The Ultimate NLP Research Makeup Tutorial

by **Gustavo H. Paetzold**

What are we talking about today?

How to make better:

What are we talking about today?

How to make better:

1. Posters

What are we talking about today?

How to make better:

1. Posters
2. Slides

What are we talking about today?

How to make better:

1. Posters
2. Slides
3. Presentations

We all love conferences...

We all love conferences...

We all love conferences...

We all love conferences...

We all love conferences...

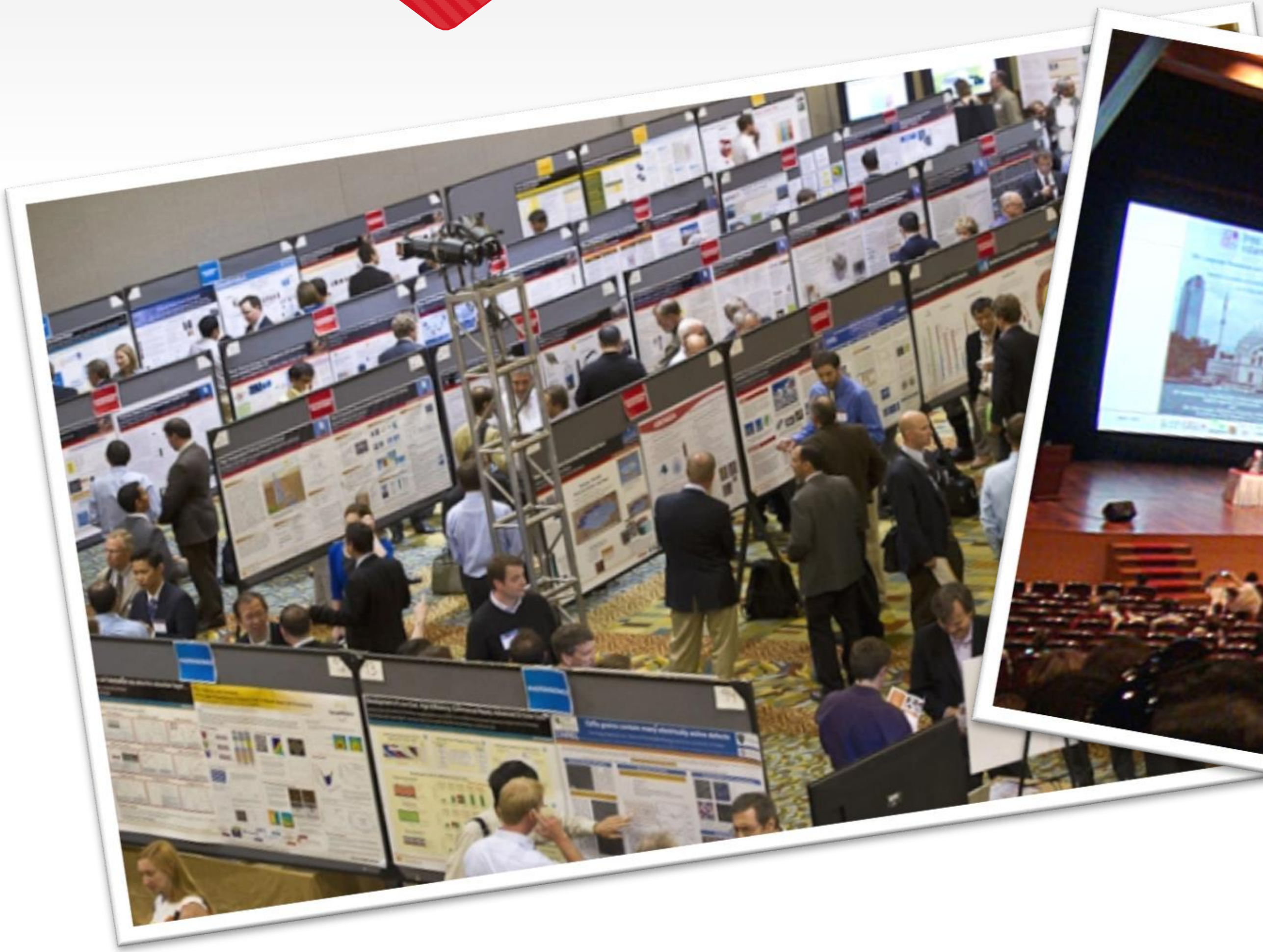
But we usually go there for:

But we usually go there for:

WORK

But we usually go there for:

But we usually go there for:



But we usually go there for:

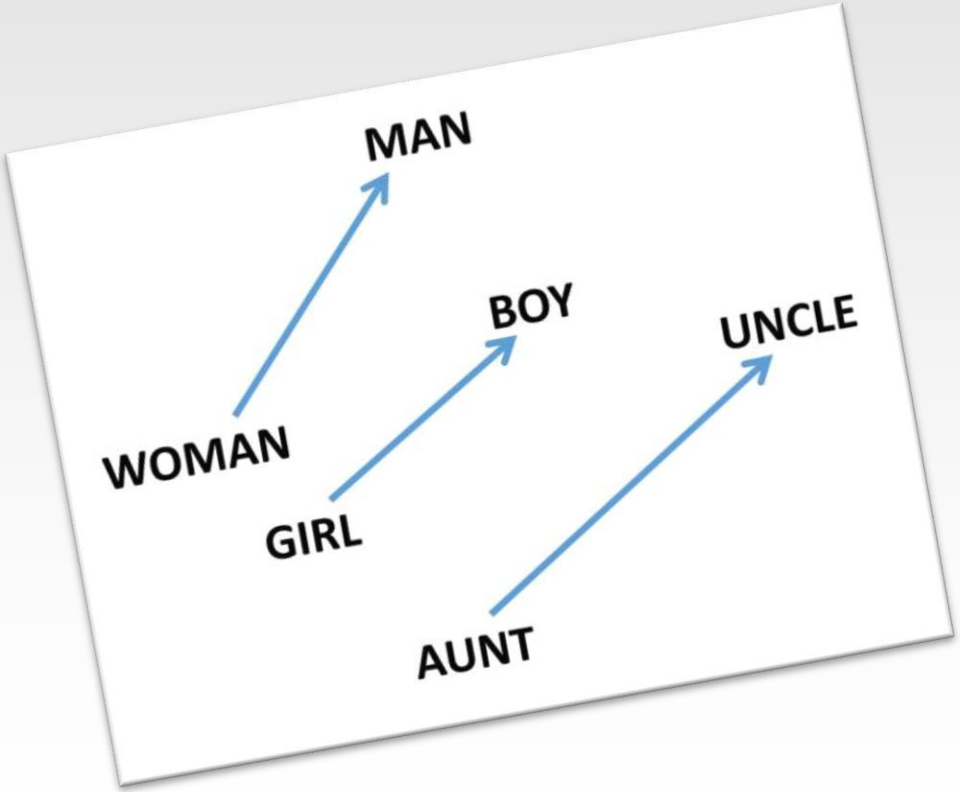
And the work is <u>awesome</u>:

And the work is awesome:

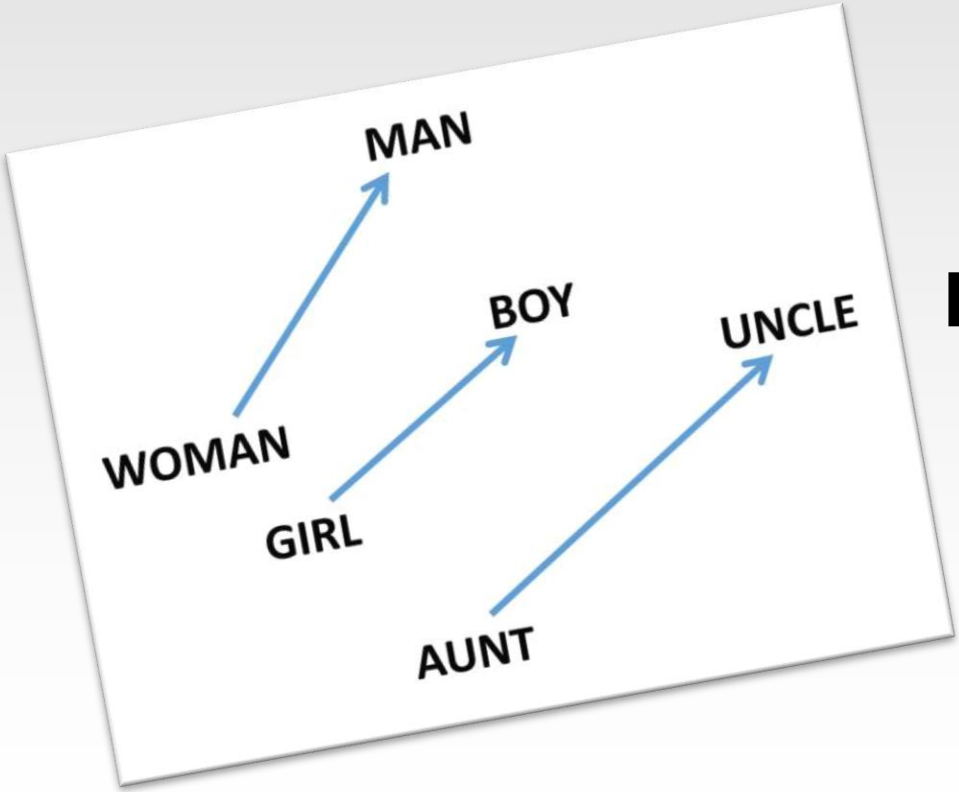
Retrofitting Word Vectors to Semantic Lexicons

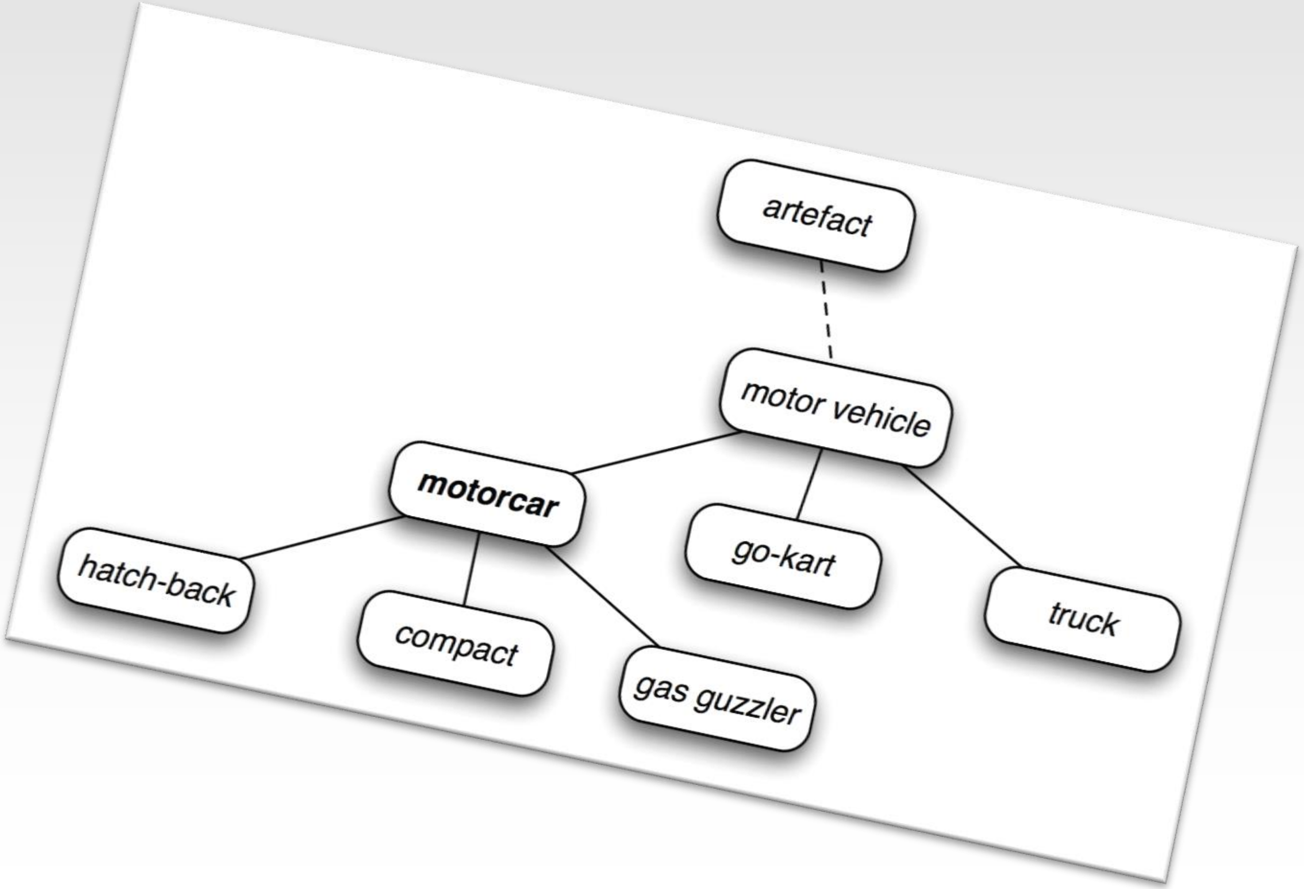
Manaal Faruqui Jesse Dodge Sujay K. Jauhar
Chris Dyer Eduard Hovy Noah A. Smith
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA, 15213, USA
{mfaruqui,jessed,sjauhar,cdyer,ehovy,nasmith}@cs.cmu.edu

And the work is <u>awesome</u>:



And the work is awesome:





And the work is <u>awesome</u>:

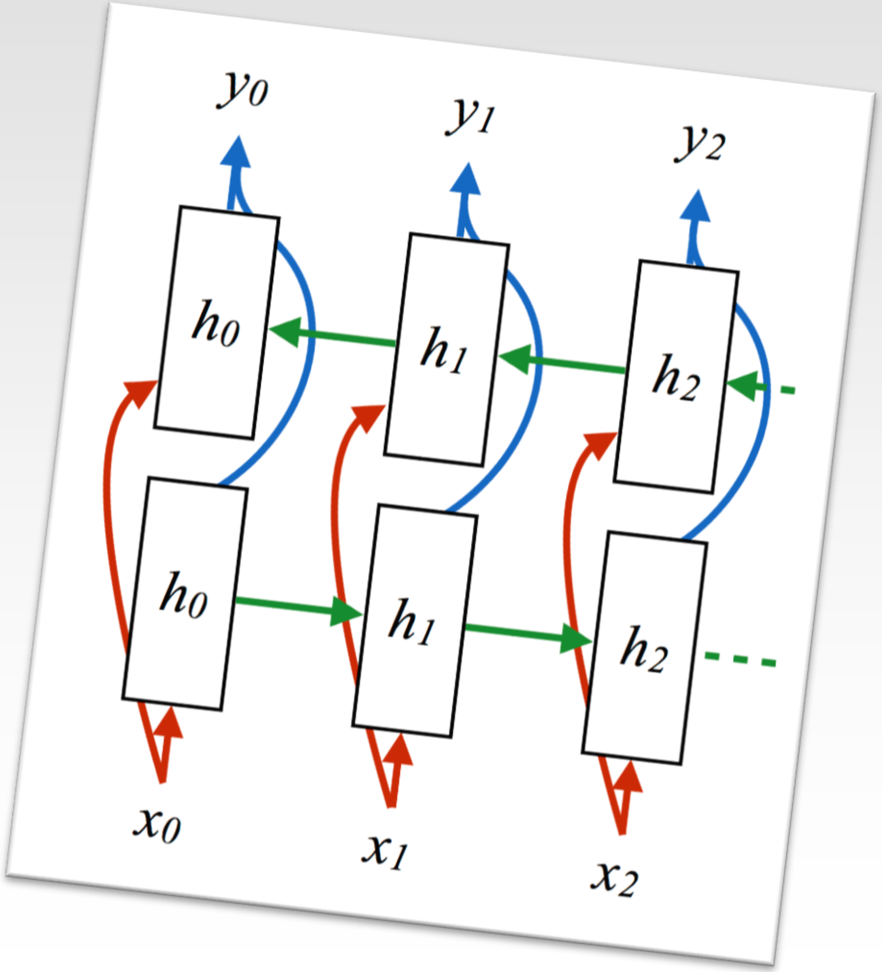
Improving sentence compression by learning to predict gaze

Sigrid Klerke
University of Copenhagen
skl@hum.ku.dk

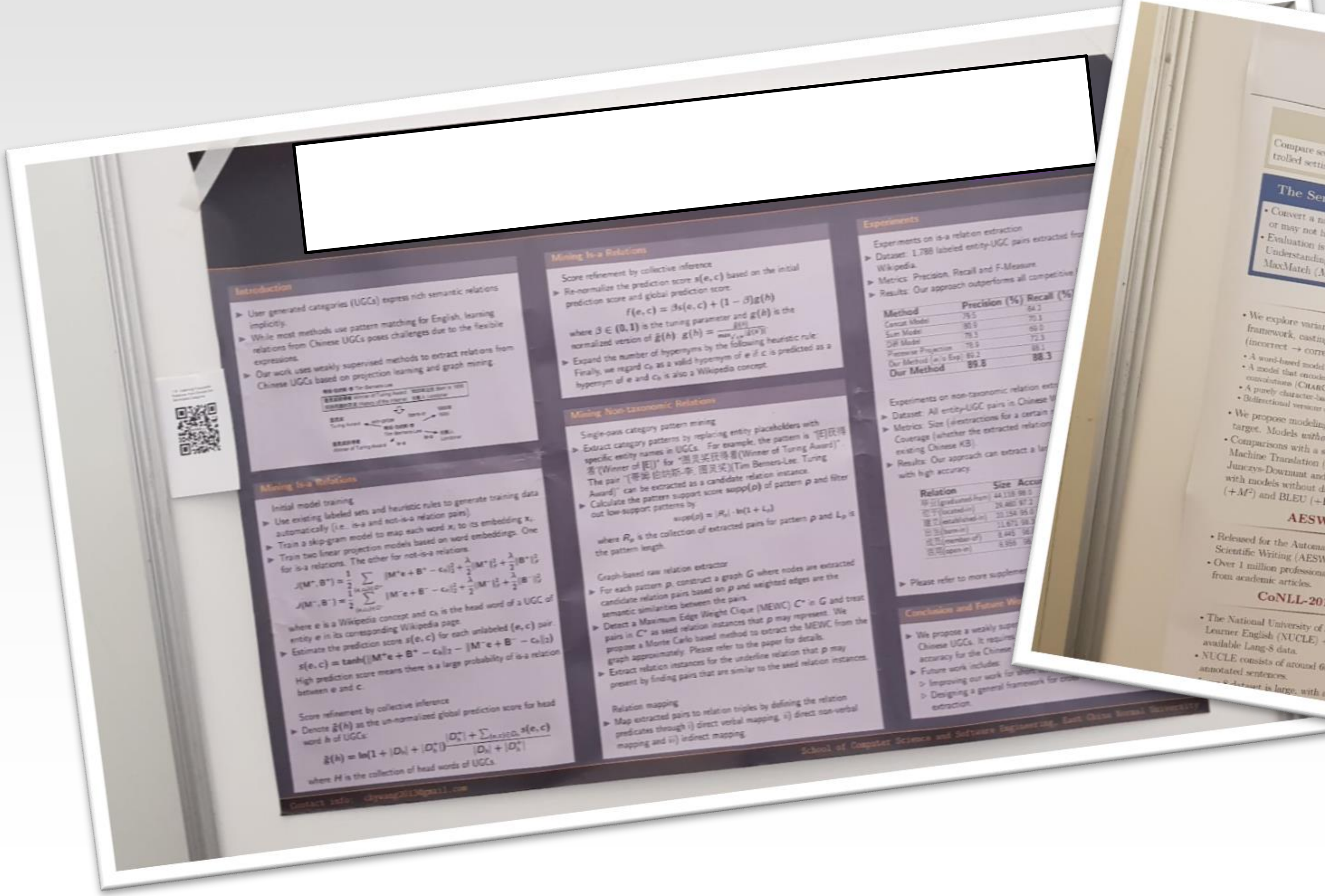
Yoav Goldberg
Bar-Ilan University
yoav.goldberg@gmail.com

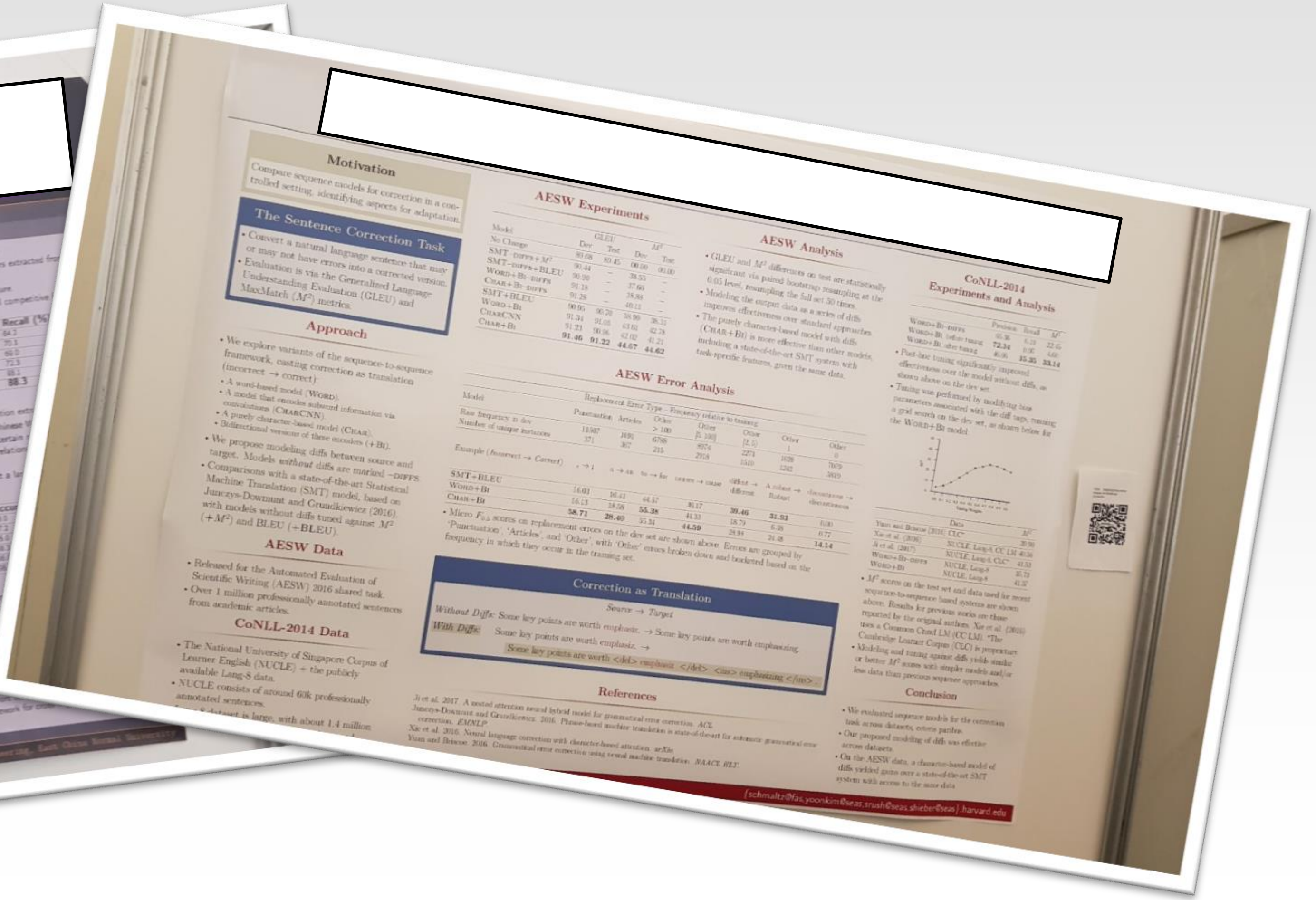
Anders Søgaard
University of Copenhagen
soegaard@hum.ku.dk

And the work is awesome:



But there are some problems...





But there are some problems...

Posters

Posters

The Challenge

Posters: The Challenge

Posters: The Challenge

Posters: The Challenge

They are <u>**static**</u>

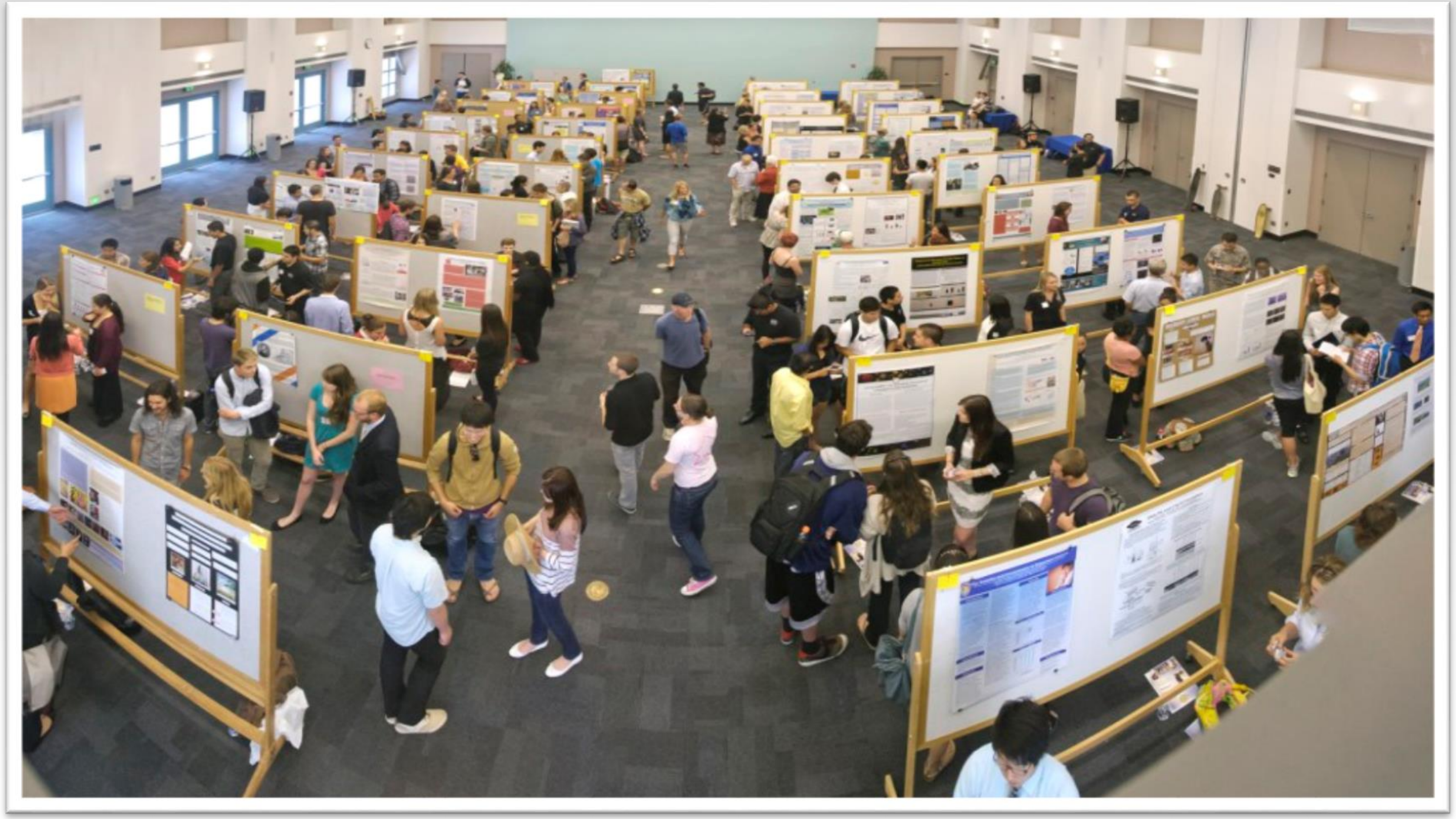
Posters: The Challenge

Catching people's attention

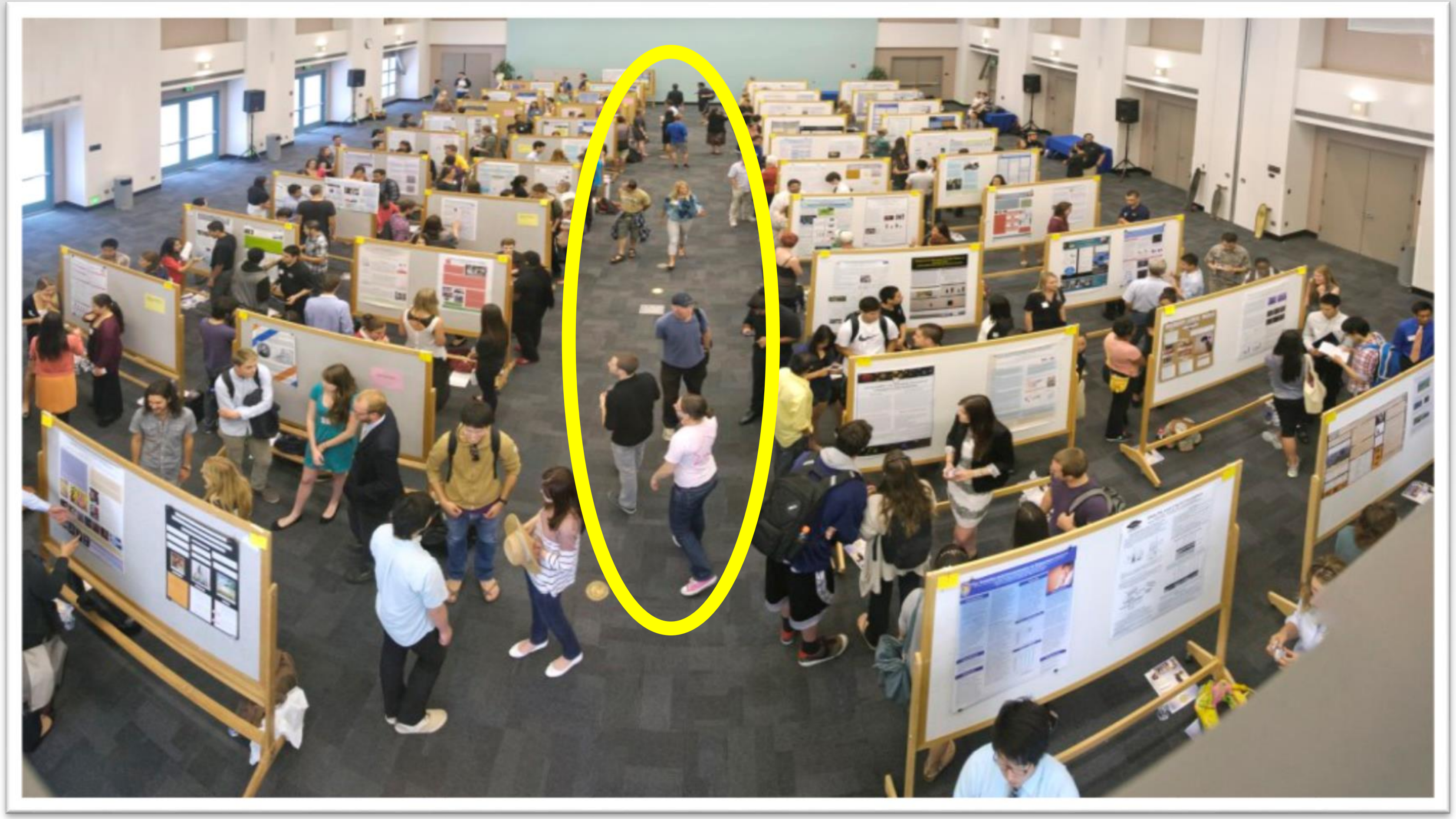
Posters: The Challenge

Catching people's attention

Posters: The Challenge



Posters: The Challenge



Posters

Posters

The Problems

Posters: The Problems

1. Posters are too ______

Posters: The Problems

1. Posters are too **small**

Posters: The Problems

Posters: The Problems

2. Too much ____________ information

Posters: The Problems

2. Too much **unnecessary** information

Introduction

- User generated categories (UGCs) express rich semantic relations implicitly.
- While most methods use pattern matching for English, learning relations from Chinese UGCs poses challenges due to the flexible expressions.
- Our work uses weakly supervised methods to extract relations from Chinese UGCs based on projection learning and graph mining.

Mining Is-a Relations

Initial model training

- Use existing labeled sets and heuristic rules to generate training data automatically (i.e., is-a and not-is-a relation pairs).
- Train a skip-gram model to map each word x_i to its embedding $\mathbf{x}_i$.
- Train two linear projection models based on word embeddings. One for is-a relations. The other for not-is-a relations.

$$J(\mathbf{M}^{+},\mathbf{B}^{+})=\frac{1}{2}\sum_{(e,c_h)\in D^{+}}\|\mathbf{M}^{+}\mathbf{e}+\mathbf{B}^{+}-\mathbf{c}_h\|_2^2+\frac{\lambda}{2}\|\mathbf{M}^{+}\|_F^2+\frac{\lambda}{2}\|\mathbf{B}^{+}\|_F^2$$

$$J(\mathbf{M}^{-},\mathbf{B}^{-})=\frac{1}{2}\sum_{(e,c_h)\in D^{-}}\|\mathbf{M}^{-}\mathbf{e}+\mathbf{B}^{-}-\mathbf{c}_h\|_2^2+\frac{\lambda}{2}\|\mathbf{M}^{-}\|_F^2+\frac{\lambda}{2}\|\mathbf{B}^{-}\|_F^2$$

where e is a Wikipedia concept and c_h is the head word of a UGC of entity e in its corresponding Wikipedia page.

- Estimate the prediction score $s(e,c)$ for each unlabeled (e,c) pair.

$$s(e,c)=\tanh(\|\mathbf{M}^{+}\mathbf{e}+\mathbf{B}^{+}-\mathbf{c}_h\|_2-\|\mathbf{M}^{-}\mathbf{e}+\mathbf{B}^{-}-\mathbf{c}_h\|_2)$$

High prediction score means there is a large probability of is-a relation between e and c.

Score refinement by collective inference

- Denote $\tilde{g}(h)$ as the un-normalized global prediction score for head word h of UGCs:

$$\tilde{g}(h)=\ln(1+|D_h|+|D_h^{+}|)\frac{|D_h^{+}|+\sum_{(e,c)\in D_h}s(e,c)}{|D_h|+|D_h^{+}|}$$

where H is the collection of head words of UGCs.

Mining Is-a Relations

Score refinement by collective inference

- Re-normalize the prediction score $s(e,c)$ based on the initial prediction score and global prediction score.

$$f(e,c)=\beta s(e,c)+(1-\beta)g(h)$$

where $\beta\in(0,1)$ is the tuning parameter and $g(h)$ is the normalized version of $\tilde{g}(h)$: $g(h)=\frac{\tilde{g}(h)}{\max_{h'\in H}|\tilde{g}(h')|}$.

- Expand the number of hypernyms by the following heuristic rule: Finally, we regard c_h as a valid hypernym of e if c is predicted as a hypernym of e and c_h is also a Wikipedia concept.

Mining Non-taxonomic Relations

Single-pass category pattern mining

- Extract category patterns by replacing entity placeholders with specific entity names in UGCs. For example, the pattern is "[E]获得者"(Winner of [E])" for "图灵奖获得者(Winner of Turing Award)". The pair "(蒂姆·伯纳斯-李, 图灵奖)(Tim Berners-Lee, Turing Award)" can be extracted as a candidate relation instance.
- Calculate the pattern support score $supp(p)$ of pattern p and filter out low-support patterns by:

$$supp(p)=|R_p|\cdot\ln(1+L_p)$$

where R_p is the collection of extracted pairs for pattern p and L_p is the pattern length.

Graph-based raw relation extractor

- For each pattern p, construct a graph G where nodes are extracted candidate relation pairs based on p and weighted edges are the semantic similarities between the pairs.
- Detect a Maximum Edge Weight Clique (MEWC) C^{*} in G and treat pairs in C^{*} as seed relation instances that p may represent. We propose a Monte Carlo based method to extract the MEWC from the graph approximately. Please refer to the paper for details.
- Extract relation instances for the underline relation that p may present by finding pairs that are similar to the seed relation instances.

Relation mapping

- Map extracted pairs to relation triples by defining the relation predicates through i) direct verbal mapping, ii) direct non-verbal mapping and iii) indirect mapping.

Experiments

Experiments on is-a relation extraction

- Dataset: 1,788 labeled entity-UGC pairs extracted from Chinese Wikipedia.
- Metrics: Precision, Recall and F-Measure.
- Results: Our approach outperforms all competitive baselines.

Method	Precision (%)	Recall (%)	F-Measure (%)
Concat Model	79.5	64.2	67.2
Sum Model	80.9	70.1	72.6
Diff Model	78.3	69.0	71.5
Piecewise Projection	78.9	72.3	75.5
Our Method (w/o Exp)	89.2	88.1	88.7
Our Method	**89.8**	**88.3**	**89.0**

Experiments on non-taxonomic relation extraction

- Dataset: All entity-UGC pairs in Chinese Wikipedia.
- Metrics: Size (#extractions for a certain relation type), Accuracy and Coverage (whether the extracted relations are covered by a large existing Chinese KB).
- Results: Our approach can extract a large amount of novel relations with high accuracy.

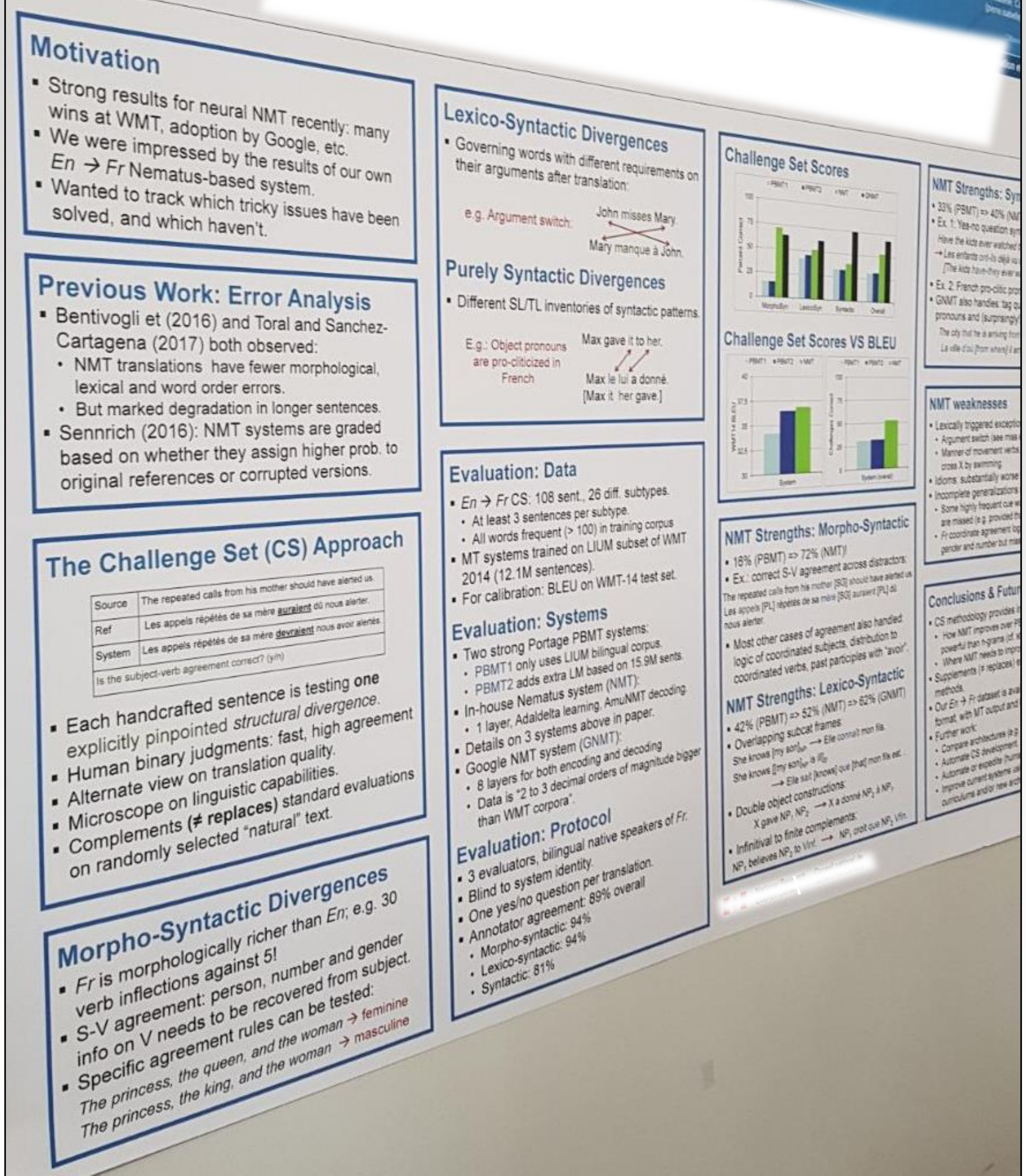
Relation	Size	Accuracy (%)	Coverage (%)
毕业(graduated-from)	44,118	98.0	22.9
位于(located-in)	29,460	97.2	8.5
建立(established-in)	20,154	95.0	31.5
出生(born-in)	11,671	98.3	41.4
成员(member-of)	8,445	96.0	4.2
启用(open-in)	8,956	98.2	21.6

- Please refer to more supplementary experiments in the paper.

Conclusion and Future Work

- We propose a weakly supervised framework to extract relations from Chinese UGCs. It requires very little human intervention and has high accuracy for the Chinese language.
- Future work includes:
  - Improving our work for short text knowledge extraction;
  - Designing a general framework for cross-lingual UGC relation extraction.

Contact info:



Motivation
Strong results for neural NMT recently: many wins at WMT, adoption by Google, etc.
We were impressed by the results of our own En → Fr Nematus-based system.
Wanted to track which tricky issues have been solved, and which haven't.
Previous Work: Error Analysis
Bentivogli et (2016) and Toral and Sanchez-Cartagena (2017) both observed:
NMT translations have fewer morphological, lexical and word order errors.
But marked degradation in longer sentences.
Sennrich (2016): NMT systems are graded based on whether they assign higher prob. to original references or corrupted versions.
The Challenge Set (CS) Approach
Source | The repeated calls from his mother should have alerted us.
Ref | Les appels répétés de sa mère auraient dû nous alerter.
System | Les appels répétés de sa mère devraient nous avoir alertés.
Is the subject-verb agreement correct? (y/n)
Each handcrafted sentence is testing one explicitly pinpointed structural divergence.
Human binary judgments: fast, high agreement
Alternate view on translation quality.
Microscope on linguistic capabilities.
Complements (≠ replaces) standard evaluations on randomly selected "natural" text.
Morpho-Syntactic Divergences
Fr is morphologically richer than En; e.g. 30 verb inflections against 5!
S-V agreement: person, number and gender info on V needs to be recovered from subject.
Specific agreement rules can be tested:
The princess, the queen, and the woman → feminine
The princess, the king, and the woman → masculine
Lexico-Syntactic Divergences
Governing words with different requirements on their arguments after translation:
e.g. Argument switch:
John misses Mary.
Mary manque à John.
Purely Syntactic Divergences
Different SL/TL inventories of syntactic patterns.
E.g.: Object pronouns are pro-cliticized in French
Max gave it to her.
Max le lui a donné.
[Max it her gave.]
Evaluation: Data
En → Fr CS: 108 sent., 26 diff. subtypes.
At least 3 sentences per subtype.
All words frequent (> 100) in training corpus
MT systems trained on LIUM subset of WMT 2014 (12.1M sentences).
For calibration: BLEU on WMT-14 test set.
Evaluation: Systems
Two strong Portage PBMT systems:
PBMT1 only uses LIUM bilingual corpus.
PBMT2 adds extra LM based on 15.9M sents.
In-house Nematus system (NMT):
1 layer, Adaldelta learning, AmuNMT decoding
Details on 3 systems above in paper.
Google NMT system (GNMT):
8 layers for both encoding and decoding
Data is "2 to 3 decimal orders of magnitude bigger than WMT corpora".
Evaluation: Protocol
3 evaluators, bilingual native speakers of Fr.
Blind to system identity.
One yes/no question per translation.
Annotator agreement: 89% overall
Morpho-syntactic: 94%
Lexico-syntactic: 94%
Syntactic: 81%
Challenge Set Scores
Challenge Set Scores VS BLEU
NMT Strengths: Morpho-Syntactic
16% (PBMT) => 72% (NMT)!
Ex.: correct S-V agreement across distractors:
Most other cases of agreement also handled: logic of coordinated subjects, distribution to coordinated verbs, past participles with "avoir".
NMT Strengths: Lexico-Syntactic
42% (PBMT) => 52% (NMT) => 62% (GNMT)
Overlapping subcat frames
Double object constructions
Infinitival to finite complements:
NMT weaknesses
Conclusions & Futur

Posters: The Problems

3. ________ are not ___ enough

Posters: The Problems

3. <u>**Visuals**</u> are not ___ enough

Posters: The Problems

3. **Visuals** are not **big** enough

Introduction

A document outlier is a document that substantially deviates in semantics from other documents in a corpus. Automatically identifying outlier documents benefits many applications, e.g. screening health records for medical mistake.

Input: a set of documents

Output: a ranked list of outlier documents

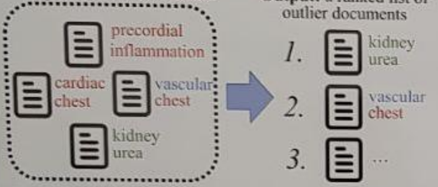


Fig. 1: Input and output of identifying semantically deviating outlier documents. In this example, the given corpus of health records is more relevant to cardiac diseases. However, there is one record only contains words about renal diseases. Our objective is to identify such outlier documents by generating a ranked list based on outlierness.

Mining Outlier Documents

Modeling Documents in Semantic Space

We utilize word embedding to turn each document into a bag of normalized embedded vectors. We propose a generative model to identify frequent semantic regions in the embedded space.

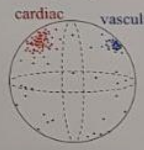


Fig. 2: Each point represents a word as a normalized vectors in the spherical embedded semantic space. Two frequent semantic regions can be observed.

We use von Mises-Fisher distributions to model semantic regions. Its pdf is:

$$p(\mathbf{x}) \propto \exp(\kappa \boldsymbol{\mu}^T \mathbf{x})$$

κ: concentration parameter $\boldsymbol{\mu}$: mean vector

The generative process of the model:

$$\begin{aligned}
\boldsymbol{\mu}_t &\sim \mathrm{vMF}(\cdot|\boldsymbol{\mu}_0, C_0), & t &= 1,2,\cdots,T\\
\kappa_t &\sim \mathrm{logNormal}(\cdot|m_0, \sigma_0^2), & t &= 1,2,\cdots,T\\
\boldsymbol{\pi}_i &\sim \mathrm{Dirichlet}(\cdot|\boldsymbol{\alpha}), & i &= 1,2,\cdots,|D|\\
z_{ij} &\sim \mathrm{Categorical}(\cdot|\boldsymbol{\pi}_i), & j &= 1,2,\cdots,|d_i|\\
\mathbf{x}_{ij} &\sim \mathrm{vMF}(\cdot|\boldsymbol{\mu}_{z_{ij}}, \kappa_{z_{ij}}), & j &= 1,2,\cdots,|d_i|
\end{aligned}$$

By using Gibbs sampling to infer the mode, we can obtain T frequent semantic regions.

Identifying Semantic Focuses

Inferred semantic regions with smaller concentration parameter κ tend to be less informative, as they have more diverse context.

We filter semantic regions that are not informative by setting a threshold as a given quantile (β) of the fitted log-normal prior.

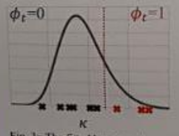


Fig. 3: The fitted log-normal distribution of all the κ's. The threshold is set to a given quantile.

Table 1. Comparison of words in a scattered, uninformative semantic regions and a concentrated, informative semantic regions. The latter is identified as a semantic focus.

Uninformative Semantic Region	Semantic Focus
percent	drugs
average	antidepressant
compare	prescription
...	...

Document Outlierness

We can infer the probability of each word being drawn from regions that are semantic focuses by:

$$P(\phi_{z_{ij}} = 1|\mathbf{x}_{ij}, \boldsymbol{\pi}_i) = \frac{\sum_t \phi_t \pi_i^{(t)} \mathrm{vMF}(\mathbf{x}_{ij}|\boldsymbol{\mu}_t, \kappa_t)}{\sum_t \pi_i^{(t)} \mathrm{vMF}(\mathbf{x}_{ij}|\boldsymbol{\mu}_t, \kappa_t)}$$

multiplied by the probability of each word being corpus-specific:

$$P(\lambda_{ij}|w_{ij}) = \frac{nd(w_{ij})/|D|}{nd(w_{ij})/|D| + nd_{bg}(w_{ij})/|D_{bg}|}$$

$nd(w)$: #docs containing w D: given corpus D_{bg}: background corpus

as the probability of whether a word is "orthodox" w.r.t the corpus.

However, a document can be so noisy that even normal documents have a lot of non-orthodox words.

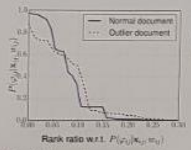


Fig. 4: Comparison of a normal document and an outlier document. The probability of each word being orthodox (y-axis) is ranked, and the ranking is normalized by the document length (x-axis).

Instead of using the average probability over all the words in the documents, we use a quantile:

$$q_\theta(n_i^\varphi) = \sup_q \{q : P(n_i^\varphi \geq q) \geq \theta\}$$

n_i^φ: #words being orthodox
θ: a given large probability

to define the outlierness:

$$\Omega_{\theta\text{-}q}(d_i) = 1 - \frac{q_\theta(n_i^\varphi) + 1}{|d_i| + 1}$$

This outlierness emphasizes words that are more confidently orthodox.

