Inf Sci Master thesis topics

2019/2020

Arianna Bisazza

a.bisazza@rug.nl

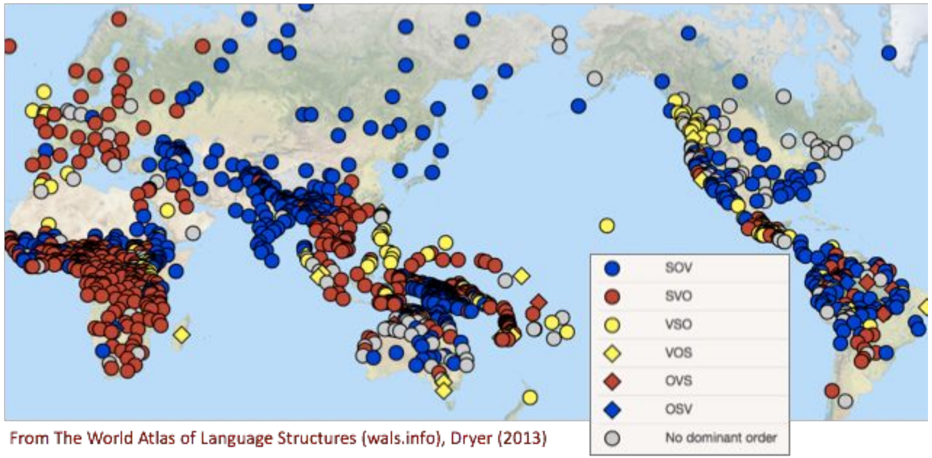
[P1] Predicting success in Neural MT

[P2] Compositionality in Neural Networks

[P3] Extracting Access Control policies from text

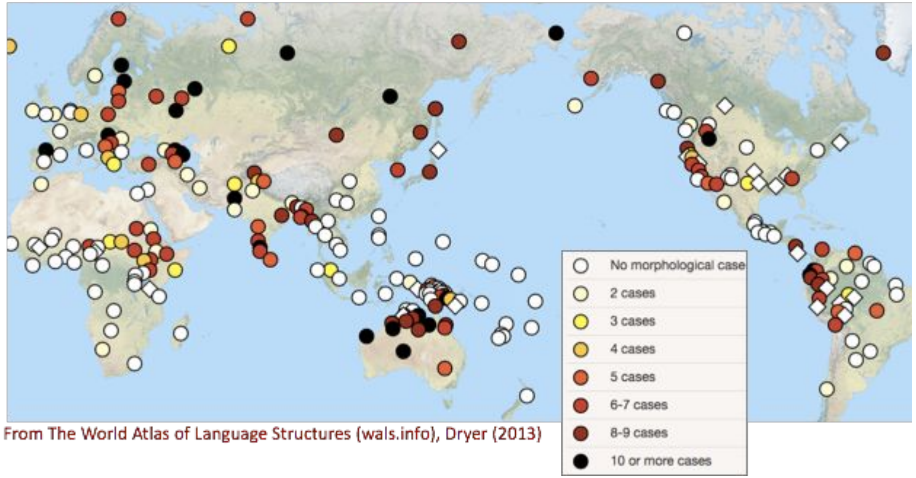
[P1] Predicting success in Neural Machine Translation (NMT)

Order of Subject, Object, Verb



From The World Atlas of Language Structures (wals.info), Dryer (2013)

Number of nominal cases (nominative, genitive, acccusative...)



From The World Atlas of Language Structures (wals.info), Dryer (2013)

NMT quality varies dramatically among language pairs. Differences in morphology and word order lead to worse quality

Which typological features make a language harder to model?

[P1] Predicting success in Neural Machine Translation (NMT)

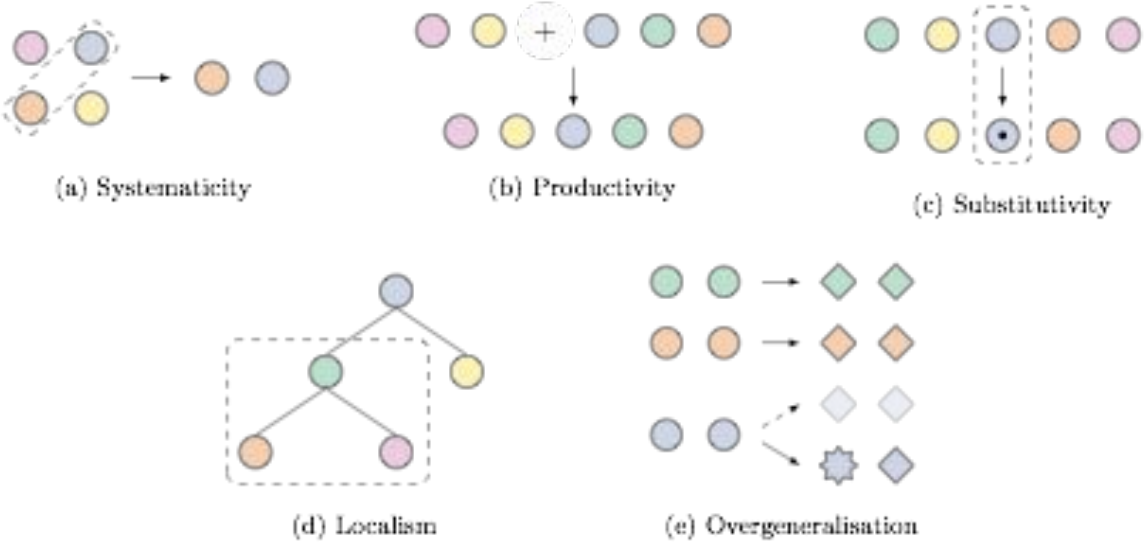
To isolate specific features we work with *synthetic languages*:

