

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 awesome:

And the work is awesome:

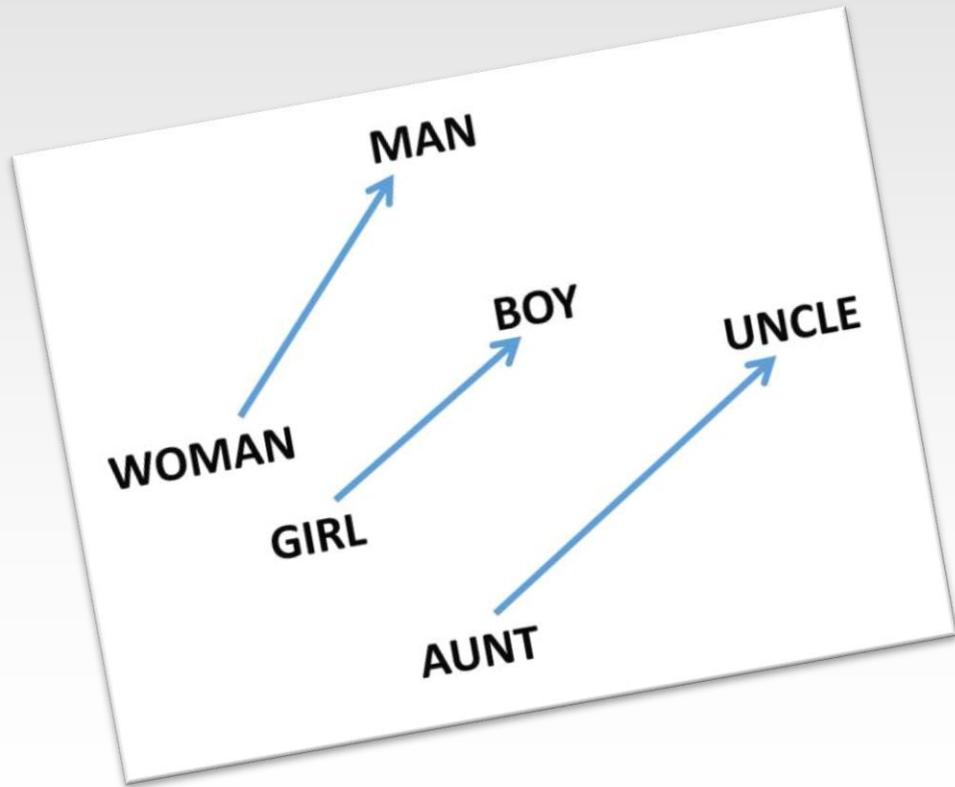
Retrofitting Word Vectors to Semantic Lexicons

Manaal Faruqui **Jesse Dodge** **Sujay K. Jauhar**
Chris Dyer **Eduard Hovy** **Noah A. Smith**