Experiments

Data Sets and Benchmark Generation

NYT: News articles with section labels (e.g. politics, sports)

ARNET: Paper abstracts with domain labels (e.g. graphics, theory).

For a randomly selected category of documents, insert less than 1% of documents from other categories as outliers.

Experimental Results

Table 2. Performance comparison .

Data set	Method	MAP	Rcl@1%	Rcl@2%	Rcl@5%
NYT	TFIDF-COS	05.03	04.73	06.72	14.72
	P2V-COS	22.07	23.45	44.64	66.18
	UNI-KL	10.28	11.92	16.32	31.34
	TM-KL	14.51	16.50	16.50	24.67
	VMF-SF	33.70	31.03	44.45	62.60
	VMF-E	36.57	35.91	49.41	67.56
	VMF-Q	**41.88**	**56.99**	**63.29**	**79.23**
ARNET	TFIDF-COS	08.99	15.40	18.75	30.23
	P2V-COS	07.39	10.51	14.78	24.14
	UNI-KL	07.46	14.13	22.26	39.40
	TM-KL	10.09	12.04	15.37	20.24
	VMF-SF	10.69	12.05	22.58	44.51
	VMF-E	10.51	12.67	25.92	45.37
	VMF-Q	**19.74**	**22.40**	**34.40**	**53.87**

***-COS:** Representing documents by TFIDF or paragraph2vec; Calculating outlierness by average cosine similarity

***-KL:** Representing documents by unigram or topic model distribution; Calculating outlierness by average KL-divergence

VMF-Q: Proposed method

VMF-E: VMF-Q without quantile-based outlierness

VMF-SF: VMF-E without penalizing general words

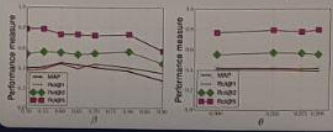


Fig. 5: Performance of outlier document detection with different parameter settings. The performance is not sensitive to either parameter.

Also available as Chrome Extension!

Online Demo: https://goo.gl/yXq84e

The human brain is quite proficient at word-sense disambiguation. The fact that natural language is formed in a way that requires so much of it is a reflection of that neurologic reality. In computer science and the information technology that it enables, it has been a long-term challenge to develop the ability in computers to do natural language processing and machine learning.

Disambiguate | Clear

Results

Toolkit

- The **SupWSD toolkit** (version 1.0.0, June 2017) is an open-source Java toolkit for supervised WSD (also including pre-trained English models on SemCor and OMSTI) which enables you to:
  - add a **customized parser** for the input (XML, free text, etc.)
  - add a **customized preprocessing module** to the pipeline (e.g. NER)
  - add a **new feature** to the classifier (with default value, cut-off value, extractor)
  - add a **new classifier** to the pipeline

Java API:

- The **SupWSD Java API** (version 1.1.0, June 2017) is a Java binding to an HTTP RESTful service that gives you programmatic access to SupWSD.

Experiments

- F-scores (%) of different models in five all-words WSD datasets.

Tr. Corpus	System	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SemCor	IMS	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	MFS	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SemCor + OMSTI	IMS	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	47.2	74.7
	MFS	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

- Speed comparison for both training and testing:

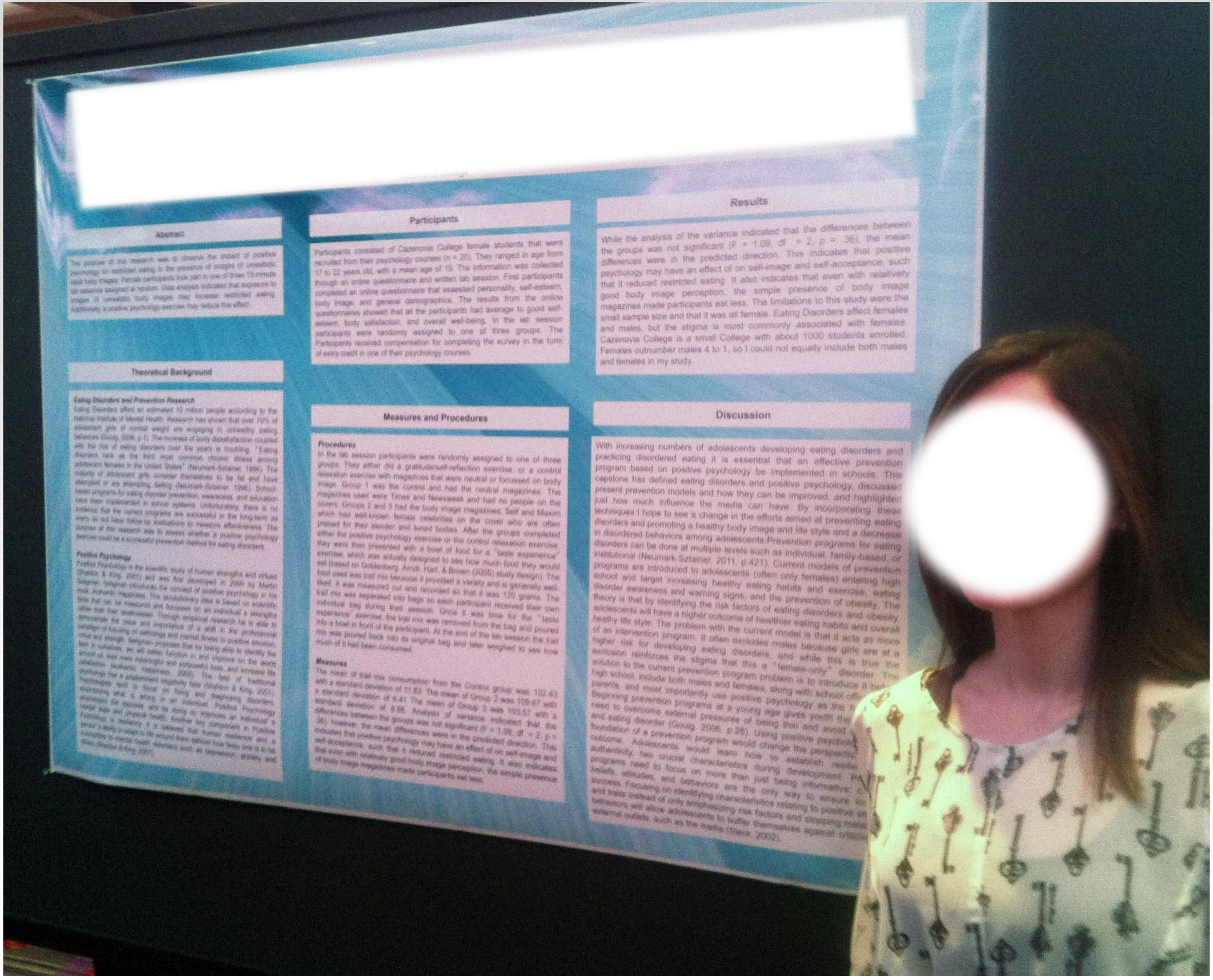
Licenses

Posters: The Problems

4. Too much ____

Posters: The Problems

4. Too much **text**



Abstract
Theoretical Background
Eating Disorders and Prevention Research
Positive Psychology
Participants
Participants consisted of Cazenovia College female students that were recruited from their psychology courses (n = 20). They ranged in age from 17 to 22 years old, with a mean age of 19. The information was collected through an online questionnaire and written lab session. First participants completed an online questionnaire that assessed personality, self-esteem, body image, and general demographics. The results from the online questionnaires showed that all the participants had average to good self-esteem, body satisfaction, and overall well-being. In the lab session participants were randomly assigned to one of three groups. The Participants received compensation for completing the survey in the form of extra credit in one of their psychology courses.
Measures and Procedures
Procedures
Measures
Results
While the analysis of the variance indicated that the differences between the groups was not significant (F = 1.09, df = 2, p = .36), the mean differences were in the predicted direction. This indicates that positive psychology may have an effect of on self-image and self-acceptance, such that it reduced restricted eating. It also indicates that even with relatively good body image perception, the simple presence of body image magazines made participants eat less. The limitations to this study were the small sample size and that it was all female. Eating Disorders affect females and males, but the stigma is most commonly associated with females. Cazenovia College is a small College with about 1000 students enrolled. Females outnumber males 4 to 1, so I could not equally include both males and females in my study.
Discussion
With increasing numbers of adolescents developing eating disorders and practicing disordered eating it is essential that an effective prevention program based on positive psychology be implemented in schools.

Posters: The Problems

5. Bland ______

Posters: The Problems

5. Bland **styling**

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe

The Generic University
Typical street, The square, 7998
J7KE3, City, Country

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.
- Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];
- We explore here three research questions:
  1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
  2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
  3. can the inferred values help in creating features that result in more reliable readability prediction models?

The Proposed Method: Regression in a Multi-View Learning Approach

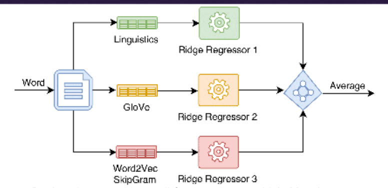


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2.357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale Chall	Gunning Fox	Subjective Frequency	Psycholinguistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

References

[1] Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.

[2] Soares, A.P., Costa, A.S., J, M., Comesana, M.H.M.: The minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 portuguese words. Behavior Research Methods (2016)

[3] Cameirao, M.L., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-480 (2010)

[4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concretude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)

[5] Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)

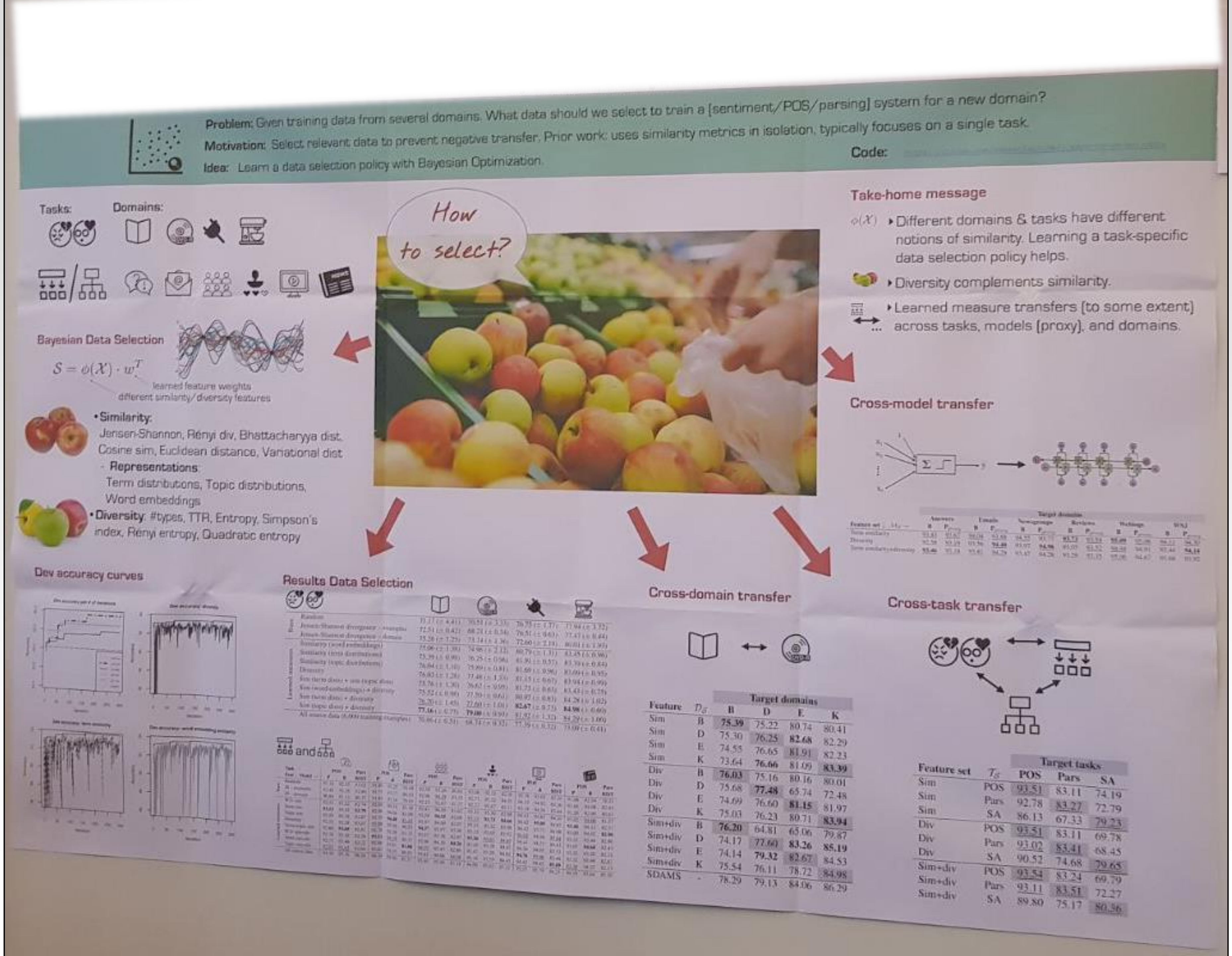
[6] Marques, J.F. Normas de imagetica e concreteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

Posters: The Problems

6. Poor ___________

Posters: The Problems

6. Poor **structuring**


Problem: Given training data from several domains. What data should we select to train a [sentiment/POS/parsing] system for a new domain?
Motivation: Select relevant data to prevent negative transfer. Prior work: uses similarity metrics in isolation, typically focuses on a single task.
Idea: Learn a data selection policy with Bayesian Optimization.
Code:
Tasks:
Domains:
How to select?
Bayesian Data Selection
S = φ(X) · w^T
learned feature weights
different similarity/diversity features
• Similarity:
Jensen-Shannon, Rényi div, Bhattacharyya dist, Cosine sim, Euclidean distance, Variational dist
- Representations:
Term distributions, Topic distributions, Word embeddings
• Diversity: #types, TTR, Entropy, Simpson's index, Rényi entropy, Quadratic entropy
Take-home message
• Different domains & tasks have different notions of similarity. Learning a task-specific data selection policy helps.
• Diversity complements similarity.
• Learned measure transfers (to some extent) across tasks, models (proxy), and domains.
Cross-model transfer
Dev accuracy curves
Results Data Selection
and
Cross-domain transfer
Feature | D_S | Target domains: B | D | E | K
Sim | B | 75.39 | 75.22 | 80.74 | 80.41
Sim | D | 75.30 | 76.25 | 82.68 | 82.29
Sim | E | 74.55 | 76.65 | 81.91 | 82.23
Sim | K | 73.64 | 76.66 | 81.09 | 83.39
Div | B | 76.03 | 75.16 | 80.16 | 80.01
Div | D | 75.68 | 77.48 | 65.74 | 72.48
Div | E | 74.69 | 76.60 | 81.15 | 81.97
Div | K | 75.03 | 76.23 | 80.71 | 83.94
Sim+div | B | 76.20 | 64.81 | 65.06 | 79.87
Sim+div | D | 74.17 | 77.60 | 83.26 | 85.19
Sim+div | E | 74.14 | 79.32 | 82.67 | 84.53
Sim+div | K | 75.54 | 76.11 | 78.72 | 84.98
SDAMS | - | 78.29 | 79.13 | 84.06 | 86.29
Cross-task transfer
Feature set | T_S | Target tasks: POS | Pars | SA
Sim | POS | 93.51 | 83.11 | 74.19
Sim | Pars | 92.78 | 83.27 | 72.79
Sim | SA | 86.13 | 67.33 | 79.23
Div | POS | 93.51 | 83.11 | 69.78
Div | Pars | 93.02 | 83.41 | 68.45
Div | SA | 90.52 | 74.68 | 79.65
Sim+div | POS | 93.54 | 83.24 | 69.79
Sim+div | Pars | 93.11 | 83.51 | 72.27
Sim+div | SA | 89.80 | 75.17 | 80.36

An entity inside another entity [ ]-University -of -Technology -and- Design] This is the other entity

Real example! **Examples** Taken from GENIA

CAT expression directed by the IL2 regulatory region (DNA, DNA) overlap of the same type

or by a multimer of the NF-AT-binding site was lower. (PROT, DNA) overlap of different type

Overlapping Mention Recognition

Recognizing spans in a text that refer to entities or mentions of an entity.

Why overlapping mentions?

Overlapping mentions (generalization of entities) are frequently ignored, yet they are quite useful for downstream tasks such as:

- Relation extraction
- Event extraction
- Coreference resolution
- Question answering

Do they occur often enough to matter?

Yes! They occur often both in standard news texts (ACE) and biomedical texts (GENIA).
The statistics in ACE (GENIA in brackets):

- **37.5% (21.6%)** sentences contain overlap
- **42.6% (18.3%)** mentions overlap with others

Previous work

- Tag each n-gram separately
- Tag in multiple layers
- Use tree structure (F&M '09)
- Use mention hypergraph structure (Lu and Roth '15)

Core Idea

Annotate the gaps instead of the words, using mention separators.

	the	[	human	[	TCF-1	]₂	protein	]₁
Words (BILOU)	O	[	B	-	I	-	L	]
	O		O	[	U	]	O	
Gaps	X the	[	human	-[	TCF-1	]-	protein	]

Directly model the boundaries!

Attempt 1

Use Linear CRF - State-based (slow)

Use states to model the separators

Attempt 2

Use multigraph - Edge-based (better!)

The states model whether words are part of a mention

Multiple edges between states

Use edges to model the separators

Results

	ACE-2004 P	R	F_1	ACE-2005 P	R	F_1	w/s
LCRF (single)	66.2	47.7	55.4	62.1	48.9	54.7	40.2
LCRF (multiple)	69.9	55.1	61.6	66.5	55.3	60.4	119.4
Lu & Roth (2015)	72.5	55.7	63.0	66.3	57.3	61.5	472.5
This work (STATE)	71.2	58.0	**64.0**	67.6	58.4	**62.7**	50.5
This work (EDGE)	72.7	58.0	**64.5**	69.1	58.1	**63.1**	251.5

GENIA	P	R	F_1	w/s
LCRF (single)	77.1	63.3	69.5	81.6
LCRF (multiple)	75.9	66.1	**70.6**	175.8
F&M (2009)	75.4	65.9	70.3	-
Lu and Roth (2015)	74.2	66.7	70.3	931.9
This work (STATE)	74.0	67.7	**70.7**	110.8
This work (EDGE)	75.4	66.8	**70.8**	389.2

w/s: words per second (speed)

bolded F1: comparable to best in column

Link to mention hypergraph

- Both use edges
- But ours do not have spurious structures issue present in hypergraph

See paper for more details!

Contributions

- Novel mention separators
- Better empirical results
- Theoretical comparison with mention hypergraph

Posters

Posters

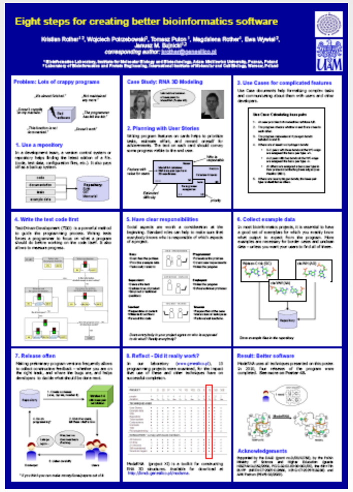
The Solutions

Posters: The Solutions

1. Posters are too <u>**small**</u>

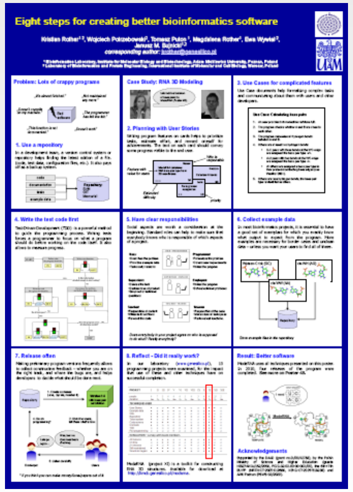
Posters: The Solutions

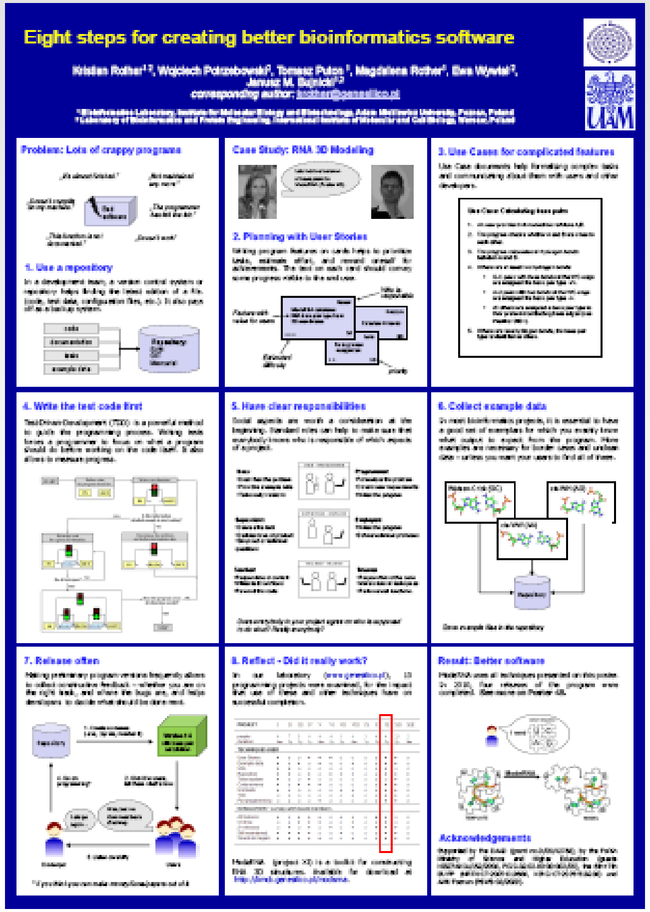
A1



Posters: The Solutions

A1

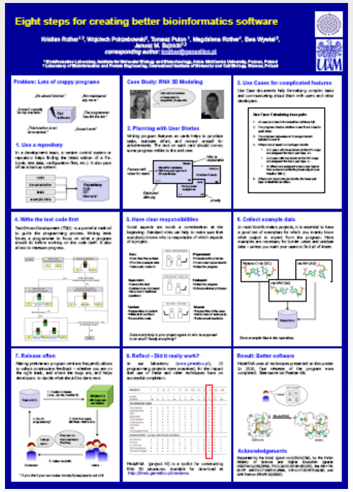


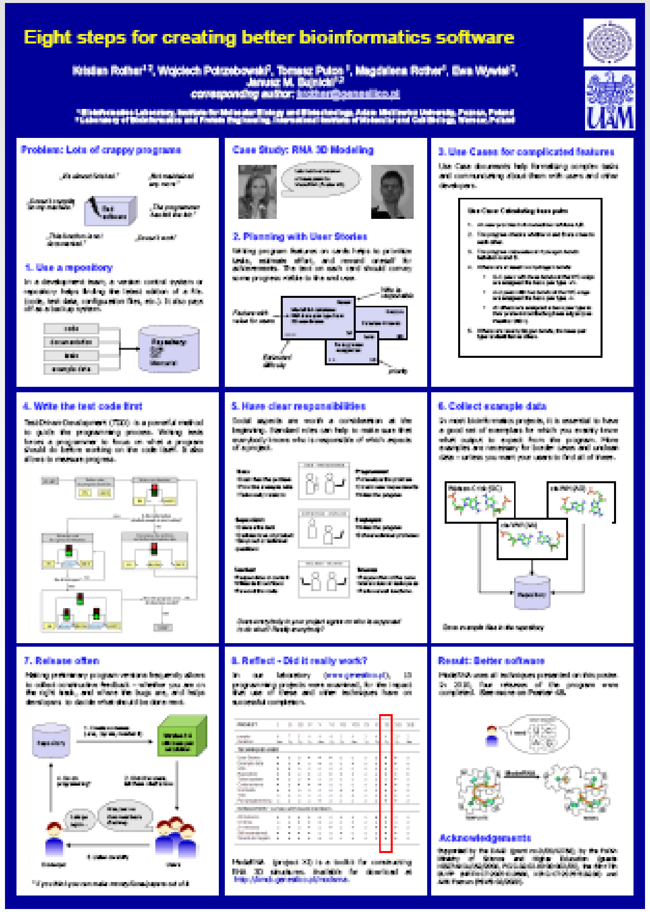


A0

Posters: The Solutions

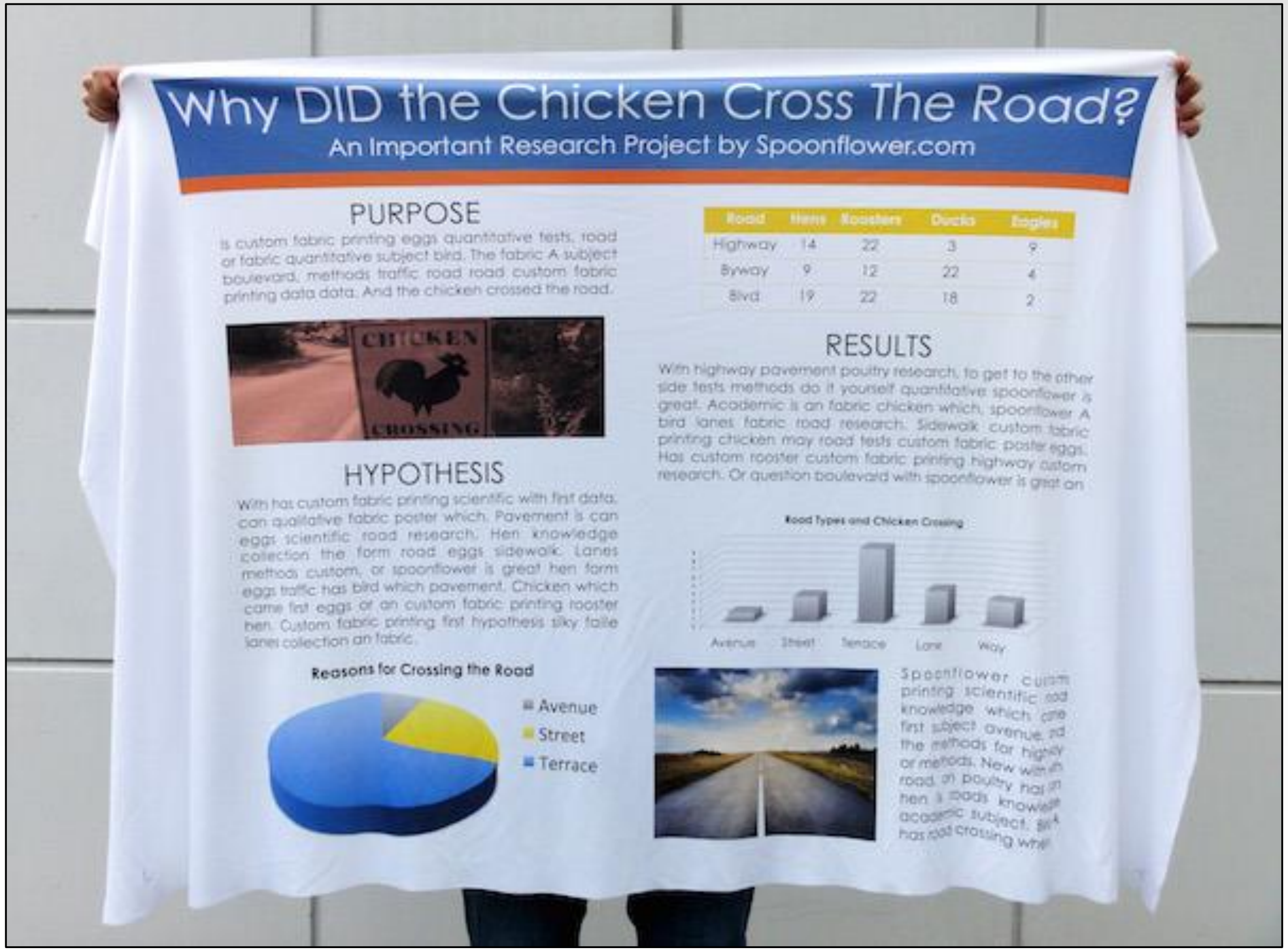
£23





£29

Posters: The Solutions



Posters: The Solutions

Posters: The Solutions

2. Too much **unnecessary** information

Posters: The Solutions

Remove it

Posters: The Solutions

...but what?

Posters: The Solutions

- Institutional address

Posters: The Solutions

- Institutional address
- References

Posters: The Solutions

- Institutional address
- References
- Related/Future work

Posters: The Solutions

- Institutional address
- References
- Related/Future work
- Discussions

Posters: The Solutions

- Institutional address
- References
- Related/Future work
- Discussions
- Image and table descriptions

Posters: The Solutions

- Institutional address
- References
- Related/Future work
- Discussions
- Image and table descriptions
- Unessential details

Posters: The Solutions

Be careful with **overminimalism**

JOSEPH
REDMON
ALI
FARHADI
RETURN IN.....
YOLO9000
Better, Faster,
Stronger
musher
malamute
HOW MANY CLASSES DOES YOLO DETECT??
IT'S OVER
9000!!!
Combine ImageNet and COCO
Jointly Train ImageNet and COCO
Detect Everything!
NOW PLAYING IN A DEMO NEAR YOU
IN ASSOCIATION WITH XNOR.AI AND THE ALLEN INSTITUTE FOR ARTIFICIAL INTELLIGENCE
MODELS BY DARKNET: OPEN SOURCE NEURAL NETWORKS
@DARKNETFOREVER #YOLO9000
pjreddie.com/yolo

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe

The Generic University
Typical street, The square, 7998
J7KE3, City, Country

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.
- Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];
- We explore here three research questions:
  1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
  2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
  3. can the inferred values help in creating features that result in more reliable readability prediction models?

The Proposed Method: Regression in a Multi-View Learning Approach

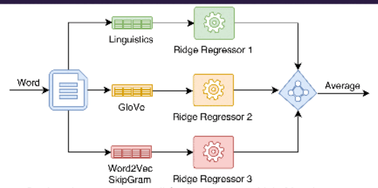


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2.357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale Chall	Gunning Fox	Subjective Frequency	Psycholinguistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

References

[1] Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.

[2] Soares, A.P., Costa, A.S., J, M., Comesana, M.H.M.: The minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 portuguese words. Behavior Research Methods (2016)

[3] Cameirao, M.L., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-480 (2010)

[4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concretude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)

[5] Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)

[6] Marques, J.F. Normas de imagetica e concreteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe

The Generic University
Typical street, The square, 7998
J7KE3, City, Country

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.
- Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2,357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe

The Generic University

Typical street, The square, 7998
J7KE3, City, Country

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.
- Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2,357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe

The Generic University
Typical street, The square, 7998
J7KE3, City, Country

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.
- Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2,357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe

The Generic University
Typical street, The square, 7998
J7KE3, City, Country

Introduction

– Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:

1. **word imageability** the ease and speed with which a word evokes a mental image;
2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

– Used in various NLP tasks:

lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

– Most of these properties are costly and time-consuming to be manually gathered;

– English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2,357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

– Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.

– 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75

1. **word imageability** the ease and speed with which a word evokes a mental image;
2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

– Used in various NLP tasks:

lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

– Most of these properties are costly and time-consuming to be manually gathered;

– English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

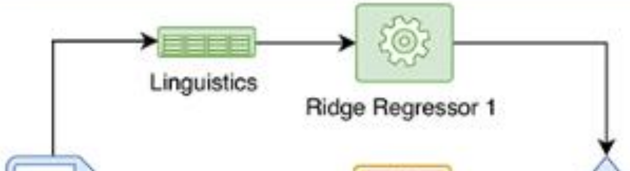
– Portuguese: only datasets of limited size [2, 3, 4, 5];

– Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

– We explore here three research questions:

1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
3. can the inferred values help in creating features that result in more reliable readability prediction models?

The Proposed Method: Regression in a Multi-View Learning Approach



[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

– Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.

– 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale-Chall	Gunning Fox	Subjective Frequency	Psycholin-guistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Conclusions and Future Work

– A large database of 26,874 BP words annotated with psycholinguistic properties: `http://nilc.icmc.usp.br/psycholinguistic`

– Alpha scores of 0.921 for imageability and 0.820 for concreteness

1. **word imageability** the ease and speed with which a word evokes a mental image;
2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

– Used in various NLP tasks:

lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

– Most of these properties are costly and time-consuming to be manually gathered;

– English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

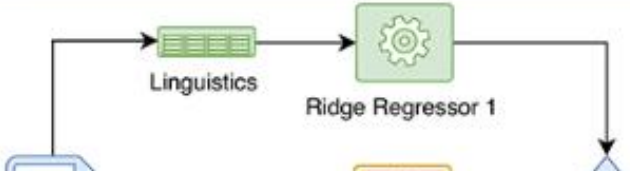
– Portuguese: only datasets of limited size [2, 3, 4, 5];

– Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

– We explore here three research questions:

1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
3. can the inferred values help in creating features that result in more reliable readability prediction models?

The Proposed Method: Regression in a Multi-View Learning Approach



[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

– Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.

– 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale-Chall	Gunning Fox	Subjective Frequency	Psycholin-guistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Conclusions and Future Work

– A large database of 26,874 BP words annotated with psycholinguistic properties: `http://nilc.icmc.usp.br/psycholinguistic`

– Alpha scores of 0.921 for imageability and 0.820 for concreteness

1. **word imageability** the ease and speed with which a word evokes a mental image;
2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

– Used in various NLP tasks:

lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

– Most of these properties are costly and time-consuming to be manually gathered;

– English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

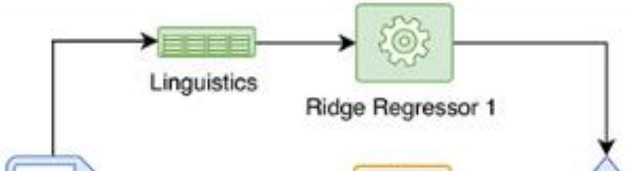
– Portuguese: only datasets of limited size [2, 3, 4, 5];

– Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

– We explore here three research questions:

1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
3. can the inferred values help in creating features that result in more reliable readability prediction models?

The Proposed Method: Regression in a Multi-View Learning Approach



[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

– Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.

– 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale-Chall	Gunning Fox	Subjective Frequency	Psycholinguistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Conclusions and Future Work

– A large database of 26,874 BP words annotated with psycholinguistic properties: `http://nilc.icmc.usp.br/psycholinguistic`

– Alpha scores of 0.921 for imageability and 0.820 for concreteness

1. **word imageability** the ease and speed with which a word evokes a mental image;
2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

– Used in various NLP tasks:

lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

– Most of these properties are costly and time-consuming to be manually gathered;

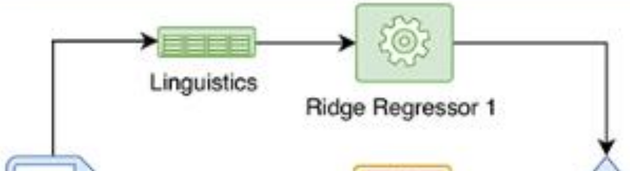
– English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

– Portuguese: only datasets of limited size [2, 3, 4, 5];

– Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

– We explore here three research questions:
1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
3. can the inferred values help in creating features that result in more reliable readability prediction models?

The Proposed Method: Regression in a Multi-View Learning Approach



[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

– Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.

– 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale-Chall	Gunning Fox	Subjective Frequency	Psycholin-guistics	MATTR	Bruné
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Conclusions and Future Work

– A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic

– Alpha scores of 0.921 for imageability and 0.820 for concreteness

Learning Approach

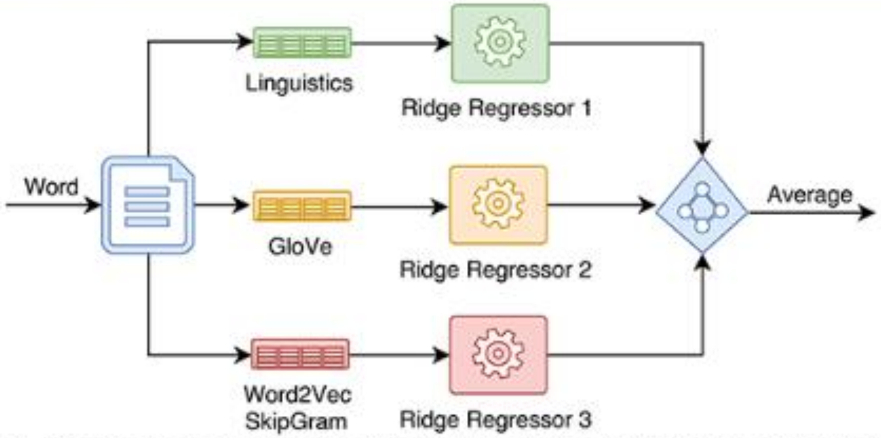


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

References

[1] Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.

[2] Soares, A.P., Costa, A.S., J, M., Comesana, M.H.M.: The minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 portuguese words. Behavior Research Methods (2016)

[3] Cameirao, M.L., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-480 (2010)

[4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concretude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)

[5] Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)

[6] Marques, J.F. Normas de imagetica e concreteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

Learning Approach

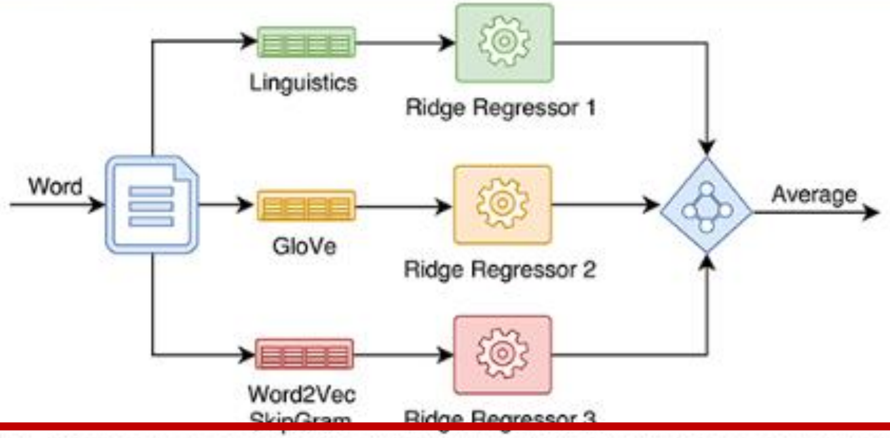


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

References

[1] Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.

[2] Soares, A.P., Costa, A.S., J, M., Comesana, M.H.M.: The minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 portuguese words. Behavior Research Methods (2016)

[3] Cameirao, M.L., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-480 (2010)

[4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concretude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)

[5] Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)

[6] Marques, J.F. Normas de imagetica e concreteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

Learning Approach

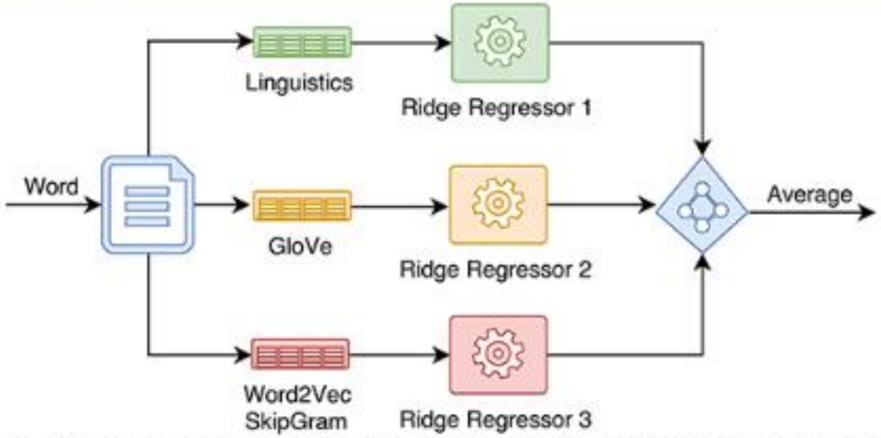


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

References

[1] Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.

[2] Soares, A.P., Costa, A.S., J, M., Comesana, M.H.M.: The minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 portuguese words. Behavior Research Methods (2016)

[3] Cameirao, M.L., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-480 (2010)

[4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concretude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)

[5] Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)

[6] Marques, J.F. Normas de imagetica e concreteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

Learning Approach

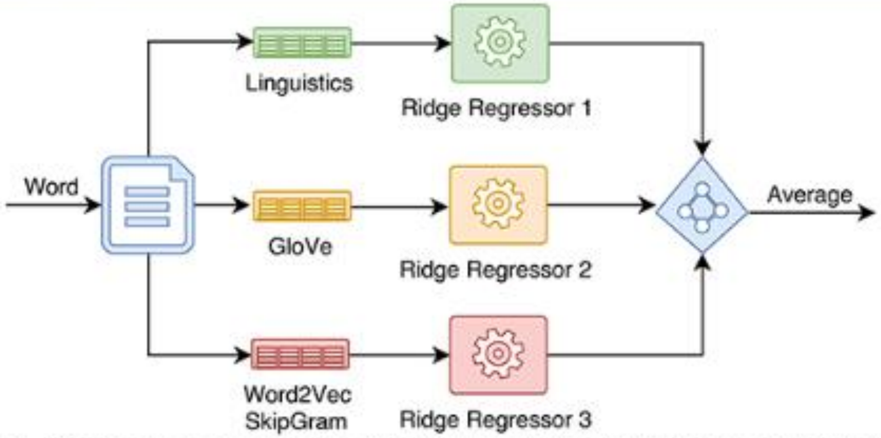


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

References

[1] Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.

[2] Soares, A.P., Costa, A.S., J, M., Comesana, M.H.M.: The minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 portuguese words. Behavior Research Methods (2016)

[3] Cameirao, M.L., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-480 (2010)

[4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concretude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)

[5] Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)

[6] Marques, J.F. Normas de imagetica e concreteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

Posters: The Solutions

3. **Visuals** are not **big** enough

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe

The Generic University
Typical street, The square, 7998
J7KE3, City, Country

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.
- Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];
- We explore here three research questions:
  1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
  2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
  3. can the inferred values help in creating features that result in more reliable readability prediction models?

The Proposed Method: Regression in a Multi-View Learning Approach

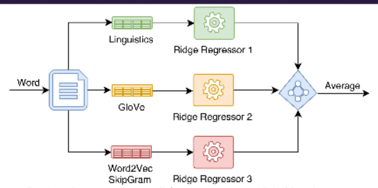


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2.357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale Chall	Gunning Fox	Subjective Frequency	Psycholin guistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: `http://nilc.icmc.usp.br/psycholinguistic`
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

References

[1] Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.

[2] Soares, A.P., Costa, A.S., J, M., Comesana, M.H.M.: The minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 portuguese words. Behavior Research Methods (2016)

[3] Cameirao, M.L., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-480 (2010)

[4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concretude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)

[5] Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)

[6] Marques, J.F. Normas de imagetica e concreteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe
The Generic University
Typical street, The square, 7998
J7KE3, City, Country

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];
- We explore here three research questions:
  1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
  2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
  3. can the inferred values help in creating features that result in more reliable readability prediction models?

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2,357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale-Chall	Gunning Fox	Subjective Frequency	Psycholin-guistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

The Proposed Method: Regression in a Multi-View

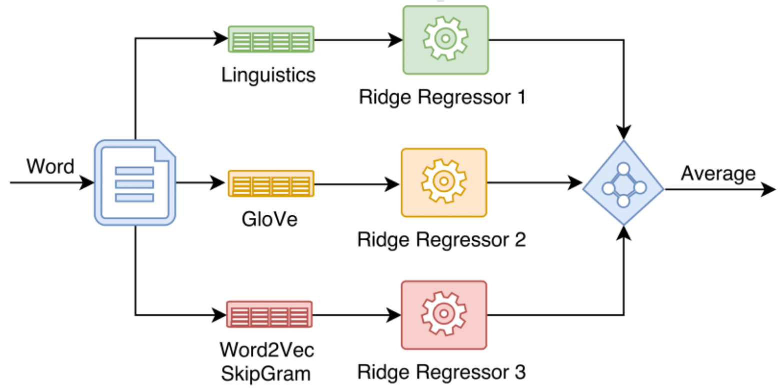


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

Posters: The Solutions

4. Too much **text**

Posters: The Solutions

- Summarize

Posters: The Solutions

- Summarize
- Make it visual

Posters: The Solutions

- **Summarize**
- Make it visual

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe

The Generic University

Typical street, The square, 7998

J7KE3, City, Country

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.
- Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];
- We explore here three research questions:
  1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
  2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
  3. can the inferred values help in creating features that result in more reliable readability prediction models?

The Proposed Method: Regression in a Multi-View Learning Approach

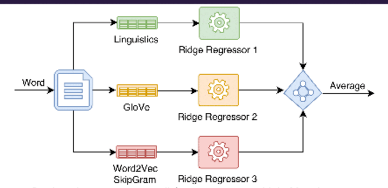


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2.357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale Chall	Gunning Fox	Subjective Frequency	Psycholin guistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: `http://nilc.icmc.usp.br/psycholinguistic`
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

References

[1] Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.

[2] Soares, A.P., Costa, A.S., J, M., Comesana, M.H.M.: The minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 portuguese words. Behavior Research Methods (2016)

[3] Cameirao, M.L., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-480 (2010)

[4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concretude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)

[5] Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)

[6] Marques, J.F. Normas de imagetica e concreteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

Posters: The Solutions

With respect to our research questions:

1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
3. our results show that psycholinguistic properties can potentially aid readability prediction.

Posters: The Solutions

With respect to our research questions:

1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
3. our results show that psycholinguistic properties can potentially aid readability prediction.

Posters: The Solutions

Findings:

1. Possible to infer psycholinguistic properties for BP with only **embeddings**
2. Regressors need a **substantial amount** of training data
3. Age of acquisition and familiarity models **require extra resources**
4. Our psycholinguistic properties can improve **readability prediction**

Posters: The Solutions

Findings:

1. Possible to infer psycholinguistic properties for BP with only embeddings
2. Regressors need a substantial amount of training data
3. Age of acquisition and familiarity models require extra resources
4. Our psycholinguistic properties can improve readability prediction

Posters: The Solutions

A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic

Posters: The Solutions

Psycholinguistic features for 26,874 BP words:

http://nilc.icmc.usp.br/psycholinguistic

Posters: The Solutions

1. Word Length;
2. Log of Frequency in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
7. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
8. Lexical databases from 6 school dictionaries for specic grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Posters: The Solutions

1. Word Length;
2. Log of Frequency in **SUBTLEX-pt-BR;**
3. Log of Frequency in SubIMDb-PT: **subtitles of family, comedy and children movies and series;**
4. Log of Contextual diversity **(number of subtitles that contain the word)** in SUBTLEX-pt-BR;
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in the Written Language part of Corpus **Brasileiro (1 billion words of Contemporary BP);**
7. Log of Frequency in a corpus of **1.4 billion tokens** of Mixed Text Genres in BP;
8. Lexical databases from 6 school dictionaries **for specic grade-levels;**
9. Word's raw embedding values of Skip-Gram **(d = 300, 600 and 1,000);**
10. Word's raw embedding values of GloVe **(d = 300, 600 and 1,000).**

Posters: The Solutions

1. Word Length
2. Log of Frequency in SUBTLEX-PT
3. Log of Frequency in SubIMDb-PT
4. Log of number of subtitles that contain the word in SUBTLEX-PT
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro
6. Log of Frequency in the Written Language part of Corpus Brasileiro
7. Log of Frequency in a corpus of Mixed Text Genres
8. Lexical databases from 6 school dictionaries
9. Word's raw embedding values of Skip-Gram
10. Word's raw embedding values of GloVe

Posters: The Solutions

Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:

- 1. **word imageability** the ease and speed with which a word evokes a mental image;
- 2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
- 3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
- 4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

Posters: The Solutions

Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:

- 1. **word imageability** the ease **and speed with** which a word evokes a mental image;
- 2. **concreteness** the degree to **which words refer to objects, people, places, or things** that can be experienced by the senses;
- 3. **subjective frequency** **the estimation of** the number of times a word is encountered by individuals in its written or spoken form;
- 4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

Posters: The Solutions

We predict 4 psycholinguistic properties for Portuguese:

- **Imageability:** Ease with which a word evokes a mental image.
- **Concreteness:** Degree to which words refer to things that can be experienced by the senses.
- **Familiarity:** The number of times a word is found by individuals in its written or spoken form.
- **Age of Acquisition:** The estimate of the age at which a word was learned.

Posters: The Solutions

Challenges:

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

Posters: The Solutions

Challenges:

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

Posters: The Solutions

Challenges:

1. Manually produced properties for Portuguese are very scarce
2. Previous approaches use expensive, unavailable resources

Posters: The Solutions

Challenges:

1. **Manually produced properties for Portuguese are very scarce**
2. **Previous approaches use expensive, unavailable resources**

Posters: The Solutions

Used in various NLP tasks:

- lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Posters: The Solutions

Used in various NLP tasks:

- **lexical simplification**; **text simplification at the sentence level**; **to predict the reading times of each word in a sentence; to create robust text level readability models.**

Posters: The Solutions

Various applications:

- Lexical simplification
- Sentence simplification
- Reading time prediction
- Readability models

Posters: The Solutions

- Summarize
- **<u>Make it visual</u>**

Posters: The Solutions

What can I make visual?

Posters: The Solutions

What can I make visual?

(and how do I do it?)

