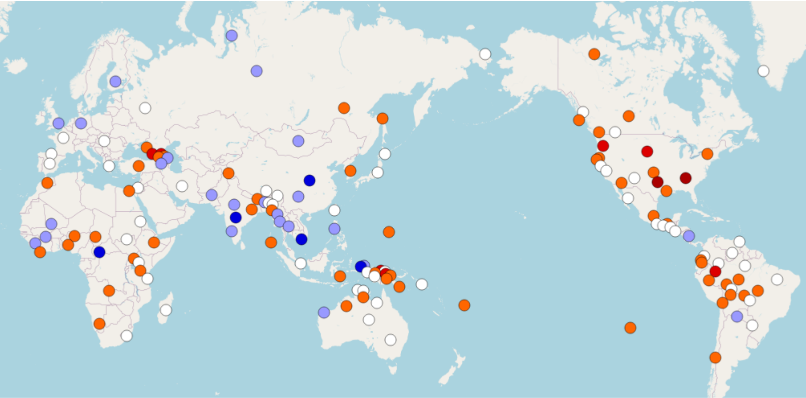
Non-linearities such as ablaut or tonal morphology can also be present

Isolating languages: the Sahel Belt in West Africa, Southeast Asia and the Pacific

Tonal–concatenative morphology can be found in Mesoamerican languages

Exclusively concatenative	125
Exclusively isolating	16
Exclusively tonal	3
Tonal/isolating	1
Tonal/concatenative	2
Ablaut/concatenative	5
Isolating/concatenative	13

Inflectional Synthesis of the Verb (WALS 22A)

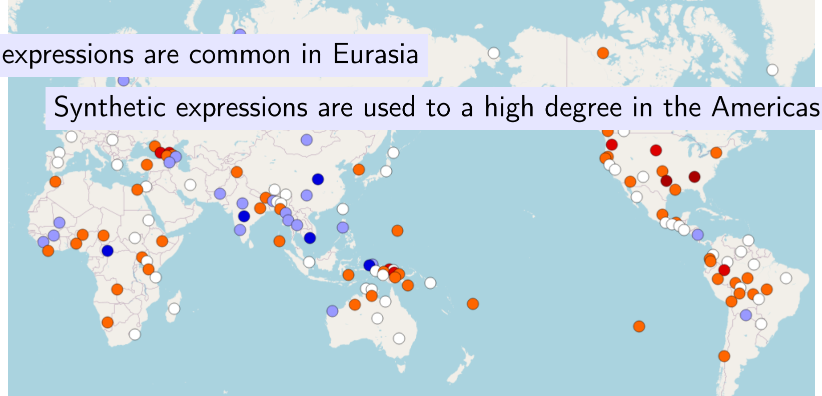


●	0-1 category per word	5
●	2-3 categories per word	24
○	4-5 categories per word	52
●	6-7 categories per word	31
●	8-9 categories per word	24
●	10-11 categories per word	7
●	12-13 categories per word	2

Inflectional Synthesis of the Verb (WALS 22A)

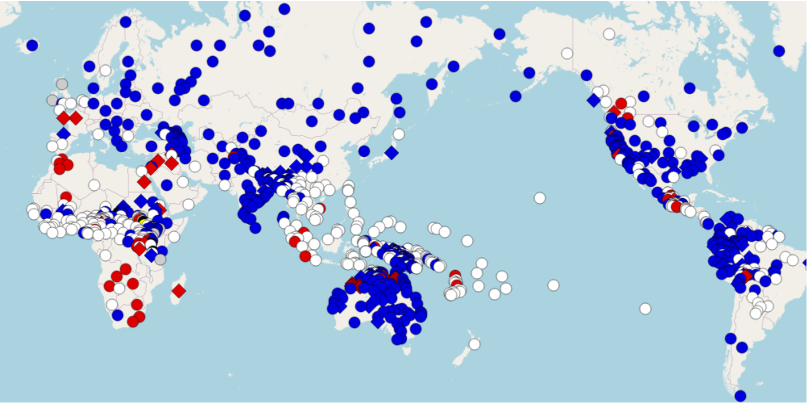
Analytic expressions are common in Eurasia

