Low-resource NLP: Lessons from Dependency Parsing

Miryam de Lhoneux
@mdlhx

UPPSALA UNIVERSITET

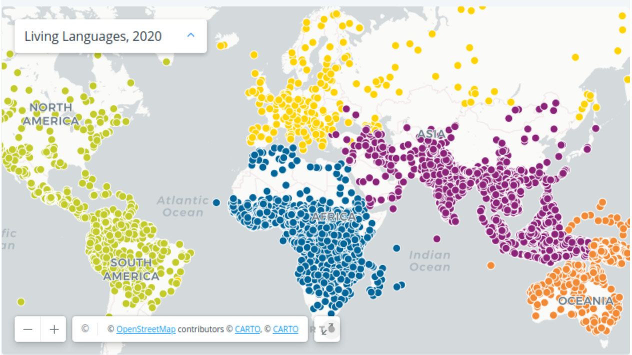
KU LEUVEN

UNIVERSITY OF COPENHAGEN

10 June 2021
SIGTYP 2021 workshop

Outline for section 1

The bad news: world's languages



https://www.ethnologue.com/guides/how-many-languages

The good news

🔥 **Cross-lingual learning is on the rise** 🔥

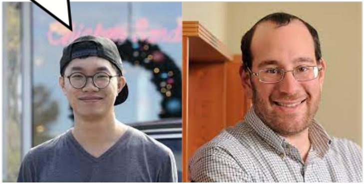
Figure from Plank (2019)

Multilingual NLP

Beto, Bentz, Becas: The Surprising Cross-Lingual Effectiveness of BERT

Shijie Wu and **Mark Dredze**
Department of Computer Science
Johns Hopkins University
shijie.wu@jhu.edu, mdredze@cs.jhu.edu

mBERT performs well for 5 NLP tasks in a zero-shot setting



Multilingual NLP

Cool. So we're done?

Multilingual NLP

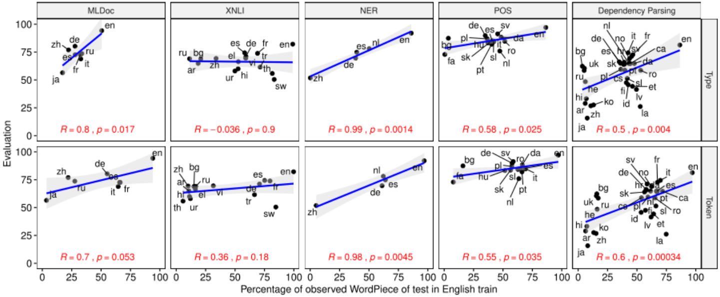
Cool. So we're done?
Not so fast. . .

Multilingual NLP

Cool. So we're done?

Not so fast. . .

strong correlation between the percentage of overlapping subwords and transfer performance



Multilingual NLP

Cool. So we're done?

Not so fast...

Transfer works best between typologically similar languages in mBERT (Pires et al., 2019)

Multilingual NLP

Cool. So we're done?

Not so fast...

Ok. Breathe. So, multilingual NLP, where are we at?

Multilingual NLP

Cool. So we're done?

Not so fast. . .

Ok. Breathe. So, multilingual NLP, where are we at?