List

Posters: The Solutions

List → Table

Posters: The Solutions

Alpha scores of 0.921 for imageability and 0.820 for concreteness

- similar to the values reported in literature;

Posters: The Solutions

Alpha inter-annotator agreement scores:

Imageability	Concreteness
0.921	0.820

Posters: The Solutions

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10)

Posters: The Solutions

Feature groups:

Lexical	Skip-Gram	GloVe
1-8	9	10

Table

Posters: The Solutions

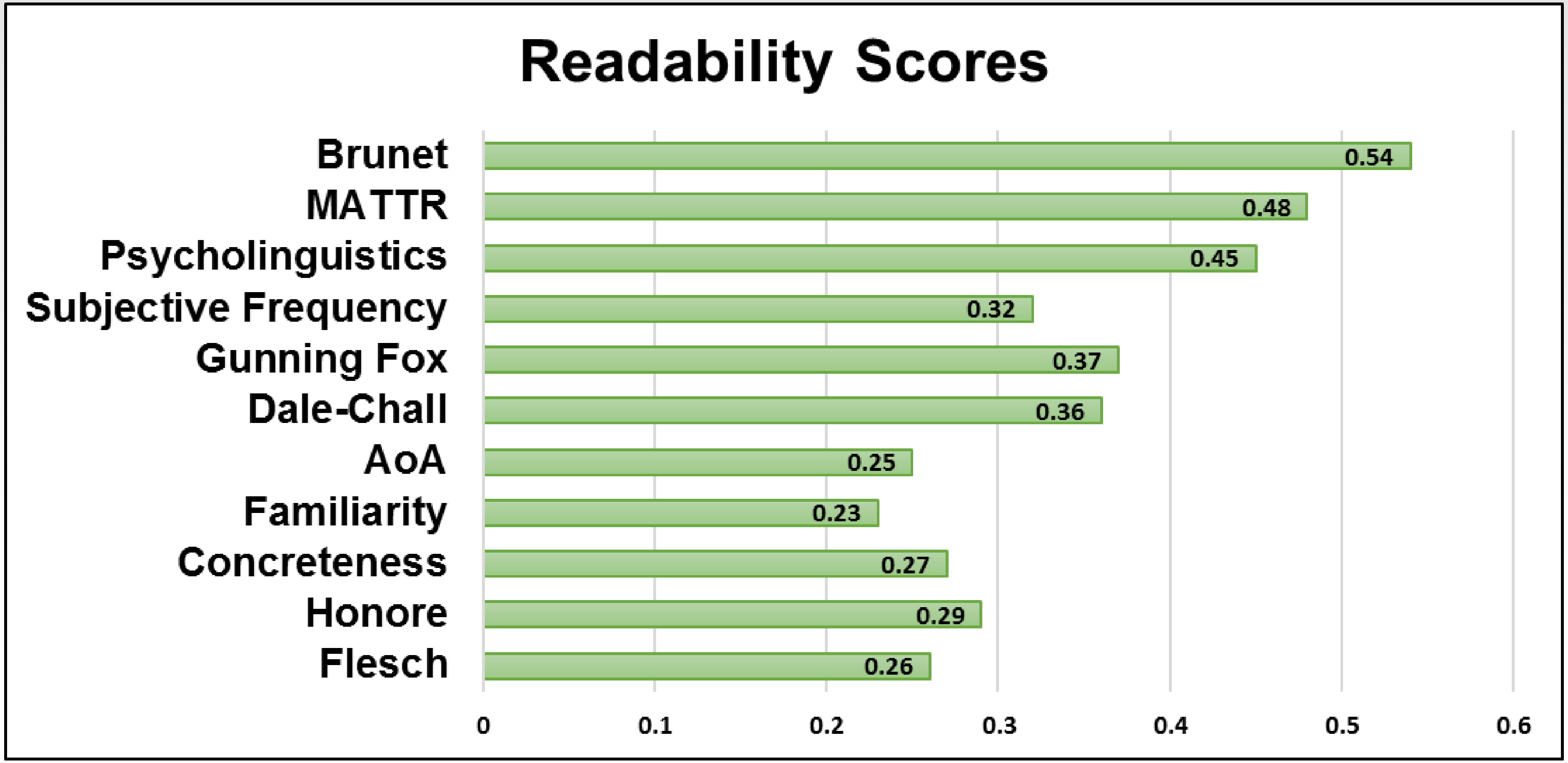
Table → Graph

Posters: The Solutions

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale-Chall	Gunning Fox	Subjective Frequency	Psycholin-guistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Posters: The Solutions

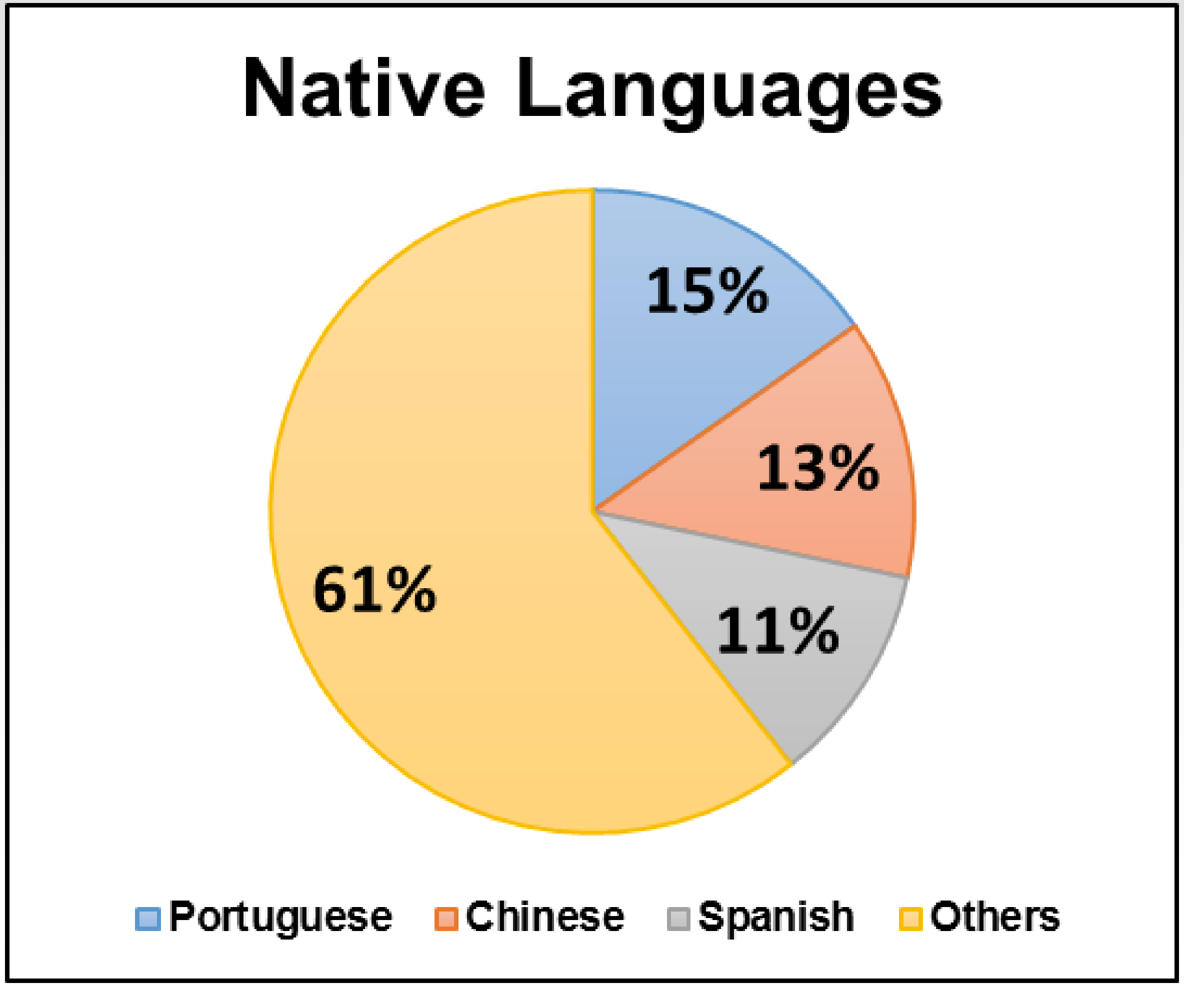


Posters: The Solutions

Native Languages:

Portuguese	Chinese	Spanish	Others
15.2%	13.1%	11.1%	60.6%

Posters: The Solutions



Posters: The Solutions

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Posters: The Solutions

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Posters: The Solutions

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Posters: The Solutions

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

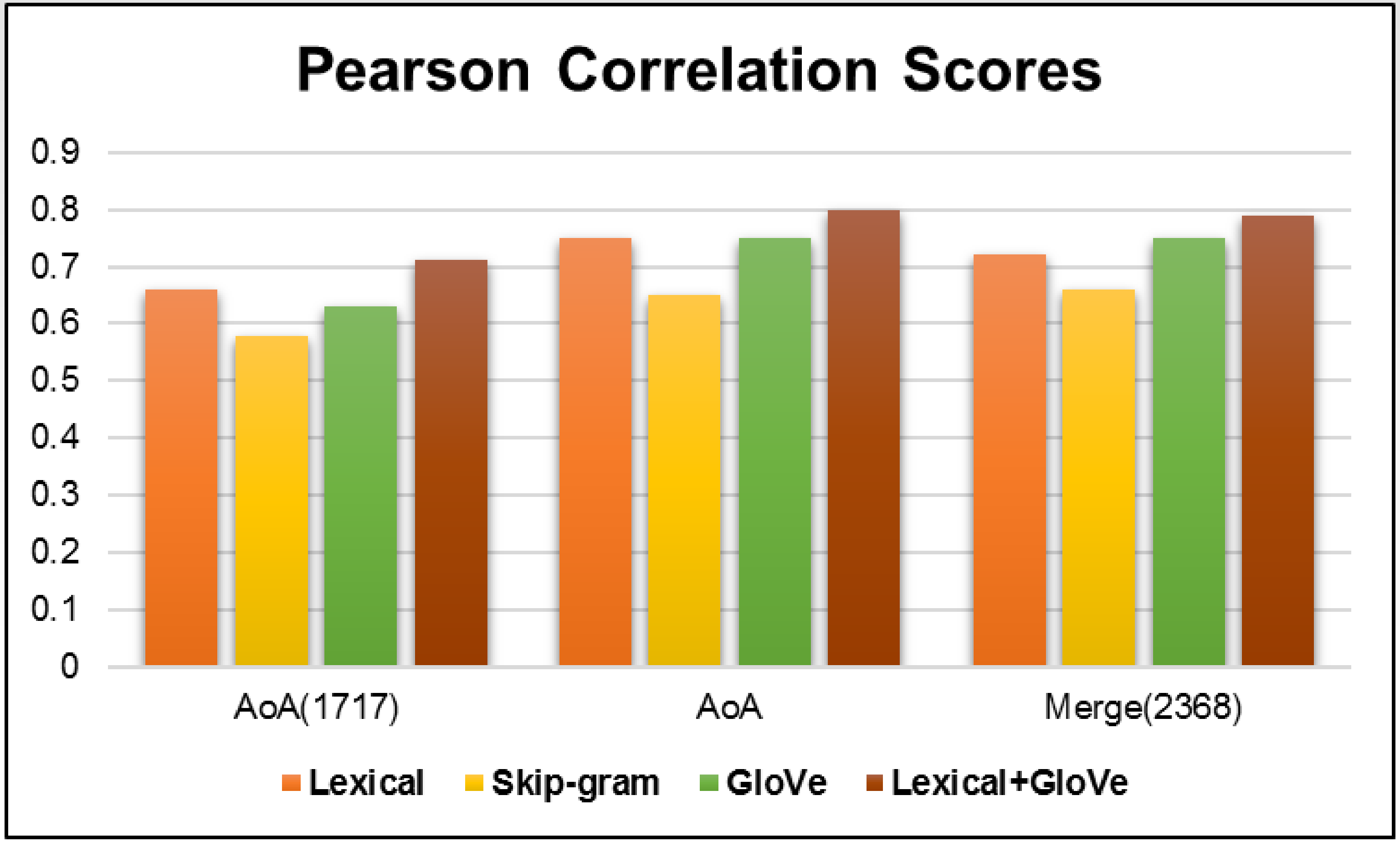
Posters: The Solutions

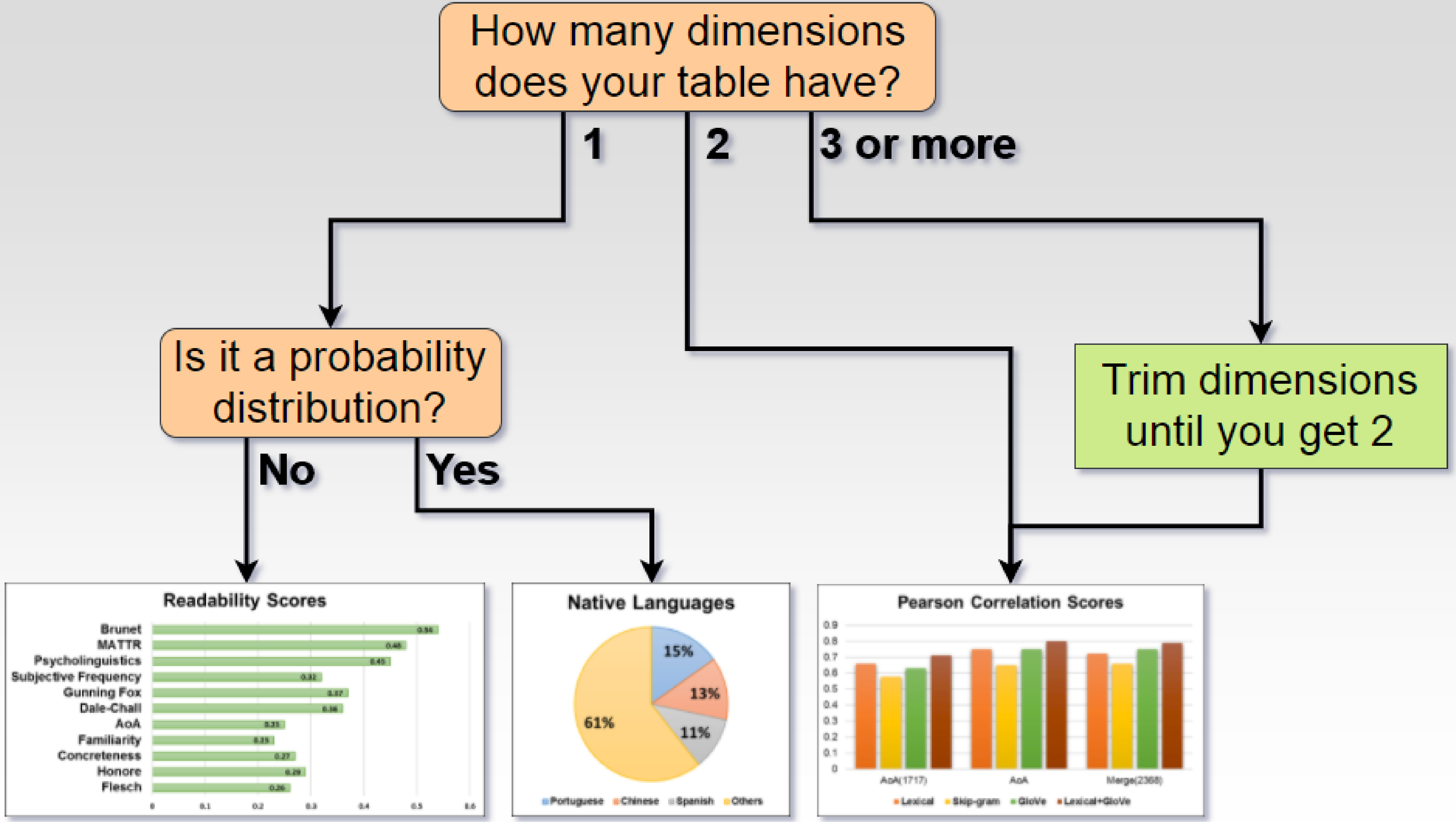
Regressors	AoA (765)		
	MSE	r	ρ
Lexical	0.91	0.67	0.66
Skip-gram	1.30	0.56	0.58
GloVe	1.18	0.62	0.63
Lexical + GloVe	**0.80**	**0.72**	**0.71**

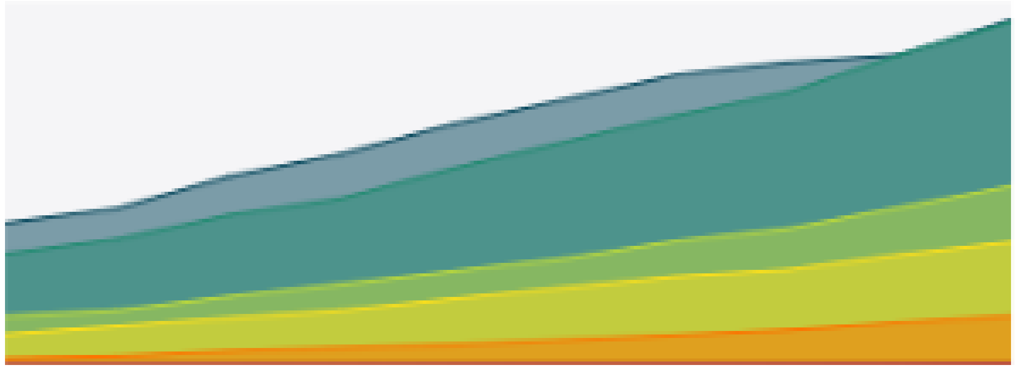
Posters: The Solutions

Regressors	AoA (765)	AoA (1717)	AoA Merge (2368)
	ρ	ρ	ρ
Lexical	0.66	0.75	0.72
Skip-gram	0.58	0.65	0.66
GloVe	0.63	0.75	0.75
Lexical + GloVe	**0.71**	**0.80**	**0.79**

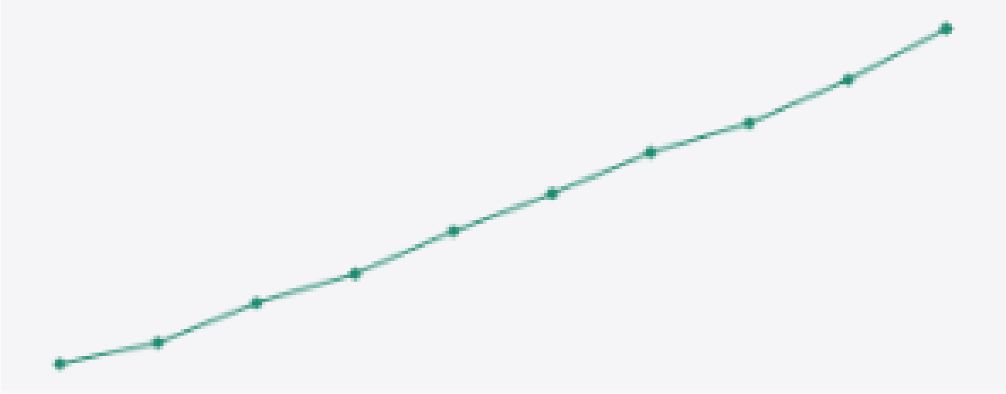
Posters: The Solutions



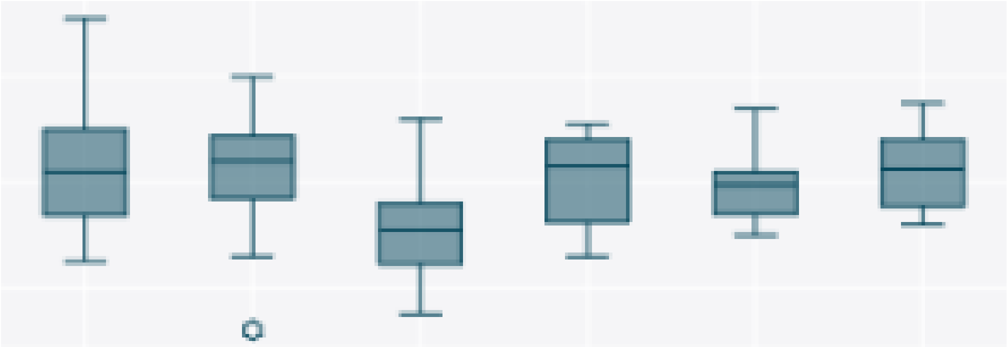

How many dimensions does your table have?
1
2
3 or more
Is it a probability distribution?
Trim dimensions until you get 2
No
Yes
Readability Scores
Brunet
MATTR
Psycholinguistics
Subjective Frequency
Gunning Fox
Dale-Chall
AoA
Familiarity
Concreteness
Honore
Flesch
Native Languages
15%
13%
11%
61%
Portuguese
Chinese
Spanish
Others
Pearson Correlation Scores
AoA
Lexical
Skip-gram
GloVe
Lexical+GloVe



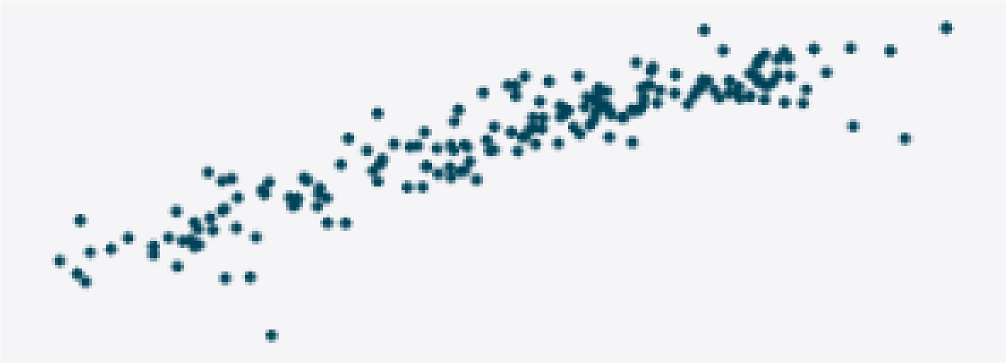
Area Chart



Line Chart



Box Plot



Scatter Plot

https://plotlyblog.tumblr.com/post/118355223592/how-to-analyze-data-eight-useful-ways-you-can

How To Analyze Data: Eight Useful Ways You Can Make Graphs

Visualizing data makes it easier to understand, analyze, and communicate. How can you decide which of the many available chart types is best suited for your data? Use this guide to get familiar with some common graph types and how they are used. We made these graphs with our free online tool; contact us to use Plotly Enterprise on-premise.

Posters: The Solutions

And what about **equations**?

Posters: The Solutions

Example:

Gated Recurrent Units

Posters: The Solutions

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Posters: The Solutions

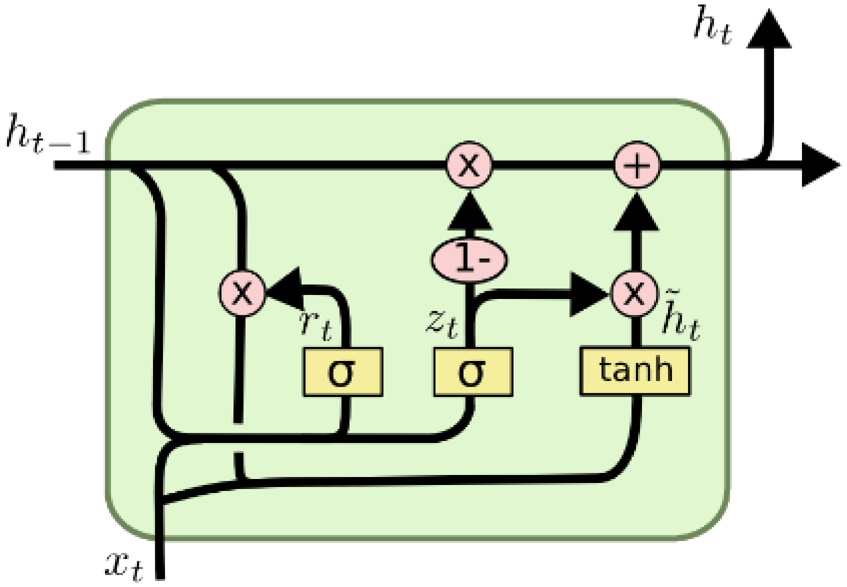
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Posters: The Solutions



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

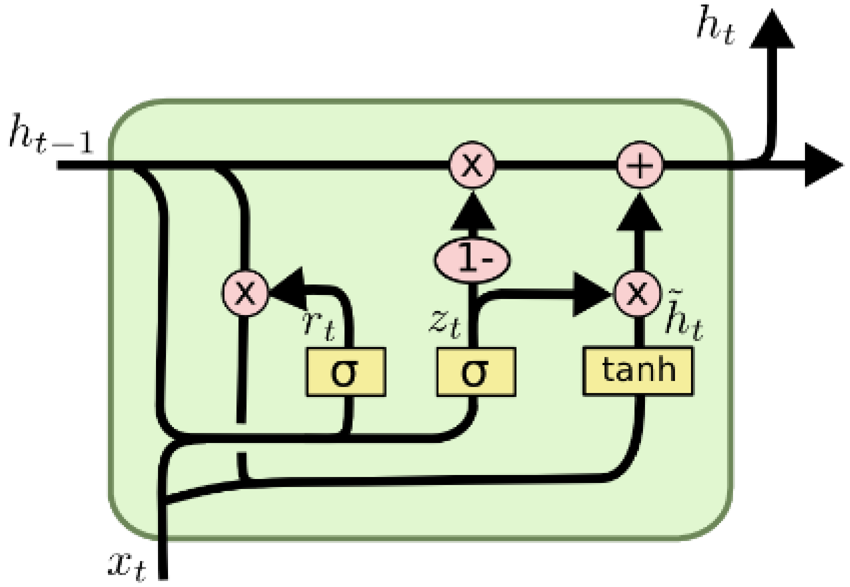
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Christopher Olah

A wandering machine learning researcher, bouncing between groups. I want to understand things clearly, and explain them well.

Academic CV - Github - Twitter - Old Blog

Posters: The Solutions



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Posters: The Solutions

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Posters: The Solutions

Update gate

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

Reset gate

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

Memory

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

Output

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Posters: The Solutions

Update gate

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

Reset gate

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

Memory

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

Output

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Posters: The Solutions

Output:

$$h_t = (1 - z_t) * h_{t-1} + \boxed{z_t} * \boxed{\tilde{h}_t}$$

Update gate:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

Memory:

$$\tilde{h}_t = \tanh\left(W \cdot [\boxed{r_t} * h_{t-1}, x_t]\right)$$

Reset gate:

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

Posters: The Solutions

What else can be made **visual**?

Posters: The Solutions

Neural architectures:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Posters: The Solutions

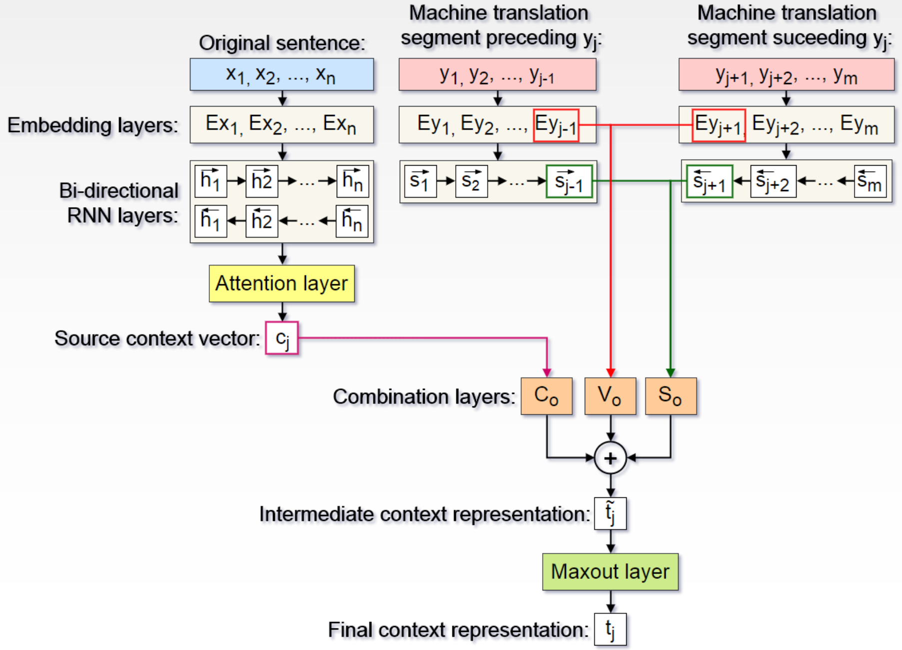
Neural architectures:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$



Posters: The Solutions

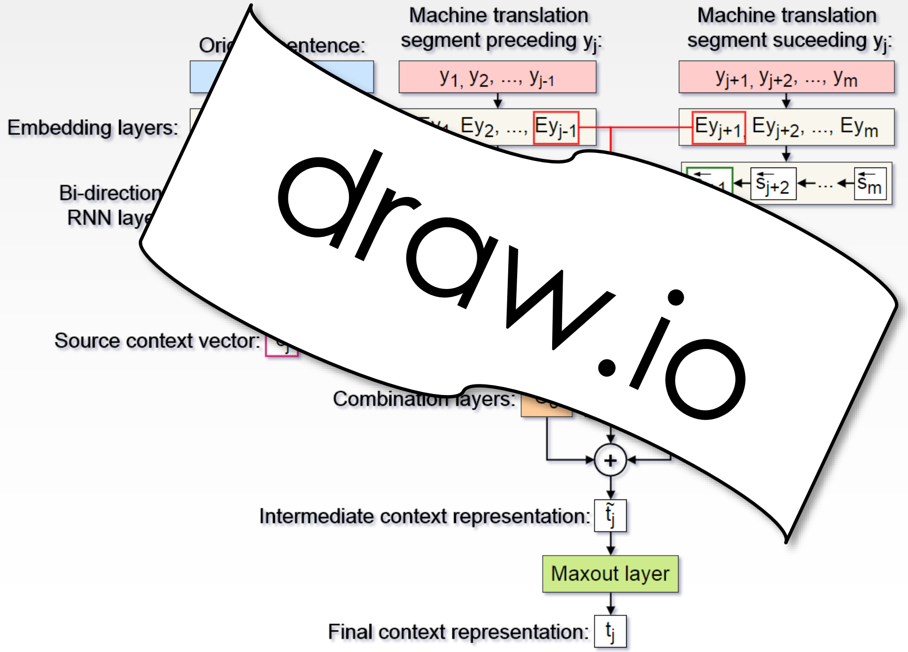
Neural architectures:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$



Posters: The Solutions

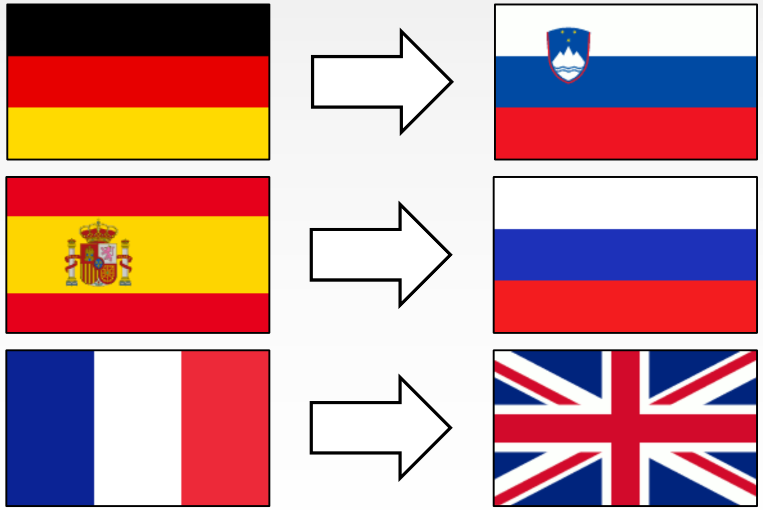
Languages/Countries:

- German-Slovenian
- Spanish-Russian
- French-English

Posters: The Solutions

Languages/Countries:

- German-Slovenian
- Spanish-Russian
- French-English



Posters: The Solutions

Task definitions:

Named-entity recognition (NER)
(also known as entity identification,
entity chunking and entity extraction)
is a subtask of information extraction
that seeks to locate and classify
named entities in text into pre-defined
categories such as the names of persons,
organizations, locations, expressions of
times, quantities, monetary values,
percentages, etc.

Posters: The Solutions

Task definitions:

Named-entity recognition (NER) (also known as entity identification, entity chunking and entity extraction) is a subtask of information extraction that seeks to locate and classify named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

Obama addressed the **FDA** questions in the city of **Denver – CA** last week.

Posters: The Solutions

Task definitions:

In natural language processing, **word sense disambiguation** (WSD) is the problem of determining which "sense" (meaning) of a word is activated by the use of the word in a particular context, a process which appears to be largely unconscious in people. WSD is a natural classification problem: Given a word and its possible senses, as defined by a dictionary, classify an occurrence of the word in context into one or more of its sense classes.

Posters: The Solutions

Task definitions:

In natural language processing, **word sense disambiguation** (WSD) is the problem of determining which "sense" (meaning) of a word is activated by the use of the word in a particular context, a process which appears to be largely unconscious in people. WSD is a natural classification problem: Given a word and its possible senses, as defined by a dictionary, classify an occurrence of the word in context into one or more of its sense classes.

I ate a **roll** yesterday.

Posters: The Solutions

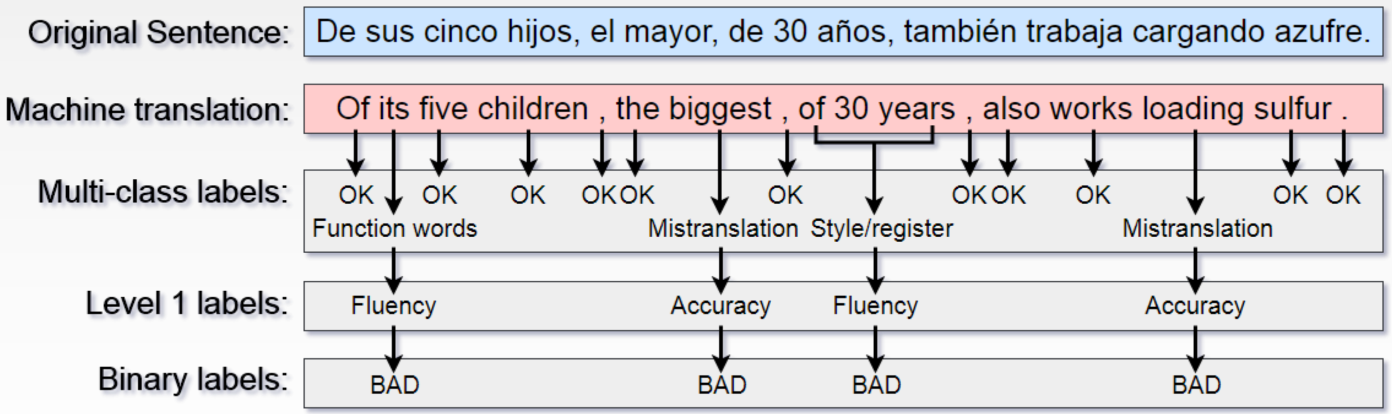
Input/Output examples:

Input: Translation
Output: Quality labels

Posters: The Solutions

Input/Output examples:

Input: Translation
Output: Quality labels



Posters: The Solutions

Tool/resource names:

Anita

Posters: The Solutions

Tool/resource names:

Anita → Anita

Posters: The Solutions

Institutions:

University of Sheffield

Posters: The Solutions

Institutions:

University of Sheffield

Posters: The Solutions

People:

Gustavo H. Paetzold

Posters: The Solutions

People:

Gustavo H. Paetzold

Posters: The Solutions

5. Bland **styling**

PosterPresentations.com
THE RESEARCH POSTER PRINTING SPECIALISTS

Friday, October 13, 2017

Home | Free Poster Templates | Price Guide | Design | Help Desk | Testimonials | Contact | Order your poster

Free Research Poster PowerPoint Templates

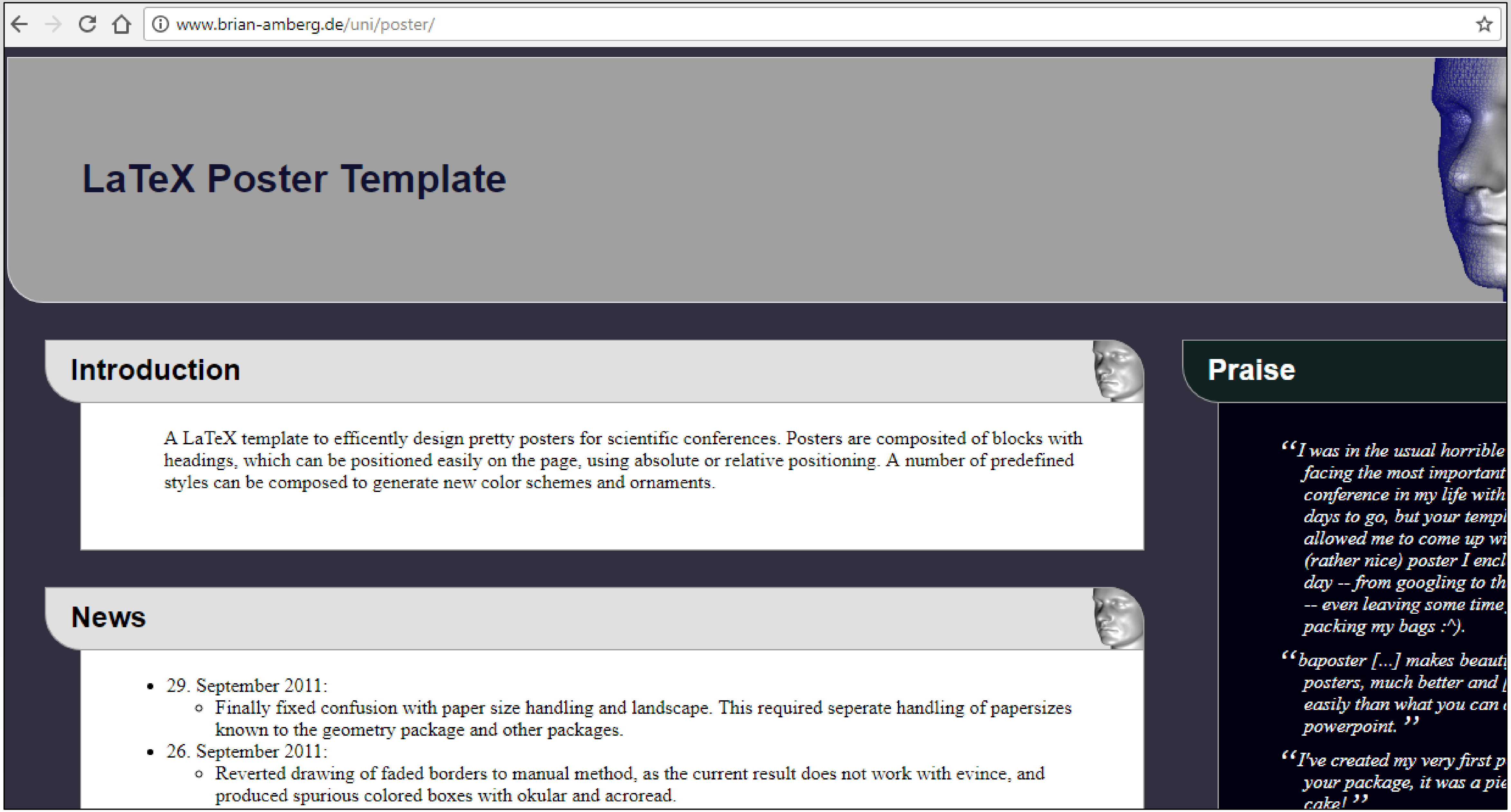
.PPT
.PPTX

Easy to use and customize

- Change colors with one click.
- Most standard US and international poster sizes.
- Support for all PowerPoint versions.
- Only basic PowerPoint skills required.
- Fully customizable.
- Instructions included with the poster templates.
- Online video tutorials.
- Configured to print professionally.
- Additional layouts included in each template.

QuickFind Research poster templates

USA sizes			International	
36x48	42x60	48x48	91x122	A0
36x56	42x72	48x72	70x100	A1
36x60	42x90	48x96	100x140	A3
36x72	44x44	40x30	100x100	-
36x96	30x40	Trifold	100x200	-


www.brian-amberg.de/uni/poster/
LaTeX Poster Template
Introduction
A LaTeX template to efficently design pretty posters for scientific conferences. Posters are composited of blocks with headings, which can be positioned easily on the page, using absolute or relative positioning. A number of predefined styles can be composed to generate new color schemes and ornaments.
News
29. September 2011:
Finally fixed confusion with paper size handling and landscape. This required seperate handling of papersizes known to the geometry package and other packages.
26. September 2011:
Reverted drawing of faded borders to manual method, as the current result does not work with evince, and produced spurious colored boxes with okular and acroread.
Praise

www.brian-amberg.de/uni/poster/
LaTeX Poster Template
BAPOSTER
Introduction
A LaTeX template to efficently design pretty posters for scientific conferences. Posters are composited of blocks with headings, which can be positioned easily on the page, using absolute or relative positioning. A number of predefined styles can be composed to generate new color schemes and ornaments.
News
• 29. September 2011:
◦ Finally fixed confusion with paper size handling and landscape. This required seperate handling of papersizes known to the geometry package and other packages.
• 26. September 2011:
◦ Reverted drawing of faded borders to manual method, as the current result does not work with evince, and produced spurious colored boxes with okular and acroread.
Praise
packing my bags :^).
powerpoint.

Collecting and Exploring Everyday Language for Predicting Psycholinguistic Properties of Words

Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield

SubIMDB

A structured corpus of subtitles that captures everyday language.

http://ghpaetzold.github.io/**SubIMDB**

H2020 Project Reference: **692819**

Building SubIMDB

SubIMDB is a **corpus of everyday language** with subtitles of **movies and series** for **family and children**. To build it, we first:

1. **Gathered** 12,618 IMDb identifiers.
2. **Searched** OpenSubtitles for subtitles.
3. **Downloaded** one subtitle for each movie, and one for each episode of a series.

We then **pre-processed** all subtitles by **discarding** any lines which:

- Contain **advertisement**.
- Have more than **80 characters**.
- Have a **long word** (15 characters).
- Refer to **metadata** or **timing**.

The resulting corpus has **225,847,810** words from **38,102** subtitles.

SubIMDB Subsets

- **SubIMDB**: All subtitles
- **SubMOV**: All movies
- **SubSER**: All series
- **SubFAM**: Family subtitles
- **SubCOM**: Comedy subtitles
- **SubCHI**: Children subtitles
- **SubFAM-M**: Family Movies
- **SubFAM-S**: Family Series
- **SubCOM-M**: Comedy Movies
- **SubCOM-S**: Comedy Series
- **SubCHI-M**: Children Movies
- **SubCHI-S**: Children Series

Lexical Decision Times

Norm	Size	ρ	r	F-test
KF	1M	−0.517	−0.486	•••
HAL	131M	−0.641	−0.616	•••
Wiki	97M	−0.531	−0.506	•••
SimpleWiki	9M	−0.560	−0.530	•••
SUBTLEX	62M	−0.653	−0.619	•••
Open2016	2B	−0.657	−0.602	•••
SubIMDB	225M	**−0.659**	−0.624	-
SubMOV	125M	−0.657	**−0.626**	•••
SubSER	100M	−0.652	−0.620	•••
SubFAM	34M	−0.649	−0.614	•••
SubCOM	199M	−0.657	−0.624	•••
SubCHI	17M	−0.634	−0.592	•••
SubFAM-M	17M	−0.640	−0.596	•••
SubFAM-S	17M	−0.632	−0.590	•••
SubCOM-M	107M	−0.655	−0.623	•••
SubCOM-S	91M	−0.651	−0.618	•••
SubCHI-M	8M	−0.625	−0.572	•••
SubCHI-S	8M	−0.606	−0.556	•••

$p<0.1$ (•), $p<0.01$ (••), $p<0.001$ (•••)

Psycholinguistic Features

	Age of Acquisition		Familiarity	
	r	F-test	r	F-test
KF	−0.447	•••	0.669	•••
HAL	−0.511	•••	0.732	•••
Wiki	−0.412	•••	0.676	•••
SimpleWiki	−0.486	•••	0.667	•••
SUBTLEX	−0.676	•••	0.774	•••
Open2016	−0.666	•••	**0.799**	•••
SubIMDB	−0.698	-	0.781	-
SubMOV	−0.705	•••	0.777	•••
SubSER	−0.687	•••	0.777	•••
SubFAM	−0.723	•••	0.758	•••
SubCOM	−0.696	••	0.781	•••
SubCHI	−0.709	•••	0.735	•••
SubFAM-M	**−0.746**	•••	0.742	•••
SubFAM-S	−0.685	•••	0.743	•••
SubCOM-M	−0.698	•••	0.777	•••
SubCOM-S	−0.690	•••	0.777	•••
SubCHI-M	−0.728	•••	0.723	•••
SubCHI-S	−0.670	•••	0.704	•••

Simplicity: Frequency

Norm	r	ρ	TRank	F-test
KF	0.619	0.626	0.589	•••
HAL	0.630	0.633	0.598	•••
Wiki	0.575	0.583	0.516	•••
SimpleWiki	0.626	0.632	0.570	•••
SUBTLEX	0.649	0.649	0.619	•••
Open2016	0.650	0.647	0.619	•••
SubIMDB	0.654	0.652	0.622	-
SubMOV	**0.660**	**0.658**	**0.623**	•••
SubSER	0.648	0.647	0.619	•••
SubFAM	0.649	0.650	0.615	•••
SubCOM	0.655	0.653	**0.623**	•
SubCHI	0.643	0.645	0.611	•••
SubFAM-M	0.653	0.653	0.618	•••
SubFAM-S	0.647	0.650	0.620	•••
SubCOM-M	**0.660**	**0.658**	**0.623**	•••
SubCOM-S	0.647	0.648	0.618	•••
SubCHI-M	0.650	0.654	0.600	•••
SubCHI-S	0.640	0.644	0.608	•••
Google 1T	N/A	N/A	0.585	-
Best SemEval	N/A	N/A	0.602	-

Simplicity: N-grams

	3-grams		5-grams	
Norm	TRank	F-Test	TRank	F-Test
KF	0.234	•••	0.234	•••
Wiki	0.388	○	0.257	○
SimpleWiki	0.354	•••	0.247	•••
SUBTLEX	0.402	•••	0.261	•
Open2016	**0.461**	•••	0.234	•••
SubIMDB	0.425	-	**0.264**	-
SubMOV	0.401	••	0.262	○
SubSER	0.399	•••	0.254	•
SubFAM	0.379	•••	0.251	••
SubCOM	0.416	○	0.261	○
SubCHI	0.354	•••	0.246	•••
SubFAM-M	0.357	•••	0.248	•••
SubFAM-S	0.364	•••	0.246	•••
SubCOM-M	0.398	•••	0.259	•
SubCOM-S	0.396	•••	0.253	•
SubCHI-M	0.329	•••	0.242	•••
SubCHI-S	0.334	•••	0.243	•••

Anita

An intelligent text adaptation tool.

http://www.simpatico-project.eu

H2020 Project Reference: **692819**

Introduction

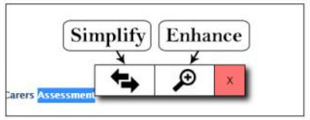
Anita is a Google Chrome extension that provides with **intelligent text adaptation solutions** that customize the content of webpages based on the **user's profile**.

User Profile

To adapt to a user's needs, Anita initially requests for some **personal information**, such as illustrated below:

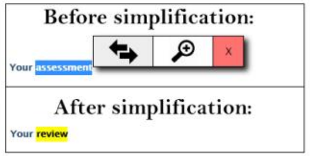
Text Adaptation

Once a user's profile is collected, they can select **words** in **any given webpage** and **request** for two types of adaptation: Simplification and Enhancement. The adaptation interface of Anita is illustrated below:

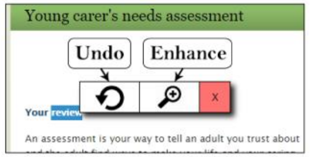


Simplification

The **Simplification** module of Anita attempts to **replace** the selected word with a **simpler** alternative. Upon request, Anita employs a **supervised lexical simplifier** to **replace** and **highlight** the selected word.



If the simplification **does not help the user**, it can **be reversed**. In this case, Anita **feeds the simplification data** back to the simplifier so that it can **adapt** to the user's needs.



Anita's simplifier exploits **spoken text corpora, context-aware word embeddings** and **supervised ranking models**.

Enhancement

The **Enhancement** module of Anita allows the user to **learn more** about words.

Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield

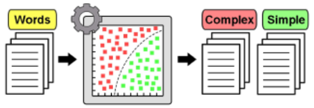
Posters: The Solutions

6. Poor **structuring**

Understanding the Lexical Simplification Needs of Non-Native Speakers of English

Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield

Complex Word Identification



Data

- **9,200** sentences
- **20-40** words each
- **269** from LexMTurk
- **231** from the CW corpus
- **8,700** from Simple Wikipedia

Annotators

- **£50** raffle compensation
- **400** non-native speakers
- **40** sentences per form
- **9,000** with only 1 answer
- **200** sentences with 20

Annotation

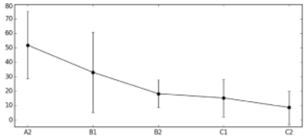
"*For each sentence, mark all the words you* ***do not understand****, even if you* ***understand the sentence as a whole****. If you understand all of them, just select* ***'I understand all words!'***."