Synthetic expressions are used to a high degree in the Americas



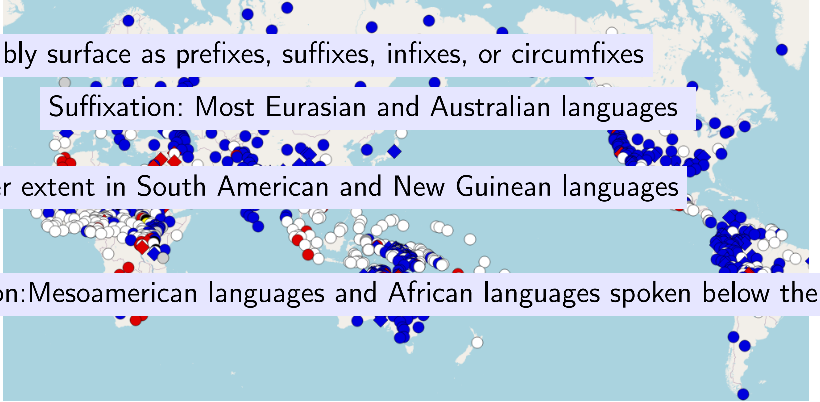
●	0-1 category per word	5
●	2-3 categories per word	24
○	4-5 categories per word	52
●	6-7 categories per word	31
●	8-9 categories per word	24
●	10-11 categories per word	7
●	12-13 categories per word	2

Position of Case Affixes (WALS 51A)



●	Case suffixes	452
●	Case prefixes	38
●	Case tone	5
●	Case stem change	1
●	Mixed morphological case	9
◆	Postpositional clitics	123
◆	Prepositional clitics	17
◆	Inpositional clitics	7

Position of Case Affixes (WALS 51A)



Can variably surface as prefixes, suffixes, infixes, or circumfixes

Suffixation: Most Eurasian and Australian languages

to a lesser extent in South American and New Guinean languages

Prefixation:Mesoamerican languages and African languages spoken below the Sahara

●	Case suffixes	452
●	Case prefixes	38
●	Case tone	5
●	Case stem change	1
●	Mixed morphological case	9
◆	Postpositional clitics	123
◆	Prepositional clitics	17
◆	Inpositional clitics	7

Language-Specific Biases

As Bender(2009, 2016) notes architectures and training and tuning algorithms still present language-specific biases

SIGMORPHON 2020 SHARED TASK 0

As Bender(2009, 2016) notes architectures and training and tuning algorithms still present language-specific biases

Let's focus on typological diversity and aim to investigate systems' ability to generalize across typologically distinct languages!

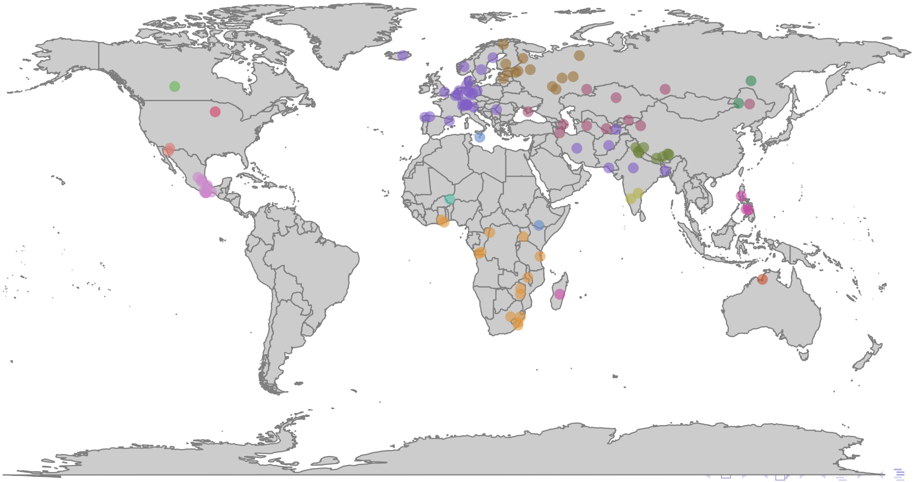
SIGMORPHON 2020 SHARED TASK 0

As Bender(2009, 2016) notes architectures and training and tuning algorithms still present language-specific biases

Let's focus on typological diversity and aim to investigate systems' ability to generalize across typologically distinct languages!