Let's look at the data

Multilingual datasets

XNLI

Multilingual datasets

XNLI

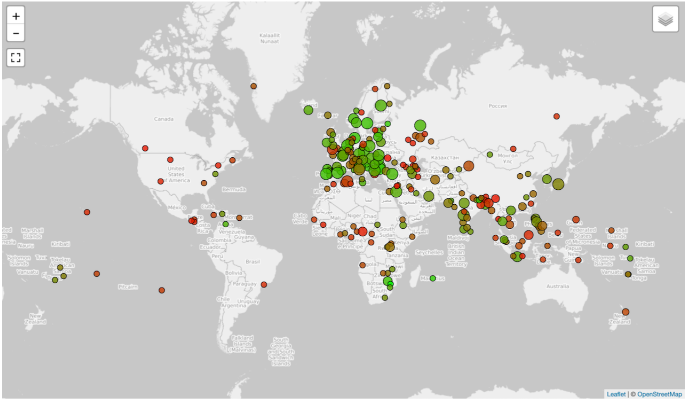
Data translated from English

Multilingual datasets

XQUAD

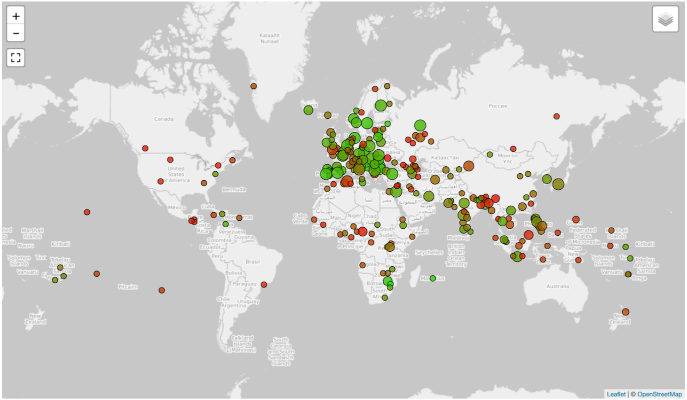
Data translated from English

Multilingual datasets



Tatoeba

Multilingual datasets



Tatoeba

Caswell et al. (2021): serious issues in web-crawled data

Universal Dependencies

UD v1.0

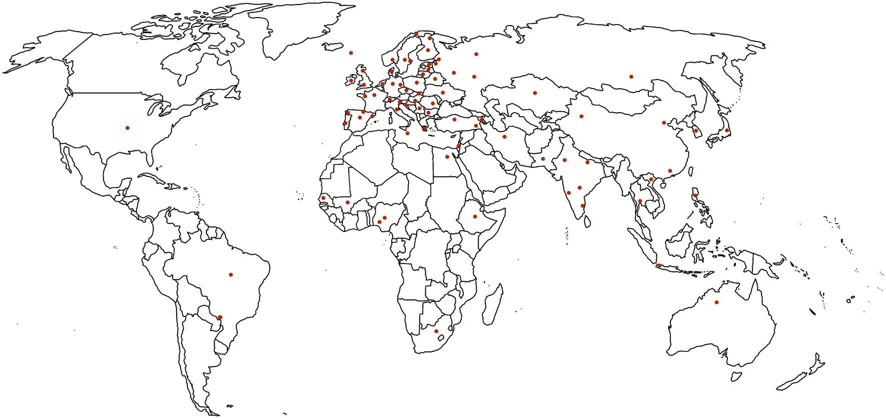
Figure adapted from Nivre et al. (2020)

Universal Dependencies

UD v2.0

Figure adapted from Nivre et al. (2020)

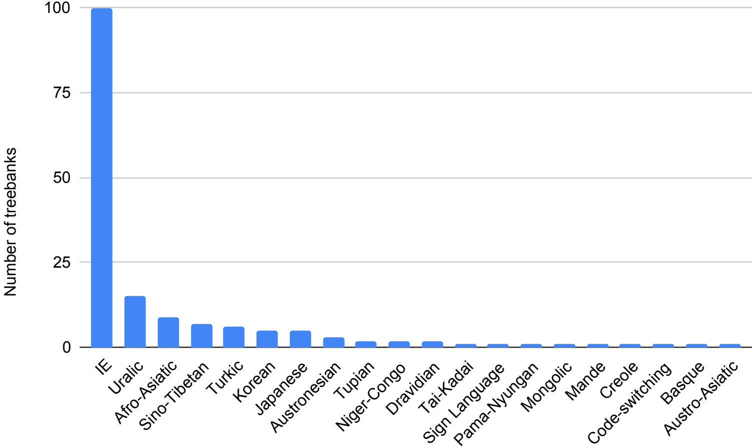
Universal Dependencies



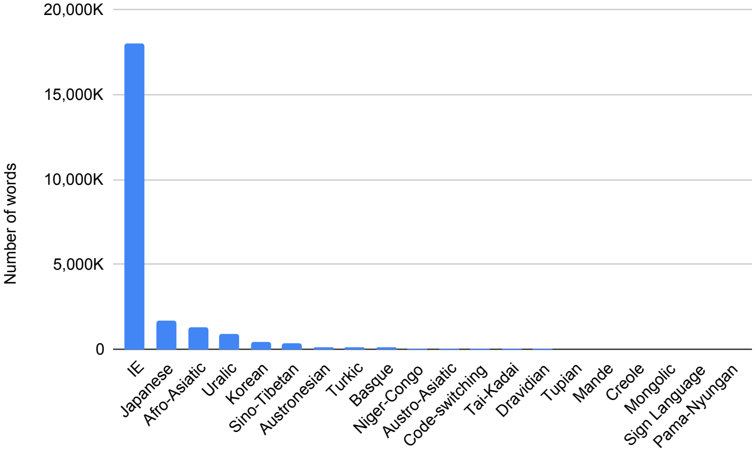
UD v2.5

Figure adapted from Nivre et al. (2020)

Universal Dependencies



Universal Dependencies



UD: opportunities

UD: opportunities

- Can transfer learning mitigate the language technology gap between high and low-resource languages?

UD: opportunities

- Can transfer learning mitigate the language technology gap between high and low-resource languages?
- Can typological features mitigate this gap?

UD: opportunities

- Can transfer learning mitigate the language technology gap between high and low-resource languages?
- Can typological features mitigate this gap?

What have we learned so far?

Other data initiatives

- Nekoto et al. (2020): MT for over 30 African languages
- Adelani et al. (2021): NER for 10 African languages
- Mager et al. (2021): MT for 10 indigenous languages from the Americas (and Spanish)
- Ramesh et al. (2021): MT for 11 indic languages (and English)

Other data initiatives

- Nekoto et al. (2020): MT for over 30 African languages
- Adelani et al. (2021): NER for 10 African languages
- Mager et al. (2021): MT for 10 indigenous languages from the Americas (and Spanish)
- Ramesh et al. (2021): MT for 11 indic languages (and English)