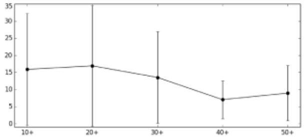
Agreement

- **0.616** between all annotators
- **0.575** within proficiencies
- **0.638** within educations
- **0.671** within age bands
- **0.718** within languages

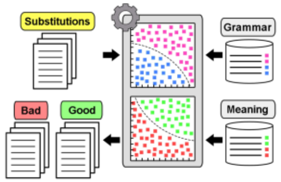
Findings

Feature	Complex	Simple	p
Length	7.1 ± 2	6.1 ± 2	○
Syllables	2.2 ± 1	1.7 ± 1	○
Senses	1.1 ± 1	8.8 ± 9	•
Synonyms	2.3 ± 3	22.7 ± 22	•
Hypernyms	0.9 ± 1	5.9 ± 7	•
Hyponyms	0.8 ± 2	32.8 ± 52	•
Subimdb[0,0]	-6.6 ± 1	-4.5 ± 1	•
Subtlex[0,0]	-51.3 ± 46	-4.4 ± 1	•
Simple[0,0]	-8.4 ± 14	-4.2 ± 1	○
Subimdb[1,1]	-13.2 ± 3	-9.7 ± 3	•
Subtlex[1,1]	-59.7 ± 52	-13.8 ± 21	•
Simple[1,1]	-10.7 ± 15	-8.1 ± 2	○





Substitution Selection



Data

- **1,471** complex words
- **10** replacements for each
- **1-3** sentences with each
- **2,554** total sentences
- **25,540** total instances

Annotators

- **£50** raffle compensation
- **400** fluent speakers
- **80** instances per form
- **23,940** with only 1 answer
- **1,600** sentences with 5

Annotation

"*Judge the following candidate substitutions of complex words with respect to their* ***grammaticality*** *and* ***meaning preservation****. When judging, please* ***ignore any grammatical errors*** *that are* ***not*** *caused by the substitution.*"

Agreement

- **0.424** for meaning
- **0.391** for grammaticality
- **0.450** for both jointly

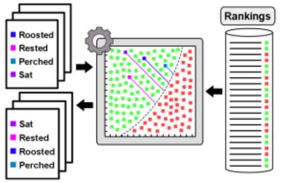
Findings

	Grammaticality		
Feature	**Good**	**Bad**	p
Prob. Subimdb	-0.9 ± 0.3	-1.0 ± 0.3	○
Prob. Subtlex	-3.1 ± 1.3	-3.2 ± 1.7	•
Prob. Simple	-4.2 ± 1.6	-4.3 ± 1.9	•
Target Sim.	0.41 ± 0.2	0.29 ± 0.2	•
Context Sim.	0.08 ± 0.1	0.06 ± 0.1	○
POS Prob.	0.62 ± 0.4	0.44 ± 0.4	•

	Meaning		
Feature	**Good**	**Bad**	p
Prob. Subimdb	-1.0 ± 0.3	-0.9 ± 0.3	○
Prob. Subtlex	-3.2 ± 1.4	-3.2 ± 1.7	•
Prob. Simple	-4.2 ± 1.8	-4.3 ± 1.9	•
Target Sim.	0.39 ± 0.2	0.28 ± 0.2	•
Context Sim.	0.08 ± 0.1	0.06 ± 0.1	○
POS Prob.	0.53 ± 0.4	0.46 ± 0.4	○

	Joint (G/M)		
Feature	**Good**	**Bad**	p
Prob. Subimdb	-0.9 ± 0.2	-1.0 ± 0.3	•
Prob. Subtlex	-3.1 ± 1.5	-3.4 ± 1.8	○
Prob. Simple	-4.2 ± 1.7	-4.4 ± 2.0	○
Target Sim.	0.34 ± 0.2	0.27 ± 0.2	•
Context Sim.	0.07 ± 0.1	0.06 ± 0.1	•
POS Prob.	0.58 ± 0.4	0.32 ± 0.4	•

Substitution Ranking



Data

- **901** sentences with a gap
- **All** with a target word
- **2-4** pool of good candidates
- Target **added to pool**
- **4,200** total candidate pairs

Annotators

- **£50** raffle compensation
- **300** fluent speakers
- **70** instances per form
- **All** with 5 annotations
- **21,000** total annotations

Annotation

"*For each of the following instances, select* ***which candidate*** *makes the sentence* ***easier to understand****. If the words are* ***equally complex/simple****, select the "****The words are equally simple****" option. Please* ***overlook*** *any grammatical or spelling errors.*"

Agreement

- **0.454** between all annotators
- **0.486** within proficiencies
- **0.468** within educations
- **0.482** within age bands
- **0.601** within languages

Findings

Feature	r	ρ	TRank
Length	0.172	0.179	0.386
Syllables	0.097	0.095	0.340
Senses	−0.345	−0.349	0.505
Synonyms	−0.288	−0.297	0.454
Hypernyms	−0.289	−0.297	0.472
Hyponyms	−0.309	−0.300	0.453
Subimdb[0,0]	−0.419	−0.436	0.539
Subtlex[0,0]	−0.465	−0.467	0.556
Simple[0,0]	−0.490	−0.468	0.578
Subimdb[1,1]	−0.463	−0.473	0.579
Subtlex[1,1]	−0.496	−0.496	0.590
Simple[1,1]	**−0.501**	**−0.475**	**0.593**

Download

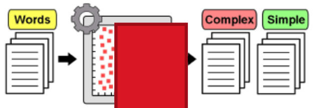
To find this data (**and much more**), visit:
http://gustavopaetzold.wordpress.com

H2020 Project Reference: **692819**

Understanding the Lexical Simplification Needs of Non-Native Speakers of English

Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield

Complex Word Identification



Data

- ences
- ds each
- LexMTurk
- he CW corpus
- Simple Wikipedia

Annotators

- compensation
- ative speakers
- es per form
- only 1 answer
- ces with 20

Annotation

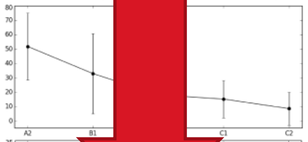
"*For each senten … the words you* ***do not understand*** *… u* ***understand the sentence as a*** *… u understand all of them, just select … and all words!*"

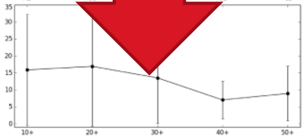
Agreement

- veen all annotators
- in proficiencies
- in educations
- in age bands
- in languages

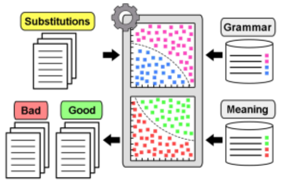
Findings

Feature		Simple	p
Length		6.1 ± 2	○
Syllables		1.7 ± 1	○
Senses		8.8 ± 9	•
Synonyms		22.7 ± 22	•
Hypernyms		5.9 ± 7	•
Hyponyms		32.8 ± 52	•
Subimdb[0,0]		-4.5 ± 1	•
Subtlex[0,0]		-4.4 ± 1	•
Simple[0,0]		-4.2 ± 1	○
Subimdb[1,1]		-9.7 ± 3	•
Subtlex[1,1]		-13.8 ± 21	•
Simple[1,1]		-8.1 ± 2	○





Substitution Selection



Data

- **1,471** complex words
- **10** replacements for each
- **1-3** sentences with each
- **2,554** total sentences
- **25,540** total instances

Annotators

- **£50** raffle compensation
- **400** fluent speakers
- **80** instances per form
- **23,940** with only 1 answer
- **1,600** sentences with 5

Annotation

"*Judge the following candidate substitutions of complex words with respect to their* ***grammaticality*** *and* ***meaning preservation.*** *When judging, please* ***ignore any grammatical errors*** *that are* ***not*** *caused by the substitution.*"

Agreement

- **0.424** for meaning
- **0.391** for grammaticality
- **0.450** for both jointly

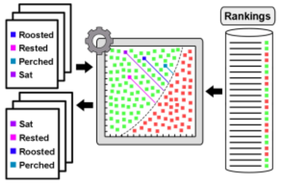
Findings

Feature	Grammaticality Good	Grammaticality Bad	p
Prob. Subimdb	-0.9 ± 0.3	-1.0 ± 0.3	○
Prob. Subtlex	-3.1 ± 1.3	-3.2 ± 1.7	•
Prob. Simple	-4.2 ± 1.6	-4.3 ± 1.9	•
Target Sim.	0.41 ± 0.2	0.29 ± 0.2	•
Context Sim.	0.08 ± 0.1	0.06 ± 0.1	○
POS Prob.	0.62 ± 0.4	0.44 ± 0.4	•

Feature	Meaning Good	Meaning Bad	p
Prob. Subimdb	-1.0 ± 0.3	-0.9 ± 0.3	○
Prob. Subtlex	-3.2 ± 1.4	-3.2 ± 1.7	•
Prob. Simple	-4.2 ± 1.8	-4.3 ± 1.9	•
Target Sim.	0.39 ± 0.2	0.28 ± 0.2	•
Context Sim.	0.08 ± 0.1	0.06 ± 0.1	○
POS Prob.	0.53 ± 0.4	0.46 ± 0.4	○

Feature	Joint (G/M) Good	Joint (G/M) Bad	p
Prob. Subimdb	-0.9 ± 0.2	-1.0 ± 0.3	•
Prob. Subtlex	-3.1 ± 1.5	-3.4 ± 1.8	○
Prob. Simple	-4.2 ± 1.7	-4.4 ± 2.0	○
Target Sim.	0.34 ± 0.2	0.27 ± 0.2	•
Context Sim.	0.07 ± 0.1	0.06 ± 0.1	•
POS Prob.	0.58 ± 0.4	0.32 ± 0.4	•

Substitution Ranking



Data

- **901** sentences with a gap
- **All** with a target word
- **2-4** pool of good candidates
- Target **added to pool**
- **4,200** total candidate pairs

Annotators

- **£50** raffle compensation
- **300** fluent speakers
- **70** instances per form
- **All** with 5 annotations
- **21,000** total annotations

Annotation

"*For each of the following instances, select* ***which candidate*** *makes the sentence* ***easier to understand.*** *If the words are* ***equally complex/simple****, select the "****The words are equally simple****" option. Please* ***overlook*** *any grammatical or spelling errors.*"

Agreement

- **0.454** between all annotators
- **0.486** within proficiencies
- **0.468** within educations
- **0.482** within age bands
- **0.601** within languages

Findings

Feature	r	ρ	TRank
Length	0.172	0.179	0.386
Syllables	0.097	0.095	0.340
Senses	−0.345	−0.349	0.505
Synonyms	−0.288	−0.297	0.454
Hypernyms	−0.289	−0.297	0.472
Hyponyms	−0.309	−0.300	0.453
Subimdb[0,0]	−0.419	−0.436	0.539
Subtlex[0,0]	−0.465	−0.467	0.556
Simple[0,0]	−0.490	−0.468	0.578
Subimdb[1,1]	−0.463	−0.473	0.579
Subtlex[1,1]	−0.496	−0.496	0.590
Simple[1,1]	**−0.501**	**−0.475**	**0.593**

Download

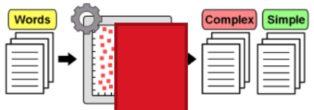
To find this data (**and much more**), visit:
http://gustavopaetzold.wordpress.com

H2020 Project Reference: **692819**

UNDERSTANDING THE LEXICAL SIMPLIFICATION NEEDS OF NON-NATIVE SPEAKERS OF ENGLISH

Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield

Complex Word Identification



Data

- ences
- ds each
- exMTurk
- he CW corpus
- Simple Wikipedia

Annotators

- compensation
- ative speakers
- es per form
- only 1 answer
- ces with 20

Annotation

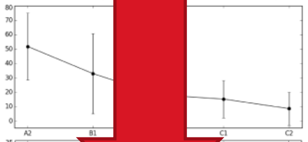
"*For each senten … l the words you do* ***not understand*** … ***u understand the sentence as a*** … *u understand all of them, just select* … ***and all words!***"

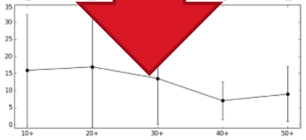
Agreement

- een all annotators
- in proficiencies
- in educations
- in age bands
- in languages

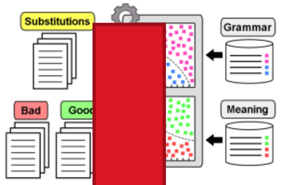
Findings

Feature		Simple	p
Length		6.1 ± 2	○
Syllables		1.7 ± 1	○
Senses		8.8 ± 9	●
Synonyms		22.7 ± 22	●
Hypernyms		5.9 ± 7	●
Hyponyms		32.8 ± 52	●
Subimdb[0,0]		-4.5 ± 1	●
Subtlex[0,0]		-4.4 ± 1	●
Simple[0,0]		-4.2 ± 1	○
Subimdb[1,1]		-9.7 ± 3	●
Subtlex[1,1]		-13.8 ± 21	●
Simple[1,1]		-8.1 ± 2	○





Substitution Selection



Data

- plex words
- ements for each
- ces with each
- l sentences
- tal instances

Annotators

- compensation
- speakers
- es per form
- th only 1 answer
- tences with 5

Annotation

"*Judge the follo … late substitutions of complex words … to their* ***grammaticality*** *and* ***me*** … ***servation.*** *When judging, please … grammatical errors that are n … the substitution.*"

Agreement

- meaning
- grammaticality
- both jointly

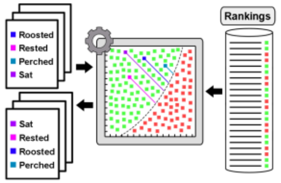
Findings

Feature		Bad	p
		nmaticality	
Prob. Subimdb		-1.0 ± 0.3	○
Prob. Subtlex		-3.2 ± 1.7	●
Prob. Simple		-4.3 ± 1.9	●
Target Sim.		0.29 ± 0.2	●
Context Sim.		0.06 ± 0.1	○
POS Prob.		0.44 ± 0.4	●

Feature		Bad	p
		leaning	
Prob. Subimdb		-0.9 ± 0.3	○
Prob. Subtlex		-3.2 ± 1.7	●
Prob. Simple		-4.3 ± 1.9	●
Target Sim.		0.28 ± 0.2	●
Context Sim.		0.06 ± 0.1	○
POS Prob.		0.46 ± 0.4	○

Feature		Bad	p
		G/M)	
Prob. Subimdb		-1.0 ± 0.3	●
Prob. Subtlex	1.5	-3.4 ± 1.8	○
Prob. Simple	-4.2 ± 1.7	-4.4 ± 2.0	○
Target Sim.	0.34 ± 0.2	0.27 ± 0.2	●
Context Sim.	0.07 ± 0.1	0.06 ± 0.1	●
POS Prob.	0.58 ± 0.4	0.32 ± 0.4	●

Substitution Ranking



Data

- **901** sentences with a gap
- **All** with a target word
- **2-4** pool of good candidates
- Target **added to pool**
- **4,200** total candidate pairs

Annotators

- **£50** raffle compensation
- **300** fluent speakers
- **70** instances per form
- **All** with 5 annotations
- **21,000** total annotations

Annotation

"*For each of the following instances, select* ***which candidate*** *makes the sentence* ***easier to understand.*** *If the words are* ***equally complex/simple,*** *select the "****The words are equally simple****" option. Please* ***overlook*** *any grammatical or spelling errors.*"

Agreement

- **0.454** between all annotators
- **0.486** within proficiencies
- **0.468** within educations
- **0.482** within age bands
- **0.601** within languages

Findings

Feature	r	ρ	TRank
Length	0.172	0.179	0.386
Syllables	0.097	0.095	0.340
Senses	−0.345	−0.349	0.505
Synonyms	−0.288	−0.297	0.454
Hypernyms	−0.289	−0.297	0.472
Hyponyms	−0.309	−0.300	0.453
Subimdb[0,0]	−0.419	−0.436	0.539
Subtlex[0,0]	−0.465	−0.467	0.556
Simple[0,0]	−0.490	−0.468	0.578
Subimdb[1,1]	−0.463	−0.473	0.579
Subtlex[1,1]	−0.496	−0.496	0.590
Simple[1,1]	**−0.501**	**−0.475**	**0.593**

Download

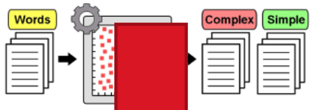
To find this data (**and much more**), visit:
http://gustavopaetzold.wordpress.com

H2020 Project Reference: **692819**

Understanding the Lexical Simplification Needs of Non-Native Speakers of English

Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield

Complex Word Identification



Data

- ences
- ls each
- exMTurk
- he CW corpus
- Simple Wikipedia

Annotators

- compensation
- tive speakers
- es per form
- only 1 answer
- ces with 20

Annotation

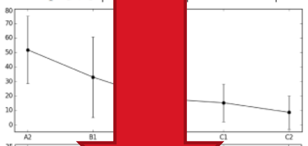
"*For each senten… l the words you* ***do not understand*** *… u* ***understand the sentence as a*** *… u understand all of them, just select …* ***and all words!***"

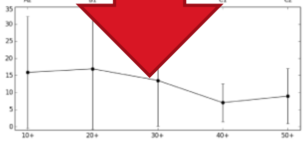
Agreement

- veen all annotators
- in proficiencies
- in educations
- in age bands
- in languages

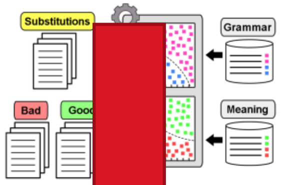
Findings

Feature	Simple	p
Length	6.1 ± 2	○
Syllables	1.7 ± 1	○
Senses	8.8 ± 9	•
Synonyms	22.7 ± 22	•
Hypernyms	5.9 ± 7	•
Hyponyms	32.8 ± 52	•
Subimdb[0,0]	-4.5 ± 1	•
Subtlex[0,0]	-4.4 ± 1	•
Simple[0,0]	-4.2 ± 1	○
Subimdb[1,1]	-9.7 ± 3	•
Subtlex[1,1]	-13.8 ± 21	•
Simple[1,1]	-8.1 ± 2	○





Substitution Selection



Data

- nplex words
- ements for each
- ces with each
- l sentences
- tal instances

Annotators

- compensation
- speakers
- es per form
- th only 1 answer
- ences with 5

Annotation

"*Judge the follo… late substitutions of complex words … to their* ***grammaticality*** *and* ***me… servation.*** *When judging, please … y* ***grammatical errors*** *that are n… the substitution.*"

Agreement

- meaning
- grammaticality
- both jointly

Findings

mmaticality

Feature	Bad	p
Prob. Subimdb	-1.0 ± 0.3	○
Prob. Subtlex	-3.2 ± 1.7	•
Prob. Simple	-4.3 ± 1.9	•
Target Sim.	0.29 ± 0.2	•
Context Sim.	0.06 ± 0.1	○
POS Prob.	0.44 ± 0.4	•

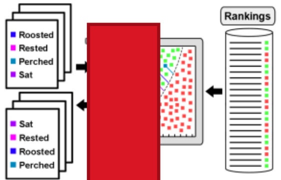
leaning

Feature	Bad	p
Prob. Subimdb	-0.9 ± 0.3	○
Prob. Subtlex	-3.2 ± 1.7	•
Prob. Simple	-4.3 ± 1.9	•
Target Sim.	0.28 ± 0.2	•
Context Sim.	0.06 ± 0.1	○
POS Prob.	0.46 ± 0.4	○

G/M)

Feature		Bad	p
Prob. Subimdb		-1.0 ± 0.3	•
Prob. Subtlex	1.5	-3.4 ± 1.8	○
Prob. Simple	-4.2 ± 1.7	-4.4 ± 2.0	○
Target Sim.	0.34 ± 0.2	0.27 ± 0.2	•
Context Sim.	0.07 ± 0.1	0.06 ± 0.1	•
POS Prob.	0.58 ± 0.4	0.32 ± 0.4	•

Substitution Ranking



Data

- nces with a gap
- a target word
- of good candidates
- **dded to pool**
- al candidate pairs

Annotators

- compensation
- t speakers
- ces per form
- 5 annotations
- otal annotations

Annotation

"*For each of … ng instances, select* ***which candid… the sentence easier to understan…*** *words are* ***equally complex/simp… he "The words are equally simpl… Please overlook*** *any grammatical or … ors.*"

Agreement

- tween all annotators
- thin proficiencies
- thin educations
- thin age bands
- thin languages

Findings

Feature	ρ	TRank
Length	0.179	0.386
Syllables	0.095	0.340
Senses	-0.349	0.505
Synonyms	-0.297	0.454
Hypernyms	-0.297	0.472
Hyponyms	-0.300	0.453
Subimdb[0,0]	-0.436	0.539
Subtlex[0,0]	-0.467	0.556
Simple[0,0]	-0.468	0.578
Subimdb[1,1]	-0.473	0.579
Subtlex[1,1]	-0.496	0.590
Simple[1,1]	-0.475	0.593

Download

To find this data (**and much more**), visit:
http://gustavopaetzold.wordpress.com

H2020 Project Reference: **692819**

Anita

An intelligent text adaptation tool.

http://www.simpatico-project.eu

H2020 Project Reference: **692819**

Introduction

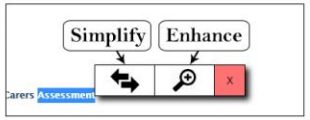
Anita is a Google Chrome extension that provides with **intelligent text adaptation solutions** that customize the content of webpages based on the **user's profile**.

User Profile

To adapt to a user's needs, **Anita** initially requests for some **personal information**, such as illustrated below:

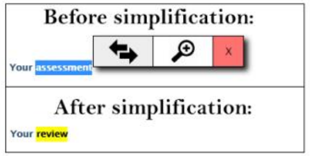
Text Adaptation

Once a user's profile is collected, they can **select words** in **any given webpage** and **request** for two types of adaptation: **Simplification** and **Enhancement**. The adaptation interface of **Anita** is illustrated below:

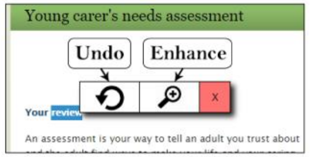


Simplification

The **Simplification** module of **Anita** attempts to **replace** the selected word with a **simpler** alternative. Upon request, Anita employs a **supervised lexical simplifier** to **replace** and **highlight** the selected word.



If the simplification **does not help the user**, it can **be reversed**. In this case, **Anita feeds the simplification data** back to the simplifier so that it can **adapt** to the user's needs.



Anita's simplifier exploits **spoken text corpora, context-aware word embeddings** and **supervised ranking models**.

Enhancement

The **Enhancement** module of **Anita** allows the user to **learn more** about words.

Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield

Anita

An intelligent text adaptation tool.

http://www.simpatico-project.eu

H2020 Project Reference: **692819**

Introduction

Anita is a Googl xtension that provides with **intelli** **adaptation solutions** that custo ntent of webpages based on the **user**

User Profil

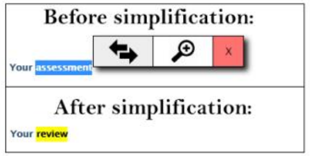
To adapt to a us **Anita** initially requests for some **p** **ormation**, such as illustrated below:

Text Adapt

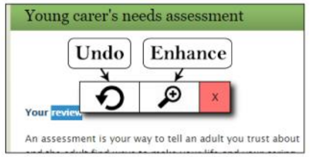
Once a user's pr cted, they can **select words** in **a** **webpage** and **request** for two typ ation: **Simplification** and **Enhan** The adaptation interface of **Anita** l below:

Simplification

The **Simplification** module of **Anita** attempts to **replace** the selected word with a **simpler** alternative. Upon request, Anita employs a **supervised lexical simplifier** to **replace** and **highlight** the selected word.



If the simplification **does not help the user**, it can **be reversed**. In this case, **Anita feeds the simplification data** back to the simplifier so that it can **adapt** to the user's needs.



Anita's simplifier exploits **spoken text corpora, context-aware word embeddings** and **supervised ranking models**.

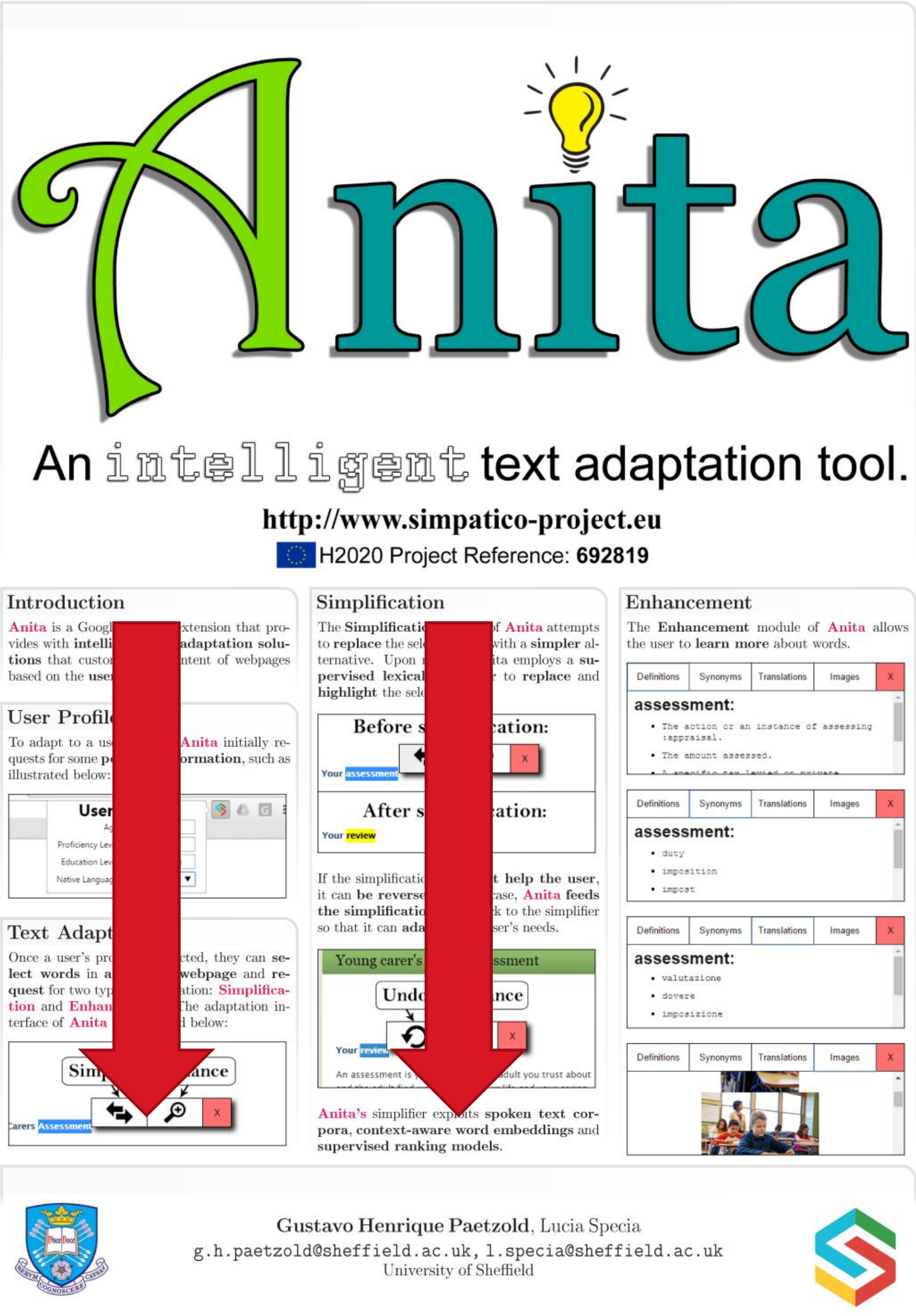
Enhancement

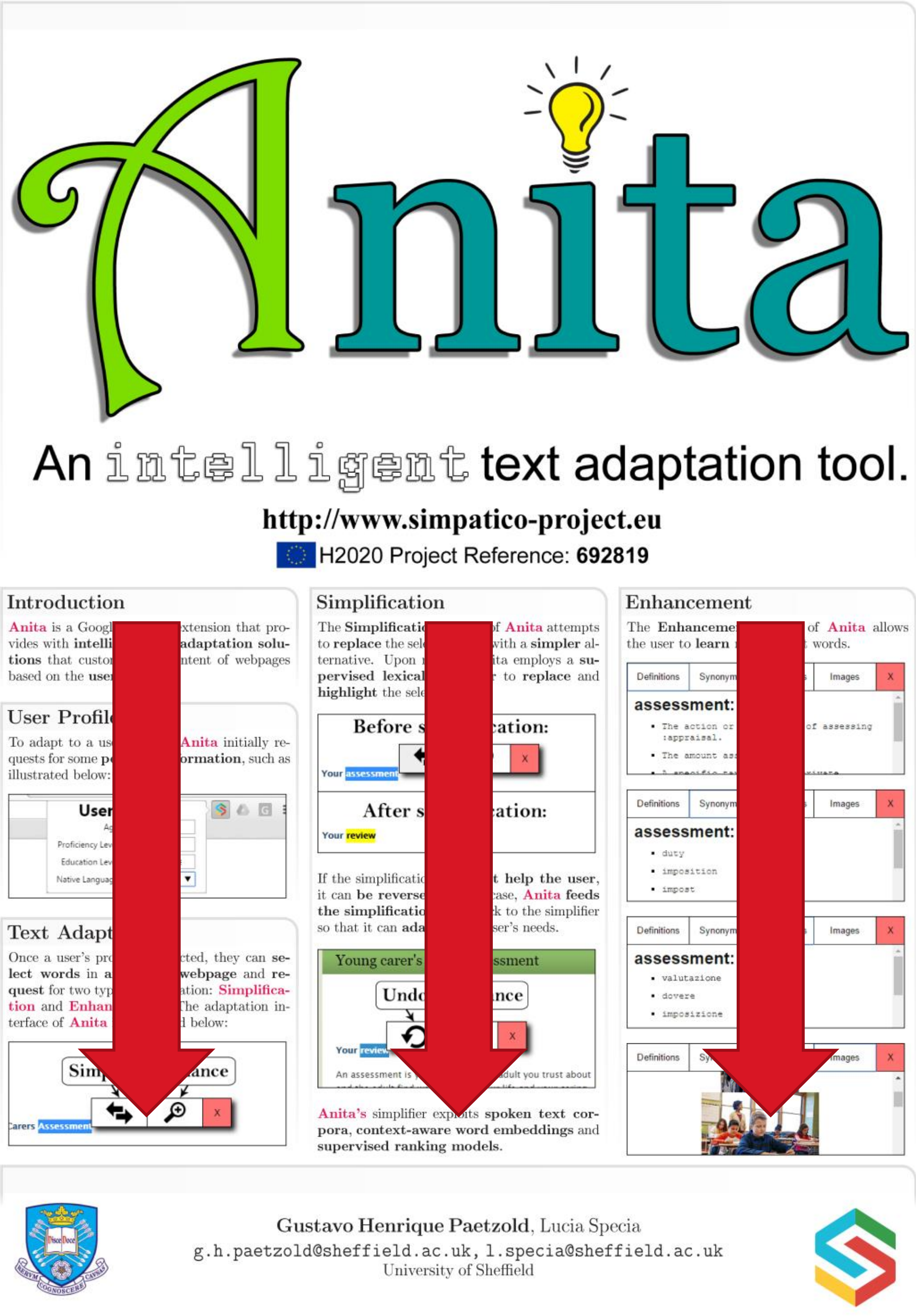
The **Enhancement** module of **Anita** allows the user to **learn more** about words.

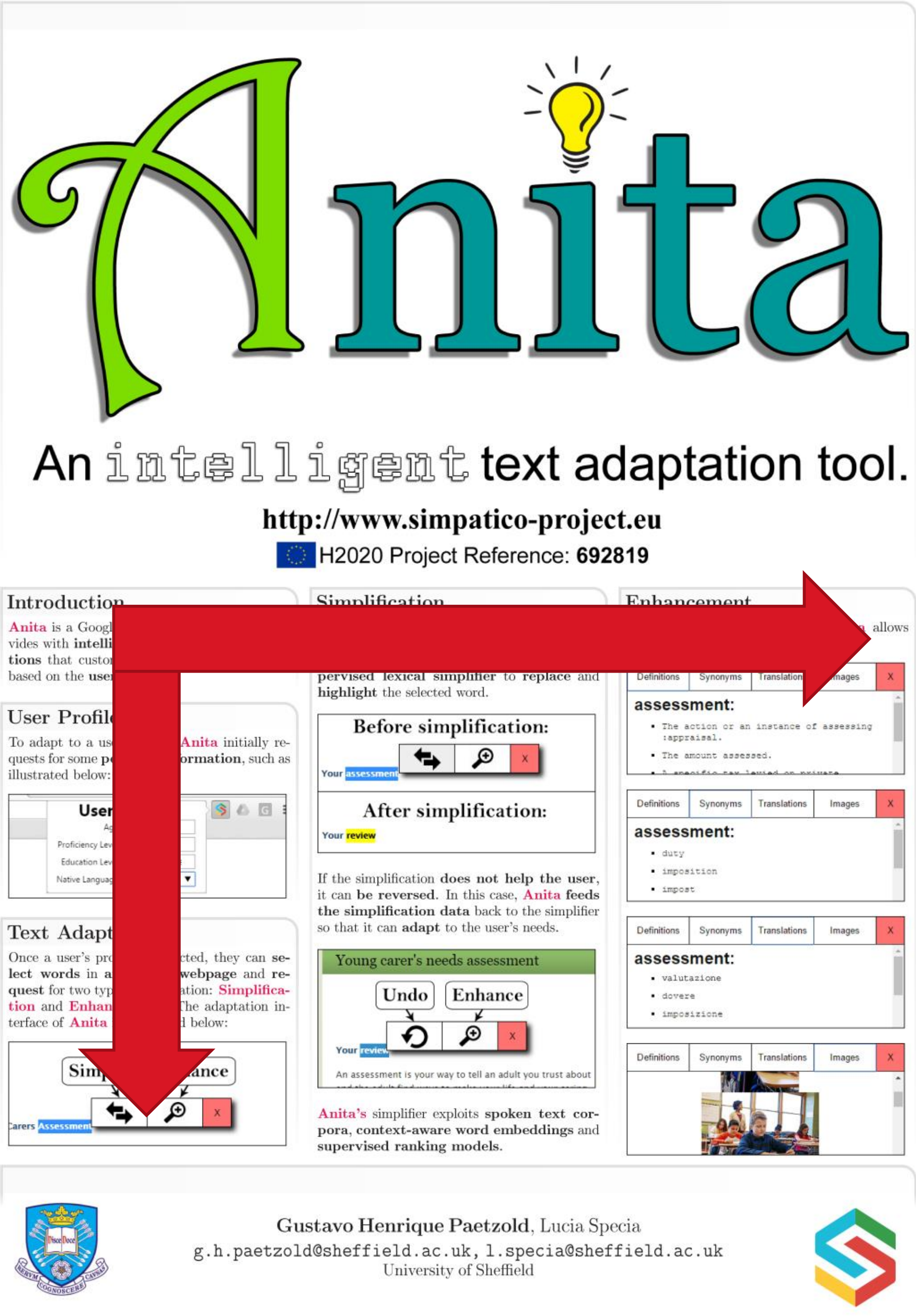
Gustavo Henrique Paetzold, Lucia Specia

g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk

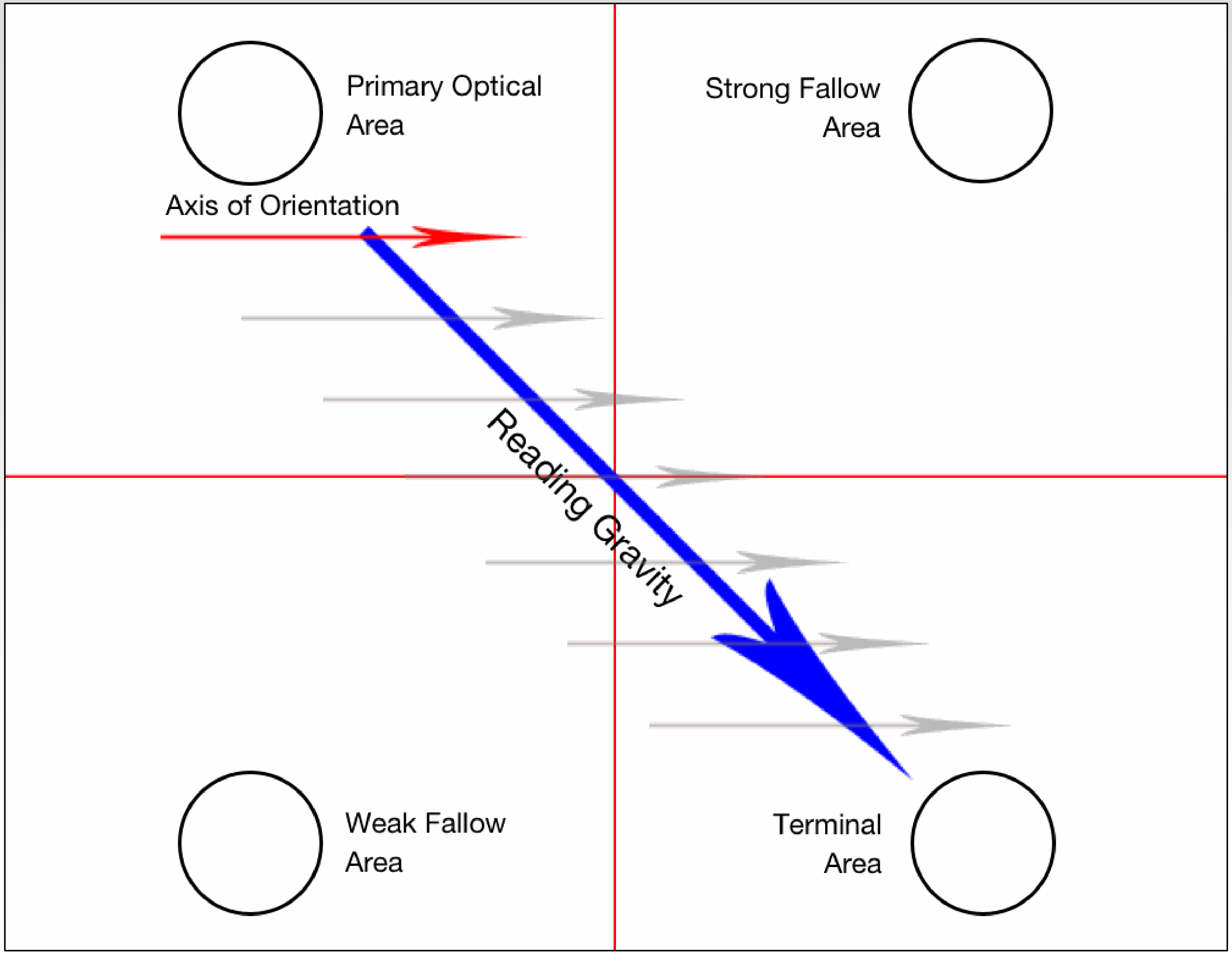
University of Sheffield


Anita
An intelligent text adaptation tool.
http://www.simpatico-project.eu
H2020 Project Reference: 692819
Introduction
Anita is a Googl xtension that provides with intelli adaptation solutions that custo ntent of webpages based on the use
User Profil
To adapt to a us Anita initially requests for some p ormation, such as illustrated below:
Text Adapt
Once a user's pr cted, they can select words in a webpage and request for two typ ation: Simplification and Enhan The adaptation interface of Anita l below:
Simplification
The Simplificati f Anita attempts to replace the sel with a simpler alternative. Upon ita employs a supervised lexical to replace and highlight the sel
Before s cation:
Your assessment
After s cation:
Your review
If the simplificati t help the user, it can be reverse case, Anita feeds the simplificatio k to the simplifier so that it can ada ser's needs.
Young carer's ssment
Anita's simplifier exploits spoken text corpora, context-aware word embeddings and supervised ranking models.
Enhancement
The Enhancement module of Anita allows the user to learn more about words.
Definitions Synonyms Translations Images
assessment:
The action or an instance of assessing
The amount assessed.
duty
imposition
impost
valutazione
dovere
imposizione
Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield


Anita
An intelligent text adaptation tool.
http://www.simpatico-project.eu
H2020 Project Reference: 692819
Introduction
Anita is a Googl xtension that provides with intelli adaptation solutions that custor ntent of webpages based on the user
User Profile
To adapt to a us Anita initially requests for some p ormation, such as illustrated below:
Proficiency Lev
Education Lev
Native Languag
Text Adapt
Once a user's pr cted, they can select words in a webpage and request for two typ ation: Simplification and Enhan The adaptation interface of Anita l below:
Simplification
The Simplificati f Anita attempts to replace the sel vith a simpler alternative. Upon ita employs a supervised lexical to replace and highlight the sel
Before s cation:
Your assessment
After s cation:
Your review
If the simplificati t help the user, it can be reverse case, Anita feeds the simplificatio k to the simplifier so that it can ada ser's needs.
Young carer's ssment
Your review
Anita's simplifier exploits spoken text corpora, context-aware word embeddings and supervised ranking models.
Enhancement
The Enhanceme of Anita allows the user to learn words.
Definitions
Images
assessment:
duty
imposition
impost
valutazione
dovere
imposizione
Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield


Anita
An intelligent text adaptation tool.
http://www.simpatico-project.eu
H2020 Project Reference: 692819
Introduction
Anita is a Googl
vides with intelli
tions that custo
based on the use
User Profil
To adapt to a us
Anita initially re-
quests for some p
ormation, such as
illustrated below:
Proficiency Lev
Education Lev
Native Languag
Text Adapt
Once a user's pr
cted, they can se-
lect words in a
webpage and re-
quest for two typ
ation: Simplifica-
tion and Enhan
The adaptation in-
terface of Anita
d below:
Carers Assessment
Simplification
pervised lexical simplifier to replace and
highlight the selected word.
Before simplification:
Your assessment
After simplification:
Your review
If the simplification does not help the user, it can be reversed. In this case, Anita feeds the simplification data back to the simplifier so that it can adapt to the user's needs.
Young carer's needs assessment
Undo
Enhance
Your review
An assessment is your way to tell an adult you trust about
Anita's simplifier exploits spoken text corpora, context-aware word embeddings and supervised ranking models.
Enhancement
allows
Definitions
Synonyms
Translations
Images
assessment:
The action or an instance of assessing appraisal.
The amount assessed.
duty
imposition
impost
valutazione
dovere
imposizione
Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield

Compositional Flow



Grammatical error correction

Grammatical error correction (GEC) in non-native text attempts to automatically detect and correct errors that are typical of those found in learner writing:

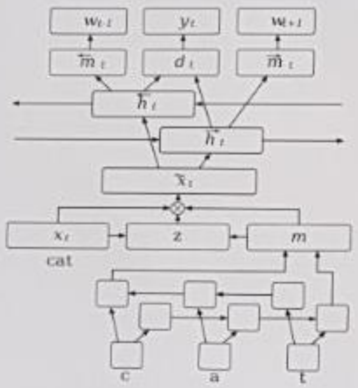
- If you need ~~futher~~ further information do not hesitate to contact us
- I am glad to help~~ing~~ you ~~for~~ with the organisation of the international student conference.
- I am ~~piece~~ pleased to ~~tell~~ provide the information ~~do~~ you need for the group.

Our approach

We propose an approach to **N-best list re-ranking** using **neural sequence-labelling models**:

- We train a compositional model for **error detection** that calculates the probability of each token being *correct* or *incorrect*.
- We then re-rank the hypotheses generated by statistical machine translation (SMT) systems.
- Our approach achieves state-of-the-art results on three different GEC datasets:
  - First Certificate in English dataset (FCE)
  - CoNLL 2014 dataset
  - JHU FLuency-Extended GUG corpus (JFLEG)

Neural sequence labelling



Statistical machine translation

- We employ two SMT systems: **CAMB16$_{SMT}$** and **AMU16$_{SMT}$**.
- For each SMT system, we generate the list of all the 10 best candidate hypotheses.
- We then determine a new ranking using features from the detection model:
  - Sentence probability
  - Levenshtein distance
  - True and false positives
- We use a linear combination of the above three scores together with the overall score given by the SMT in an unsupervised way.

Evaluation

	FCE		CoNLL		JFLEG	
	$F_{0.5}$	GLEU	$F_{0.5}$	GLEU	$F_{0.5}$	GLEU
CAMB16$_{SMT}$	52.90	70.15	37.33	64.90	52.44	46.10
CAMB16$_{SMT}$ + LSTM$_{camb}$	55.60	71.76	42.44	66.42	54.66	47.72
Oracle	71.60	78.54	58.13	70.42	61.92	50.64
AMU16$_{SMT(replicated)}$	31.66	63.73	49.34	68.23	44.77	41.98
AMU16$_{SMT(replicated)}$ + LSTM$_{camb}$	35.07	64.78	51.08	68.69	48.88	43.26
Oracle	53.54	69.52	62.41	71.18	57.49	45.00

Error type performance ($F_{0.5}$)

	CAMB16$_{SMT}$	CAMB16$_{SMT}$ + LSTM$_{camb}$
R:NOUN:POSS	35.71	**55.56**
R:VERB:SVA	58.38	**69.40**
U:ADV	13.51	**22.73**
U:DET	46.27	**55.30**
U:PRON	30.77	**39.33**
U:VERB:TENSE	28.41	**41.67**
M:PREP	**43.69**	39.43
M:VERB:FORM	**50.00**	38.46

Conclusion

- To the best of our knowledge, no prior work has investigated **the impact of detection models on correction performance**.
- Detection models can be more fine-tuned to finer nuances of grammaticality, and therefore better able to distinguish between correct and incorrect versions of a sentence.
- Our approach can be applied to any GEC system that produces multiple alternative hypotheses.
- Our results demonstrate **the benefits of integrating detection approaches with correction systems**, and how one can complement the other.

Introduction

- User generated categories (UGCs) express rich semantic relations implicitly.
- While most methods use pattern matching for English, learning relations from Chinese UGCs poses challenges due to the flexible expressions.
- Our work uses weakly supervised methods to extract relations from Chinese UGCs based on projection learning and graph mining.

Mining Is-a Relations

Initial model training

- Use existing labeled sets and heuristic rules to generate training data automatically (i.e., is-a and not-is-a relation pairs).
- Train a skip-gram model to map each word x_i to its embedding $\mathbf{x}_i$.
- Train two linear projection models based on word embeddings. One for is-a relations. The other for not-is-a relations.

$$J(\mathbf{M}^+,\mathbf{B}^+)=\frac{1}{2}\sum_{(e,c_h)\in D^+}\|\mathbf{M}^+\mathbf{e}+\mathbf{B}^+-\mathbf{c}_h\|_2^2+\frac{\lambda}{2}\|\mathbf{M}^+\|_F^2+\frac{\lambda}{2}\|\mathbf{B}^+\|_F^2$$

$$J(\mathbf{M}^-,\mathbf{B}^-)=\frac{1}{2}\sum_{(e,c_h)\in D^-}\|\mathbf{M}^-\mathbf{e}+\mathbf{B}^--\mathbf{c}_h\|_2^2+\frac{\lambda}{2}\|\mathbf{M}^-\|_F^2+\frac{\lambda}{2}\|\mathbf{B}^-\|_F^2$$

where e is a Wikipedia concept and c_h is the head word of a UGC of entity e in its corresponding Wikipedia page.

- Estimate the prediction score $s(e,c)$ for each unlabeled (e,c) pair.

$$s(e,c)=\tanh(\|\mathbf{M}^+\mathbf{e}+\mathbf{B}^+-\mathbf{c}_h\|_2-\|\mathbf{M}^-\mathbf{e}+\mathbf{B}^--\mathbf{c}_h\|_2)$$

High prediction score means there is a large probability of is-a relation between e and c.

Score refinement by collective inference

- Denote $\tilde{g}(h)$ as the un-normalized global prediction score for head word h of UGCs:

$$\tilde{g}(h)=\ln(1+|D_h|+|D_h^+|)\frac{|D_h^+|+\sum_{(e,c)\in D_h}s(e,c)}{|D_h|+|D_h^+|}$$

where H is the collection of head words of UGCs.

Mining Is-a Relations

Score refinement by collective inference

- Re-normalize the prediction score $s(e,c)$ based on the initial prediction score and global prediction score.

$$f(e,c)=\beta s(e,c)+(1-\beta)g(h)$$

where $\beta\in(0,1)$ is the tuning parameter and $g(h)$ is the normalized version of $\tilde{g}(h)$: $g(h)=\frac{\tilde{g}(h)}{\max_{h'\in H}|\tilde{g}(h')|}$.

- Expand the number of hypernyms by the following heuristic rule: Finally, we regard c_h as a valid hypernym of e if c is predicted as a hypernym of e and c_h is also a Wikipedia concept.

Mining Non-taxonomic Relations

Single-pass category pattern mining

- Extract category patterns by replacing entity placeholders with specific entity names in UGCs. For example, the pattern is "[E]获得者"(Winner of [E])" for "图灵奖获得者(Winner of Turing Award)". The pair "(蒂姆·伯纳斯-李, 图灵奖)(Tim Berners-Lee, Turing Award)" can be extracted as a candidate relation instance.
- Calculate the pattern support score $supp(p)$ of pattern p and filter out low-support patterns by:

$$supp(p)=|R_p|\cdot\ln(1+L_p)$$

where R_p is the collection of extracted pairs for pattern p and L_p is the pattern length.

Graph-based raw relation extractor

- For each pattern p, construct a graph G where nodes are extracted candidate relation pairs based on p and weighted edges are the semantic similarities between the pairs.
- Detect a Maximum Edge Weight Clique (MEWC) C^* in G and treat pairs in C^* as seed relation instances that p may represent. We propose a Monte Carlo based method to extract the MEWC from the graph approximately. Please refer to the paper for details.
- Extract relation instances for the underline relation that p may present by finding pairs that are similar to the seed relation instances.

Relation mapping

- Map extracted pairs to relation triples by defining the relation predicates through i) direct verbal mapping, ii) direct non-verbal mapping and iii) indirect mapping.