If a model works well for a sample of IE languages, should the same model also work well for Tupi–Guarani languages?

90 Languages from 13 languages families



Three Phases

Development

2 months; train & dev: 45 languages from 5 families (Austronesian, Niger-Congo, Oto-Manguean, Uralic, IE)

Generalization

1 week; train & dev: 45 languages from 10 families (Afro-Asiatic, Algic, Dravidian, Indo-European, Niger-Congo, Sino-Tibetan, Siouan, Songhay, Southern Daly, Tungusic, Turkic, Uralic, and Uto-Aztecan)

Evaluation

1 week; test: all 90 languages

Data

Preparation

✓Manually converted their features (tags) into the UniMorph format
✓Canonicalized (`https://github.com/unimorph/um-canonicalize`) the converted language data

Splitting

✓Used only noun, verb, and adjective forms to construct training, development, and evaluation sets.
✓Randomly sampled 70%, 10%, and 20% for train, development, and test, respectively.
✓Zarma, Tajik, Lingala, Ludian, Māori, Sotho, Võro, Anglo-Norman, and Zulu contain less than 400 training samples

Systems: Baselines

Non-neural

Simple alignment-based as in previous years (Cotterell et al., 2017;2018)

Systems: Baselines

Neural

✓Neural transducer (Wu et al, 2019), which is essentially a hard monotonic attention model (`mono-*`)

✓Transformer adopted for character-level tasks Wu et al, (2020; `trm-*`), SoTA on ST 2017

+ data augmentation technique used by Anastasopoulos et al. (2019;`-aug-`)

+ family-wise shared parameters (`*-shared`)

Team	Description	System	Model Features			
			Neural	Ensemble	Multilingual	Hallucination
Baseline	wu2019exact	mono-single	✓			
		mono-aug-single	✓			✓
		mono-shared	✓		✓	
		mono-aug-shared	✓		✓	✓
	wu2020applying	trm-single	✓			
		trm-aug-single	✓			✓
		trm-shared	✓		✓	
		trm-aug-shared	✓		✓	✓

Systems: Teams

10 teams submitted 22 systems in total, out of which 19 were neural

Team	Description	System	Model Features			
			Neural	Ensemble	Multilingual	Hallucination
CMU Tartan	Jayarao et al.(2020)	cmu_tartan_00-0	✓			✓
		cmu_tartan_00-1	✓		✓	✓
		cmu_tartan_01-0	✓			✓
		cmu_tartan_01-1	✓		✓	✓
		cmu_tartan_02-1	✓		✓	✓
CU7565	Beemer et al. (2020)	CU7565-01-0				
		CU7565-02-0				
CULing	Liu et al. (2020)	CULing-01-0	✓	✓		
DeepSpin	Peterset al. (2020)	deepspin-01-1	✓		✓	
		deepspin-02-1	✓		✓	
ETH Zurich	Forster et al. (2020)	ETHZ00-1	✓		✓	
		ETHZ02-1	✓		✓	
Flexica	Scherbakov (2020)	flexica-01-0				
		flexica-02-1	✓		✓	
		flexica-03-1	✓		✓	✓
IMS	Yu et al. (2020)	IMS-00-0	✓	✓		✓
LTI	Murikinati et al. (2020)	LTI-00-1	✓		✓	✓
NYU-CUBoulder	Singer et al. (2020)	NYU-CUBoulder-01-0	✓	✓		✓
		NYU-CUBoulder-02-0	✓			✓
		NYU-CUBoulder-03-0	✓	✓		✓
		NYU-CUBoulder-04-0	✓			✓
UIUC	Canby et al. (2020)	uiuc-01-0	✓			

Systems: Description (* – winning system)

Improving neural baselines

✓***UIUC**: transformers with synchronous bidirectional decoding technique (Zhou et al.,2019) and family-wise fine-tuning
✓**ETH Zurich**: exact decoding strategy that uses Dijkstra's search algorithm

Improving previous years' models: Hard Monotonic Attention

✓**IMS**: L2R+R2L models with a genetic algorithm for ensemble search and data hallucination
✓**Flexica**:multilingual (family-wise) model with improved alignment strategy
+ new data hallucination technique based on phonotactic modelling