Huge community efforts that are starting to fill dataset gaps

Outline for section 2

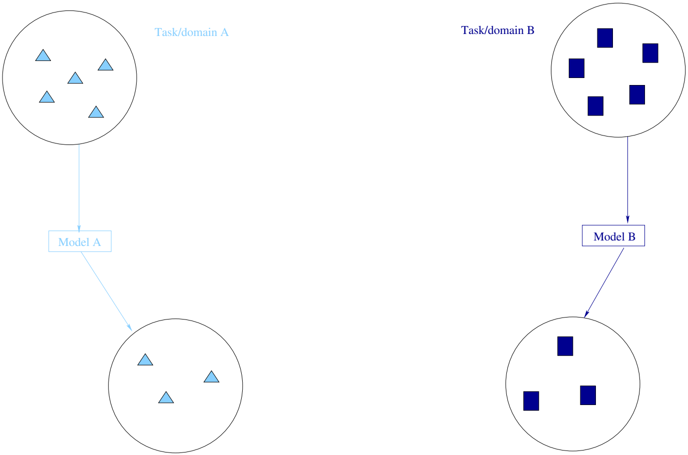
Parsing low-resource languages

Can transfer learning mitigate the language technology gap between high and low-resource languages?

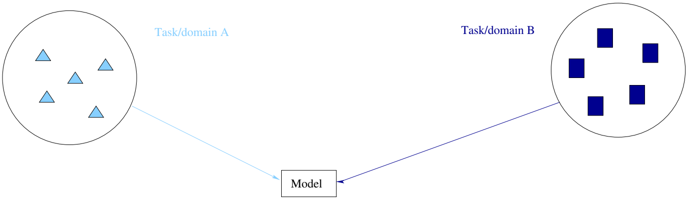
Transfer learning

Figure inspired by `https://ruder.io/transfer-learning/`

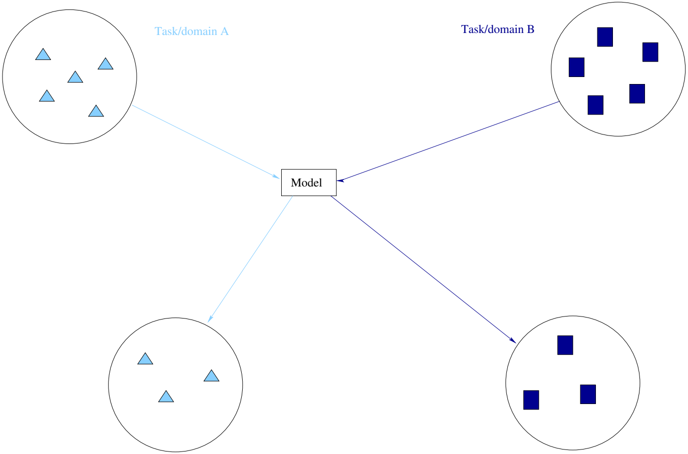
Transfer learning



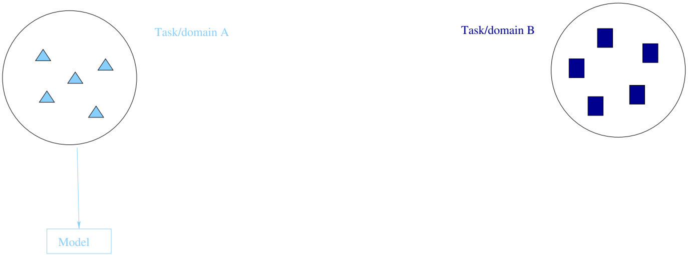
Transfer learning



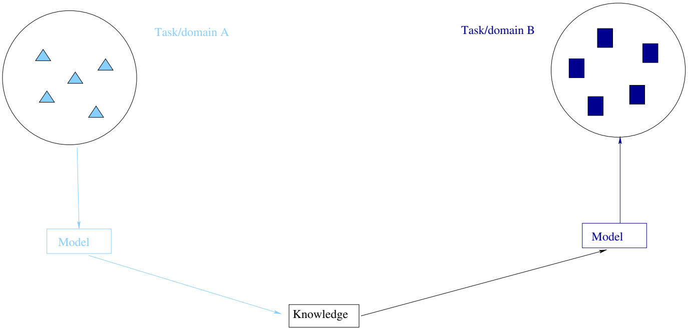
Transfer learning



Transfer learning



Transfer learning



Transfer learning for parsing

One model, two languages: training bilingual parsers with harmonized treebanks

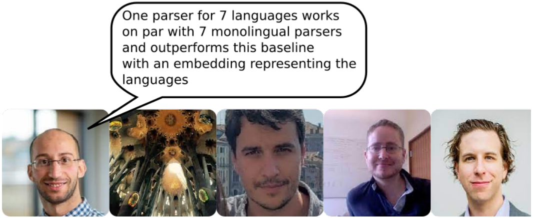
David Vilares, Carlos Gómez-Rodríguez and Miguel A. Alonso
Grupo LyS, Departamento de Computación, Universidade da Coruña
Campus de A Coruña s/n, 15071, A Coruña, Spain

Concatenate pairs of UD treebanks. No drop in LAS, increase in LAS for some pairs.