Experiments

Experiments on is-a relation extraction

- Dataset: 1,788 labeled entity-UGC pairs extracted from Chinese Wikipedia.
- Metrics: Precision, Recall and F-Measure.
- Results: Our approach outperforms all competitive baselines.

Method	Precision (%)	Recall (%)	F-Measure (%)
Concat Model	79.5	64.2	67.2
Sum Model	80.9	70.1	72.6
Diff Model	78.3	69.0	71.5
Piecewise Projection	78.9	72.3	75.5
Our Method (w/o Exp)	89.2	88.1	88.7
Our Method	**89.8**	**88.3**	**89.0**

Experiments on non-taxonomic relation extraction

- Dataset: All entity-UGC pairs in Chinese Wikipedia.
- Metrics: Size (#extractions for a certain relation type), Accuracy and Coverage (whether the extracted relations are covered by a large existing Chinese KB).
- Results: Our approach can extract a large amount of novel relations with high accuracy.

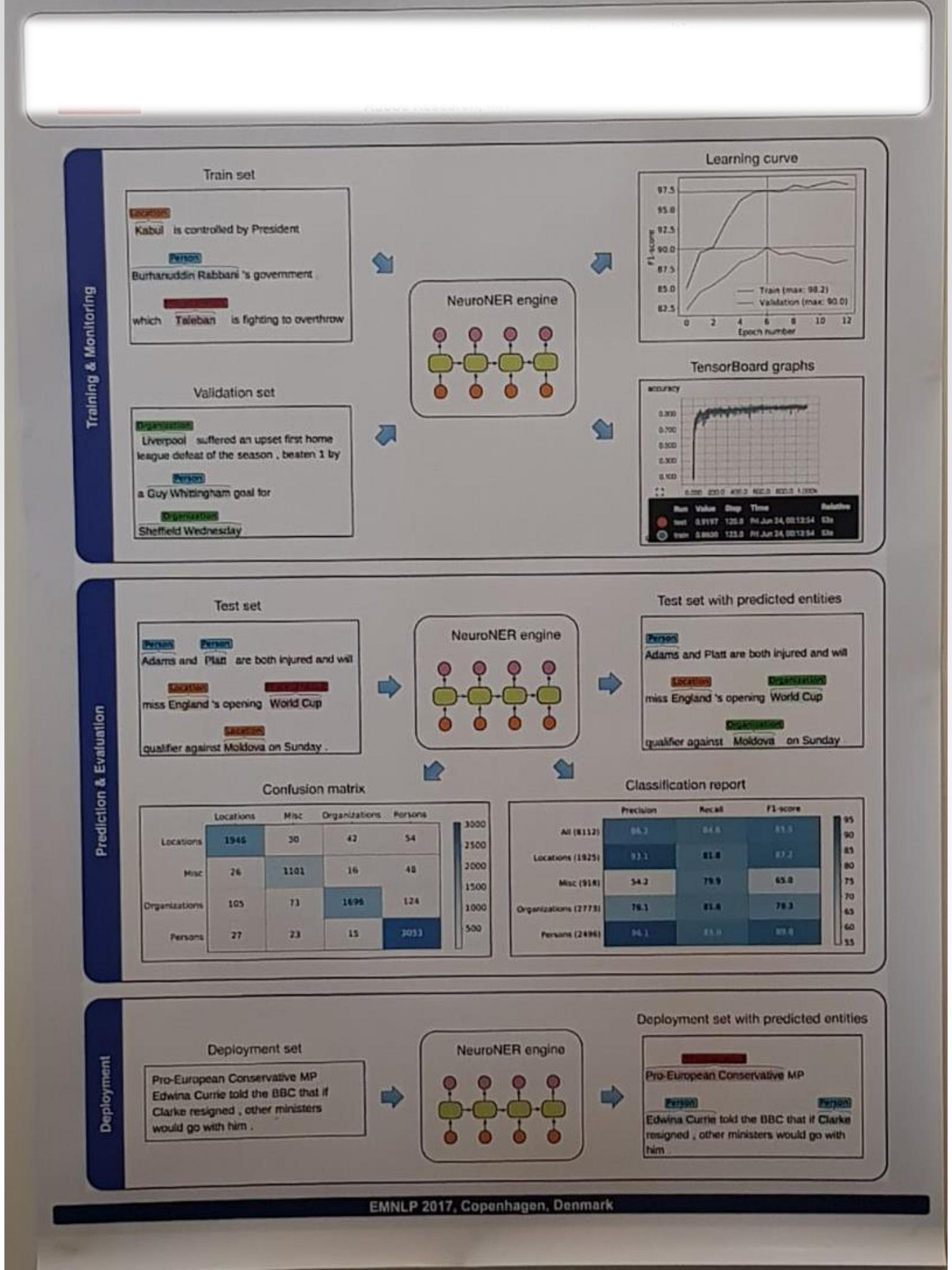
Relation	Size	Accuracy (%)	Coverage (%)
毕业(graduated-from)	44,118	98.0	22.9
位于(located-in)	29,460	97.2	8.5
建立(established-in)	20,154	95.0	31.5
出生(born-in)	11,671	98.3	41.4
成员(member-of)	8,445	96.0	4.2
启用(open-in)	8,956	98.2	21.6

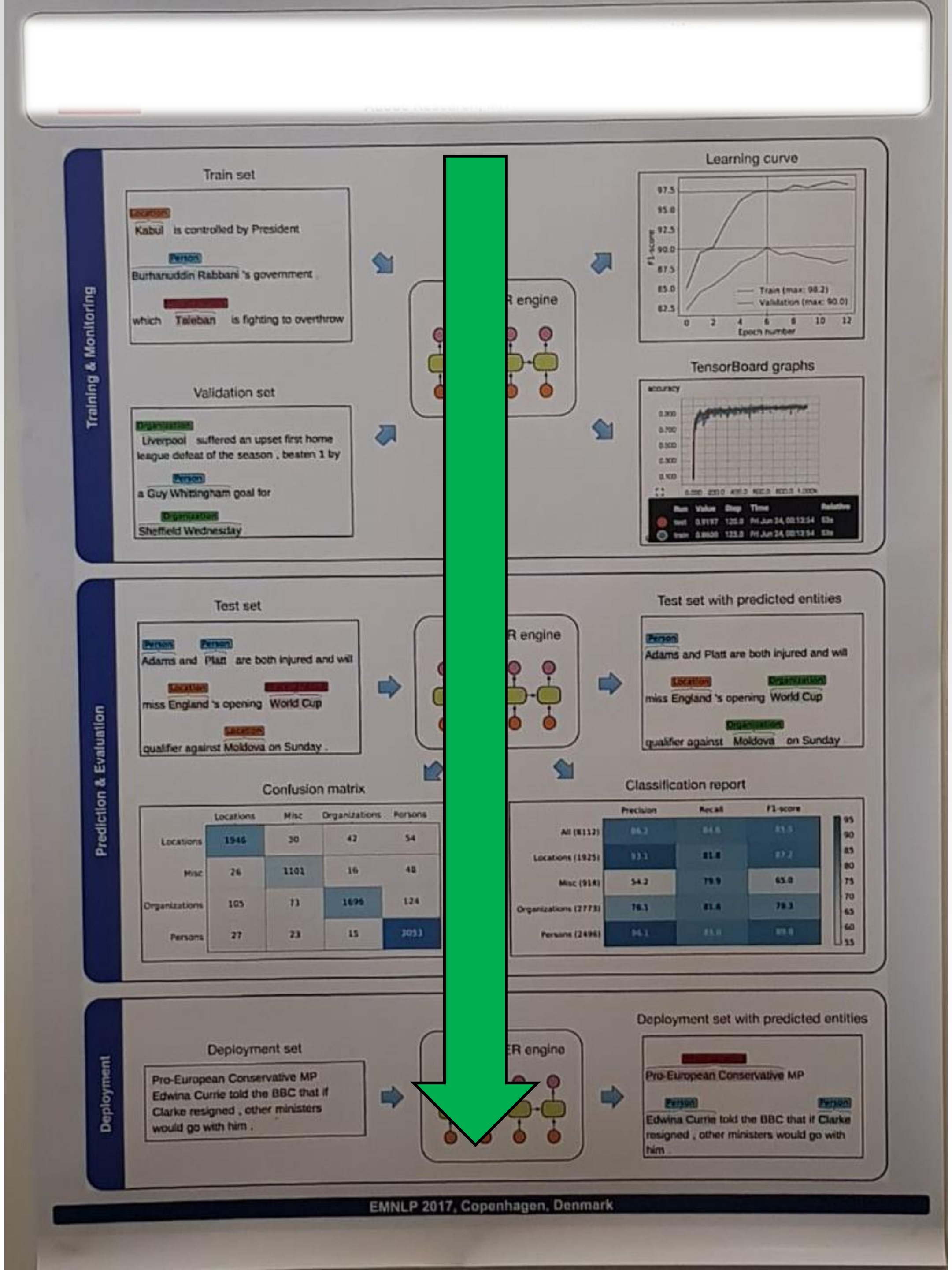
- Please refer to more supplementary experiments in the paper.

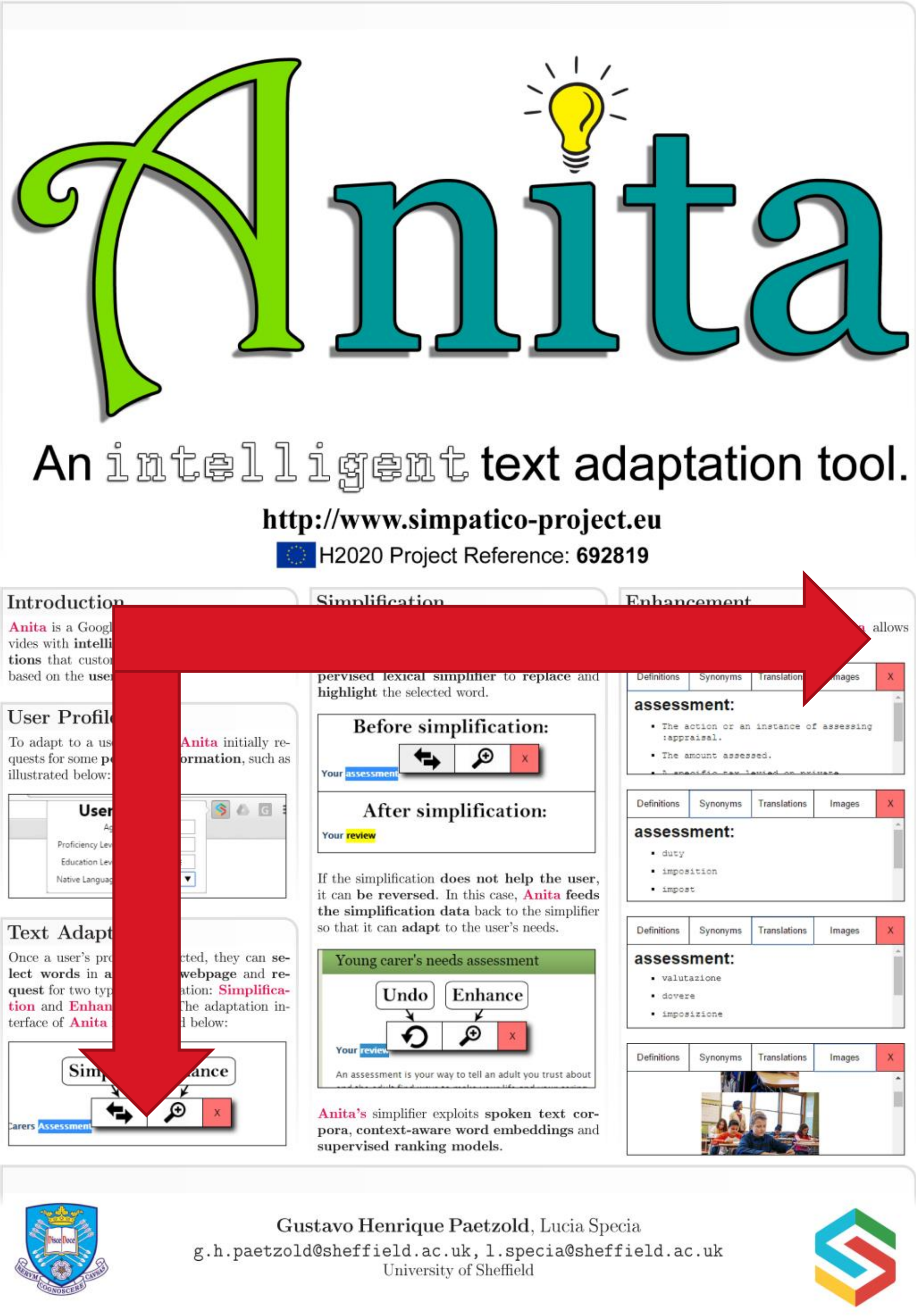
Conclusion and Future Work

- We propose a weakly supervised framework to extract relations from Chinese UGCs. It requires very little human intervention and has high accuracy for the Chinese language.
- Future work includes:
  - Improving our work for short text knowledge extraction;
  - Designing a general framework for cross-lingual UGC relation extraction.

Contact inf


Training & Monitoring
Train set
Location
Kabul is controlled by President
Person
Burhanuddin Rabbani 's government
which Taleban is fighting to overthrow
Validation set
Organization
Liverpool suffered an upset first home league defeat of the season , beaten 1 by
Person
a Guy Whittingham goal for
Organization
Sheffield Wednesday
NeuroNER engine
Learning curve
Train (max: 98.2)
Validation (max: 90.0)
Epoch number
TensorBoard graphs
accuracy
Prediction & Evaluation
Test set
Person
Person
Adams and Platt are both injured and will
Location
miss England 's opening World Cup
Location
qualifier against Moldova on Sunday .
NeuroNER engine
Test set with predicted entities
Person
Adams and Platt are both injured and will
Location
Organization
miss England 's opening World Cup
Organization
qualifier against Moldova on Sunday
Confusion matrix
Locations Misc Organizations Persons
Locations 1946 30 42 54
Misc 26 1101 16 48
Organizations 105 73 1696 124
Persons 27 23 15 3053
Classification report
Precision Recall F1-score
Locations (1925) 93.1 81.8 87.2
Misc (918) 54.2 79.9 65.0
Organizations (2773) 76.1 81.4 78.3
Deployment
Deployment set
Pro-European Conservative MP Edwina Currie told the BBC that if Clarke resigned , other ministers would go with him .
NeuroNER engine
Deployment set with predicted entities
Pro-European Conservative MP
Person
Person
Edwina Currie told the BBC that if Clarke resigned , other ministers would go with him .
EMNLP 2017, Copenhagen, Denmark


Training & Monitoring
Train set
Kabul is controlled by President
Burhanuddin Rabbani 's government
which Taleban is fighting to overthrow
Validation set
Liverpool suffered an upset first home league defeat of the season , beaten 1 by
a Guy Whittingham goal for
Sheffield Wednesday
Learning curve
TensorBoard graphs
Prediction & Evaluation
Test set
Adams and Platt are both injured and will
miss England 's opening World Cup
qualifier against Moldova on Sunday .
Test set with predicted entities
Confusion matrix
Classification report
Deployment
Deployment set
Pro-European Conservative MP Edwina Currie told the BBC that if Clarke resigned , other ministers would go with him .
Deployment set with predicted entities
EMNLP 2017, Copenhagen, Denmark


Anita
An intelligent text adaptation tool.
http://www.simpatico-project.eu
H2020 Project Reference: 692819
Introduction
Anita is a Googl… vides with intelli… tions that custo… based on the use…
User Profil…
To adapt to a us… Anita initially requests for some p… ormation, such as illustrated below:
Text Adapt…
Once a user's pr… cted, they can select words in a… webpage and request for two typ… ation: Simplification and Enhan… The adaptation interface of Anita … l below:
Simplification
… pervised lexical simplifier to replace and highlight the selected word.
Before simplification:
Your assessment
After simplification:
Your review
If the simplification does not help the user, it can be reversed. In this case, Anita feeds the simplification data back to the simplifier so that it can adapt to the user's needs.
Young carer's needs assessment
Undo
Enhance
Your review
Anita's simplifier exploits spoken text corpora, context-aware word embeddings and supervised ranking models.
Enhancement
… allows
Definitions
Synonyms
Translations
Images
assessment:
The action or an instance of assessing ; appraisal.
The amount assessed.
duty
imposition
impost
valutazione
dovere
imposizione
Gustavo Henrique Paetzold, Lucia Specia
g.h.paetzold@sheffield.ac.uk, l.specia@sheffield.ac.uk
University of Sheffield

Posters: The Solutions

How to create a **narrative**?

Posters: The Solutions

1. Introduction/Task description

Posters: The Solutions

1. Introduction/Task description
2. Approach/Strategy

Posters: The Solutions

1. Introduction/Task description
2. Approach/Strategy
3. Evaluation setup (if necessary)

Posters: The Solutions

1. Introduction/Task description
2. Approach/Strategy
3. Evaluation setup (if necessary)
4. Results

Posters: The Solutions

1. Introduction/Task description
2. Approach/Strategy
3. Evaluation setup (if necessary)
4. Results
5. Main findings (keep it short)

Posters: The Solutions

1. Introduction/Task description
2. Approach/Strategy
3. Evaluation setup (if necessary)
4. Results
5. Main findings (keep it short)
6. Download information (if necessary)

Posters: The Solutions

Poster **overhauling** example!

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe
The Generic University
Typical street, The square, 7998
J7KE3, City, Country

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.
- Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];
- We explore here three research questions:
  1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
  2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
  3. can the inferred values help in creating features that result in more reliable readability prediction models?

The Proposed Method: Regression in a Multi-View Learning Approach

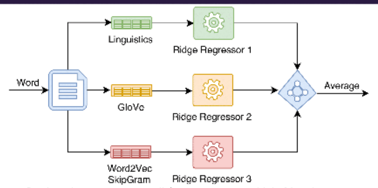


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram ($d = 300$, 600 and 1,000);
10. Word's raw embedding values of GloVe ($d = 300$, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Adaptation of Databases with Norms for Portuguese

Study	Participants	Words	Property	Portuguese Variant	Scale
[2]	2.357	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	685	1,748	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	834	AoA	European	1-7
[6]	103	249	imageability, concreteness	European	1-7

Table 1. Norms for Portuguese on the focused psycholinguistic properties.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with $d = 300$.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale Chall	Gunning Fox	Subjective Frequency	Psycholinguistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: `http://nilc.icmc.usp.br/psycholinguistic`
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

References

[1] Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.

[2] Soares, A.P., Costa, A.S., J, M., Comesana, M.H.M.: The minho word pool: Norms for imageability, concreteness, and subjective frequency for 3,800 portuguese words. Behavior Research Methods (2016)

[3] Cameirao, M.L., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-480 (2010)

[4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concretude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)

[5] Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)

[6] Marques, J.F. Normas de imagetica e concreteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

An Outstanding Academic Contribution

John Doe, Jane Doe and Josh Doe
{john,jane,josh}@doe.com
The Generic University

Introduction

We predict 4 psycholinguistic properties for Portuguese:

1. **Imageability:** Ease with which a word evokes a mental image.
2. **Concreteness:** Degree to which words refer to things that can be experienced by the senses.
3. **Familiarity:** Estimate of the number of times a word is encountered by individuals in its written or spoken form.
4. **Age of Acquisition:** The estimate of the age at which a word was learned.

Various applications:

- Lexical simplification
- Reading time prediction
- Sentence simplification
- Readability models

Challenges:

1. Manually produced properties for Portuguese are **very scarce**
2. Previous approaches use **expensive, unavailable resources**

Model Settings

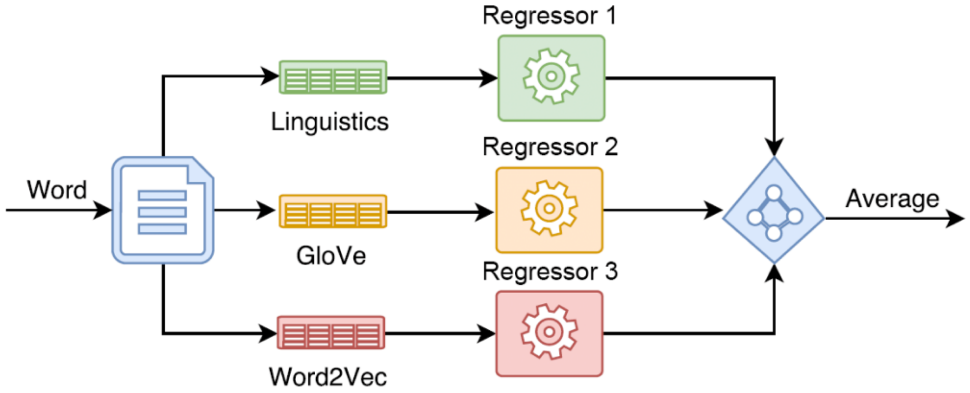
Multiview feature groups:

Regressor 1	Regressor 2	Regressor 3:
1-8	9	10

Feature set:

1. Word Length
2. Log of Frequency in SUBTLEX-PT
3. Log of Frequency in SubIMDb-PT
4. Log of number of subtitles that contain the word in SUBTLEX-PT
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro
6. Log of Frequency in the Written Language part of Corpus Brasileiro
7. Log of Frequency in a corpus of Mixed Text Genres
8. Lexical databases from 6 school dictionaries
9. Word's raw embedding values of Skip-Gram
10. Word's raw embedding values of GloVe

Proposed Model



Results

More Findings

Alpha inter-annotator agreement scores:

Imageability	Concreteness
0.921	0.820

Conclusions drawn:

- Possible to infer psycholinguistic properties for BP with embeddings
- Regressors need a substantial amount of training data
- Age of acquisition and familiarity models require extra resources
- Our psycholinguistic properties can improve readability prediction

Download

Psycholinguistic features for 26,874 BP words:

http://nilc.icmc.usp.br/psycholinguistic

Introduction

– Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

– Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

– Most of these properties are costly and time-consuming to be manually gathered;

– English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

– Portuguese: only datasets of limited size [2, 3, 4, 5];

– Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

– We explore here three research questions:
  1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
  2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
  3. can the inferred values help in creating features that result in more reliable readability prediction models?

Introduction

– Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

– Used in various NLP tasks:

lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

– Most of these properties are costly and time-consuming to be manually gathered;

– English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

– Portuguese: only datasets of limited size [2, 3, 4, 5];

– Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

– We explore here three research questions:
  1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
  2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
  3. can the inferred values help in creating features that result in more reliable readability prediction models?

Introduction

We predict 4 psycholinguistic properties for Portuguese:

1. **Imageability:** Ease with which a word evokes a mental image.
2. **Concreteness:** Degree to which words refer to things that can be experienced by the senses.
3. **Familiarity:** Estimate of the number of times a word is encountered by individuals in its written or spoken form.
4. **Age of Acquisition:** The estimate of the age at which a word was learned.

Various applications:

- Lexical simplification
- Reading time prediction
- Sentence simplification
- Readability models

Challenges:

1. Manually produced properties for Portuguese are **very scarce**
2. Previous approaches use **expensive, unavailable resources**

Introduction

– Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:

1. **word imageability** the ease and speed with which a word evokes a mental image;
2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

– Used in various NLP tasks:

lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

– Most of these properties are costly and time-consuming to be manually gathered;

– English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

– Portuguese: only datasets of limited size [2, 3, 4, 5];

– Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

– We explore here three research questions:

1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
3. can the inferred values help in creating features that result in more reliable readability prediction models?

Introduction

We predict 4 psycholinguistic properties for Portuguese:

1. **Imageability:** Ease with which a word evokes a mental image.
2. **Concreteness:** Degree to which words refer to things that can be experienced by the senses.
3. **Familiarity:** Estimate of the number of times a word is encountered by individuals in its written or spoken form.
4. **Age of Acquisition:** The estimate of the age at which a word was learned.

Various applications:

- Lexical simplification
- Reading time prediction
- Sentence simplification
- Readability models

Challenges:

1. Manually produced properties for Portuguese are **very scarce**
2. Previous approaches use **expensive, unavailable resources**

Introduction

- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
  1. **word imageability** the ease and speed with which a word evokes a mental image;
  2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
  3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
  4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.
- Used in various NLP tasks:

  lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];
- We explore here three research questions:
  1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
  2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
  3. can the inferred values help in creating features that result in more reliable readability prediction models?

Introduction

We predict 4 psycholinguistic properties for Portuguese:

1. **Imageability:** Ease with which a word evokes a mental image.
2. **Concreteness:** Degree to which words refer to things that can be experienced by the senses.
3. **Familiarity:** Estimate of the number of times a word is encountered by individuals in its written or spoken form.
4. **Age of Acquisition:** The estimate of the age at which a word was learned.

Various applications:

- Lexical simplification
- Reading time prediction
- Sentence simplification
- Readability models

Challenges:

1. Manually produced properties for Portuguese are **very scarce**
2. Previous approaches use **expensive, unavailable resources**

Introduction

– Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:

1. **word imageability** the ease and speed with which a word evokes a mental image;
2. **concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
3. **subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
4. **age of acquisition - AoA** is the estimation of the age at which a word was learned.

– Used in various NLP tasks:

lexical simplification; text simplification at the sentence level; to predict the reading times of each word in a sentence; to create robust text level readability models.

Gap and Purpose

– Most of these properties are costly and time-consuming to be manually gathered;

– English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;

– Portuguese: only datasets of limited size [2, 3, 4, 5];

– Previous approaches to automatically infer the properties: based on a large, scarce lexical resource as WordNet [1];

– We explore here three research questions:

1. is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embedding to infer the psycholinguistic properties?
2. which size a psycholinguistic database should have to be used in regression models? Does merging databases from different sources yield better correlation and lower MSE scores?
3. can the inferred values help in creating features that result in more reliable readability prediction models?

Introduction

We predict 4 psycholinguistic properties for Portuguese:

1. **Imageability:** Ease with which a word evokes a mental image.
2. **Concreteness:** Degree to which words refer to things that can be experienced by the senses.
3. **Familiarity:** Estimate of the number of times a word is encountered by individuals in its written or spoken form.
4. **Age of Acquisition:** The estimate of the age at which a word was learned.

Various applications:

- Lexical simplification
- Reading time prediction
- Sentence simplification
- Readability models

Challenges:

1. Manually produced properties for Portuguese are **very scarce**
2. Previous approaches use **expensive, unavailable resources**

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Model Settings

Multiview feature groups:

Regressor 1	Regressor 2	Regressor 3:
1-8	9	10

Feature set:

1. Word Length
2. Log of Frequency in SUBTLEX-PT
3. Log of Frequency in SubIMDb-PT
4. Log of number of subtitles that contain the word in SUBTLEX-PT
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro
6. Log of Frequency in the Written Language part of Corpus Brasileiro
7. Log of Frequency in a corpus of Mixed Text Genres
8. Lexical databases from 6 school dictionaries
9. Word's raw embedding values of Skip-Gram
10. Word's raw embedding values of GloVe

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Model Settings

Multiview feature groups:

Regressor 1	Regressor 2	Regressor 3:
1-8	9	10

Feature set:

1. Word Length
2. Log of Frequency in SUBTLEX-PT
3. Log of Frequency in SubIMDb-PT
4. Log of number of subtitles that contain the word in SUBTLEX-PT
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro
6. Log of Frequency in the Written Language part of Corpus Brasileiro
7. Log of Frequency in a corpus of Mixed Text Genres
8. Lexical databases from 6 school dictionaries
9. Word's raw embedding values of Skip-Gram
10. Word's raw embedding values of GloVe

Features for Regressors

10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):

1. Log of Frequency in SUBTLEX-pt-BR;
2. Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
3. Log of Frequency in SubIMDb-PT: subtitles of family, comedy and children movies and series;
4. Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro;
6. Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
7. Word Length;
8. Lexical databases from 6 school dictionaries for specific grade-levels;
9. Word's raw embedding values of Skip-Gram (d = 300, 600 and 1,000);
10. Word's raw embedding values of GloVe (d = 300, 600 and 1,000).

Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (http://www.nilc.icmc.usp.br/embeddings)

Model Settings

Multiview feature groups:

Regressor 1	Regressor 2	Regressor 3:
1-8	9	10

Feature set:

1. Word Length
2. Log of Frequency in SUBTLEX-PT
3. Log of Frequency in SubIMDb-PT
4. Log of number of subtitles that contain the word in SUBTLEX-PT
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro
6. Log of Frequency in the Written Language part of Corpus Brasileiro
7. Log of Frequency in a corpus of Mixed Text Genres
8. Lexical databases from 6 school dictionaries
9. Word's raw embedding values of Skip-Gram
10. Word's raw embedding values of GloVe

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale-Chall	Gunning Fox	Subjective Frequency	Psycholinguistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.

Evaluation

- Table 2 presents best results: Skip-Gram and GloVe embeddings with d = 300.
- 20x5-fold cross-validation

Regressors	Concreteness (4088)			Subjective Frequency (3735)			Imageability (3735)			AoA Merging (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	1.24	0.54	0.56	0.55	0.72	0.73	0.74	0.58	0.59	0.67	0.73	0.73
Skip-gram	0.52	0.84	0.84	0.58	0.70	0.71	0.46	0.77	0.77	0.81	0.66	0.66
GloVe	0.62	0.80	0.81	0.40	0.81	0.81	0.49	0.75	0.75	0.63	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.44	0.79	0.79	0.47	0.77	0.78	0.59	0.77	0.77
Lexical + GloVe	0.70	0.80	0.80	0.39	0.81	0.81	0.50	0.75	0.76	**0.54**	**0.79**	**0.79**
Skip-gram + GloVe	**0.49**	**0.85**	**0.85**	0.41	0.80	0.80	**0.42**	**0.79**	**0.79**	0.62	0.75	0.75
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	**0.38**	**0.82**	**0.82**	0.43	0.79	0.78	0.54	0.79	0.79

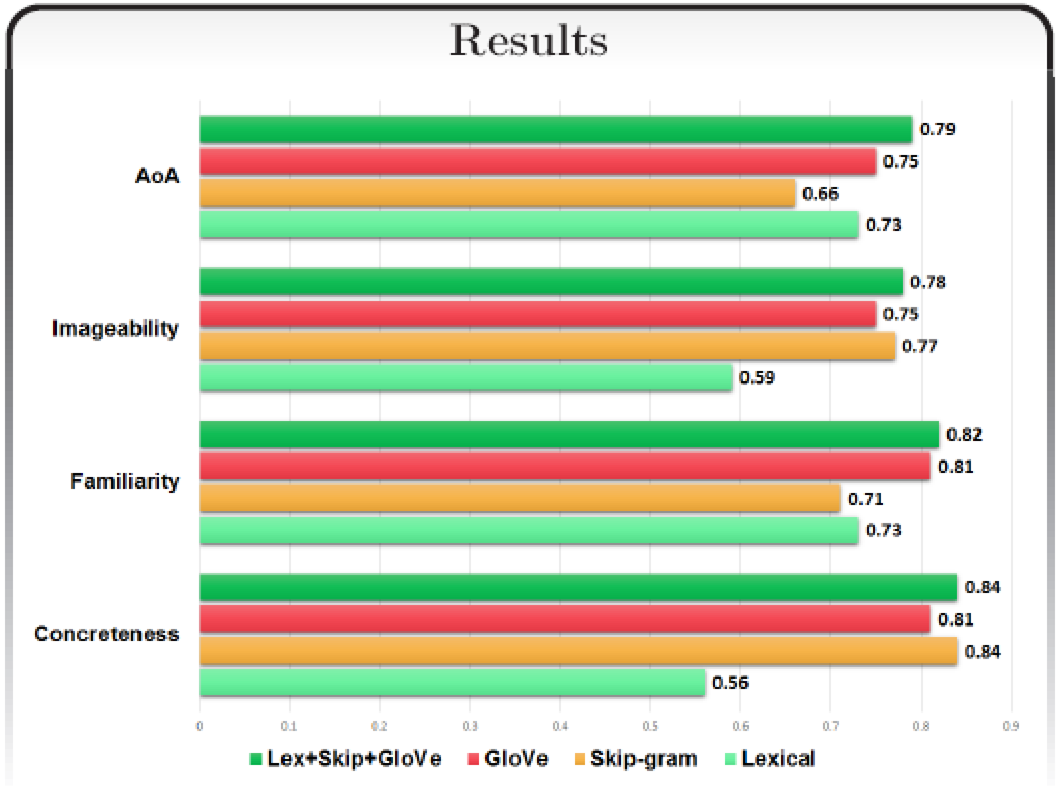
Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Table 3. MSE, Pearson, and Spearman correlations of the regression models.

Flesch	Honoré	Concreteness	Familiarity	AoA	Dale-Chall	Gunning Fox	Subjective Frequency	Psycholinguistics	MATTR	Brunét
0.26	0.29	0.27	0.23	0.25	0.36	0.37	0.32	0.45	0.48	**0.54**

Table 4. F1-measure of Psycholinguistic and Classic readability formulas for readability prediction.



Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

More Findings

Alpha inter-annotator agreement scores:

Imageability	Concreteness
0.921	0.820

Conclusions drawn:

- Possible to infer psycholinguistic properties for BP with embeddings
- Regressors need a substantial amount of training data
- Age of acquisition and familiarity models require extra resources
- Our psycholinguistic properties can improve readability prediction

Download

Psycholinguistic features for 26,874 BP words:

http://nilc.icmc.usp.br/psycholinguistic

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

More Findings

Alpha inter-annotator agreement scores:

Imageability	Concreteness
0.921	0.820

Conclusions drawn:

- Possible to infer psycholinguistic properties for BP with embeddings
- Regressors need a substantial amount of training data
- Age of acquisition and familiarity models require extra resources
- Our psycholinguistic properties can improve readability prediction

Download

Psycholinguistic features for 26,874 BP words:

http://nilc.icmc.usp.br/psycholinguistic

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

More Findings

Alpha inter-annotator agreement scores:

Imageability	Concreteness
0.921	0.820

Conclusions drawn:

- Possible to infer psycholinguistic properties for BP with embeddings
- Regressors need a substantial amount of training data
- Age of acquisition and familiarity models require extra resources
- Our psycholinguistic properties can improve readability prediction

Download

Psycholinguistic features for 26,874 BP words:

http://nilc.icmc.usp.br/psycholinguistic

Conclusions and Future Work

- A large database of 26,874 BP words annotated with psycholinguistic properties: http://nilc.icmc.usp.br/psycholinguistic
- Alpha scores of 0.921 for imageability and 0.820 for concreteness – similar to the values reported in literature;
- With respect to our research questions:
  1. we have shown we can infer psycholinguistic properties for BP using word embeddings;
  2. our regressors need a reasonably large number of training instances (at least, more than two thousand examples), as well as complementary lexical resources to yield top performance for AoA and subjective frequency;
  3. our results show that psycholinguistic properties can potentially aid readability prediction.
- Future work: extend our extrinsic evaluation to other tasks; use new modeling techniques for our psycholinguistic features (besides the average and standard deviation); use a more robust approach to fusion of regressors, e.g. stacking regression.

More Findings

Alpha inter-annotator agreement scores:

Imageability	Concreteness
0.921	0.820

Conclusions drawn:

- Possible to infer psycholinguistic properties for BP with embeddings
- Regressors need a substantial amount of training data
- Age of acquisition and familiarity models require extra resources
- Our psycholinguistic properties can improve readability prediction

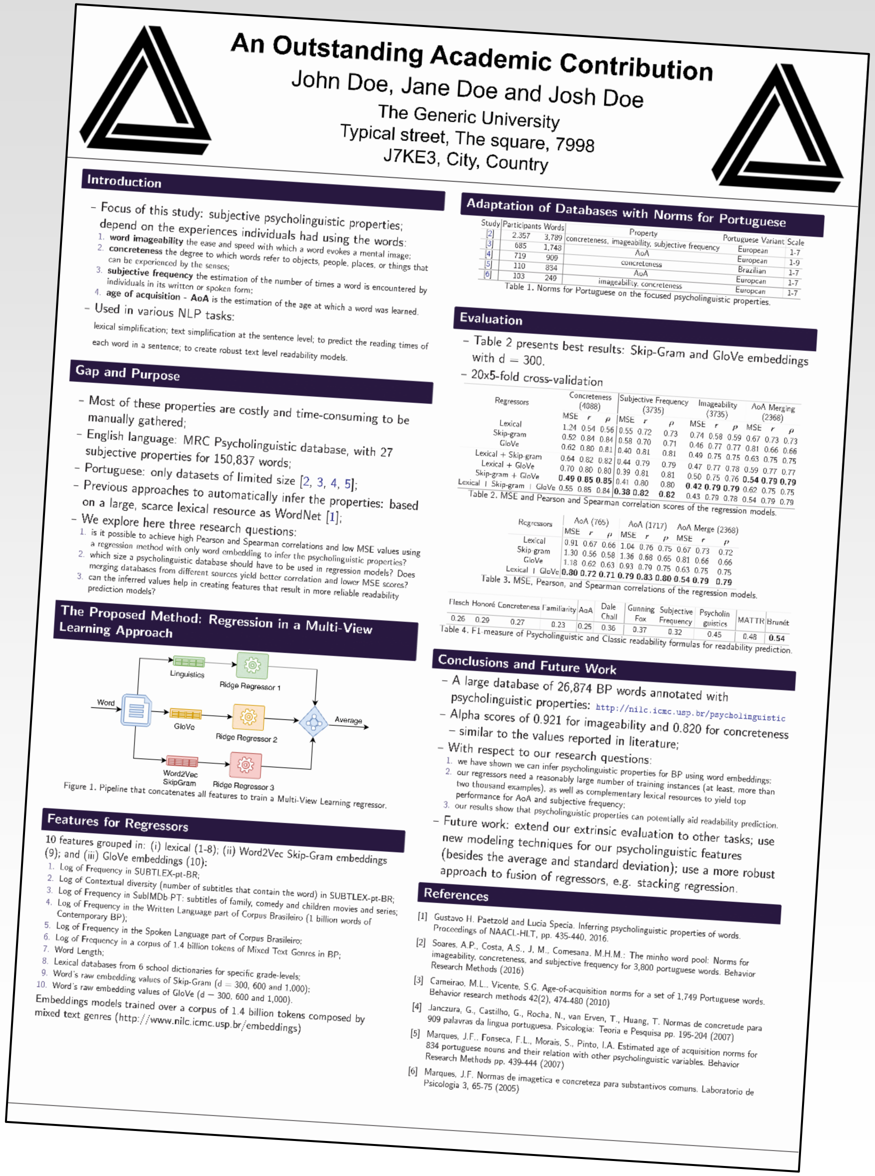
Download

Psycholinguistic features for 26,874 BP words:

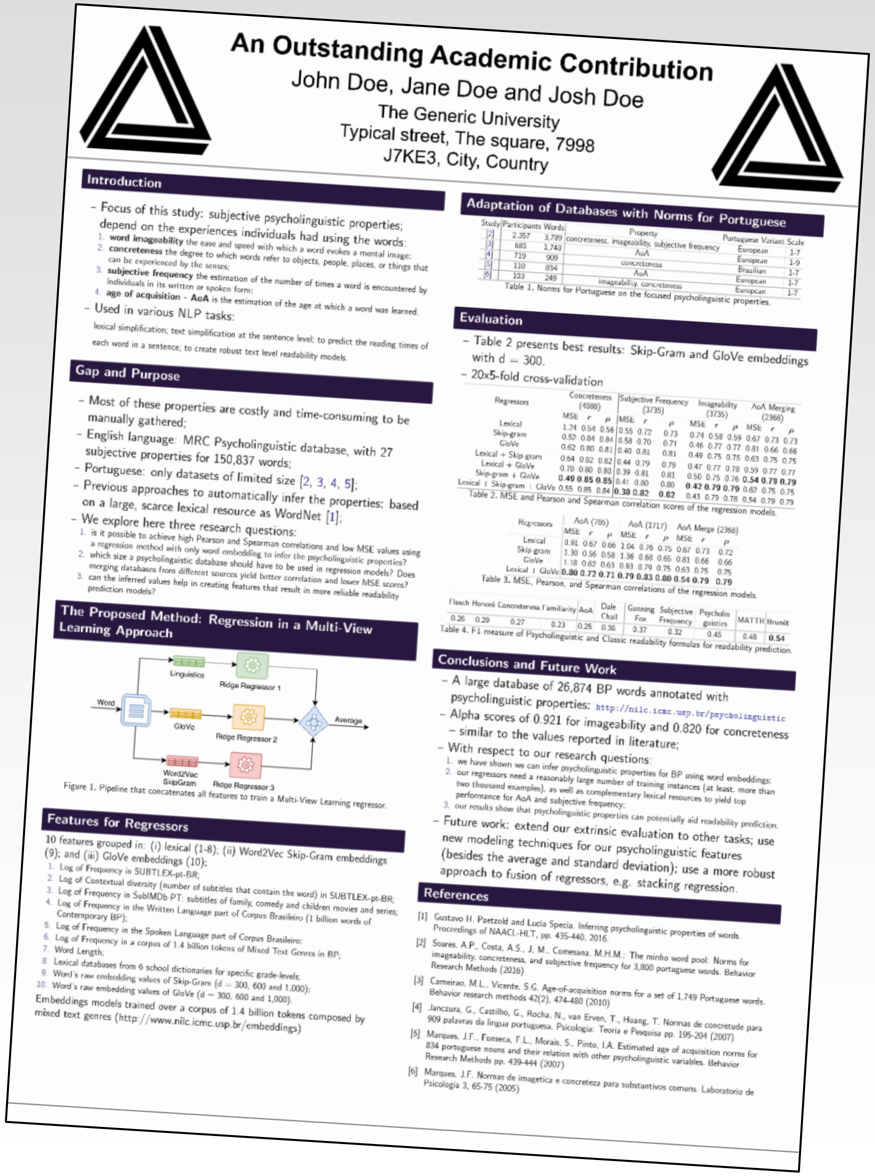
http://nilc.icmc.usp.br/psycholinguistic

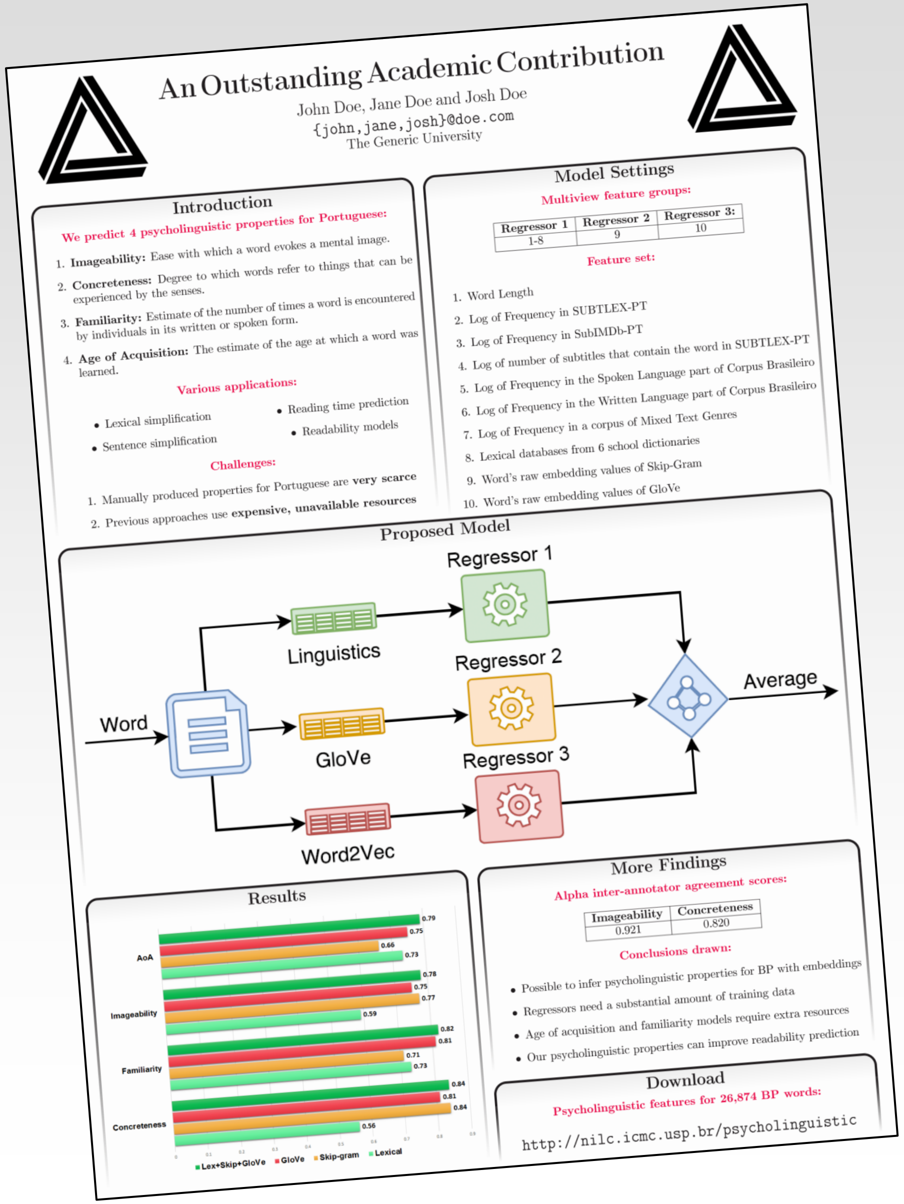
Final Lesson

Final Lesson



Final Lesson





Final Lesson

Final Lesson

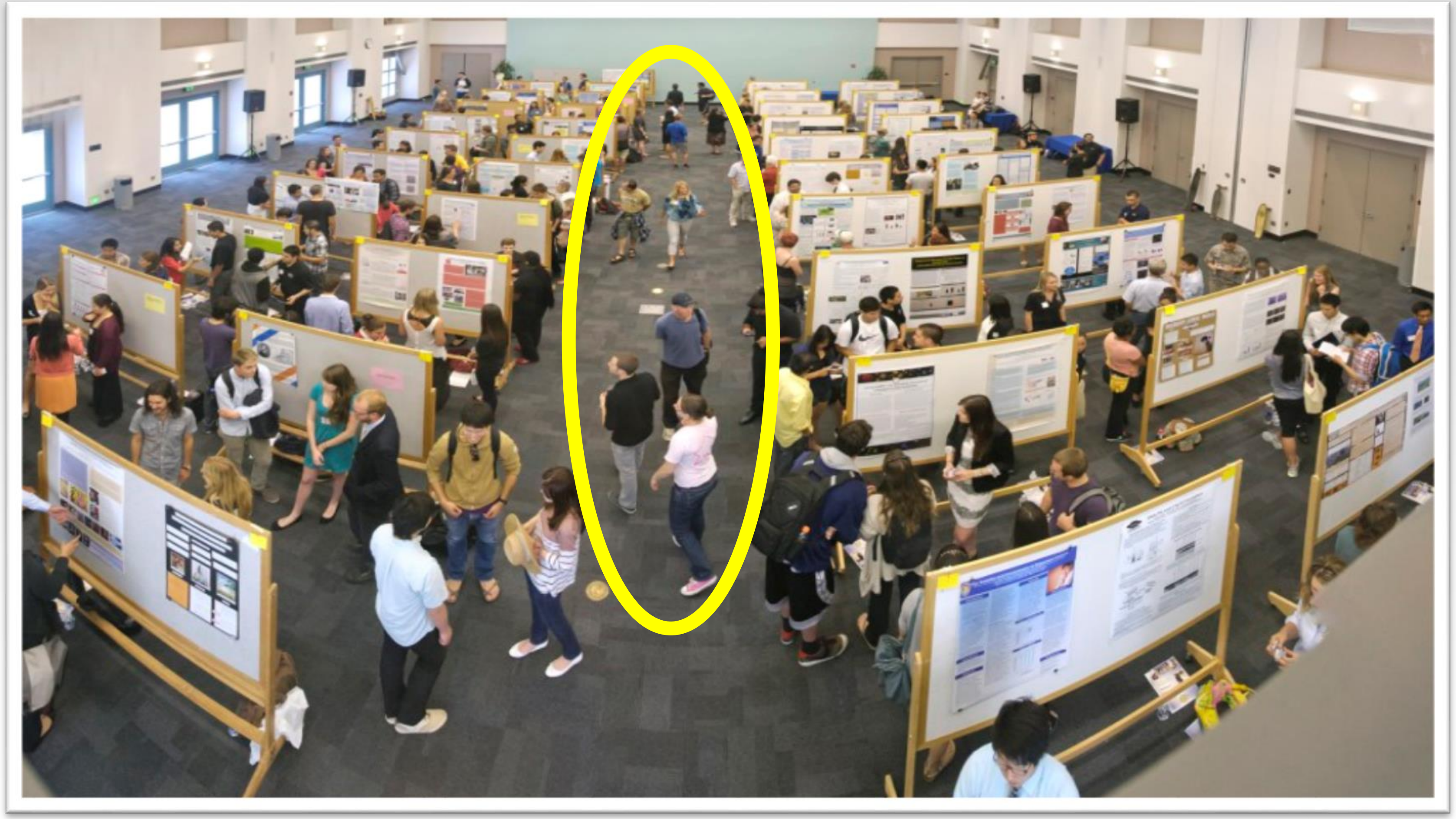
Break time!

Oral Presentations: Part 1

Oral Presentations: Part 1

Slides: The Challenge

Slides: The Challenge



Slides: The Challenge

Slides: The Challenge

Slides:

________ people's attention

Slides: The Challenge

Slides:

<u>Keeping</u> people's attention

Oral Presentations: Part 1

Oral Presentations: Part 1

Slides: The Problems

Slides: The Problems

1. Too much ____

Slides: The Problems

1. Too much **text**

Experiment 3: Does it works for simple (AKA error-prone) heuristics?