Systems: Description (* – winning system)

Improving their 2019 models

✓**LTI**: multi-source encoder–decoder with two-step attention architecture + cross-lingual transfer+ data hallucination + romanization of scripts
✓***DeepSpin**: massively multilingual (all languages) gated sparse two-headed attention model with sparsemax
+ 1.5-entmax

Transformer vs. LSTMs

✓**CMU Tartan**: compared trasformer- and LSTM-based encoder–decoders trained mono- and multilingually with data hallucination

Systems: Description (* – winning system)

Ensembles of Transformers

✓**NYU-CUBoulder**: compared vanilla and pointer-generator (monolingual) transformers + ensembles of three and five pointer-generator transformers + data hallucination (less than 1,000 samples)

✓***CULing**: ensemble of three (monolingual) transformers + augmented the initial input (that only used the lemma as a source form) with entries corresponding to other (non-lemma) slots (reinflection) to improve learning of principal parts of paradigm

Systems: Description (* – winning system)

Non-neural systems

✓**CU7565**: manually developed finite-state grammars for 25 languages
+ hierarchical paradigm clustering (based on similarity of string transformation rules)
✓**Flexica**: a method similar to Hulden (2014) but with transformation rules treated independently and assigned a score based on their frequency, specificity and diversity of surrounding characters

Evaluation

✓Per-language accuracy
✓Per-language Levenstein distance
✓Takes into account the statistical significance of differences between systems

Ranking

Any system which is the same (as assessed via statistical significance) as the best performing one is also ranked 1st for that language.
For genus/family:
We aggregate the systems' ranks and re-rank them based on the amount of times they ranked 1st, 2nd, etc.

Results: 4 winning systems (outperform baselines)

uiuc-01-0	**2.4**	**90.5**
deepspin-02-1	2.9	**90.9**
BASE: trm-single	2.8	90.1
CULing-01-0	3.2	**91.2**
deepspin-01-1	3.8	**90.5**
BASE: trm-aug-single	3.7	90.3
NYU-CUBoulder-04-0	7.1	88.8
NYU-CUBoulder-03-0	8.9	88.8
NYU-CUBoulder-02-0	8.9	88.7
IMS-00-0	10.6	89.2
NYU-CUBoulder-01-0	9.6	88.6
BASE: trm-shared	10.3	85.9
BASE: mono-aug-single	7.5	88.8
cmu_tartan_00-0	8.7	87.1
BASE: mono-single	7.9	85.8
cmu_tartan_01-1	9.0	87.1
BASE: trm-aug-shared	12.5	86.5
BASE: mono-shared	10.8	86.0
cmu_tartan_00-1	9.4	86.5
LTI-00-1	12.0	86.6
BASE: mono-aug-shared	12.8	86.8
cmu_tartan_02-1	10.6	86.1
cmu_tartan_01-0	10.9	86.6
flexica-03-1	16.7	79.6
ETHZ-00-1	20.1	75.6
**CU7565-01-0*	*24.1*	*90.7*
flexica-02-1	17.1	78.5
**CU7565-02-0*	*19.2*	*83.6*
ETHZ-02-1	17.0	80.9
flexica-01-0	24.4	70.8
Oracle (Baselines)		96.1
Oracle (Submissions)		97.7
Oracle (All)		97.9

Results: 4 winning systems (outperform baselines)

uiuc-01-0	**2.4**	**90.5**

The baselines and the submissions are complementary

deepspin-01-1	3.8	**90.5**

adding them together increases the oracle scored

NYU-CUBoulder-03-0	8.9	88.8
NYU-CUBoulder-02-0	8.9	88.7
IMS-00-0	10.6	89.2
NYU-CUBoulder-01-0	9.6	88.6
BASE: trm-shared	10.3	85.9
BASE: mono-aug-single	7.5	88.8
cmu_tartan_00-0	8.7	87.1