Transfer learning for parsing

Many Languages, One Parser

Waleed Ammar[◊] **George Mulcaire**[♡] **Miguel Ballesteros**[♠◊] **Chris Dyer**[◊] **Noah A. Smith**[♡]
[◊]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
[♡]Computer Science & Engineering, University of Washington, Seattle, WA, USA
[♠]NLP Group, Pompeu Fabra University, Barcelona, Spain



Transfer learning for parsing

Parameter sharing between dependency parsers for related languages

Miryam de Lhoneux[1*] **Johannes Bjerva**[2] **Isabelle Augenstein**[2] **Anders Søgaard**[2]
[1]Department of Linguistics and Philology
Uppsala University
Uppsala, Sweden
[2] Department of Computer Science
University of Copenhagen
Copenhagen, Denmark

Sharing parameters is beneficial for related languages. Sharing too much can hurt unrelated languages

Transfer learning for parsing

82 Treebanks, 34 Models: Universal Dependency Parsing with Multi-Treebank Models

Aaron Smith* **Bernd Bohnet†** **Miryam de Lhoneux***
Joakim Nivre* **Yan Shao*** **Sara Stymne***

*Department of Linguistics and Philology
Uppsala University
Uppsala, Sweden

†Google Research
London, UK

It is beneficial to train parsers of clusters of related languages

Transfer learning for parsing

Cross-lingual Parsing with Polyglot Training and Multi-treebank Learning: A Faroese Case Study

James Barry and **Joachim Wagner** and **Jennifer Foster**
ADAPT Centre
School of Computing, Dublin City University, Ireland

How to Parse Low-Resource Languages: Cross-Lingual Parsing, Target Language Annotation, or Both?

Ailsa Meechan-Maddon
Uppsala University
Department of Linguistics and Philology

Joakim Nivre
Uppsala University
Department of Linguistics and Philology

With clever techniques, we can improve low-resource dependency parsing using related high-resource languages

A systematic comparison of methods for low-resource dependency parsing on genuinely low-resource languages

Clara Vania[1] **Yova Kementchedjhieva**[2] **Anders Søgaard**[2] **Adam Lopez**[1]
[1]School of Informatics, University of Edinburgh, UK
[2]University of Copenhagen, Copenhagen, Denmark

Transfer learning for parsing

Polyglot Parsing for One Thousand and One Languages (And Then Some)

Ali Basirat* **Miryam de Lhoneux*** **Artur Kulmizev***
Murathan Kurfalı† **Joakim Nivre*** **Robert Östling†**

*Department of Linguistics and Philology
Uppsala University

†Department of Linguistics
Stockholm University

We train and test a parser on disjoint sets of languages using pretrained language embeddings and cross-lingual word embeddings. Very poor parser performance.

Transfer learning for parsing

Low-Resource Parsing with Crosslingual Contextualized Representations

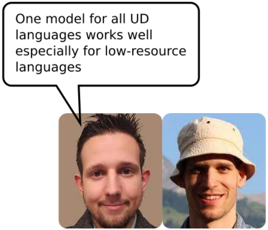
Phoebe Mulcaire[♡*] **Jungo Kasai**[♡*] **Noah A. Smith**[♡◇]
[♡]Paul G. Allen School of Computer Science & Engineering,
University of Washington, Seattle, WA, USA
[◇]Allen Institute for Artificial Intelligence, Seattle, WA, USA
{pmulc,jkasai,nasmith}@cs.washington.edu

Polyglot language models are beneficial for low-resource dependency parsing

Transfer learning for parsing

75 Languages, 1 Model: Parsing Universal Dependencies Universally

Dan Kondratyuk[1,2] and **Milan Straka**[1]
[1]Charles University, Institute of Formal and Applied Linguistics
[2]Saarland University, Department of Computational Linguistics



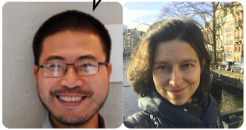
Transfer learning for parsing

Zero-shot Dependency Parsing with Pre-trained Multilingual Sentence Representations

Ke Tran[*]
Amazon Alexa AI
trnke@amazon.com

Arianna Bisazza[†]
University of Groningen
a.bisazza@rug.nl

Using mBERT for zero-shot parsing results in SOTA results for some languages but remains very poor for others



What does zero-shot mean?

Zero-shot settings

What does zero-shot mean?

Zero-shot settings

- Unseen language task [required]

What does zero-shot mean?

Zero-shot settings

- Unseen language task [required]
- Unseen language pretrained

What does zero-shot mean?

Zero-shot settings

- Unseen language task [required]
    - Unseen genus task
- Unseen language pretrained

What does zero-shot mean?

Zero-shot settings

- Unseen language task [required]
  - Unseen genus task
- Unseen language pretrained
  - Unseen genus pretrained

What does zero-shot mean?

Zero-shot settings

- Unseen language task [required]
  - Unseen genus task
    - Unseen language family task
- Unseen language pretrained
  - Unseen genus pretrained

What does zero-shot mean?

Zero-shot settings

- Unseen language task [required]
    - Unseen genus task
        - Unseen language family task
- Unseen language pretrained
    - Unseen genus pretrained
        - Unseen language family pretrained

What does zero-shot mean?

Zero-shot settings

- Unseen language task [required]
  - Unseen genus task
    - Unseen language family task
- Unseen language pretrained
  - Unseen genus pretrained
    - Unseen language family pretrained
- Unseen script task

What does zero-shot mean?

Zero-shot settings

- Unseen language task [required]
  - Unseen genus task
    - Unseen language family task
- Unseen language pretrained
  - Unseen genus pretrained
    - Unseen language family pretrained
- Unseen script task
  - Unseen script pretrained

What does zero-shot mean?