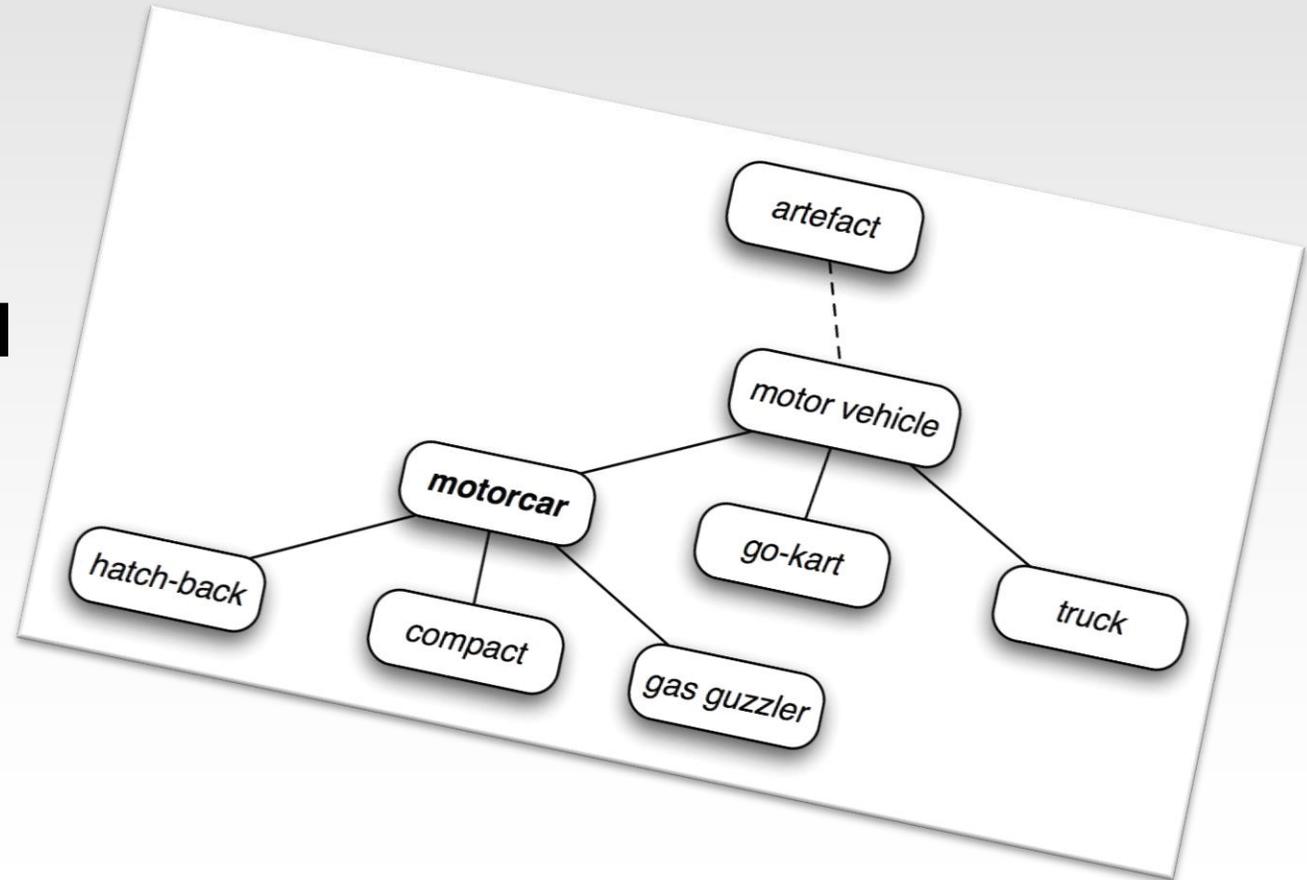
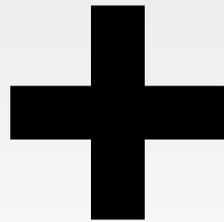
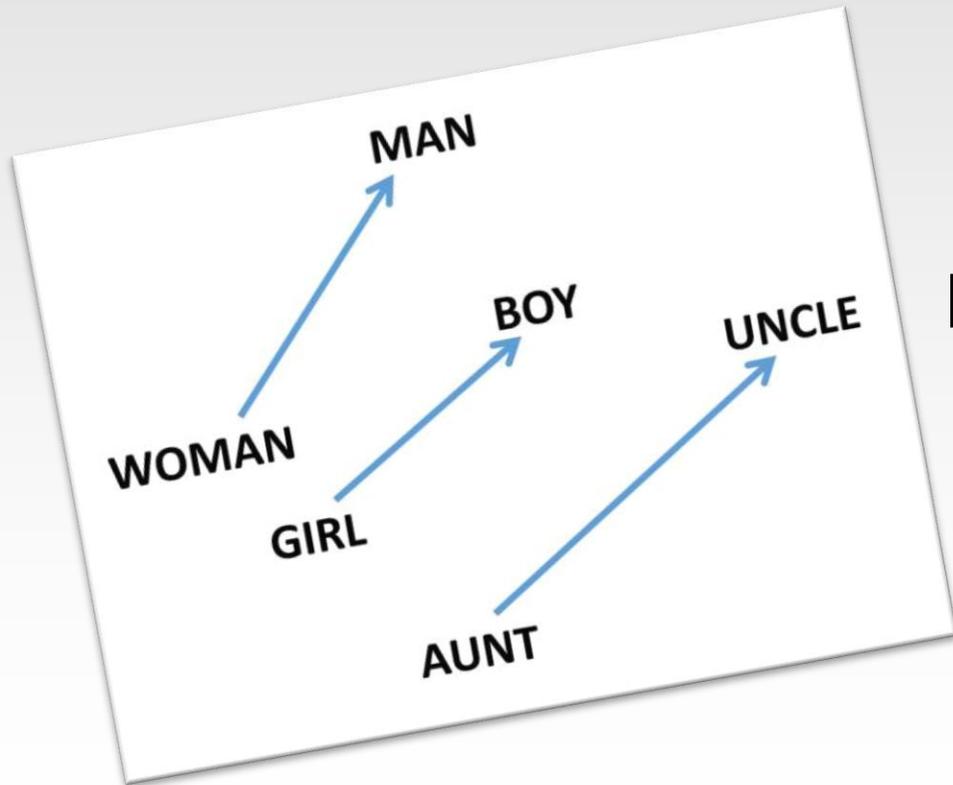
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA, 15213, USA