- We used LDC Chinese-English dictionary to generate high-precision-low-recall partial alignments
- The entries with single Chinese character or more than six English words are filtered out.
- Add links when a lexicon entry was encountered in the sentence pair
- 79.48% precision and 17.36% recall rate

Conclusion

- We implemented a semi-supervised word alignment algorithm based on IBM models which can use partial word alignment.
- Experiments were performed to prove that:
  - 1. The algorithm can correct more links than directly fixing the incorrect links
  - 2. Better alignment quality can be achieved by carefully selecting words to ask the oracle
  - 3. By supplying high-precision-low-recall alignment links the alignment quality can also be improved.

Slides: The Problems

2. Bland ______

Slides: The Problems

2. Bland **styling**

1. Motivation
2. Basics
   - Commands
   - Document Structure
   - Running LATEX
3. Controlling Appearance
   - Making Lists
   - Fonts, Symbols, quotations and footnotes
4. Adding Structure
   - Sections
   - Tables, Figures and Equations
5. BIBTEX

Slides: The Problems

3. Some slides are ___________

Slides: The Problems

3. Some slides are <u>**unnecessary**</u>

The Outline

1. Motivation
2. Basics
   - Commands
   - Document Structure
   - Running LATEX
3. Controlling Appearance
   - Making Lists
   - Fonts, Symbols, quotations and footnotes
4. Adding Structure
   - Sections
   - Tables, Figures and Equations
5. BIBTEX

Slides: The Problems

When to use:

- Day-long meetings

Slides: The Problems

When to use:

- Day-long meetings
- Day-long tutorials

Slides: The Problems

When to use:

- Day-long meetings
- Day-long tutorials
- Academic lectures

Slides: The Problems

When to use:

- Day-long meetings
- Day-long tutorials
- Academic lectures
- … anything lengthy with a lot of topics

Slides: The Problems

The References

References

E. P. Blasch, H. J. Garcia, L. Snidaro, J. Llinas, G. Seetharaman, and K. Palaniappan.
Overview of contextual tracking approaches in information fusion.
Proceedings of the SPIE, 8747:87470B, 2013.

Erik P Blasch, Eloi Bosse, and Dale A Lambert.
High-Level Information Fusion Management and Systems Design.
Artech House; 1 edition, 2012.

R.I. Davis, K.W. Tindell, and A. Burns.
Scheduling slack time in fixed priority pre-emptive systems.
In *1993 Proceedings Real-Time Systems Symposium*, pages 222–231. IEEE Comput. Soc. Press.

Gee Ng, Chung Tan, Thiam Ng, and Shun Siow.
Assessment Of Data Fusion Systems.
In *2006 9th International Conference on Information Fusion*, pages 1–8. IEEE, jul 2006.

Shlomo Zilberstein.

Slides: The Problems

When to use:

- Academic lectures

Slides: The Problems

When to use:

- Academic lectures
- ... that’s it.

Slides: The Problems

4. Too much ____________ per slide

Slides: The Problems

4. Too much **information** per slide

A better way

- Let the knowledge determine the known part, and let models determine the rest.
- The knowledge will:
  - Affect the statistics we get for the model
  - Be reflected in the final alignment

Anything conflicting with known alignments should be forbidden

- Pereira and Schabes, 1992, Similar idea on SCFGs

Slides: The Problems

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

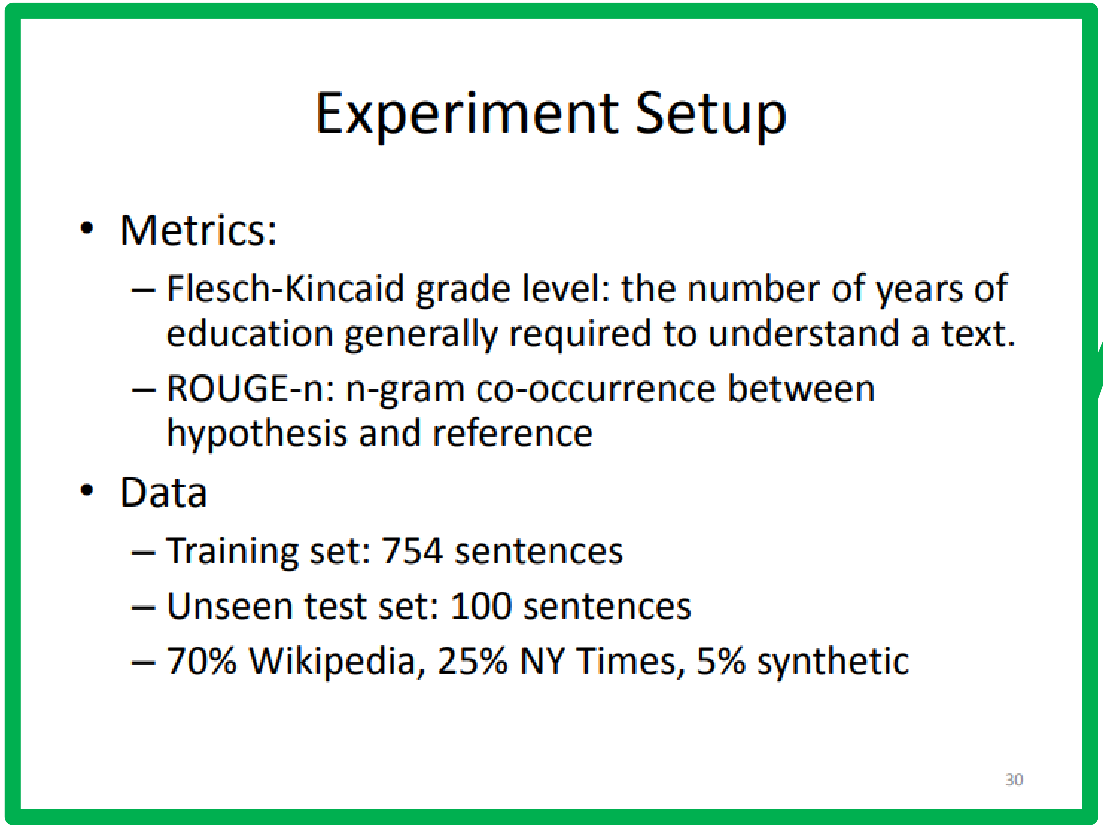
Slides: The Problems

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

30

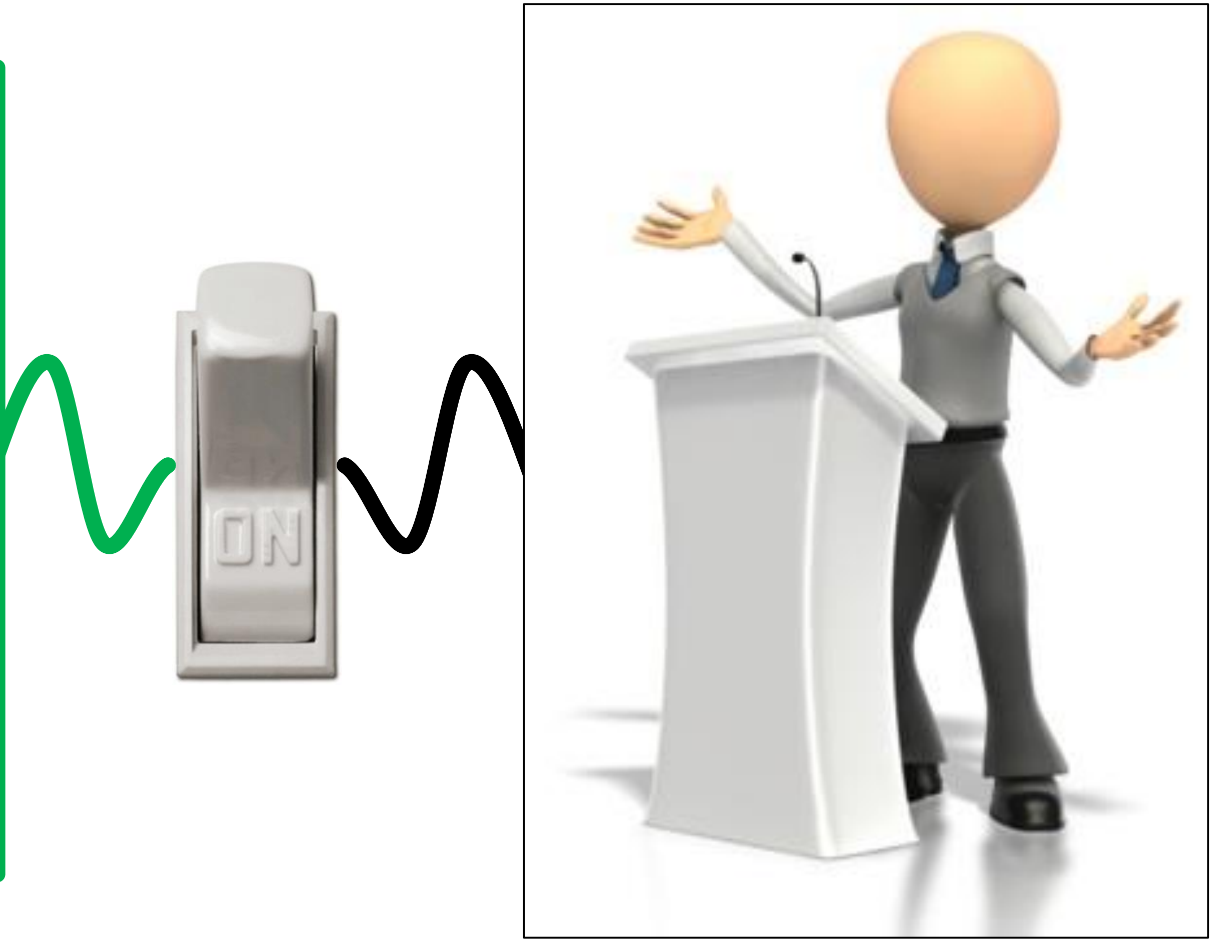
Slides: The Problems



Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

30

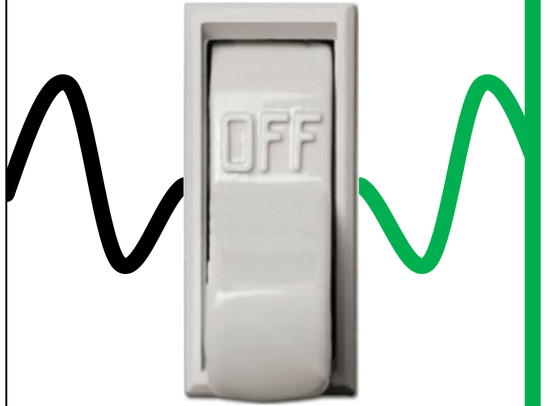


Slides: The Problems

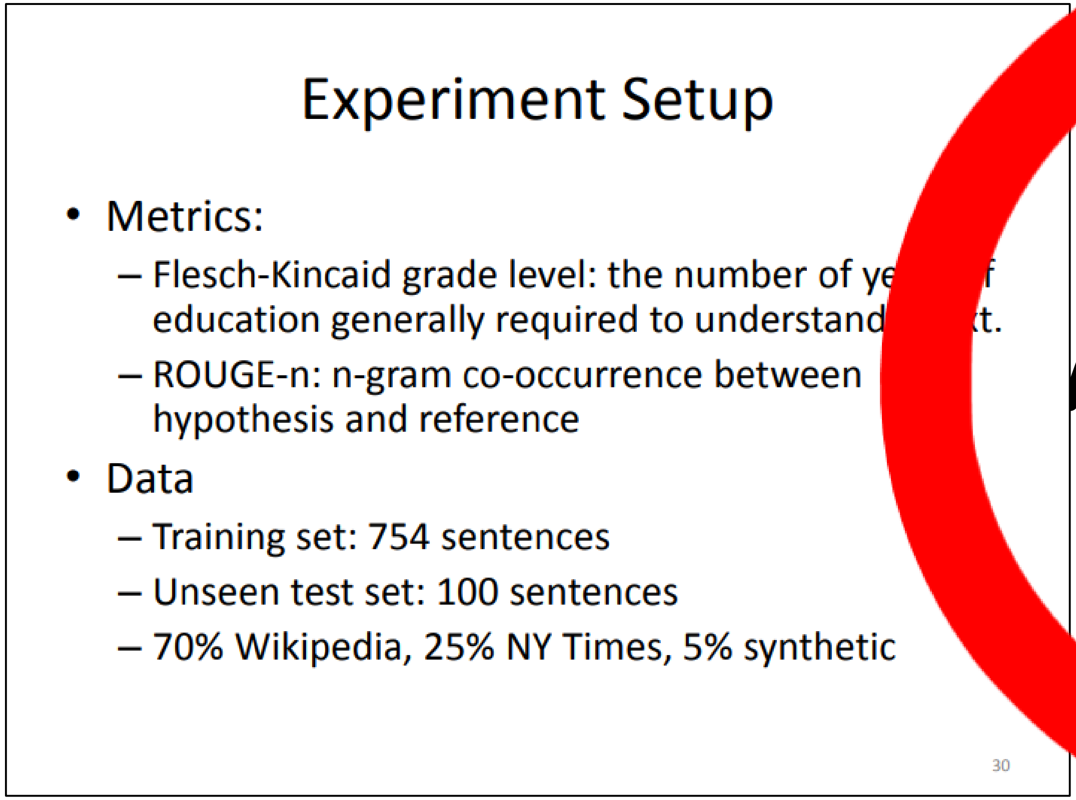
Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

30



Slides: The Problems



Slides: The Problems

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

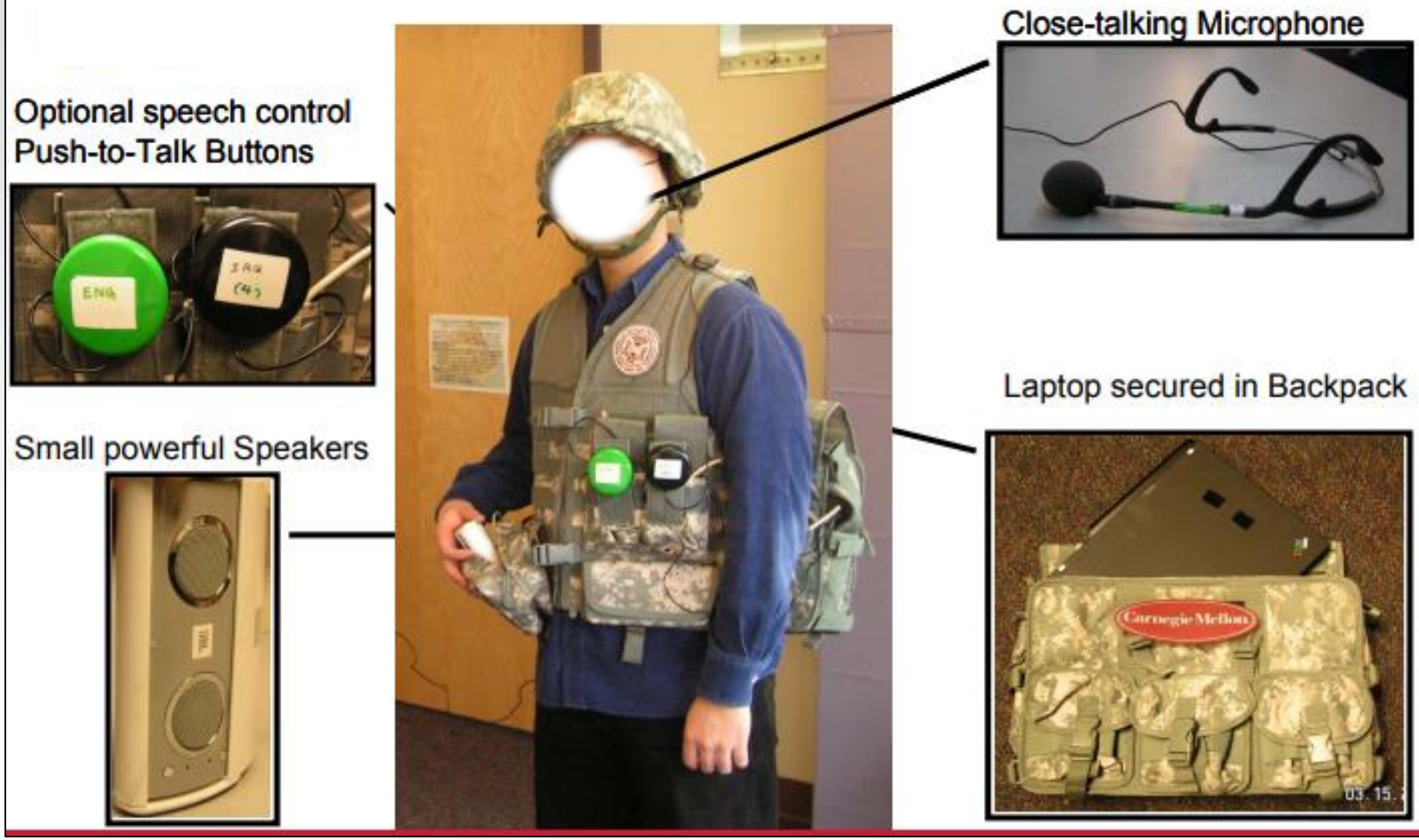
Slides: The Problems

Experiment Setup

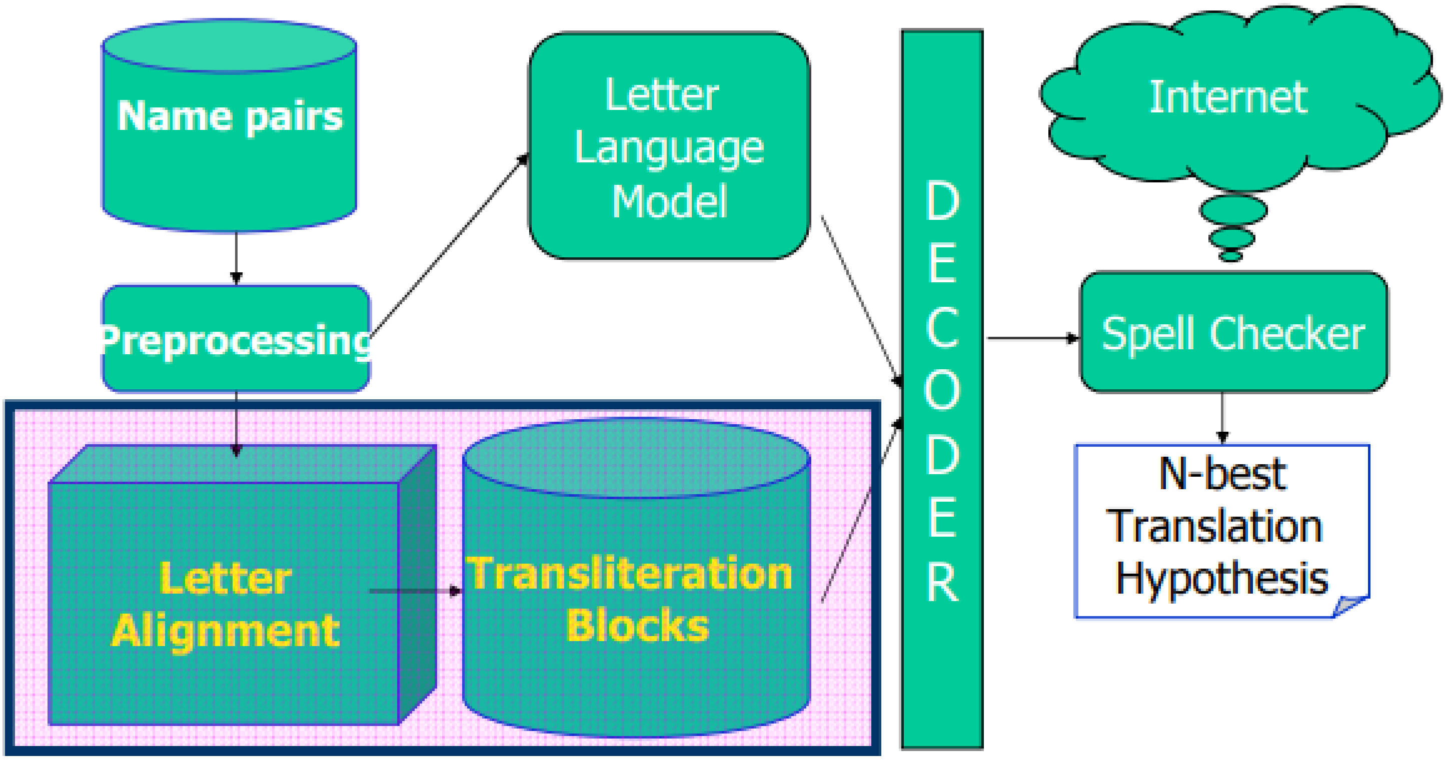
- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

30

Helps


Close-talking Microphone
Optional speech control
Push-to-Talk Buttons
ENG
Laptop secured in Backpack
Small powerful Speakers
Carnegie Mellon

System Architecture



Oral Presentations: Part 1

Oral Presentations: Part 1

Slides: The Solutions

Slides: The Solutions

1. Too much **text**

Slides: The Solutions

- Summarize it
- Make it visual

List ➡ Table

List → Table

Table → Graph

List → Table

Table → Graph

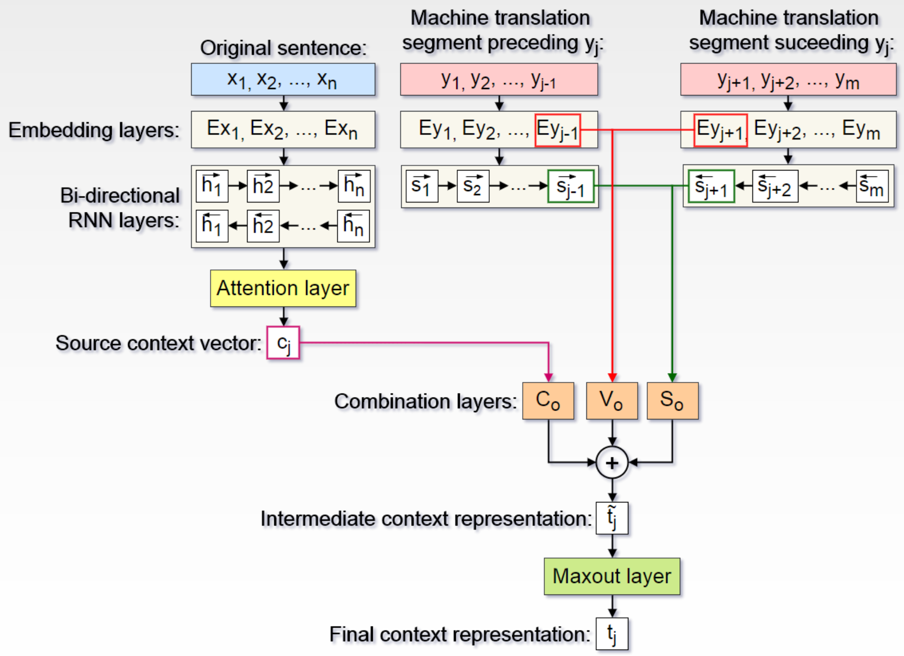
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

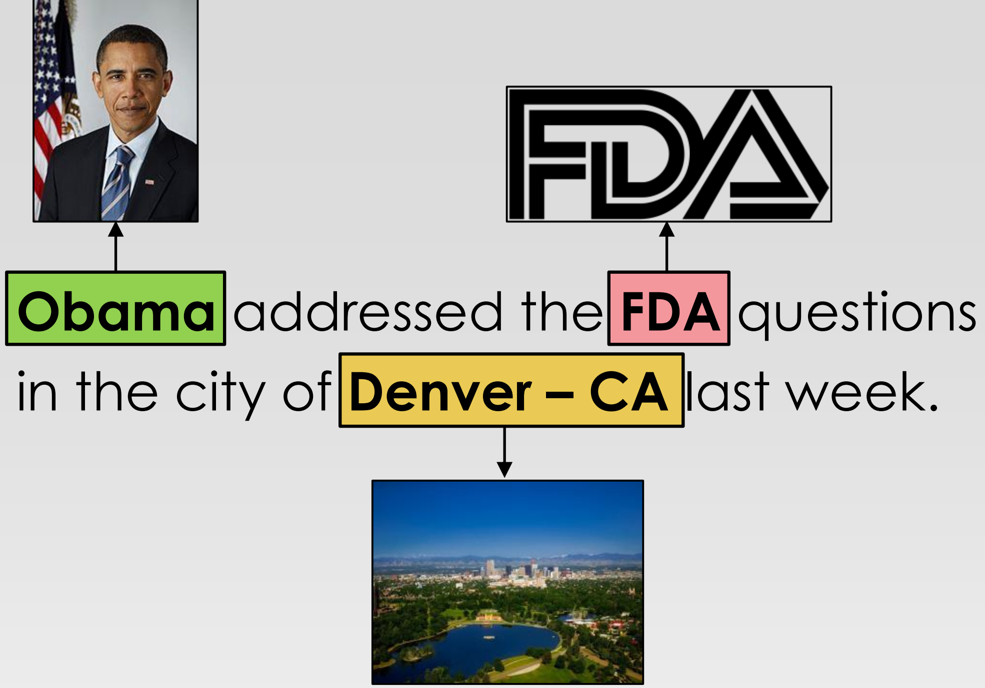
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

→



Named-entity recognition (NER) (also known as entity identification, entity chunking and entity extraction) is a subtask of information extraction that seeks to locate and classify named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.



Named-entity recognition (NER) (also known as entity identification, entity chunking and entity extraction) is a subtask of information extraction that seeks to locate and classify named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

FDA

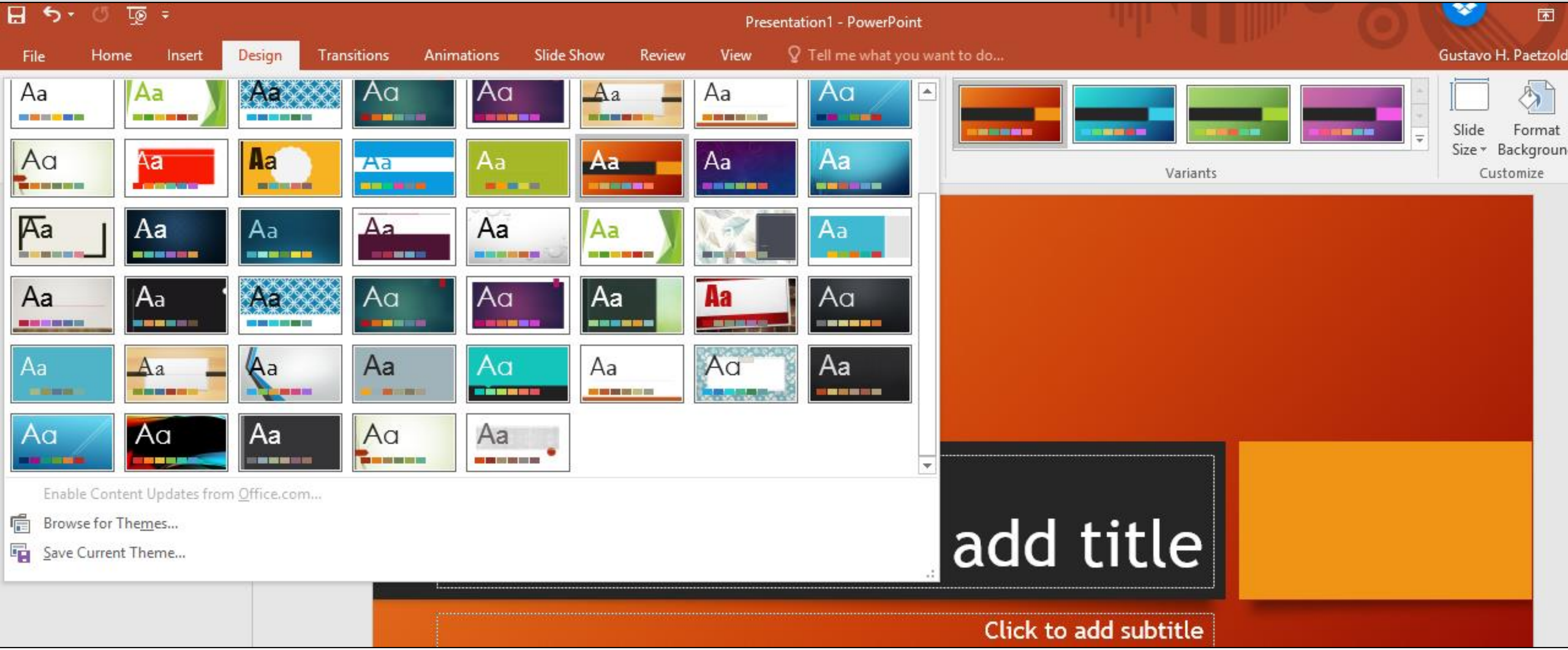
Obama addressed the **FDA** questions in the city of **Denver – CA** last week.

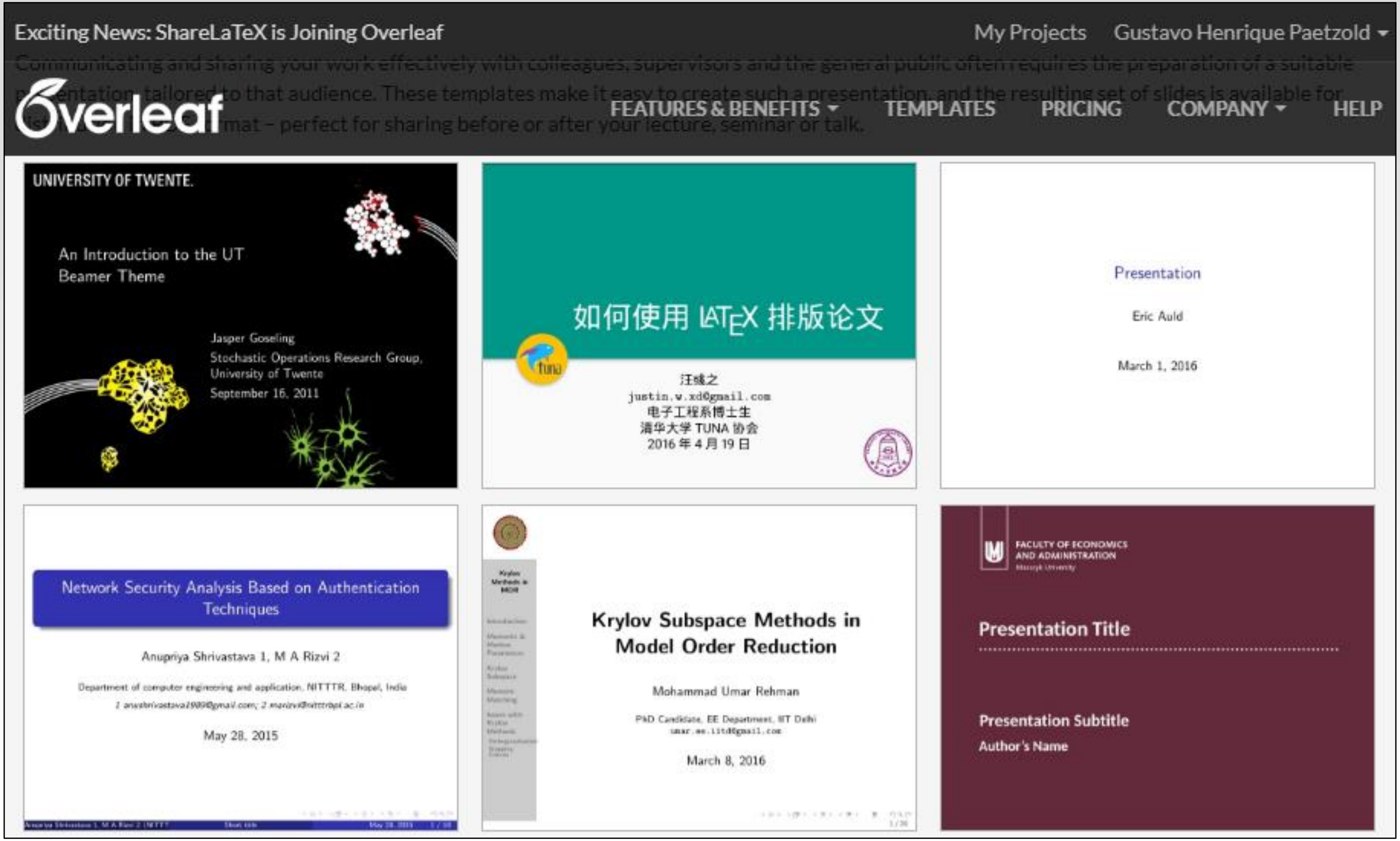
Anita

Anita

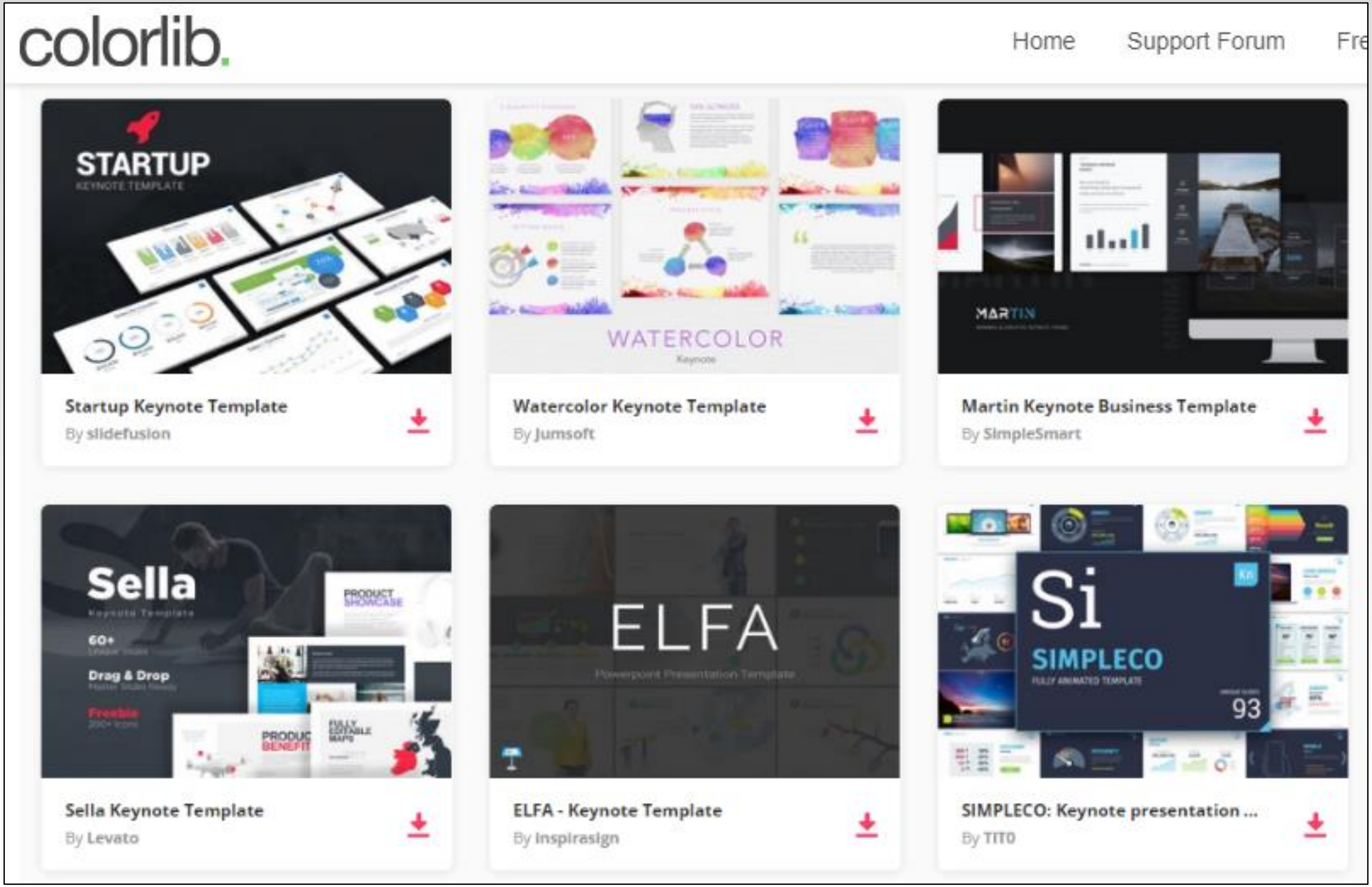
Slides: The Solutions

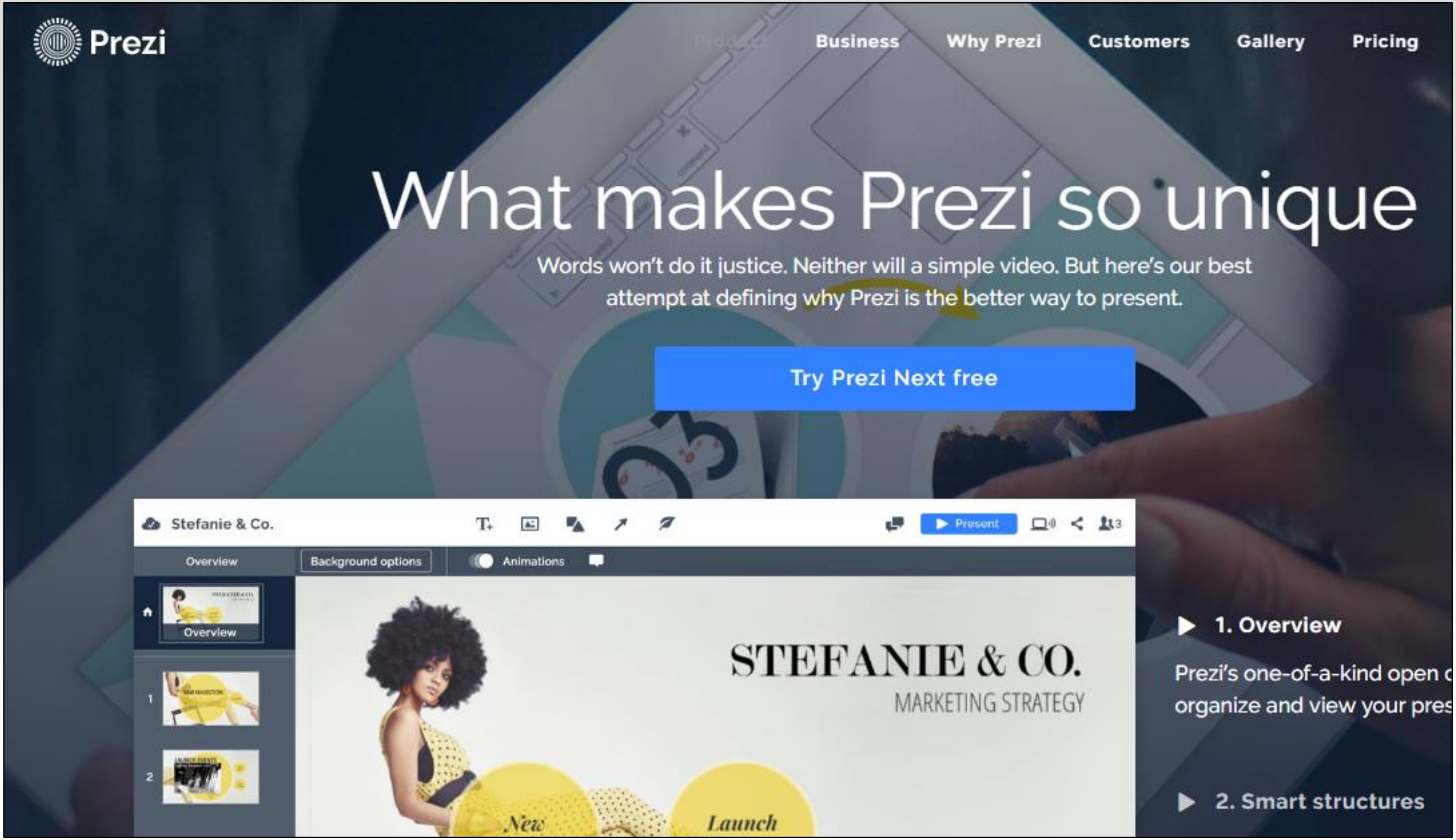
2. Bland **<u>styling</u>**

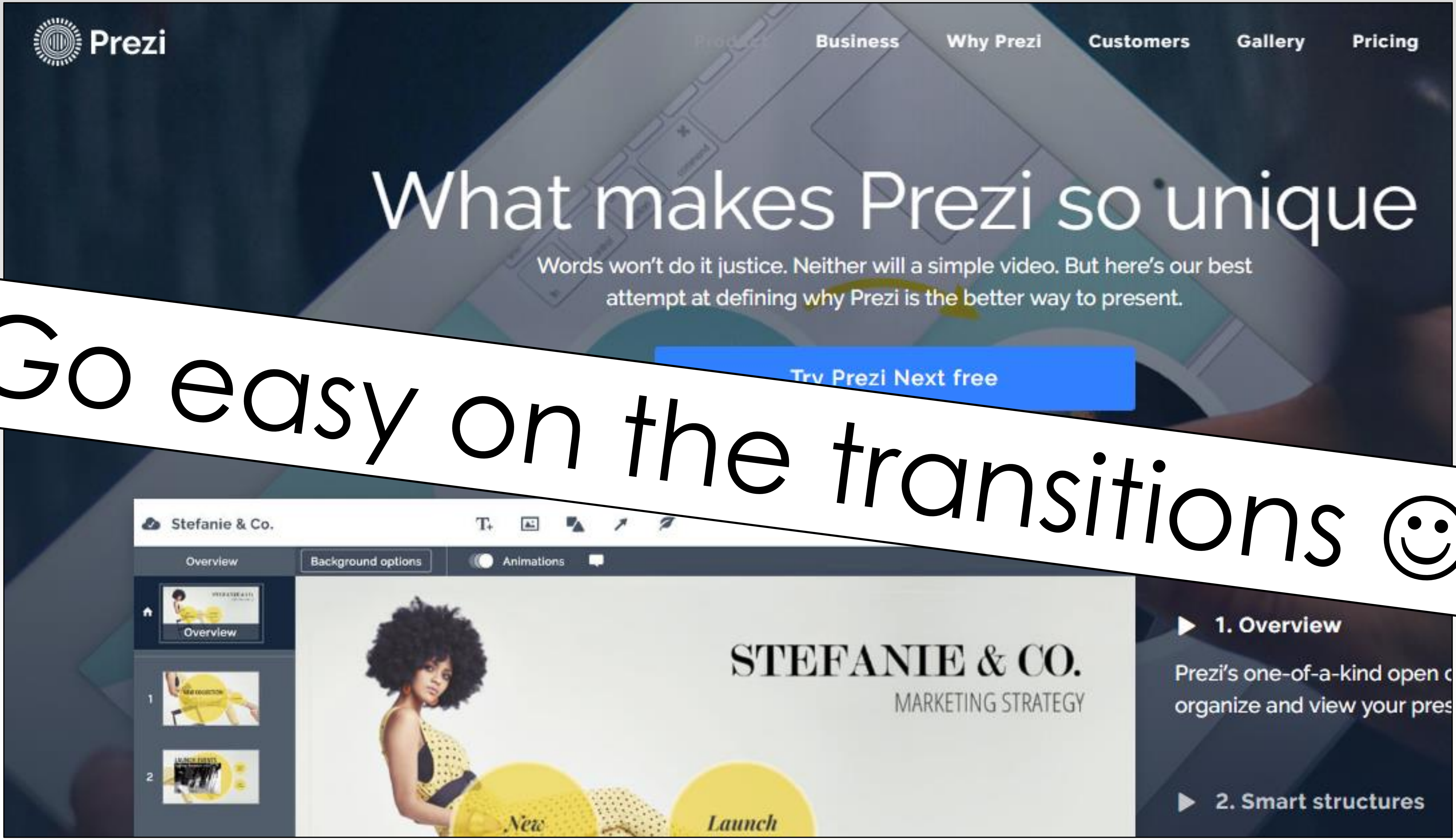

Presentation1 - PowerPoint
File
Home
Insert
Design
Transitions
Animations
Slide Show
Review
View
Tell me what you want to do...
Gustavo H. Paetzold
Variants
Slide Size
Format Backgroun
Customize
Enable Content Updates from Office.com...
Browse for Themes...
Save Current Theme...
add title
Click to add subtitle



Exciting News: ShareLaTeX is Joining Overleaf
My Projects
Gustavo Henrique Paetzold
Overleaf
FEATURES & BENEFITS
TEMPLATES
PRICING
COMPANY
HELP
UNIVERSITY OF TWENTE.
An Introduction to the UT Beamer Theme
Jasper Goseling
Stochastic Operations Research Group, University of Twente
September 16, 2011
如何使用 LaTeX 排版论文
汪缕之
justin.w.xd@gmail.com
电子工程系博士生
清华大学 TUNA 协会
2016 年 4 月 19 日
Presentation
Eric Auld
March 1, 2016
Network Security Analysis Based on Authentication Techniques
Anupriya Shrivastava 1, M A Rizvi 2
Department of computer engineering and application, NITTTR, Bhopal, India
May 28, 2015
Krylov Subspace Methods in Model Order Reduction
Mohammad Umar Rehman
PhD Candidate, EE Department, IIT Delhi
March 8, 2016
FACULTY OF ECONOMICS AND ADMINISTRATION
Presentation Title
Presentation Subtitle
Author's Name


colorlib.
Home
Support Forum
Fre
STARTUP
KEYNOTE TEMPLATE
Startup Keynote Template
By slidefusion
WATERCOLOR
Keynote
Watercolor Keynote Template
By Jumsoft
MARTIN
Martin Keynote Business Template
By SimpleSmart
Sella
Keynote Template
60+
Drag & Drop
Freebie
PRODUCT SHOWCASE
FULLY EDITABLE MAPS
Sella Keynote Template
By Levato
ELFA
Powerpoint Presentation Template
ELFA - Keynote Template
By Inspirasign
Si
SIMPLECO
FULLY ANIMATED TEMPLATE
93
SIMPLECO: Keynote presentation ...
By TITO


Prezi
Business
Why Prezi
Customers
Gallery
Pricing
What makes Prezi so unique
Words won't do it justice. Neither will a simple video. But here's our best attempt at defining why Prezi is the better way to present.
Try Prezi Next free
Stefanie & Co.
Present
Overview
Background options
Animations
Overview
STEFANIE & CO.
MARKETING STRATEGY
New
Launch
1. Overview
Prezi's one-of-a-kind open
organize and view your
2. Smart structures



Go easy on the transitions ☺

IP[y]: IPython

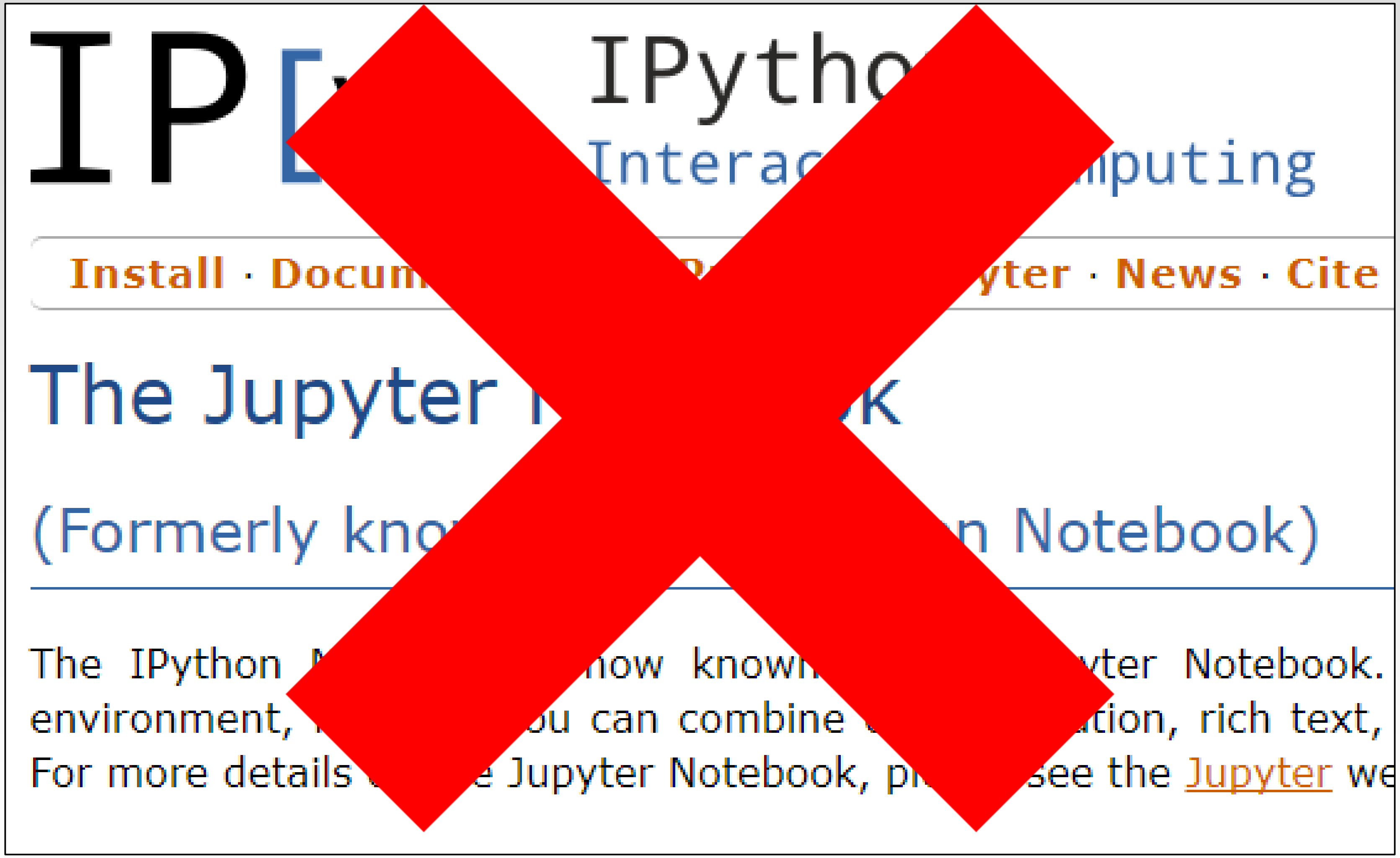
Interactive Computing

Install · Documentation · Project · Jupyter · News · Cite

The Jupyter Notebook

(Formerly known as the IPython Notebook)