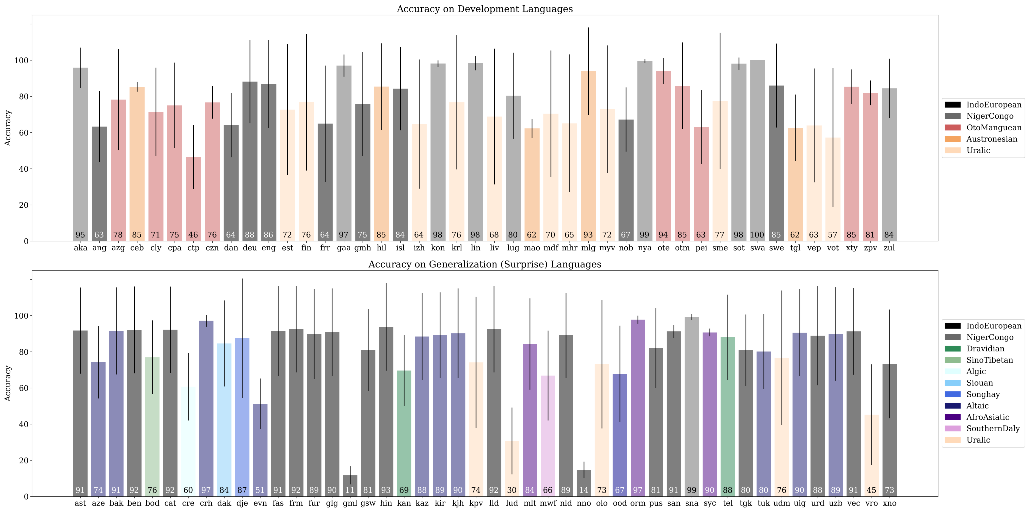
The largest gaps in oracle systems are observed in Algic, Oto-Manguean

BASE: mono-shared	10.8	86.0

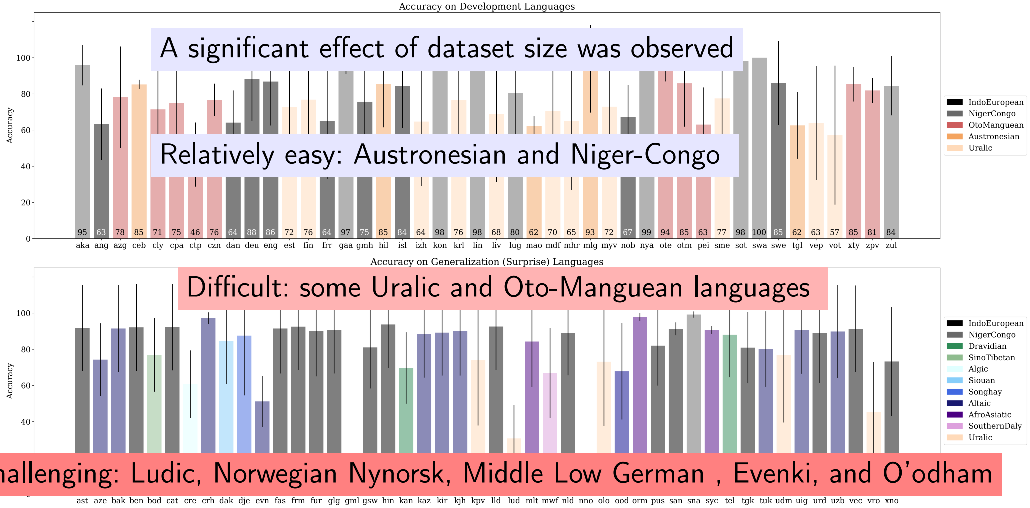
Sino-Tibetan, Southern Daly, Tungusic, and Uto-Aztecan families

BASE: mono-aug-shared	12.8	86.8
cmu_tartan_02-1	10.6	86.1
cmu_tartan_01-0	10.9	86.6
flexica-03-1	16.7	79.6
ETHZ-00-1	20.1	75.6
**CU7565-01-0*	*24.1*	*90.7*
flexica-02-1	17.1	78.5
**CU7565-02-0*	*19.2*	*83.6*
ETHZ-02-1	17.0	80.9
flexica-01-0	24.4	70.8
Oracle (Baselines)		96.1
Oracle (Submissions)		97.7
Oracle (All)		97.9

Accuracy by language averaged across all submissions



Accuracy by language averaged across all submissions



A significant effect of dataset size was observed

Relatively easy: Austronesian and Niger-Congo

Difficult: some Uralic and Oto-Manguean languages

Challenging: Ludic, Norwegian Nynorsk, Middle Low German , Evenki, and O’odham

Questions Addressed in Papers

Is developing morphological grammars manually worthwhile?