Zero-shot settings

- Unseen language task [required]
  - Unseen genus task
    - Unseen language family task
- Unseen language pretrained
  - Unseen genus pretrained
    - Unseen language family pretrained
- Unseen script task
  - Unseen script pretrained

Not all zero-shot settings are created equal!

Zero-shot settings

Zero-shot setting in Üstün et al. (2020):
mBERT + 13 treebanks

Zero-shot settings

Zero-shot setting in Üstün et al. (2020):
mBERT + 13 treebanks

language	genus	language family
Arabic	Semitic	Afro-Asiatic
Basque	Basque	Basque
Chinese	Sino-Tibetan	Sino-Tibetan
English	Germanic	IE
Finnish	Finnic	Uralic
Hebrew	Afro-Asiatic	Semitic
Hindi	Indic	IE
Italian	Romance	IE
Japanese	Japanese	Japanese
Korean	Korean	Korean
Russian	Slavic	IE
Swedish	Germanic	IE
Turkish	Southwestern Turkic	Turkic

Zero-shot settings

	language	LAS
unseen language task		
unseen genus task		
unseen language family task		
unseen language pretrained		
unseen genus pretrained		
unseen language family pretrained		
unseen script task & pretrained		

Zero-shot results of baselines in Üstün et al. (2020)

Zero-shot settings

	language	LAS
unseen language task	Belarusian	
unseen genus task	Kazakh	
unseen language family task	Yoruba	
unseen language pretrained	Faroese	
unseen genus pretrained	Komi Permyak	
unseen language family pretrained	Buryat	
unseen script task & pretrained	Amharic	

Zero-shot results of baselines in Üstün et al. (2020)

Zero-shot settings

	language	LAS
unseen language task	Belarusian	80.1
unseen genus task	Kazakh	61.9
unseen language family task	Yoruba	42.7
unseen language pretrained	Faroese	68.6
unseen genus pretrained	Komi Permyak	23.1
unseen language family pretrained	Buryat	18.9
unseen script task & pretrained	Amharic	5.9

Zero-shot results of baselines in Üstün et al. (2020)

Zero-shot settings

	language	LAS
unseen language task	Belarusian	80.1
unseen genus task	Kazakh	61.9
unseen language family task	Yoruba	42.7
unseen language pretrained	Faroese	68.6
unseen genus pretrained	Komi Permyak	23.1
unseen language family pretrained	Buryat	18.9
unseen script task & pretrained	Amharic	5.9

Zero-shot results of baselines in Üstün et al. (2020)

Muller et al. (2021) Transfer learning failures largely related to script.

Transfer learning for dependency parsing

Transfer learning for dependency parsing

So what have we learned?

Transfer learning for dependency parsing

So what have we learned?

- Surprisingly easy to transfer knowledge to related languages

Transfer learning for dependency parsing

So what have we learned?

- Surprisingly easy to transfer knowledge to related languages
- Pretrained language data helps a lot

Transfer learning for dependency parsing

So what have we learned?

- Surprisingly easy to transfer knowledge to related languages
- Pretrained language data helps a lot
- No language family data pretrained or task: very poor

Transfer learning for dependency parsing

So what have we learned?

- Surprisingly easy to transfer knowledge to related languages
- Pretrained language data helps a lot
- No language family data pretrained or task: very poor
- Unseen script: very very poor !!!

Transfer learning for dependency parsing

So what have we learned?

- Surprisingly easy to transfer knowledge to related languages
- Pretrained language data helps a lot
- No language family data pretrained or task: very poor
- Unseen script: very very poor !!!
- Have we been overestimating the benefits of transfer learning?

Parsing low-resource languages

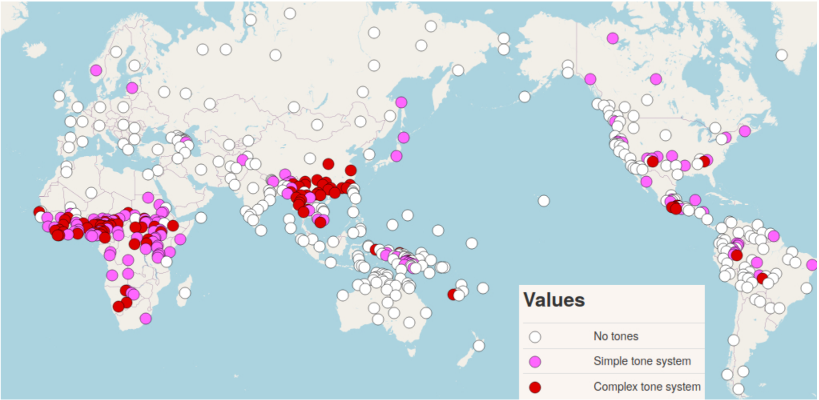
Can typological features mitigate the language technology gap between high and low-resource languages?