- Toy languages created with synchronous grammars, OR
- Real languages with fictitious features

Original		they say the broker took them out for lunch frequently . *(subject; verb; object)*
Polypersonal agreement		they saykon the broker tookkarker them out for lunch frequently . *(kon: plural subject; kar: singular subject; ker: plural object)*
Word order variation	SVO	they say the broker took out frequently them for lunch .
	SOV	they the broker them took out frequently for lunch say .
	VOS	say took out frequently them the broker for lunch they.
	VSO	say they took out frequently the broker them for lunch .
	OSV	them the broker took out frequently for lunch they say .
	OVS	them took out frequently the broker for lunch say they .
Case systems	Unambiguous	theykon saykon the brokerkar tookkarker theyker out for lunch frequently . *(kon: plural subject; kar: singular subject; ker: plural object)*
	Syncretic	theykon saykon the brokerkar tookkarkar theykar out for lunch frequently . *(kon: plural subject; kar: plural object/singular subject)*
	Argument marking	theyker sayker the brokerkin tookkerkin theyker out for lunch frequently . *(ker: plural argument; kin: singular argument)*

[Ravfogel et al. 2019] Studying the Inductive Biases of RNNs with Synthetic Variations of Natural Languages

[P2] Measuring compositionality in neural networks
(with D. Hupkes, UvA)



Do neural language models (NLMs) systematically recombine known parts and rules? Do NLMs favour rules or exceptions during training?

[Hupkes et al. 2019] The compositionality of neural networks: integrating symbolism and connectionism

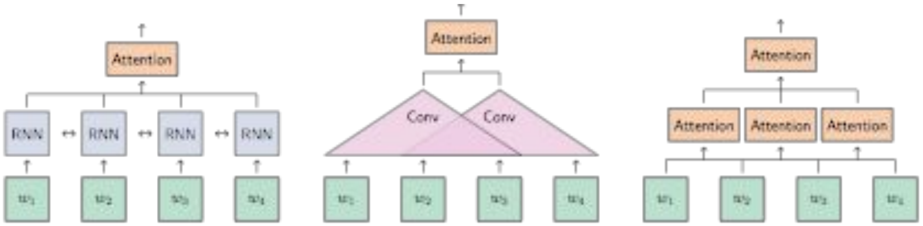
[P2] Measuring compositionality in neural networks
(with D. Hupkes, UvA)

Extend [Hupkes et al. 2019] by:

- Monitoring accuracy of various tests over training time
- Strengthening the comparison of different architectures with better hyperparameter tuning
- Extending the dataset .. ?

Benchmark: PCFG SET

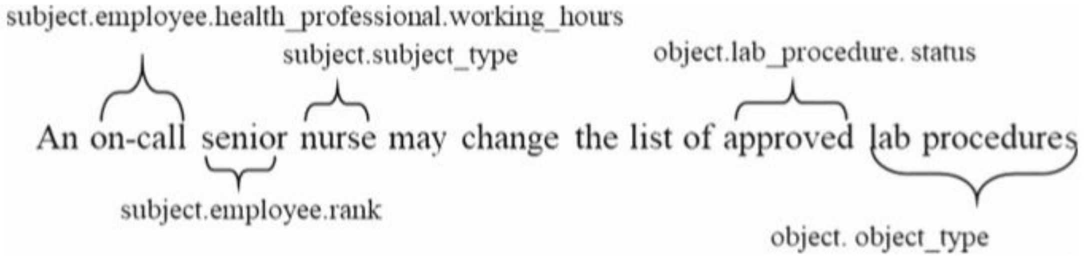
Unary functions F_U:		Binary functions F_B:	
copy $x_1 \cdots x_n$	$\rightarrow x_1 \cdots x_n$	append x, y	$\rightarrow$ x y
reverse $x_1 \cdots x_n$	$\rightarrow x_n \cdots x_1$	prepend x, y	$\rightarrow$ y x
shift $x_1 \cdots x_n$	$\rightarrow x_2 \cdots x_n\ x_1$	remove_first x, y	$\rightarrow$ y
swap $x_1 \cdots x_n$	$\rightarrow x_n\ x_2 \cdots x_{n-1}\ x_1$	remove_second x, y	$\rightarrow$ x
repeat $x_1 \cdots x_n$	$\rightarrow x_1 \cdots x_n\ x_1 \cdots x_n$		
echo $x_1 \cdots x_n$	$\rightarrow x_1 \cdots x_n\ x_n$		



(a) LSTMS2S (b) ConvS2S (c) Transformer

[Hupkes et al. 2019] The compositionality of neural networks: integrating symbolism and connectionism

[P3] Extracting machine-readable access control policies from text (with F. Turkmen, RuG-CS)




```
namespace subject{
attribute subject_type{
category = subject_cat
id = "subject_type"
type = string}
namespace employee{
attribute rank{
category = subject_cat
id = "rank"
type = string}
namespace health_professional
    {
attribute working_hours{
category = subject_cat
id = "working_hours"
type = string }}}}
namespace object{
attribute object_type{
category = object_cat
id = "object_type"
type = string}
namespace lab_procedure{
attribute status{
category = object_cat
id = status
type = string}}}
```

- Natural language is the preferred way to express *security policies* within real-world organizations
- NLP techniques (word embedding, information extraction) can help automate this process
- We'll look at ways to better evaluate state-of-the-art methods and improve their NLP components

[Alhoaly et al. 2019] Automated extraction of attributes from Natural Language ABAC Policies