CU7565 manually designed finite-state grammars for 25 languages

✓Paradigms of some languages were relatively easy to describe but neural networks also performed quite well

✓For Ingrian and Tagalog (LRL) grammars demonstrate superior performance but this comes at the expense of a significant amount of person-hours

Questions Addressed in Papers

What is the best training strategy for low-resource languages?

✓Hallucinated data highlighted its utility for LRLs.

✓Augmenting the data with tuples where lemmas are replaced with non-lemma forms and their tags

✓Multilingual training

✓Ensembles

Error Analysis

Systematic Errors:

Data Inconsistency

The train, development and test sets contain 2%, 0.3%, and 0.6% inconsistent entries

Highest rates: Azerbaijani, Old English, Cree, Danish, Middle Low German , Kannada, Norwegian Bokmål, Chichimec, and Veps

Dialectal variations in Finno-Ugric and Tungusic

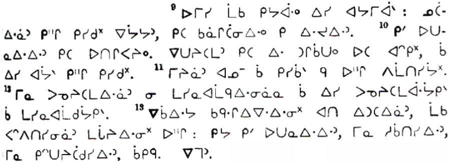
Language-Specific Errors

Algic (Cree)

Mean accuracy across systems was 65.1% (41.5% to 73%)

✓Struggled with the choice of preverbal auxiliary (‘kitta’ could refer to future, imperfective, or imperative)

✓The paradigms were very large, there were very few lemmas (28 impersonal verbs and 14 transitive verbs



Niger-Congo

Mean accuracy across systems was very good at 96.4 (62.8% to 100%)

✓Most languages in this family are considered low resource, and the resources used for data gathering may have been biased towards the languages' regular forms

Language-Specific Errors

Sino–Tibetan (Tibetan)

Mean accuracy across systems was average at 82.1%(67.9% to 85.1%)

✓Majority errors are related to allomoprhy

✓Nonce words and impossible combinations of component units (Di et al., 2019)

Language-Specific Errors

Tungusic (Evenki)

Mean accuracy across systems was average at 53.8% (43.5% to 59.0%)

✓The dataset was created from oral speech samples in various dialects of the language; there was little attempt at any standardization in the oral speech transcription

✓Annotation: various past tense forms are all annotated as PST, or there are several comitative suffixes all annotated as COM

✓Annotation: some features are present in the word form but they receive no annotation at all

Conclusion

AND.....TO CONCLUDE:

✓Submissions were able to make productive use of multilingual training
✓Data augmentation techniques such as hallucination helped
✓Combined with architecture tweaks like sparsemax, it resulted in excellent overall performance on many languages
✓Some morphology types and language families (Tungusic, Oto-Manguean, Southern Daly) are still challenging
✓In some languages (Ingrian, Tajik, Tagalog, Zarma, and Lingala) hand-encoding linguist knowledge in finite state grammars resulted in best performance

AN Some Linguistic Analysis:

Modeling morphological learning, typology, and change:

What can the neural sequence-to-sequence framework contribute?

(Elsner, Sims et al, 2019)

✓Submi

✓Data a ch as hallucination helped

✓Combined with architecture tweaks like sparsemax, it resulted in excellent overall performance on many languages

✓Some morphology types and language families (Tungusic, Oto-Manguean, Southern Daly) are still

Some Linguistic Analysis:

Lexical databases for computational analyses:

A linguistic perspective

(Malouf, Ackerman, Semenuks, 2020)

✓In a, and Lingala) hand-encoding linguist knov mmars resulted in best performance

Conclusion

SIGMORPHON–UNIMORPH 2021 Shared Task! Please Join!

✓Part 1: Generalization Across Languages (focusing on under-resourced languages; extracting data fom FSTs (e.g., Apertium) and grammar books)

✓Part 2 (with Ben Ambridge): Human-like Generalization: "wug" tests across languages

Thank you! Questions?