Typology

Typological features in WALS: cover many languages

Typology

Typological features in WALS: cover many languages



World Atlas of Language Structures

Typology

Pre-UD: Naseem et al. (2012)

Typology

UD & neural: mixed results

- Ammar et al. (2016)
- Scholivet et al. (2019)
- Fisch et al. (2019)

Typology

More promising

Üstün et al. (2020)

	be	br*	bxr*	cy	fo*	gsw*	hsb*	kk	koi*	krl*	mdf*	mr	olo*	pcm*	sa*	tl	yo*	yue*	AVG
multi-udify	**80.1**	**60.5**	26.1	53.6	68.6	43.6	53.2	**61.9**	20.8	**49.2**	24.8	**46.4**	42.1	36.1	19.4	62.7	41.2	30.5	45.2
udapter-proxy	69.9	-	-	-	64.1	23.7	44.4	45.1	-	45.6	-	29.6	41.1	-	15.1	-	-	24.5	-
udapter	79.3	58.5	**28.9**	**54.4**	**69.2**	**45.5**	**54.2**	60.7	**23.1**	48.4	**26.6**	44.4	**43.3**	**36.7**	**22.2**	**69.5**	**42.7**	**32.8**	**46.2**

Outline for section 3

Findings of the AmericasNLP 2021 Shared Task on Open Machine Translation for Indigenous Languages of the Americas

Manuel Mager[♠*] **Arturo Oncevay**[♡*] **Abteen Ebrahimi**[◇*] **John Ortega**[Ω]
Annette Rios[ψ] **Angela Fan**[∇] **Ximena Gutierrez-Vasques**[ψ] **Luis Chiruzzo**[△]
Gustavo A. Giménez-Lugo[♣] **Ricardo Ramos**[η] **Ivan Vladimir Meza Ruiz**[♯]
Rolando Coto-Solano[℧] **Alexis Palmer**[◇] **Elisabeth Mager**[♯] **Vishrav Chaudhary**[∇]
Graham Neubig[⋈] **Ngoc Thang Vu**[♠] **Katharina Kann**[◇]

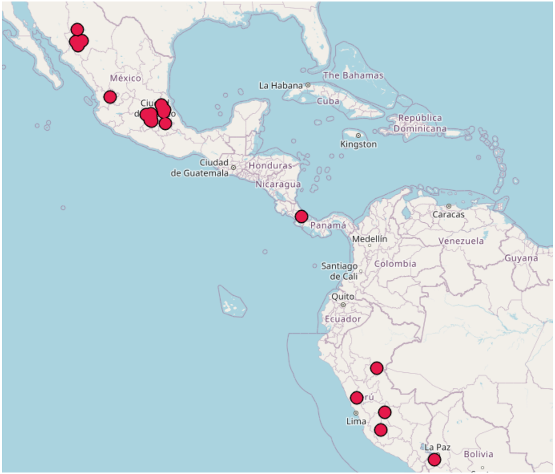
[⋈]Carnegie Mellon University [℧]Dartmouth College [∇]Facebook AI Research
[Ω]New York University [△]Universidad de la República, Uruguay
[η]Universidad Tecnológica de Tlaxcala [♯]Universidad Nacional Autónoma de México
[♣]Universidade Tecnológica Federal do Paraná [◇]University of Colorado Boulder
[♡]University of Edinburgh [♠]University of Stuttgart [ψ]University of Zurich

AmericasNLP 2021 Shared Task

Language	ISO	Family	Train	Dev	Test
Asháninka	cni	Arawak	4K	883	1K
Aymara	aym	Aymaran	7K	996	1K
Bribri	bzd	Chibchan	8K	996	1K
Guarani	gn	Tupi-Guarani	26K	995	1K
Nahuatl	nah	Uto-Aztecan	16K	672	996
Otomí	oto	Oto-Manguean	5K	599	1K
Quechua	quy	Quechuan	125K	996	1K
Rarámuri	tar	Uto-Aztecan	15K	995	1K
Shipibo-Konibo	shp	Panoan	15K	996	1K
Wixarika	hch	Uto-Aztecan	9K	994	1K

Shared task datasets

AmericasNLP 2021 Shared Task



Map representing languages of the shared task

Coastal at AmericasNLP 2021

Moses and the Character-Based Random Babbling Baseline: CoAStaL at AmericasNLP 2021 Shared Task

Marcel Bollmann **Rahul Aralikatte** **Héctor Ricardo Murrieta Bello**
Daniel Hershcovich **Miryam de Lhoneux** **Anders Søgaard**
Department of Computer Science
University of Copenhagen

Coastal at AmericasNLP 2021

What we tried

- Pre-trained transformers
- Back-translation
- Character-level NMT

Coastal at AmericasNLP 2021

What we tried

- Pre-trained transformers
- Back-translation
- Character-level NMT

What we submitted

- Phrase-Based MT (Moses) with *white space tokenization*
- Character-Based Random Babbling

Coastal at AmericasNLP 2021

What we tried

- Pre-trained transformers
- Back-translation
- Character-level NMT

What we submitted

- Phrase-Based MT (Moses) with *white space tokenization*
- Character-Based Random Babbling

Did we do well?

Coastal at AmericasNLP 2021

What we tried

- Pre-trained transformers
- Back-translation
- Character-level NMT

What we submitted

- Phrase-Based MT (Moses) with *white space tokenization*
- Character-Based Random Babbling

Did we do well? Of course not.