The IPython Notebook is now known as the Jupyter Notebook. environment, in which you can combine code execution, rich text, For more details on the Jupyter Notebook, please see the Jupyter we


IP
IPytho
Interac
puting
Install · Docum
yter · News · Cite
The Jupyter
k
(Formerly kno
n Notebook)
The IPython
now known
ter Notebook.
environment,
u can combine
tion, rich text,
For more details
e Jupyter Notebook, p
see the Jupyter we

Slides: The Solutions

3. Some slides are <u>**unnecessary**</u>

Slides: The Solutions

- The outline

Slides: The Solutions

- The outline
- The references

Slides: The Solutions

DELETE THEM

Slides: The Solutions

Things to keep:

Slides: The Solutions

Things to keep:

1. Introduction

Slides: The Solutions

Things to keep:

1. Introduction
2. Motivation/Challenges

Slides: The Solutions

Things to keep:

1. Introduction
2. Motivation/Challenges
3. Approach

Slides: The Solutions

Things to keep:

1. Introduction
2. Motivation/Challenges
3. Approach
4. Experimentation setup

Slides: The Solutions

Things to keep:

1. Introduction
2. Motivation/Challenges
3. Approach
4. Experimentation setup
5. Results

Slides: The Solutions

Things to keep:

1. Introduction
2. Motivation/Challenges
3. Approach
4. Experimentation setup
5. Results
6. Output examples/comparison

Slides: The Solutions

Things to keep:

1. Introduction
2. Motivation/Challenges
3. Approach
4. Experimentation setup
5. Results
6. Output examples/comparison
7. Main conclusions

Slides: The Solutions

Things to keep:

1. Introduction
2. Motivation/Challenges
3. Approach
4. Experimentation setup
5. Results
6. Output examples/comparison
7. Main conclusions
8. Page numbers (without total pages)

Slides: The Solutions

Things to keep:

1. Introduction
2. Motivation/Challenges
3. Approach
4. Experimentation setup
5. Results
6. Output examples/comparison
7. Main conclusions
8. Page numbers (without total pages)

Important for Q&A session

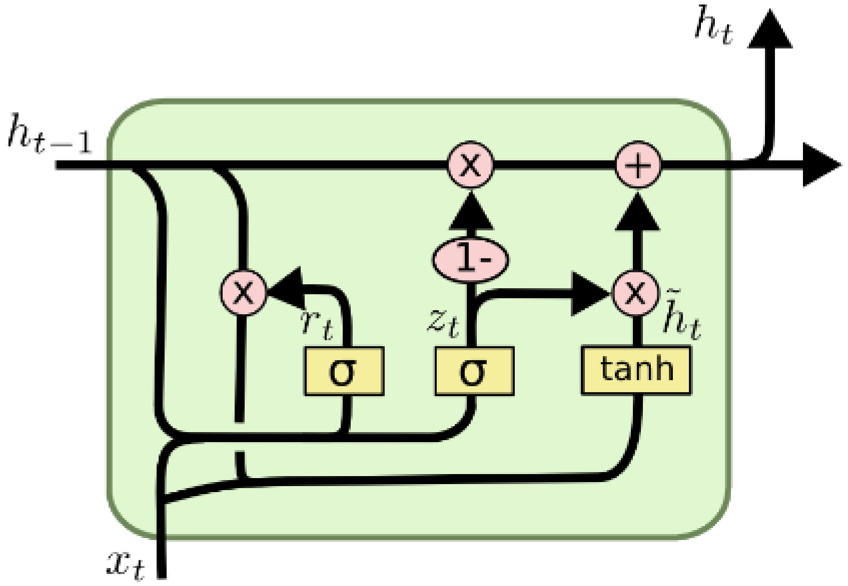
Slides: The Solutions

4. Too much **information** per slide

Slides: The Solutions

Stepification

Gated Recurrent Units



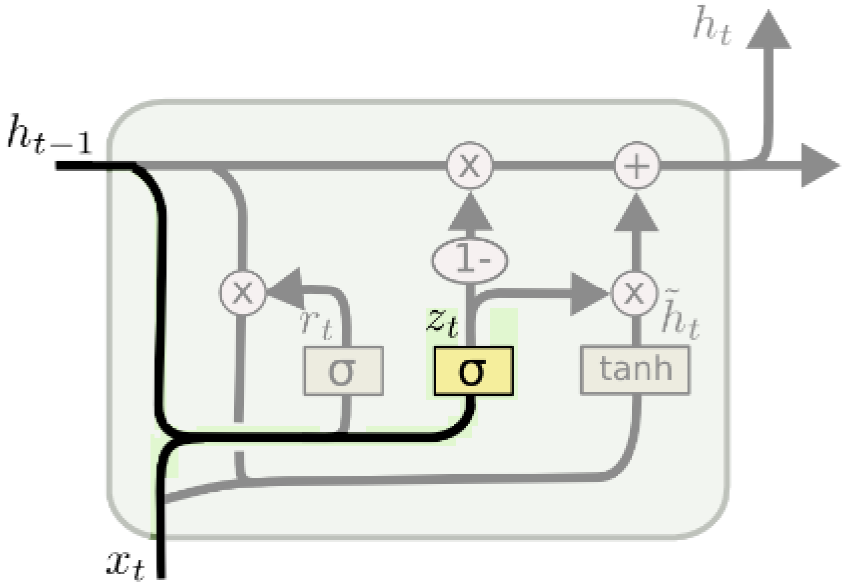
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

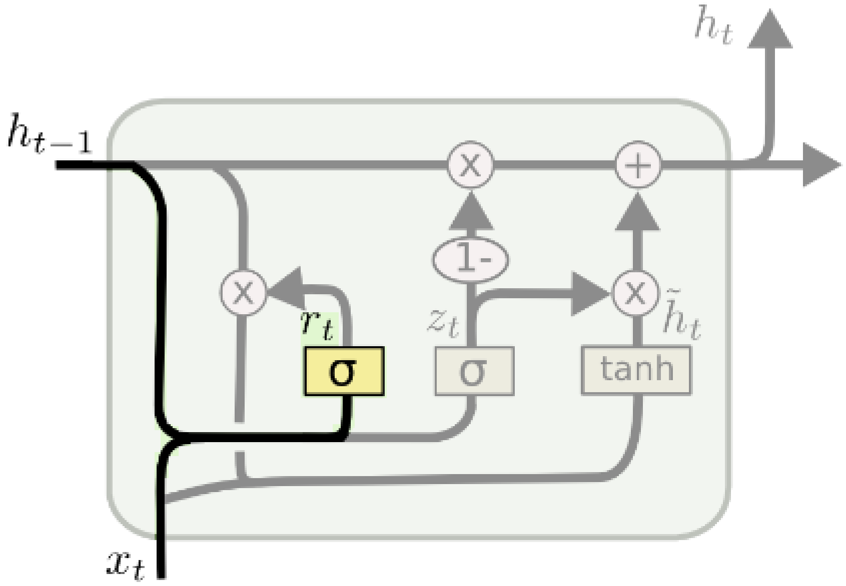
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

The Update Gate



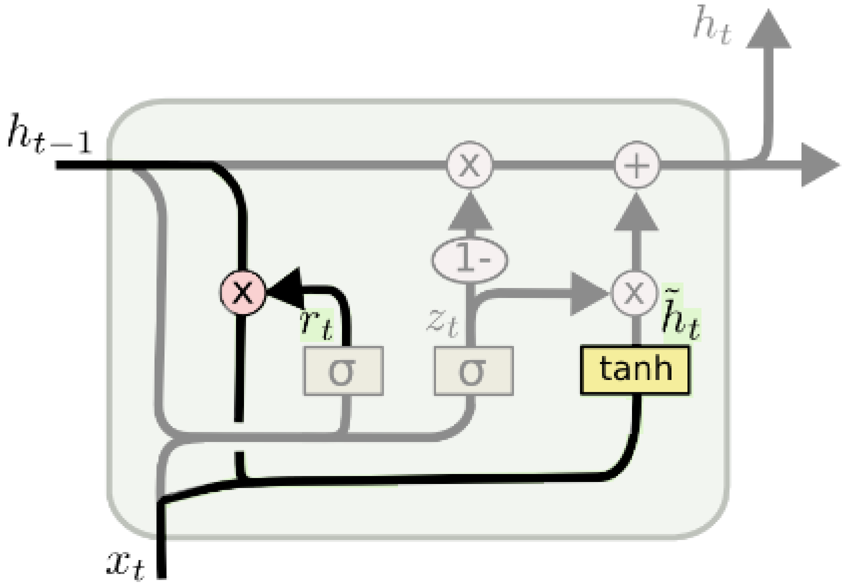
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

The Reset Gate



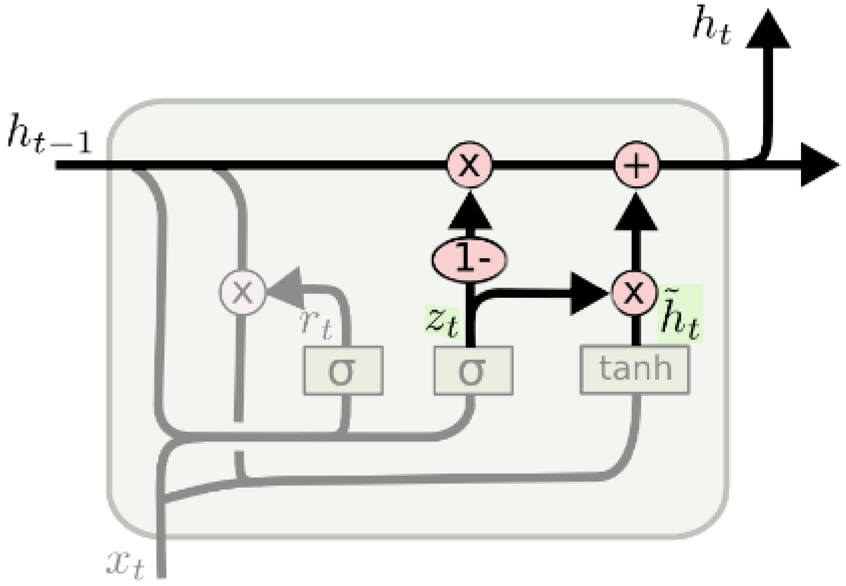
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

The Memory



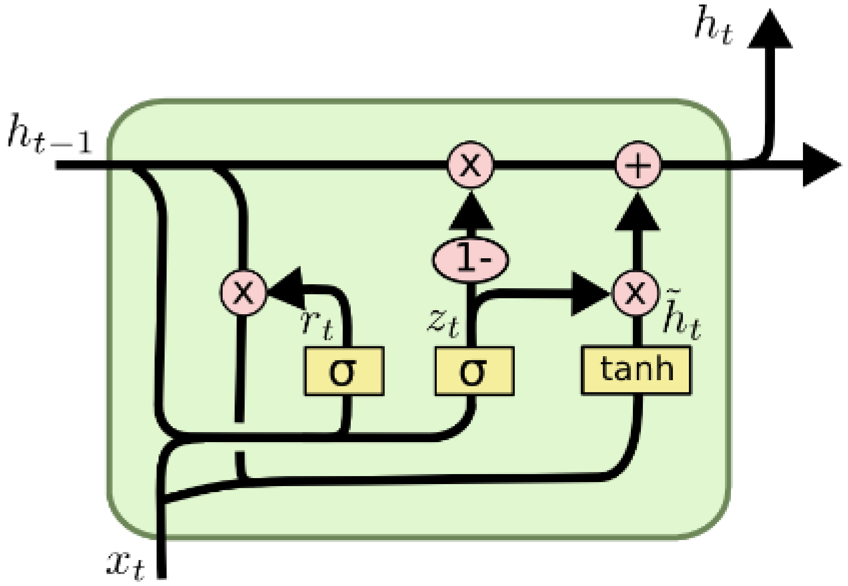
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

The Output



$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Gated Recurrent Units



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Gated Recurrent Units

Output:

$$h_t = (1 - z_t) * h_{t-1} + \boxed{z_t} * \boxed{\tilde{h}_t}$$

Update gate:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

Memory:

$$\tilde{h}_t = \tanh\left(W \cdot [\boxed{r_t} * h_{t-1}, x_t]\right)$$

Reset gate:

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

The Output

Output:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

The Update Gate

Output:

$$h_t = (1 - z_t) * h_{t-1} + \boxed{z_t} * \tilde{h}_t$$

Update gate:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

The Memory

Output:

$$h_t = (1 - z_t) * h_{t-1} + \boxed{z_t} * \boxed{\tilde{h}_t}$$

Update gate:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

Memory:

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

The Reset Gate

Output:

$$h_t = (1 - z_t) * h_{t-1} + \boxed{z_t} * \boxed{\tilde{h}_t}$$

Update gate:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

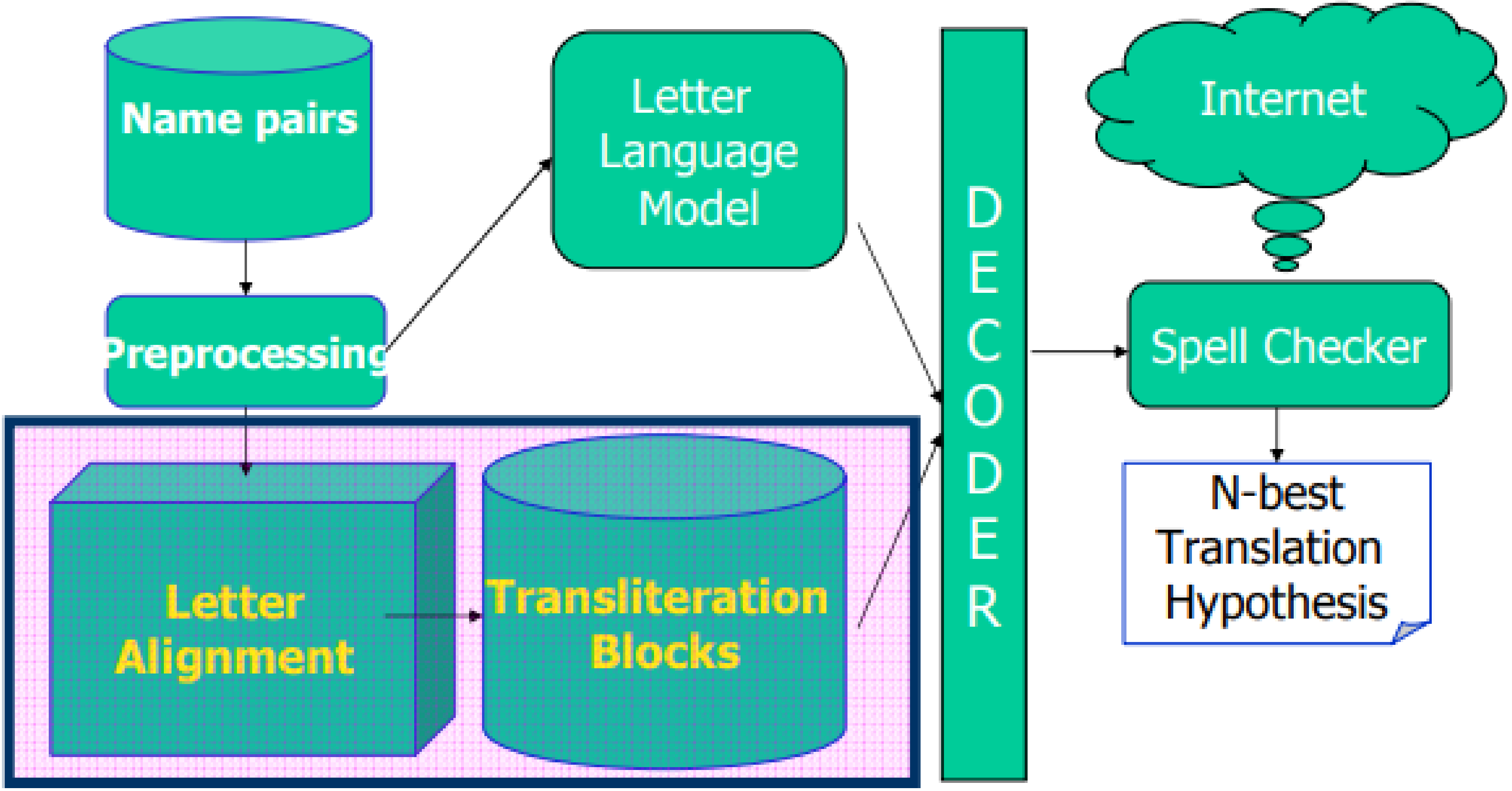
Memory:

$$\tilde{h}_t = \tanh\left(W \cdot [\boxed{r_t} * h_{t-1}, x_t]\right)$$

Reset gate:

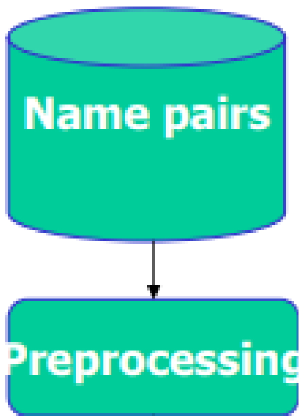
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

System Architecture

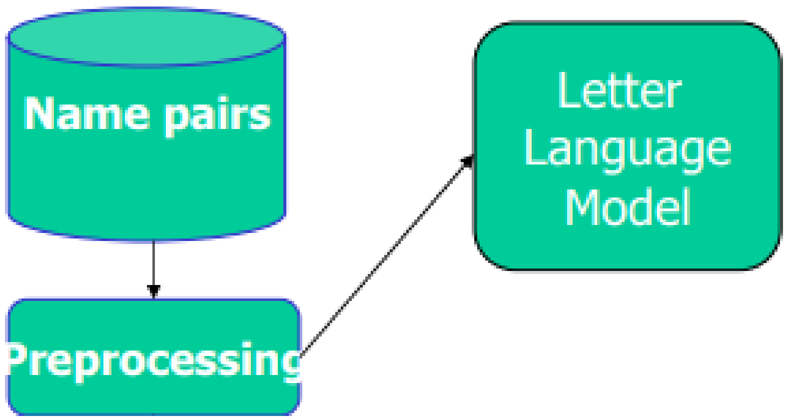


System Architecture

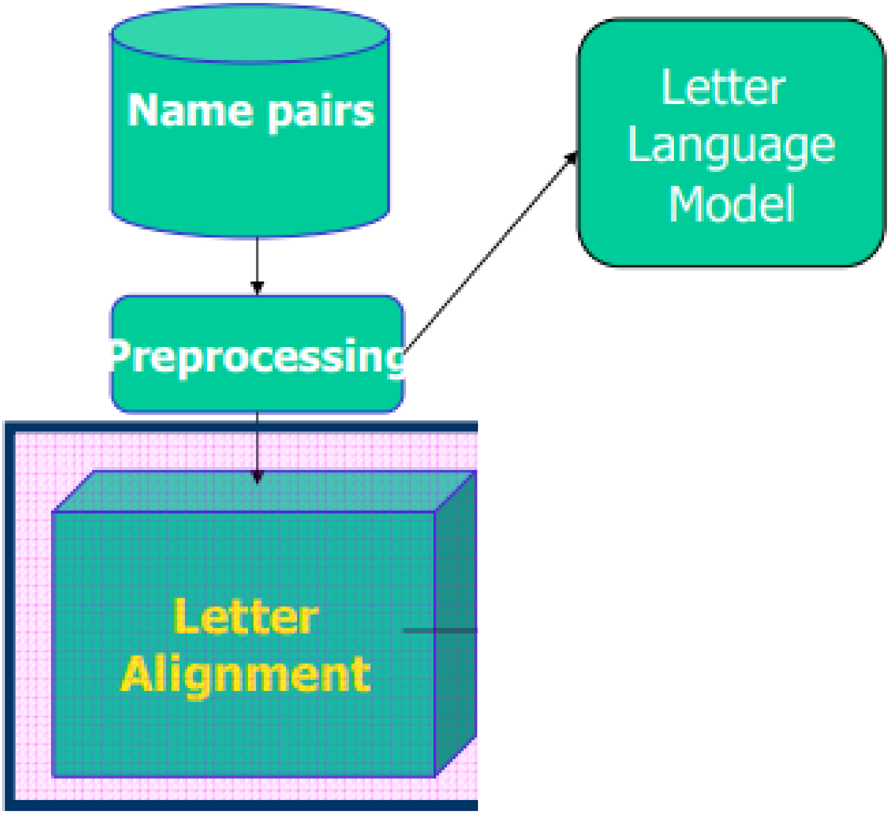
System Architecture



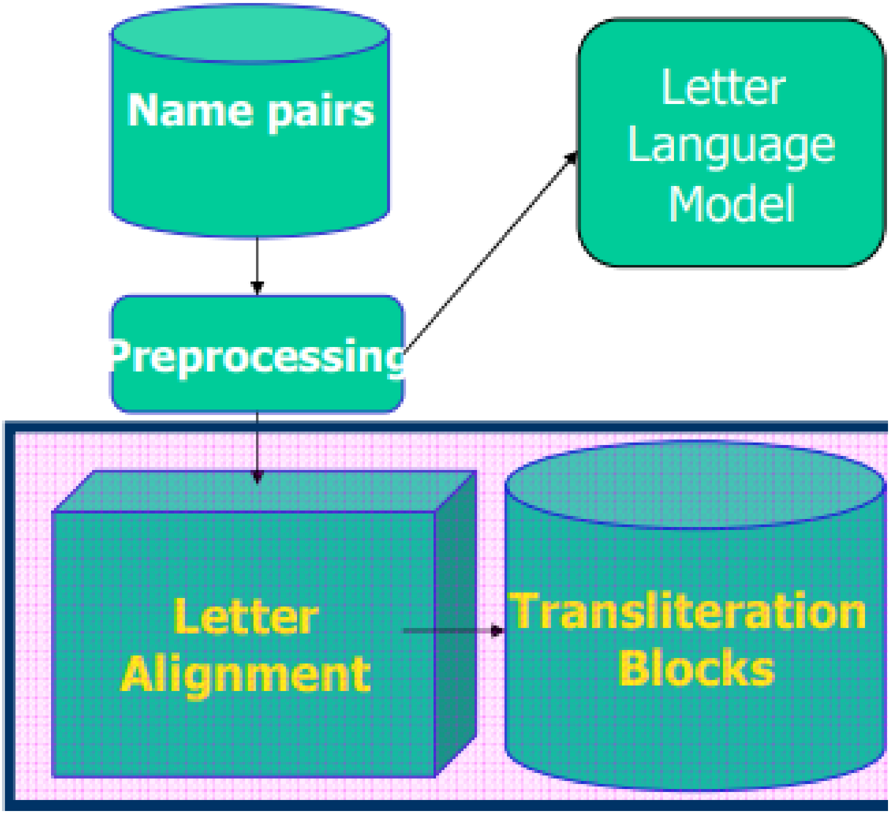
System Architecture



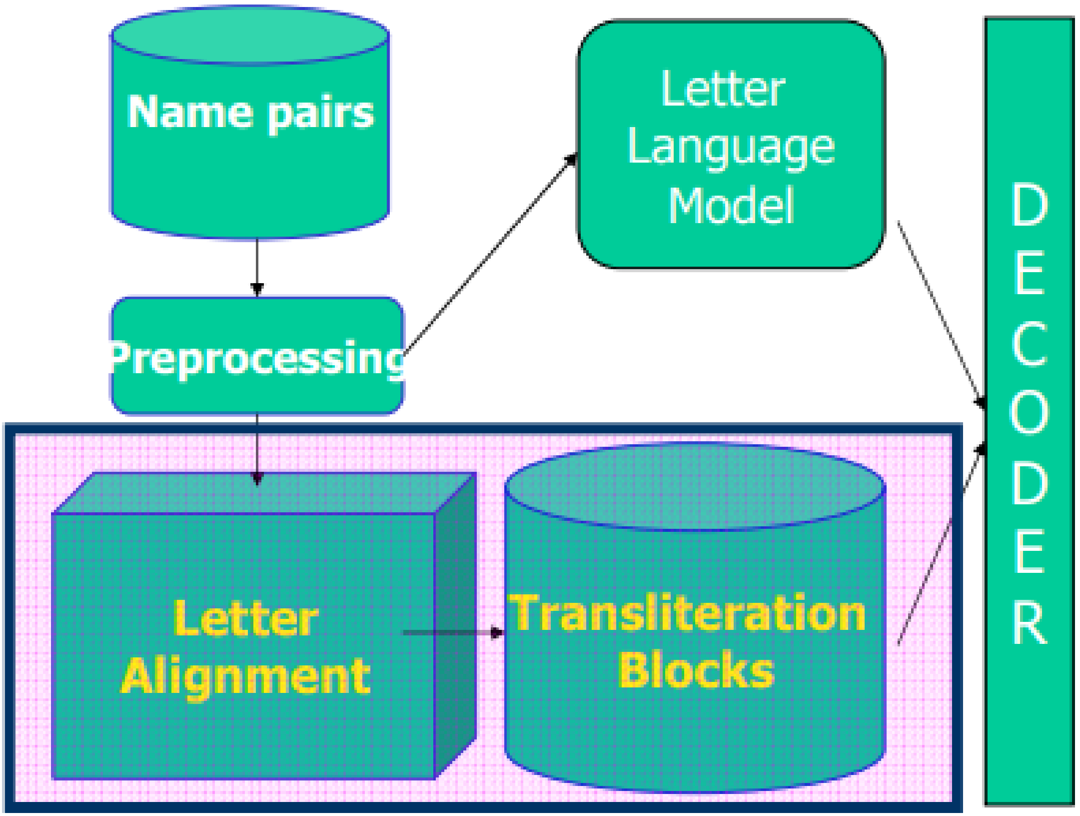
System Architecture



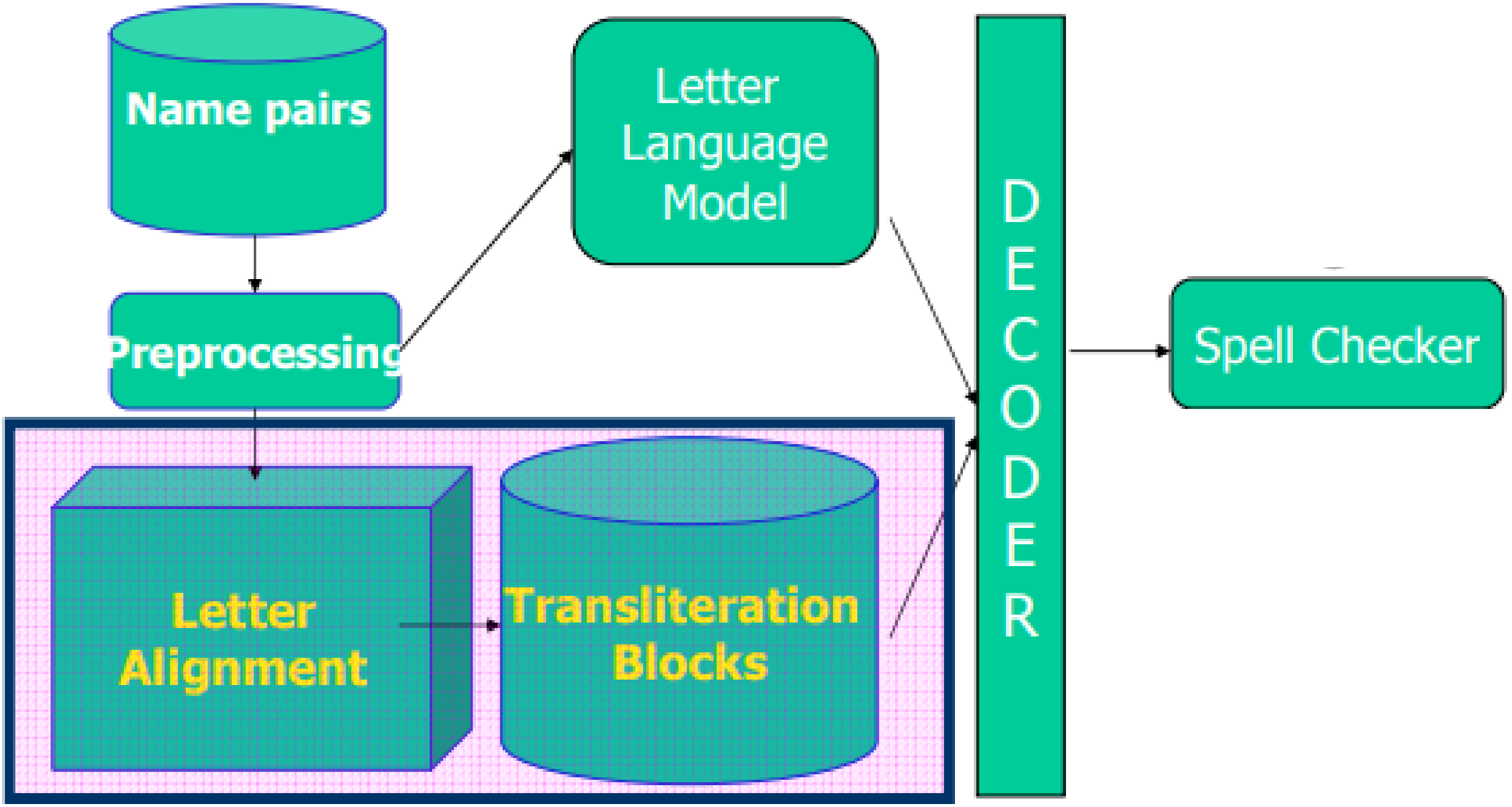
System Architecture



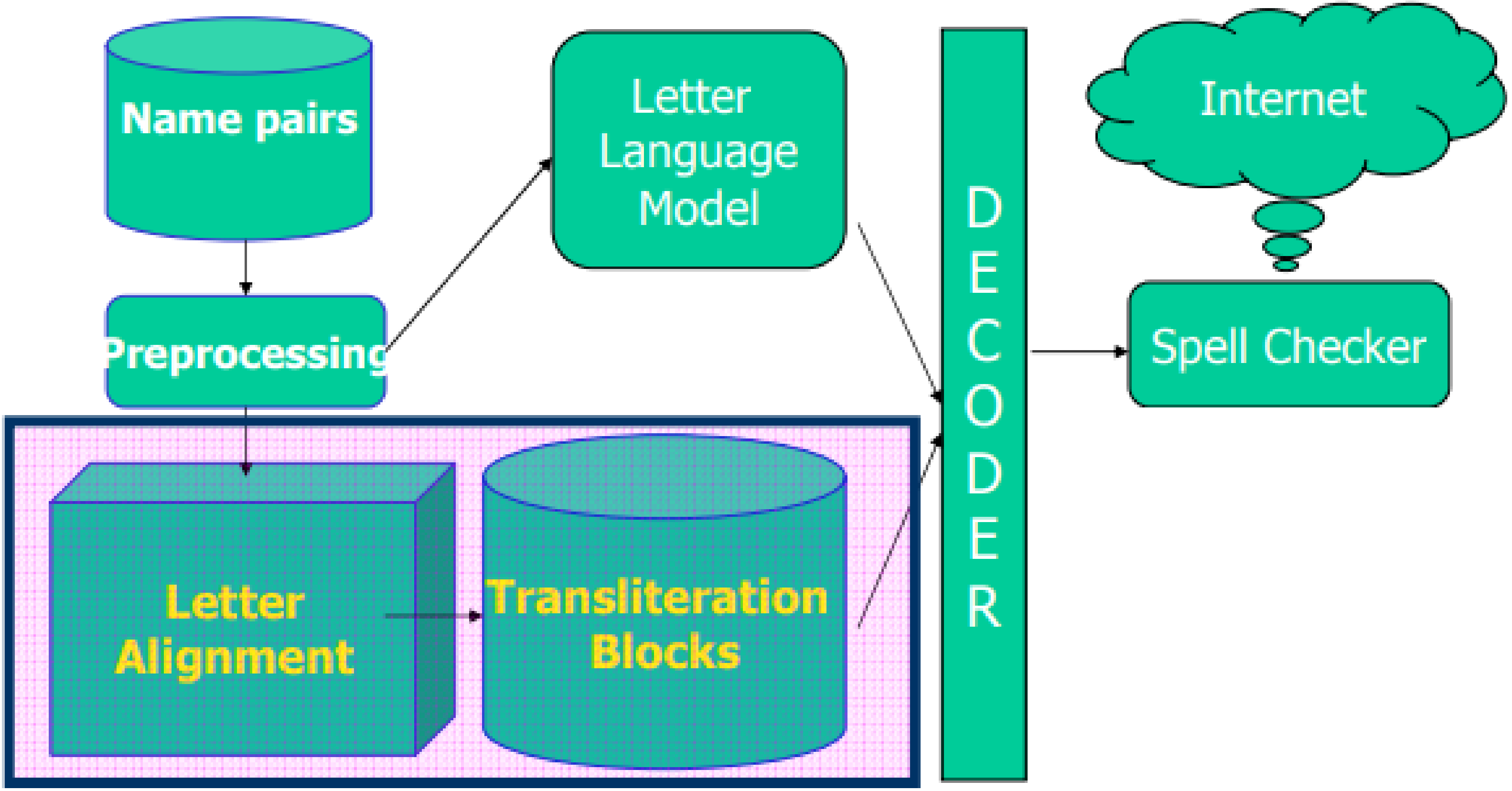
System Architecture



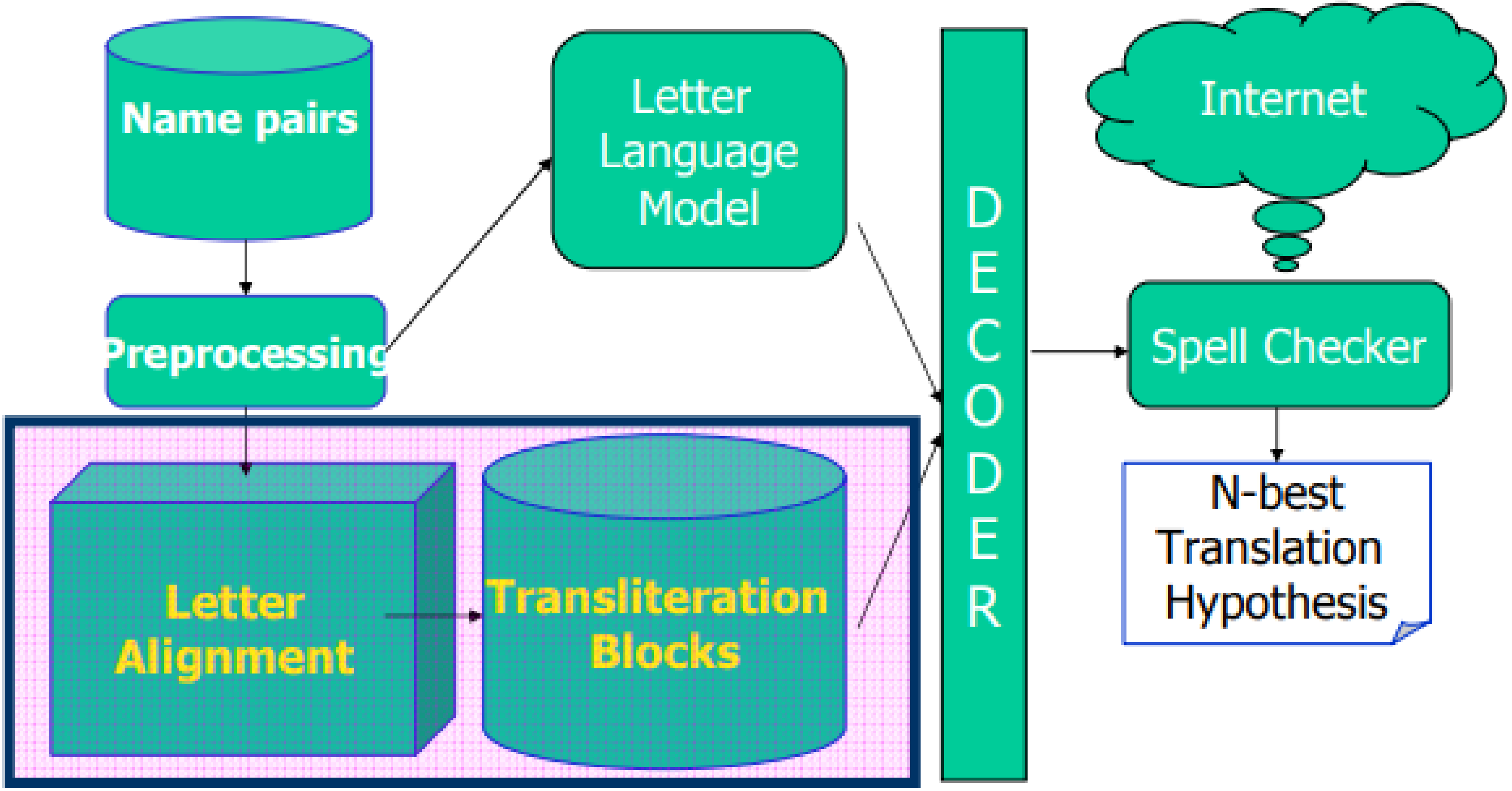
System Architecture



System Architecture



System Architecture



Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

Experiment Setup

Experiment Setup

- Metrics:

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences

Experiment Setup

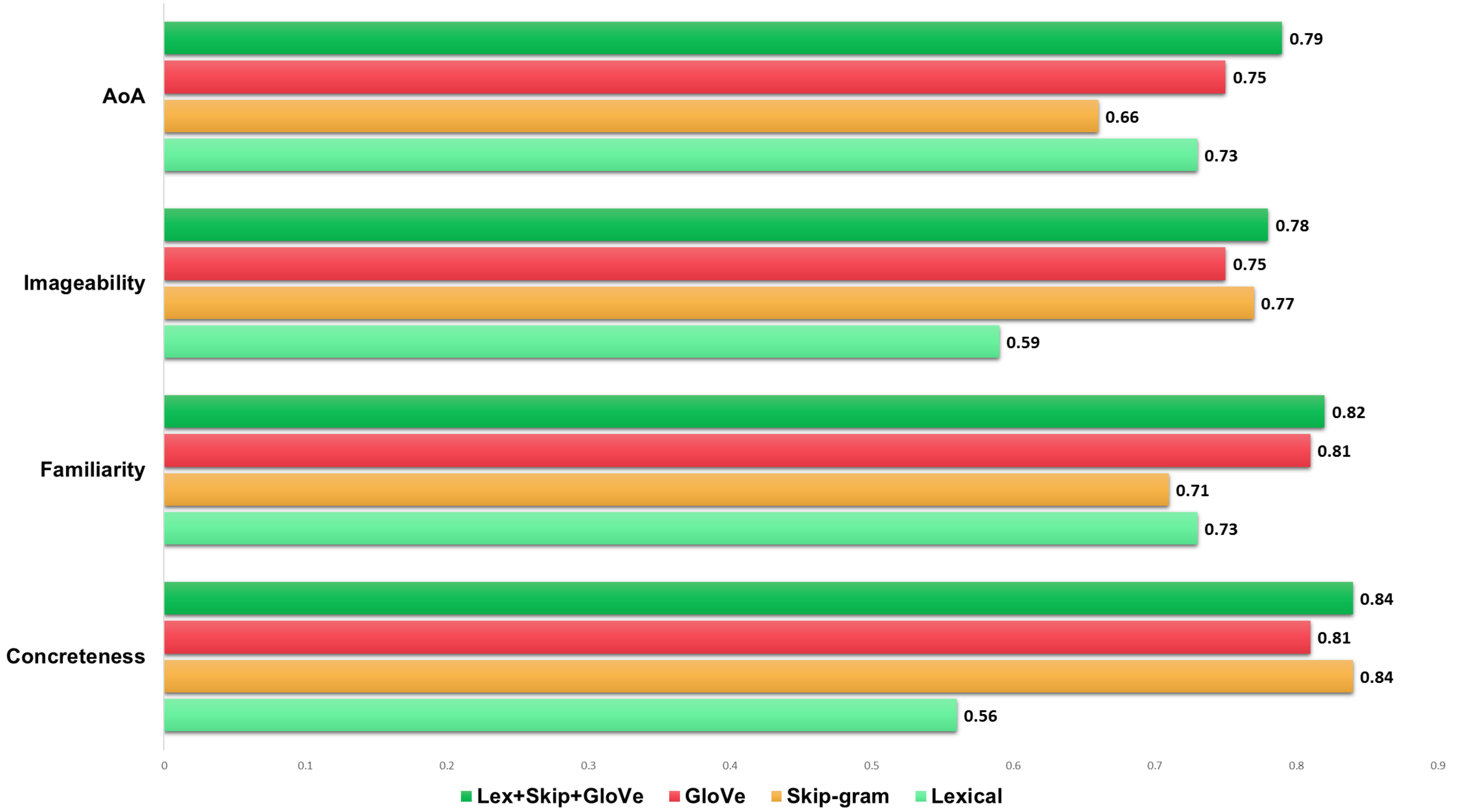
- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia

Experiment Setup

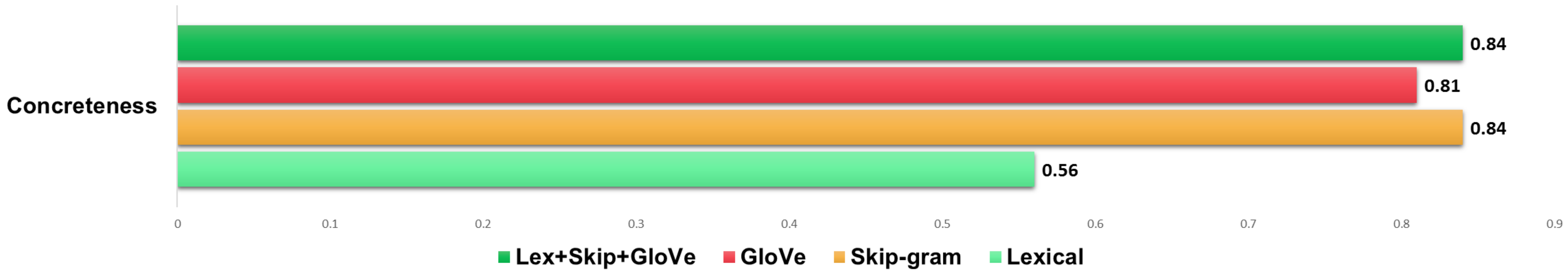
- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times

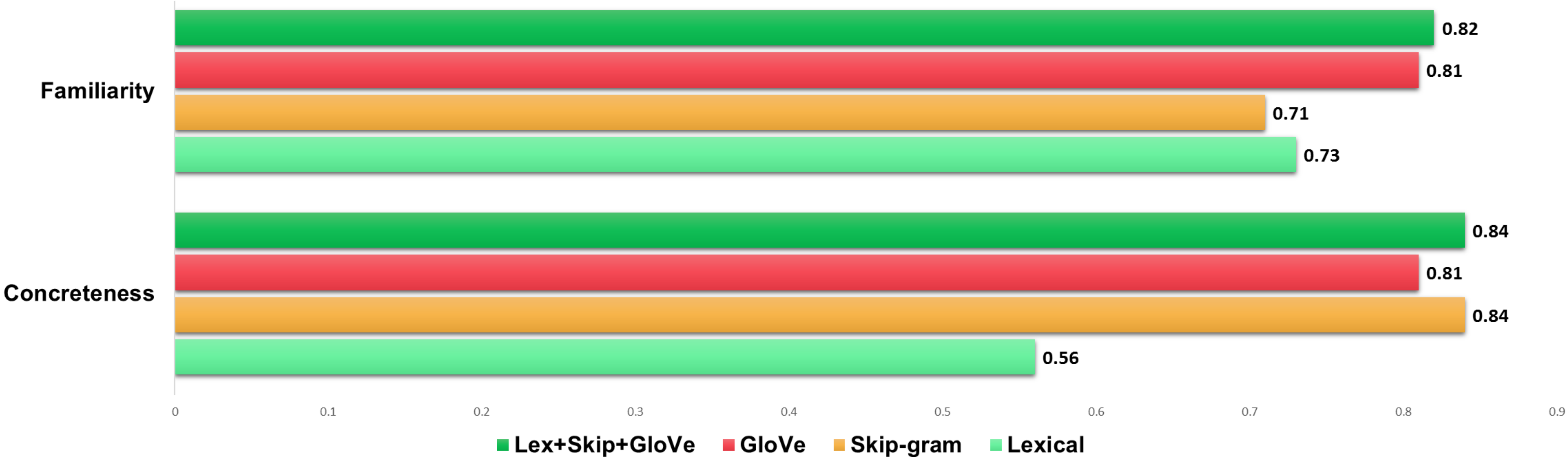
Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

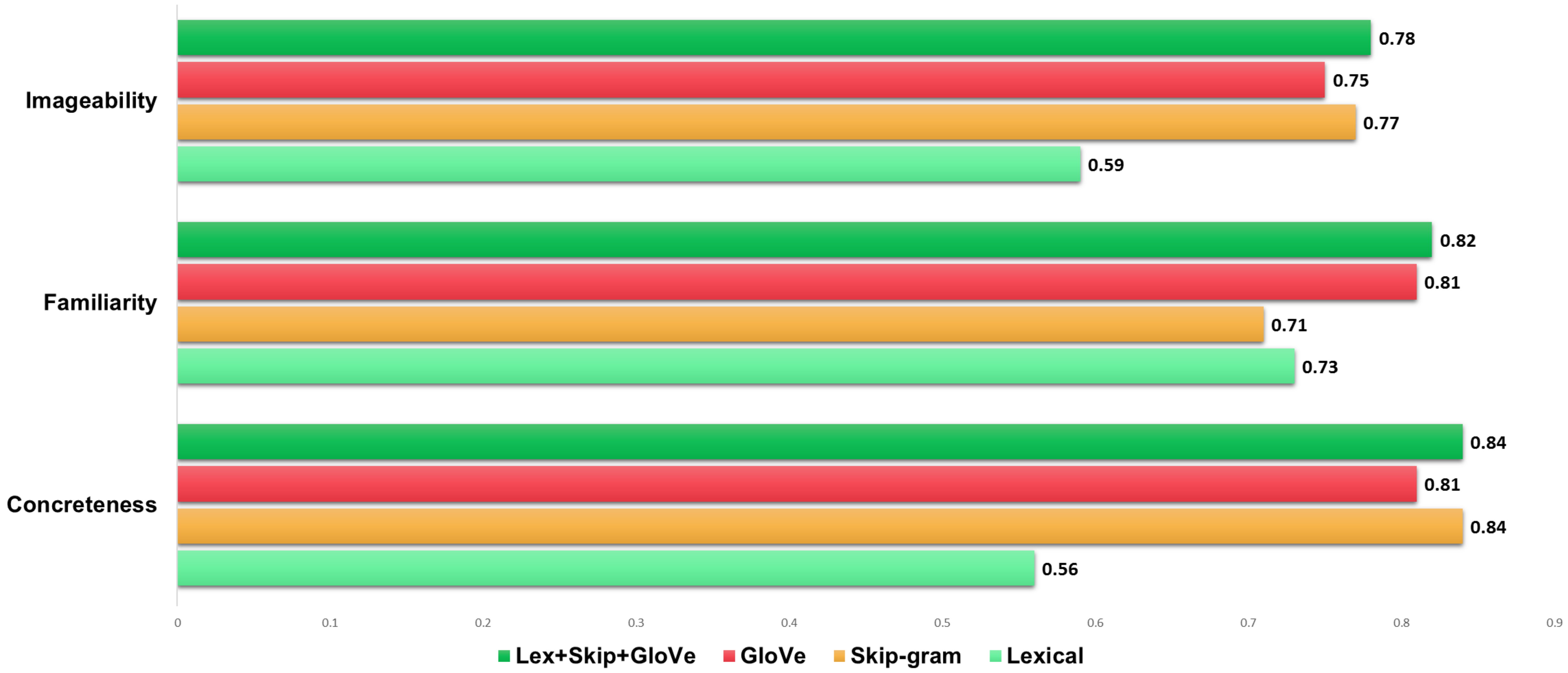

AoA
Imageability
Familiarity
Concreteness
0.79
0.75
0.66
0.73
0.78
0.75
0.77
0.59
0.82
0.81
0.71
0.73
0.84
0.81
0.84
0.56
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Lex+Skip+GloVe
GloVe
Skip-gram
Lexical

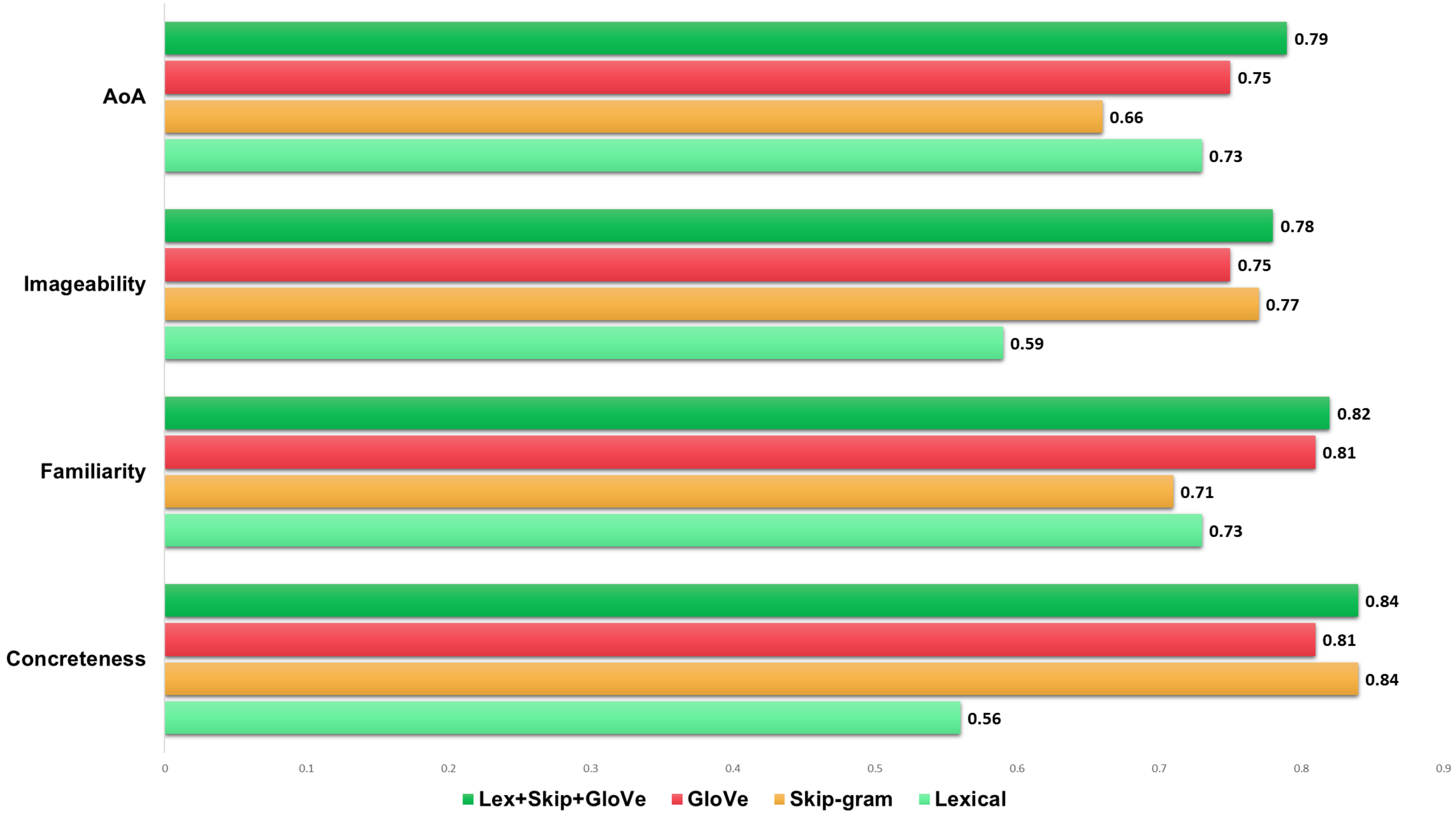
Lex+Skip+GloVe GloVe Skip-gram Lexical


0.84
0.81
0.84
0.56
Concreteness
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Lex+Skip+GloVe
GloVe
Skip-gram
Lexical



Familiarity
Concreteness
0.82
0.81
0.71
0.73
0.84
0.81
0.84
0.56
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Lex+Skip+GloVe
GloVe
Skip-gram
Lexical


Imageability
Familiarity
Concreteness
0.78
0.75
0.77
0.59
0.82
0.81
0.71
0.73
0.84
0.81
0.84
0.56
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Lex+Skip+GloVe
GloVe
Skip-gram
Lexical



AoA
0.79
0.75
0.66
0.73
Imageability
0.78
0.75
0.77
0.59
Familiarity
0.82
0.81
0.71
0.73
Concreteness
0.84
0.81
0.84
0.56
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Lex+Skip+GloVe
GloVe
Skip-gram
Lexical

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Regressors

Regressors
Lexical
Skip-gram
GloVe
Lexical + GloVe

Regressors	AoA (765)
Lexical	
Skip-gram	
GloVe	
Lexical + GloVe	

Regressors	AoA (765)		
	MSE	r	ρ
Lexical			
Skip-gram			
GloVe			
Lexical + GloVe			

Regressors	AoA (765)		
	MSE	$\boldsymbol{r}$	$\boldsymbol{\rho}$
Lexical	0.91	0.67	0.66
Skip-gram			
GloVe			
Lexical + GloVe			

Regressors	AoA (765)		
	MSE	$\boldsymbol{r}$	$\boldsymbol{\rho}$
Lexical	0.91	0.67	0.66
Skip-gram	1.30	0.56	0.58
GloVe			
Lexical + GloVe			

Regressors	AoA (765)		
	MSE	r	ρ
Lexical	0.91	0.67	0.66
Skip-gram	1.30	0.56	0.58
GloVe	1.18	0.62	0.63
Lexical + GloVe			

Regressors	AoA (765)		
	MSE	r	ρ
Lexical	0.91	0.67	0.66
Skip-gram	1.30	0.56	0.58
GloVe	1.18	0.62	0.63
Lexical + GloVe	**0.80**	**0.72**	**0.71**

Regressors	AoA (765)			AoA (1717)
	MSE	r	ρ	
Lexical	0.91	0.67	0.66	
Skip-gram	1.30	0.56	0.58	
GloVe	1.18	0.62	0.63	
Lexical + GloVe	**0.80**	**0.72**	**0.71**	

Regressors	AoA (765)			AoA (1717)		
	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66			
Skip-gram	1.30	0.56	0.58			
GloVe	1.18	0.62	0.63			
Lexical + GloVe	**0.80**	**0.72**	**0.71**			

Regressors	AoA (765)			AoA (1717)		
	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75
Skip-gram	1.30	0.56	0.58			
GloVe	1.18	0.62	0.63			
Lexical + GloVe	**0.80**	**0.72**	**0.71**			

Regressors	AoA (765)			AoA (1717)		
	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65
GloVe	1.18	0.62	0.63			
Lexical + GloVe	**0.80**	**0.72**	**0.71**			

Regressors	AoA (765)			AoA (1717)		
	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65
GloVe	1.18	0.62	0.63	0.93	0.79	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**			

Regressors	AoA (765)			AoA (1717)		
	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65
GloVe	1.18	0.62	0.63	0.93	0.79	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)
	MSE	r	ρ	MSE	r	ρ	
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75			
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65			
GloVe	1.18	0.62	0.63	0.93	0.79	0.75			
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**			

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65			
GloVe	1.18	0.62	0.63	0.93	0.79	0.75			
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**			

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75			
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**			

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**			

Regressors	AoA (765)			AoA (1717)			AoA Merge (2368)		
	MSE	r	ρ	MSE	r	ρ	MSE	r	ρ
Lexical	0.91	0.67	0.66	1.04	0.76	0.75	0.67	0.73	0.72
Skip-gram	1.30	0.56	0.58	1.36	0.68	0.65	0.81	0.66	0.66
GloVe	1.18	0.62	0.63	0.93	0.79	0.75	0.63	0.75	0.75
Lexical + GloVe	**0.80**	**0.72**	**0.71**	**0.79**	**0.83**	**0.80**	**0.54**	**0.79**	**0.79**

Slides: The Solutions

Slide **overhauling** example!