Coastal at AmericasNLP 2021

What we tried

- Pre-trained transformers
- Back-translation
- Character-level NMT

What we submitted

- Phrase-Based MT (Moses) with *white space tokenization*
- Character-Based Random Babbling

Did we do well? Of course not.
But not catastrophically *in comparison*

Coastal at AmericasNLP 2021

	AYM	BZD	CNI	GN	HCH	NAH	OTO	QUY	SHP	TAR
Base	0.01	0.01	0.01	0.12	2.20	0.01	0.00	0.05	0.01	0.00

BLEU

Coastal at AmericasNLP 2021

	AYM	BZD	CNI	GN	HCH	NAH	OTO	QUY	SHP	TAR
Rand	0.05	0.06	0.03	0.03	2.07	0.03	0.03	0.02	0.04	0.06
Base	0.01	0.01	0.01	0.12	2.20	0.01	0.00	0.05	0.01	0.00

BLEU

Coastal at AmericasNLP 2021

	AYM	BZD	CNI	GN	HCH	NAH	OTO	QUY	SHP	TAR
SMT	1.11	3.60	3.02	2.20	8.80	2.06	2.72	1.63	3.90	1.05
Rand	0.05	0.06	0.03	0.03	2.07	0.03	0.03	0.02	0.04	0.06
Base	0.01	0.01	0.01	0.12	2.20	0.01	0.00	0.05	0.01	0.00

BLEU

Coastal at AmericasNLP 2021

	AYM	BZD	CNI	GN	HCH	NAH	OTO	QUY	SHP	TAR
Best	2.80	5.18	6.09	8.92	15.67	3.25	5.59	5.38	10.49	3.56
SMT	1.11	3.60	3.02	2.20	8.80	2.06	2.72	1.63	3.90	1.05
Rand	0.05	0.06	0.03	0.03	2.07	0.03	0.03	0.02	0.04	0.06
Base	0.01	0.01	0.01	0.12	2.20	0.01	0.00	0.05	0.01	0.00

BLEU

Coastal at AmericasNLP 2021

Take-away

MT for low-resource polysynthetic languages is hard!

Conclusion

Take-away

Conclusion

Take-away

- We can test hypotheses about multilingual NLP with UD parsing

Conclusion

Take-away

- We can test hypotheses about multilingual NLP with UD parsing
- Transfer learning works surprisingly well between related languages

Conclusion

Take-away

- We can test hypotheses about multilingual NLP with UD parsing
- Transfer learning works surprisingly well between related languages
- For languages that are low-resource and have no related high-resource language, NLP is poor.

Conclusion

Take-away

- We can test hypotheses about multilingual NLP with UD parsing
- Transfer learning works surprisingly well between related languages
- For languages that are low-resource and have no related high-resource language, NLP is poor.
- Maybe we can use typology?

Conclusion

Take-away

- We can test hypotheses about multilingual NLP with UD parsing
- Transfer learning works surprisingly well between related languages
- For languages that are low-resource and have no related high-resource language, NLP is poor.
- Maybe we can use typology?
- Community efforts are making it possible to evaluate truly low-resource NLP

Conclusion

Take-away

- We can test hypotheses about multilingual NLP with UD parsing
- Transfer learning works surprisingly well between related languages
- For languages that are low-resource and have no related high-resource language, NLP is poor.
- Maybe we can use typology?
- Community efforts are making it possible to evaluate truly low-resource NLP
- We can start putting multilinguality at the core of NLP

Thanks

Thanks for your attention!

References I

David Ifeoluwa Adelani, Jade Z. Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen Hassan Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, Jesujoba O. Alabi, Seid Muhie Yimam, Tajuddeen Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Mukiibi, Verrah Otiende, Iroro Orife, Davis David, Samba Ngom, Tosin P. Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwuneke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akinode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane Mboup, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, Degaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima Diop, Abdoulaye Diallo, Adewale Akinfaderin, Tendai Marengereke, and Salomey Osei. 2021. Masakhaner: Named entity recognition for african languages. In *AfricaNLP workshop at EACL 2021*.

Waleed Ammar, Phoebe Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah Smith. 2016. Many Languages, One Parser. *TACL*, 4:431–444.

Mikel Artetxe, Sebastian Ruder, and Dani Yogatama. 2019. On the cross-lingual transferability of monolingual representations. *CoRR*, abs/1910.11856.

James Barry, Joachim Wagner, and Jennifer Foster. 2019. Cross-lingual parsing with polyglot training and multi-treebank learning: A Faroese case study. In *DeepLo 2019*.

Ali Basirat, Miryam de Lhoneux, Artur Kulmizev, Murathan Kurfal, Joakim Nivre, and Robert Östling. 2019. Polyglot parsing for one thousand and one languages (and then some). In *Typology for Polyglot NLP workshop*.

References II

Isaac Caswell, Julia Kreutzer, Lisa Wang, Ahsan Wahab, Daan van Esch, Nasanbayar Ulzii-Orshikh, Allahsera Tapo, Nishant Subramani, Artem Sokolov, Claytone Sikasote, Monang Setyawan, Supheakmungkol Sarin, Sokhar Samb, Benoît Sagot, Clara Rivera, Annette Rios, Isabel Papadimitriou, Salomey Osei, Pedro Javier Ortiz Suárez, Iroro Orife, Kelechi Ogueji, Rubungo Andre Niyongabo, Toan Q. Nguyen, Mathias Müller, André Müller, Shamsuddeen Hassan Muhammad, Nanda Muhammad, Ayanda Mnyakeni, Jamshidbek Mirzakhalov, Tapiwanashe Matangira, Colin Leong, Nze Lawson, Sneha Kudugunta, Yacine Jernite, Mathias Jenny, Orhan Firat, Bonaventure F. P. Dossou, Sakhile Dlamini, Nisansa de Silva, Sakine Çabuk Balli, Stella Biderman, Alessia Battisti, Ahmed Baruwa, Ankur Bapna, Pallavi Baljekar, Israel Abebe Azime, Ayodele Awokoya, Duygu Ataman, Orevaoghene Ahia, Oghenefego Ahia, Sweta Agrawal, and Mofetoluwa Adeyemi. 2021. In *AfricaNLP workshop at EACL 2021*. [link].

Alexis Conneau, Ruty Rinott, Guillaume Lample, Adina Williams, Samuel Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. XNLI: Evaluating cross-lingual sentence representations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Adam Fisch, Jiang Guo, and Regina Barzilay. 2019. Working hard or hardly working: Challenges of integrating typology into neural dependency parsers. In *EMNLP*.

Dan Kondratyuk and Milan Straka. 2019. 75 languages, 1 model: Parsing Universal Dependencies universally. In *EMNLP-IJCNLP*.

Miryam de Lhoneux, Johannes Bjerva, Isabelle Augenstein, and Anders Søgaard. 2018. Parameter sharing between dependency parsers for related languages. In *EMNLP*.

Manuel Mager, Arturo Oncevay, Abteen Ebrahimi, John Ortega, Annette Rios, Angela Fan, Ximena Gutierrez-Vasques, Luis Chiruzzo, Gustavo Giménez-Lugo, Ricardo Ramos, Ivan Vladimir Meza Ruiz, Rolando Coto-Solano, Alexis Palmer, Elisabeth Mager-Hois, Vishrav Chaudhary, Graham Neubig, Ngoc Thang Vu, and Katharina Kann. 2021. Findings of the AmericasNLP 2021 shared task on open machine translation for indigenous languages of the Americas. In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*.

Ailsa Meechan-Maddon and Joakim Nivre. 2019. How to parse low-resource languages: Cross-lingual parsing, target language annotation, or both? In *Depling, SyntaxFest*.

References III

Phoebe Mulcaire, Jungo Kasai, and Noah A. Smith. 2019. Low-resource parsing with crosslingual contextualized representations. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*.

Benjamin Muller, Antonios Anastasopoulos, Benoît Sagot, and Djamé Seddah. 2021. When being unseen from mBERT is just the beginning: Handling new languages with multilingual language models. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Tahira Naseem, Regina Barzilay, and Amir Globerson. 2012. Selective sharing for multilingual dependency parsing. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Muhammad, Salomon Kabongo Kabenamualu, Salomey Osei, Freshia Sackey, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa Berhe, Mofetoluwa Adeyemi, Masabata Mokgesi-Selinga, Lawrence Okegbemi, Laura Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkadir Dangana, Herman Kamper, Hady Elsahar, Goodness Duru, Ghollah Kioko, Murhabazi Espoir, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Chinenye Emezue, Bonaventure F. P. Dossou, Blessing Sibanda, Blessing Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *LREC*.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

Barbara Plank. 2019. Transferring NLP models across languages and domains. Invited talk at SyntaxFest.

References IV

Gowtham Ramesh, Sumanth Doddapaneni, Aravinth Bheemaraj, Mayank Jobanputra, Raghavan AK, Ajitesh Sharma, Sujit Sahoo, Harshita Diddee, Mahalakshmi J, Divyanshu Kakwani, Navneet Kumar, Aswin Pradeep, Kumar Deepak, Vivek Raghavan, Anoop Kunchukuttan, Pratyush Kumar, and Mitesh Shantadevi Khapra. 2021. Samanantar: The largest publicly available parallel corpora collection for 11 indic languages. *CoRR*, abs/2104.05596.

Manon Scholivet, Franck Dary, Alexis Nasr, Benoit Favre, and Carlos Ramisch. 2019. Typological features for multilingual delexicalised dependency parsing. In *NAACL*.

Aaron Smith, Bernd Bohnet, Miryam de Lhoneux, Joakim Nivre, Yan Shao, and Sara Stymne. 2018. 82 treebanks, 34 models: Universal dependency parsing with multi-treebank models. In *CoNLL 2018 Shared Task*.

Jörg Tiedemann. 2020. The tatoeba translation challenge – realistic data sets for low resource and multilingual MT. In *Proceedings of the Fifth Conference on Machine Translation*.

Ke Tran and Arianna Bisazza. 2019. Zero-shot dependency parsing with pre-trained multilingual sentence representations. In *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)*.

Ahmet Üstün, Arianna Bisazza, Gosse Bouma, and Gertjan van Noord. 2020. Udapter: Language adaptation for truly universal dependency parsing. In *EMNLP*.

Clara Vania, Yova Kementchedjhieva, Anders Søgaard, and Adam Lopez. 2019. A systematic comparison of methods for low-resource dependency parsing on genuinely low-resource languages. In *EMNLP*.

David Vilares, Carlos Gómez-Rodríguez, and Miguel A. Alonso. 2016. One model, two languages: training bilingual parsers with harmonized treebanks. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*.

Shijie Wu and Mark Dredze. 2019. Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.