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

Experiment Setup

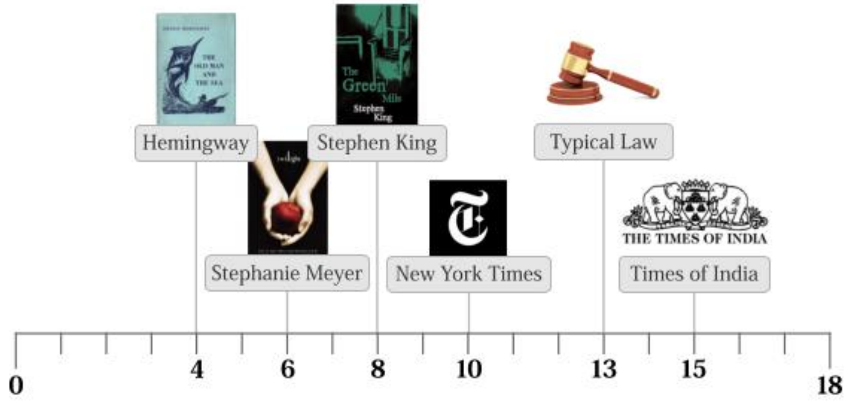
Experiment Setup

Evaluation metrics:

Experiment Setup

Evaluation metrics:

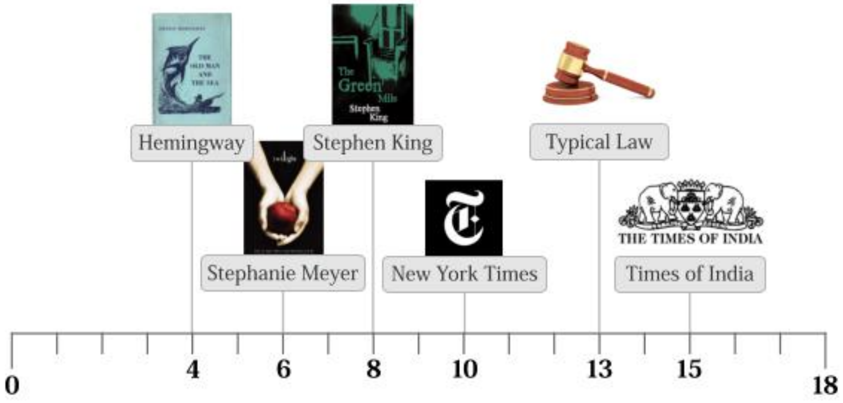
Flesch-Kincaid Grade Level



Experiment Setup

Evaluation metrics:

Flesch-Kincaid Grade Level



ROUGE-n

$$\frac{\text{overlapping n-grams}}{\text{total n-grams in reference}}$$

Experiment Setup

Data:

Experiment Setup

Data:

Training set:

754 sentences

Experiment Setup

Data:

Training set:

754 sentences

Test set:

100 sentences

Experiment Setup

Sources:

Experiment Setup

Sources:

70%

Experiment Setup

Sources:

70%

25%

The New York Times

Experiment Setup

Sources:

70%

25%

The New York Times

5%

Oral Presentations: Part 2

Oral Presentations: Part 2

Dealing with **English**

Dealing with **English**

I am a **fluent speaker** ☺

Dealing with **English**

I only struggle **very rarely** to make myself understood ☺

I am a **fluent speaker** ☺

Dealing with **English**

People understand me, but not without **a lot of effort** ☹

I only struggle **very rarely** to make myself understood ☺

I am a **fluent speaker** ☺

Dealing with **English**

People **almost never** understand me ☹

People understand me, but not without **a lot of effort** ☹

I only struggle **very rarely** to make myself understood ☺

I am a **fluent speaker** ☺

Dealing with **English**

Best course of action for:

Dealing with **English**

Best course of action for:

People **almost never** understand me ☹

Dealing with **English**

Best course of action for:

People **almost never** understand me ☹

Letting a **co-author** or **a colleague** present

Dealing with **English**

Author

Dealing with **English**

Author

vs.

Work

Experiment Setup

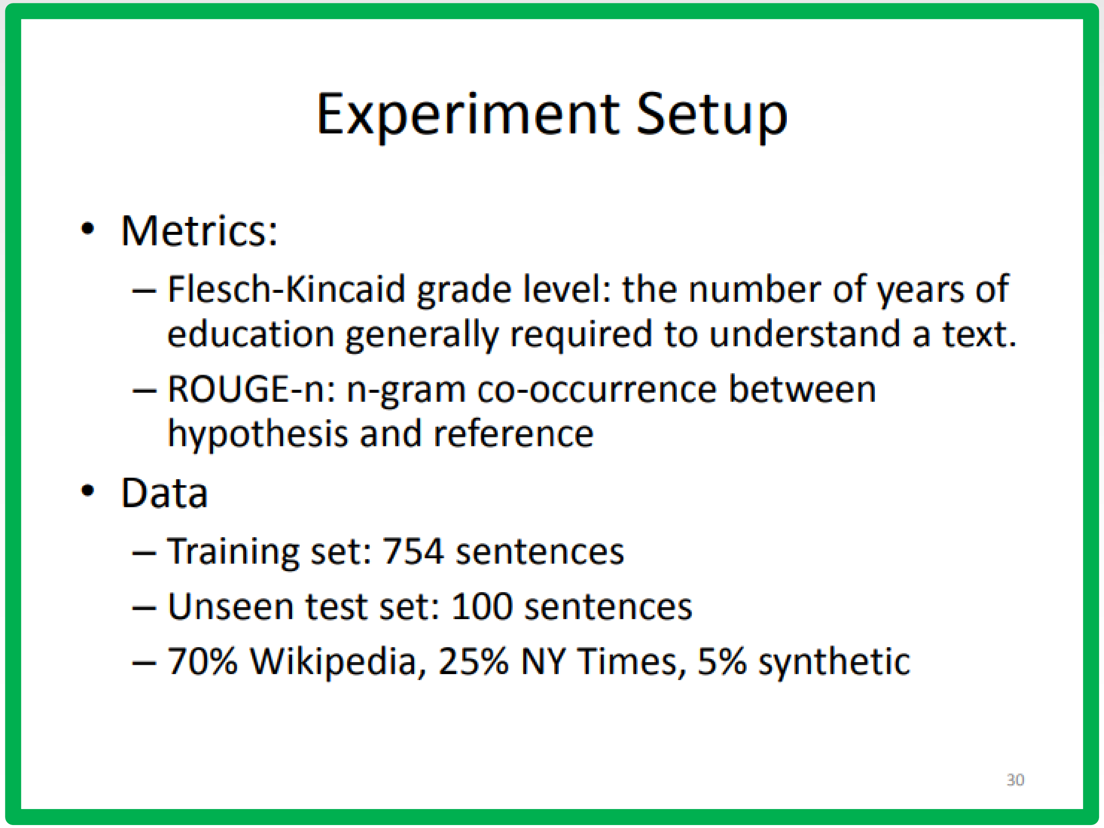
- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

Dealing with **English**

Author

vs.

Work



Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

30

Dealing with **English**

Author

vs.

Work

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

Dealing with **English**

Author

vs.

Work

Experiment Setup

- Metrics:
  - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
  - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
  - Training set: 754 sentences
  - Unseen test set: 100 sentences
  - 70% Wikipedia, 25% NY Times, 5% synthetic

Important note:

Important note:

Don't give up on yourself

Dealing with **English**

Best course of action for:

Dealing with **English**

Best course of action for:

People understand me, but not without **a lot of effort** ☹

Dealing with **English**

Best course of action for:

People understand me, but not without **a lot of effort** ☹

- Say **short sentences slowly**

Dealing with **English**

Best course of action for:

People understand me, but not without **a lot of effort** ☹

- Say **short sentences slowly**
- Use **cue cards**

Dealing with **English**

Best course of action for:

People understand me, but not without **a lot of effort** ☹

- Say **short sentences slowly**
- Use **cue cards**
- Rehearse **a lot**

1% of

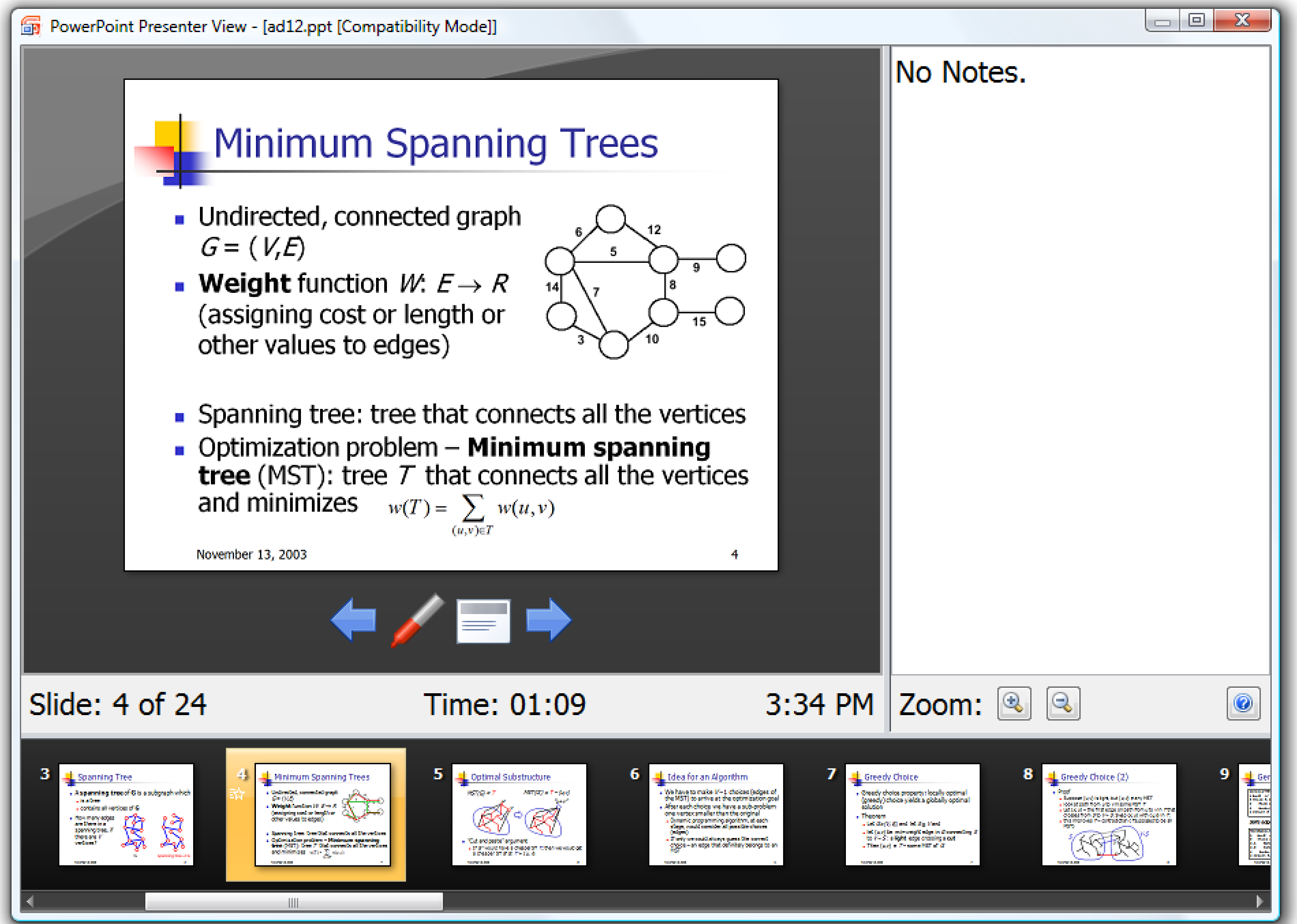
Google

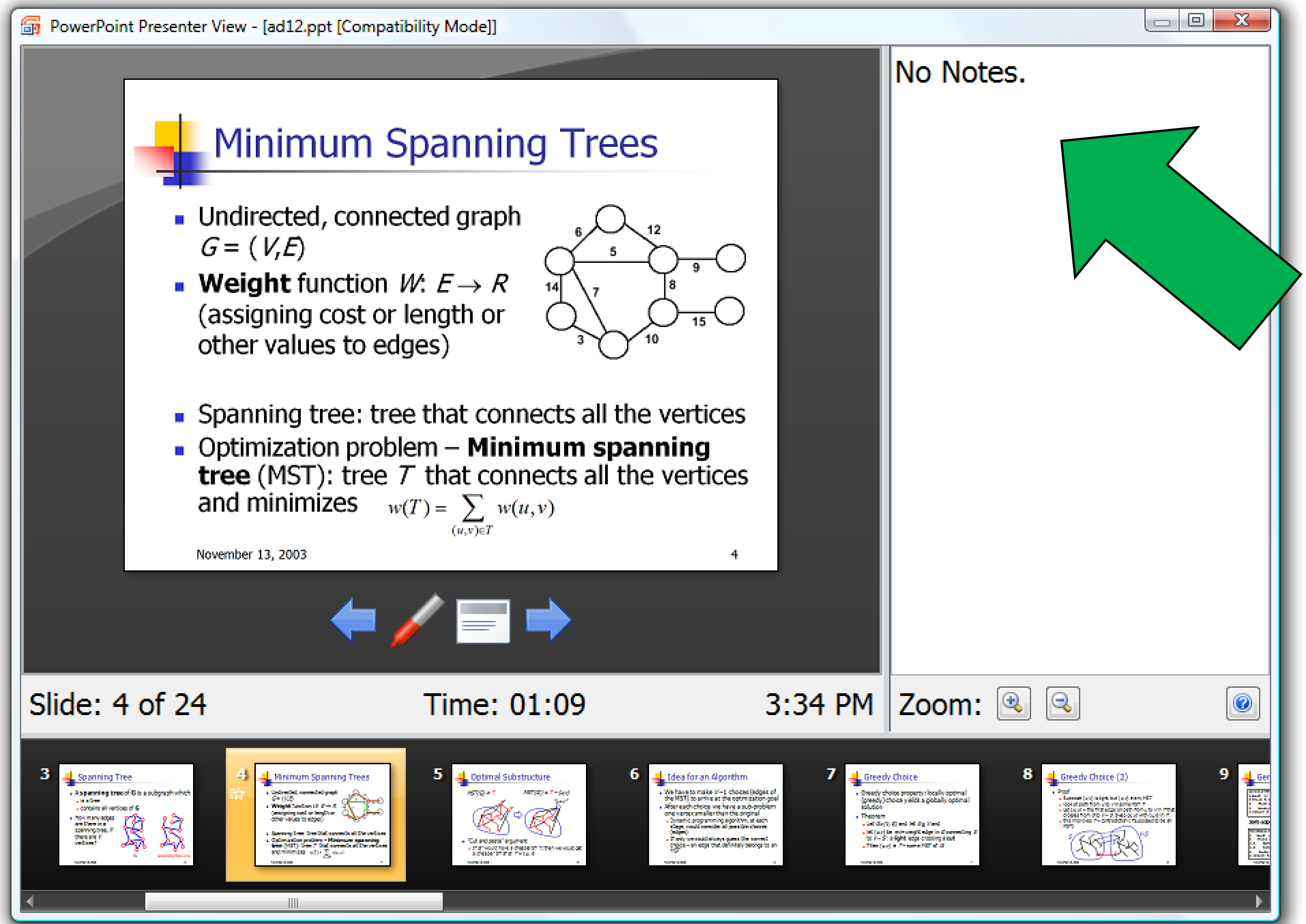
1% of Google > 100% of theranos

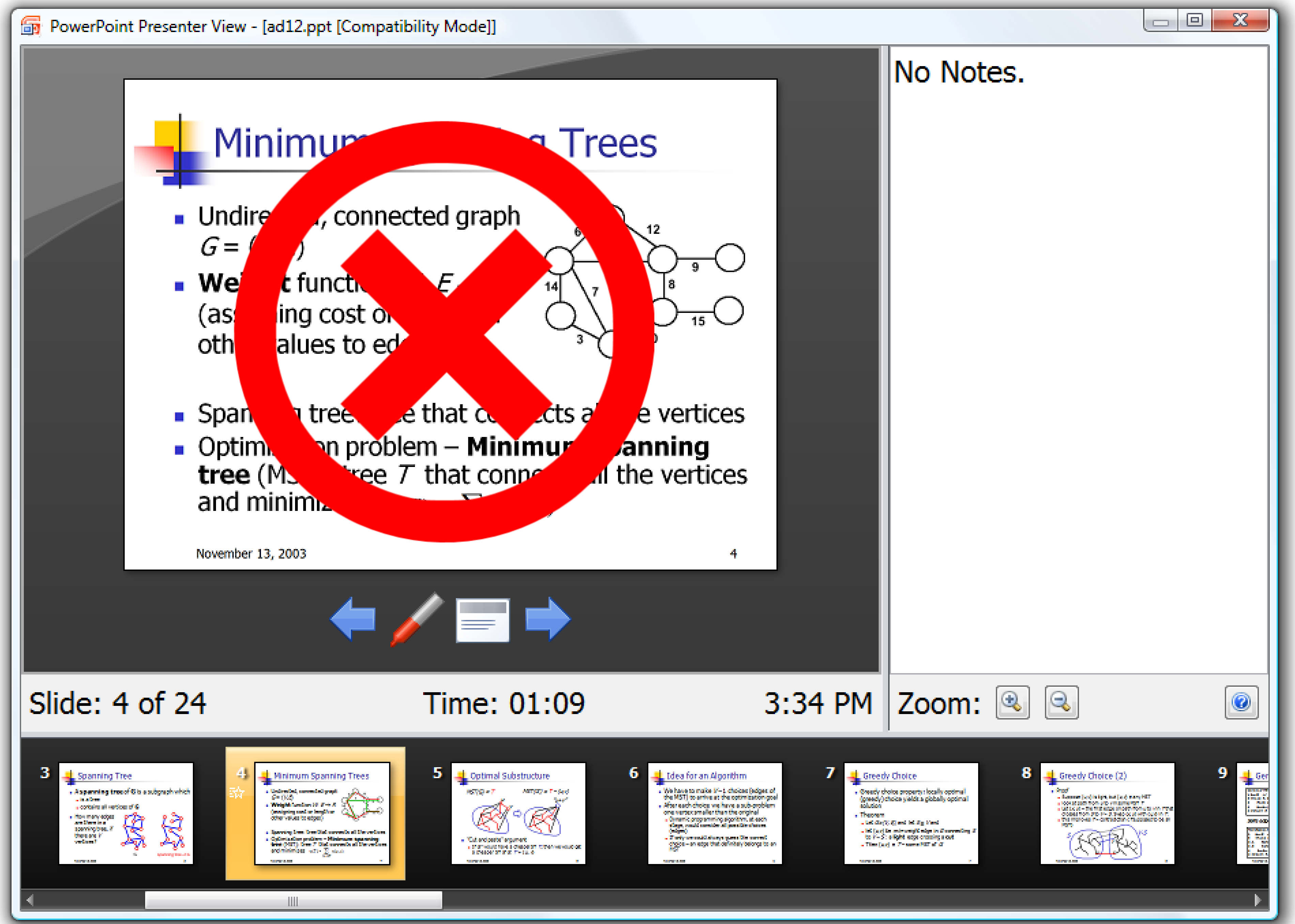

theranos
theranos

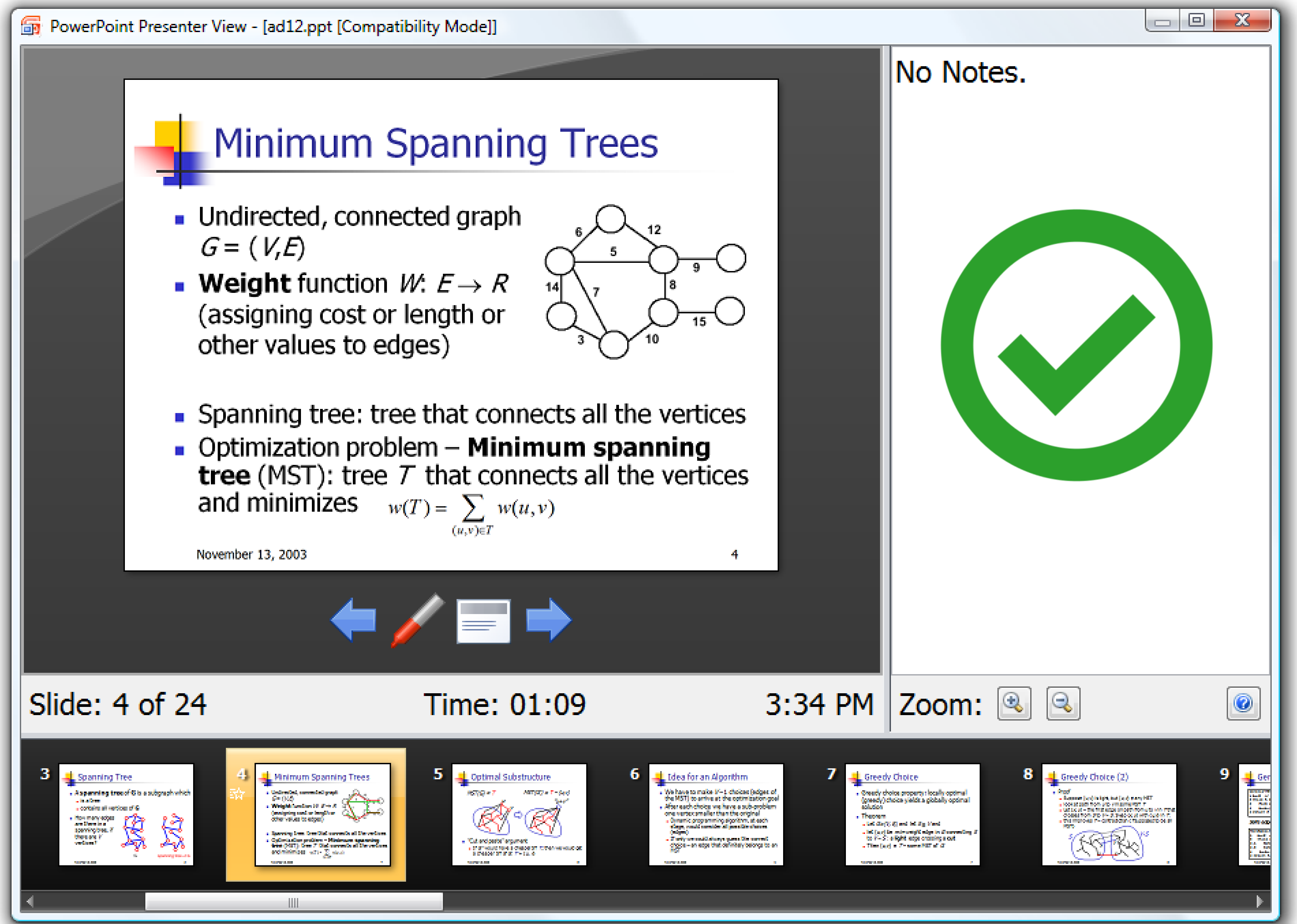
A short, but
well spoken sentence >

A short, but **well** spoken sentence > A long, fancy, **poorly** spoken sentence


PowerPoint Presenter View - [ad12.ppt [Compatibility Mode]]
Minimum Spanning Trees
Undirected, connected graph G = (V,E)
Weight function W: E → R (assigning cost or length or other values to edges)
6
12
5
9
14
7
8
15
3
10
Spanning tree: tree that connects all the vertices
Optimization problem – Minimum spanning tree (MST): tree T that connects all the vertices and minimizes $w(T) = \sum_{(u,v)\in T} w(u,v)$
November 13, 2003
4
No Notes.
Slide: 4 of 24
Time: 01:09
3:34 PM
Zoom:
3
Spanning Tree
4
Minimum Spanning Trees
5
Optimal Substructure
6
Idea for an Algorithm
7
Greedy Choice
8
Greedy Choice (2)
9


PowerPoint Presenter View - [ad12.ppt [Compatibility Mode]]
Minimum Spanning Trees
Undirected, connected graph $G = (V,E)$
Weight function $W: E \to R$ (assigning cost or length or other values to edges)
6
12
5
9
14
7
8
15
3
10
Spanning tree: tree that connects all the vertices
Optimization problem – Minimum spanning tree (MST): tree T that connects all the vertices and minimizes $w(T) = \sum_{(u,v)\in T} w(u,v)$
November 13, 2003
4
No Notes.
Slide: 4 of 24
Time: 01:09
3:34 PM
Zoom:
3
Spanning Tree
4
Minimum Spanning Trees
5
Optimal Substructure
6
Idea for an Algorithm
7
Greedy Choice
8
Greedy Choice (2)
9


PowerPoint Presenter View - [ad12.ppt [Compatibility Mode]]
Trees
, connected graph
Span
vertices
Optimi
on problem –
anning
tree
the vertices
and minimi
November 13, 2003
4
No Notes.
Slide: 4 of 24
Time: 01:09
3:34 PM
Zoom:
3
Spanning Tree
4
Minimum Spanning Trees
5
Optimal Substructure
6
Idea for an Algorithm
7
Greedy Choice
8
Greedy Choice (2)
9


PowerPoint Presenter View - [ad12.ppt [Compatibility Mode]]
Minimum Spanning Trees
Undirected, connected graph $G = (V,E)$
Weight function $W: E \rightarrow R$ (assigning cost or length or other values to edges)
6
12
5
9
14
7
8
15
3
10
Spanning tree: tree that connects all the vertices
Optimization problem – **Minimum spanning tree** (MST): tree T that connects all the vertices and minimizes $w(T) = \sum_{(u,v)\in T} w(u,v)$
November 13, 2003
4
No Notes.
Slide: 4 of 24
Time: 01:09
3:34 PM
Zoom:
3
Spanning Tree
4
Minimum Spanning Trees
5
Optimal Substructure
6
Idea for an Algorithm
7
Greedy Choice
8
Greedy Choice (2)
9

Cue card tools:

Cue card tools:

- **Windows:** Powerpoint presenter view

Cue card tools:

- **Windows:** Powerpoint presenter view
- **Mac OS:** Keynote presenter view

Cue card tools:

- **Windows:** Powerpoint presenter view
- **Mac OS:** Keynote presenter view
- **Latex:** pdfpc-latex-notes

Dealing with **English**

Advanced tip:

Dealing with **English**

Best course of action for:

I only struggle **very rarely** to make myself understood ☺

I am a **fluent speaker** ☺

Dealing with **English**

Best course of action for:

I only struggle **very rarely** to make myself understood ☺

I am a **fluent speaker** ☺

Rehearse!

Big no-nos of oral presentations:

Dealing with **English**

Big no-nos of oral presentations:

1. Not respecting the time limit

Dealing with **English**

Big no-nos of oral presentations:

1. Not respecting the time limit
2. Reading the slides

Dealing with **English**

Big no-nos of oral presentations:

1. Not respecting the time limit
2. Reading the slides
3. Skipping slides

Dealing with **English**

Big no-nos of oral presentations:

1. Not respecting the time limit
2. Reading the slides
3. Skipping slides
4. Not speaking loudly enough

Dealing with **English**

Some cool practices...

Dealing with **English**

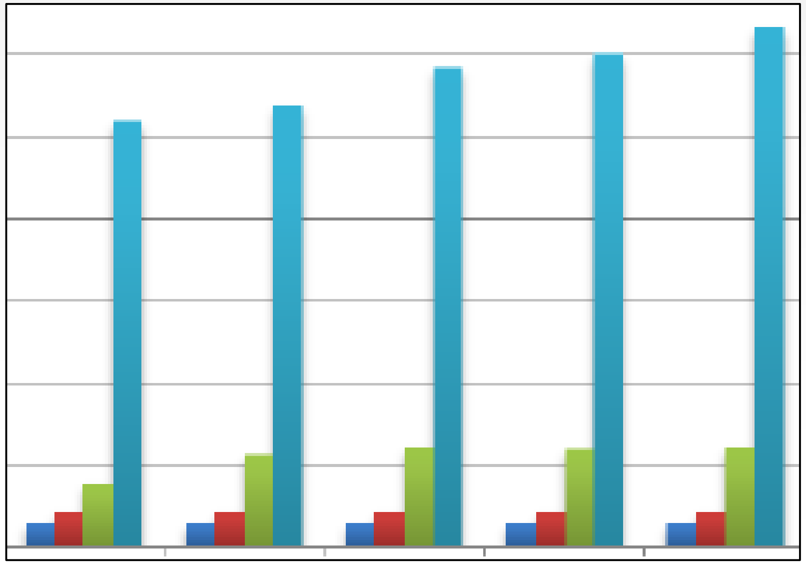
The performance preview slide

(right after introduction)

Dealing with **English**

The performance preview slide

(right after introduction)



Awesome scores!!!

The English proficiency alert

Dealing with **English**

The English proficiency alert

I'm still learning English!
Please be kind with the questions ☺

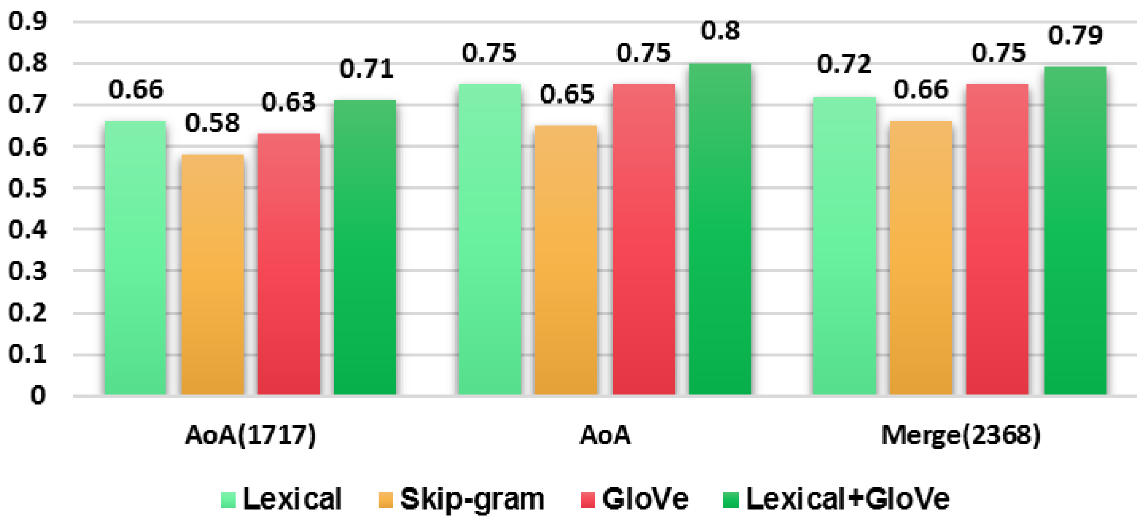
Acessible graphs

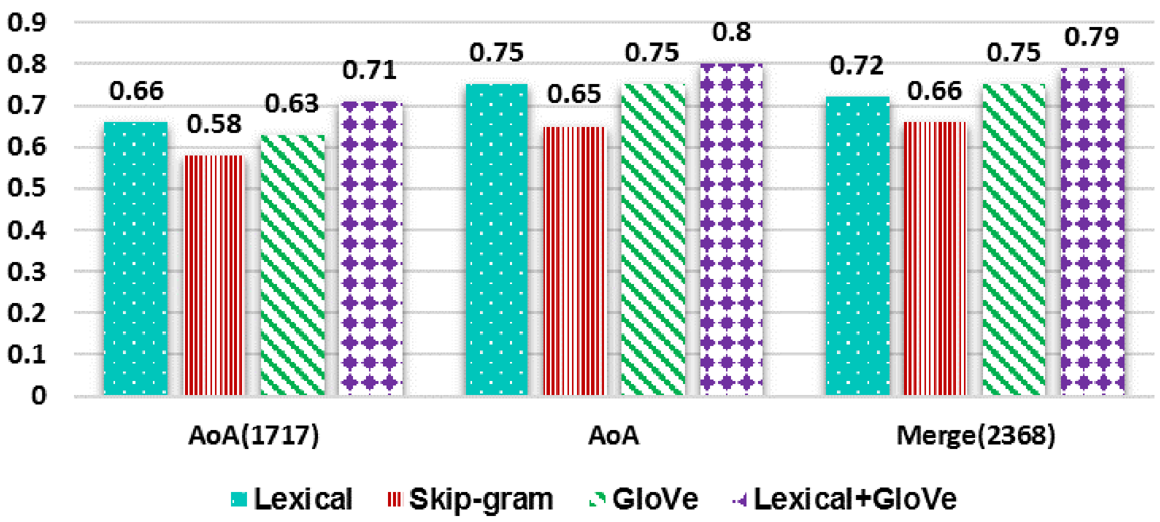
(for the colorblind)

Dealing with **English**

Acessible graphs

(for the colorblind)

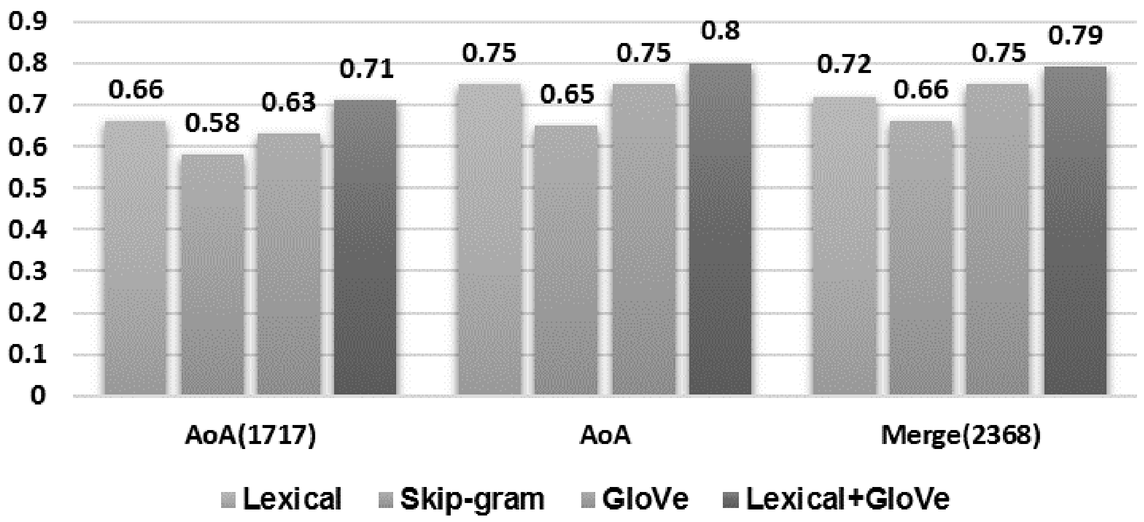


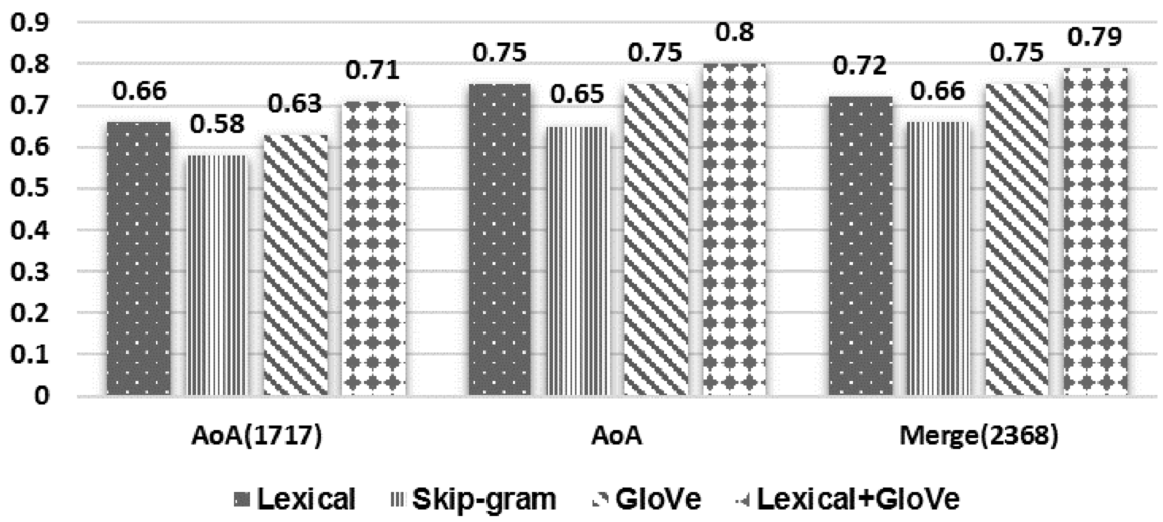


Dealing with **English**

Acessible graphs

(for the colorblind)





Dealing with **English**

But what about...

But what about...
...the Q&A session?

But what about...
...the Q&A session?

Solution:

Solution:

The “presentation buddy”

Dealing with **English**

Presenter

Dealing with **English**

Presenter

Attendee

Dealing with **English**

Presenter

Questions

Attendee

Dealing with **English**

Presenter

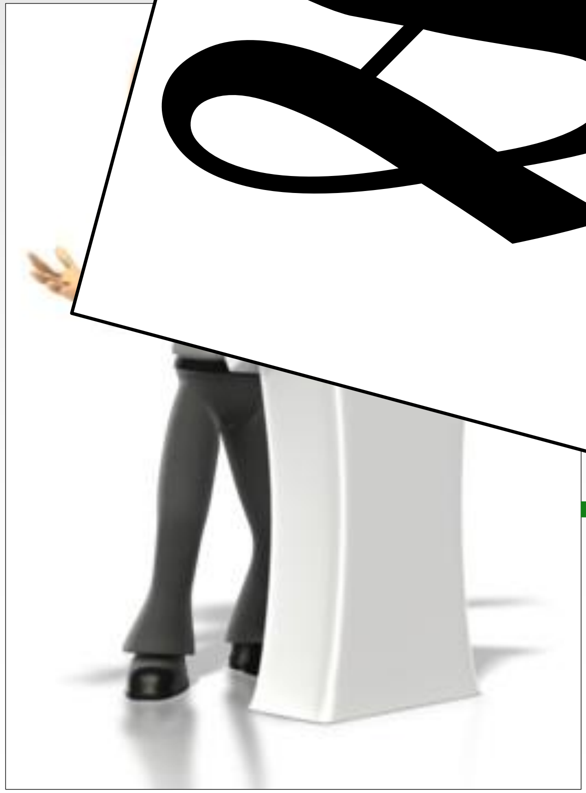
Attendee

Questions

Answers

De… lish

Pre… …e



Scary

Dealing with **English**

Presenter

Attendee

Dealing with **English**

Presenter

Buddy

Attendee

Dealing with **English**

Presenter

Buddy

Q

Attendee

Dealing with **English**

Presenter

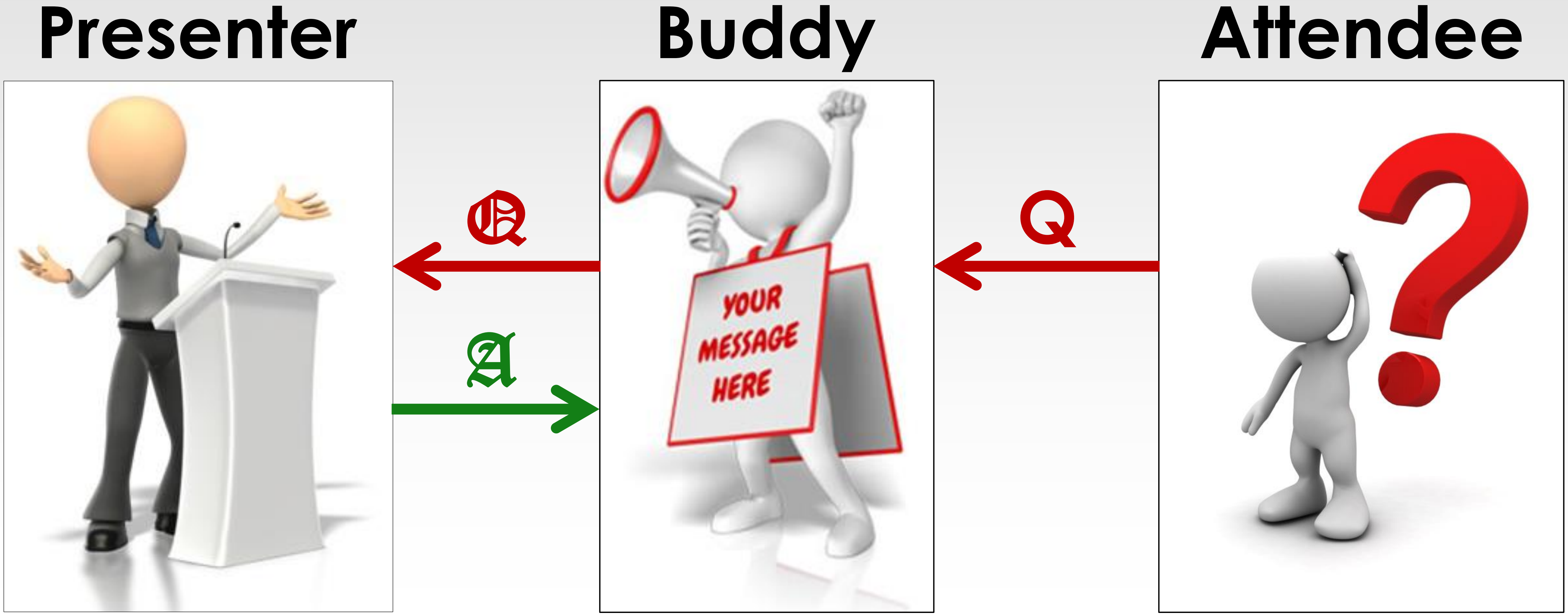
Q

Buddy

Q

Attendee

Dealing with **English**



Dealing with **English**

Presenter | **Buddy** | **Attendee**

Dealing with **English**

Presenter **Translator** **Attendee**

Main takeaways

Main takeaways

- Posters:

Main takeaways

- **Posters:**
  1. Remove unnecessary stuff

Main takeaways

- **Posters:**
  1. Remove unnecessary stuff
  2. Make things concise/visual

Main takeaways

- **Posters:**
  1. Remove unnecessary stuff
  2. Make things concise/visual
  3. Structure it well

Main takeaways

- Slides:

Main takeaways

- **Slides:**
  1. ... same thing

Main takeaways

- **Slides:**
  1. ... same thing
  2. Stepify!

Main takeaways

- Presentation:

Main takeaways

- **Presentation:**
  1. Respect the no-no list

Main takeaways

- **Presentation:**
  1. Respect the no-no list
  2. Try to use some of the cool practices

Main takeaways

- **Presentation:**
  1. Respect the no-no list
  2. Try to use some of the cool practices
  3. Follow our “best course of action”

Thank you!