{mfaruqui, jessed, sjauhar, cdyer, ehovy, nasmith}@cs.cmu.edu

And the work is awesome:



And the work is awesome:



And the work is awesome:

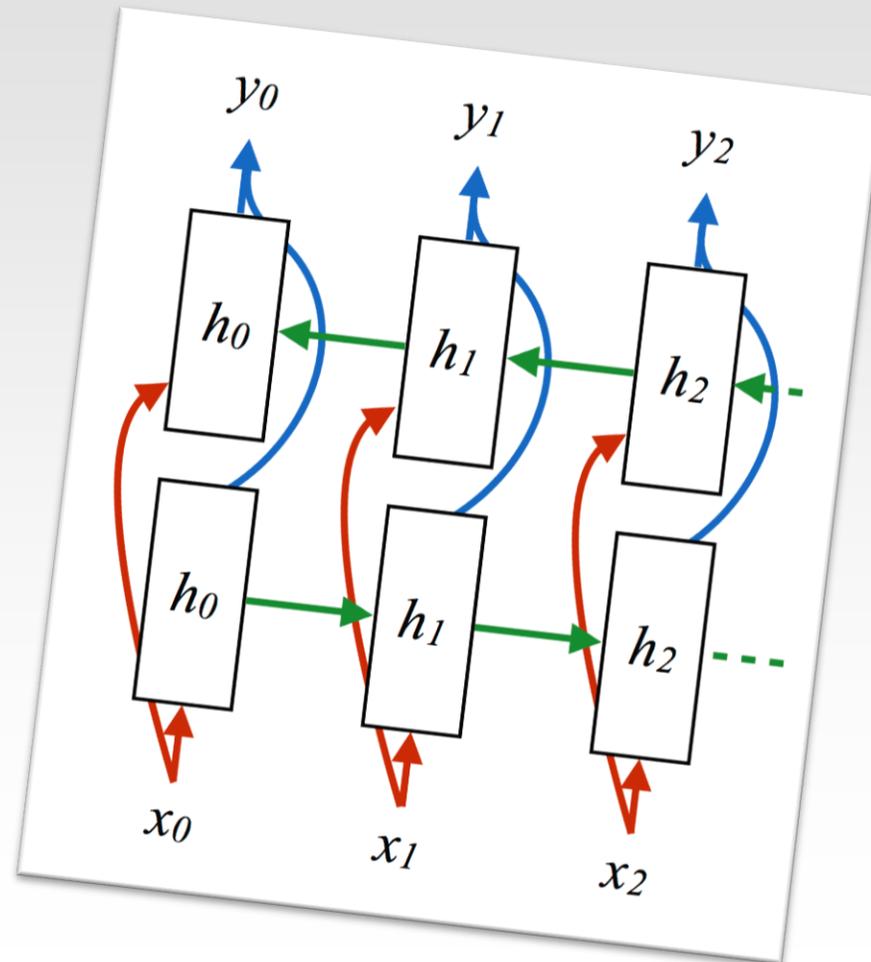
Improving sentence compression by learning to predict gaze

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

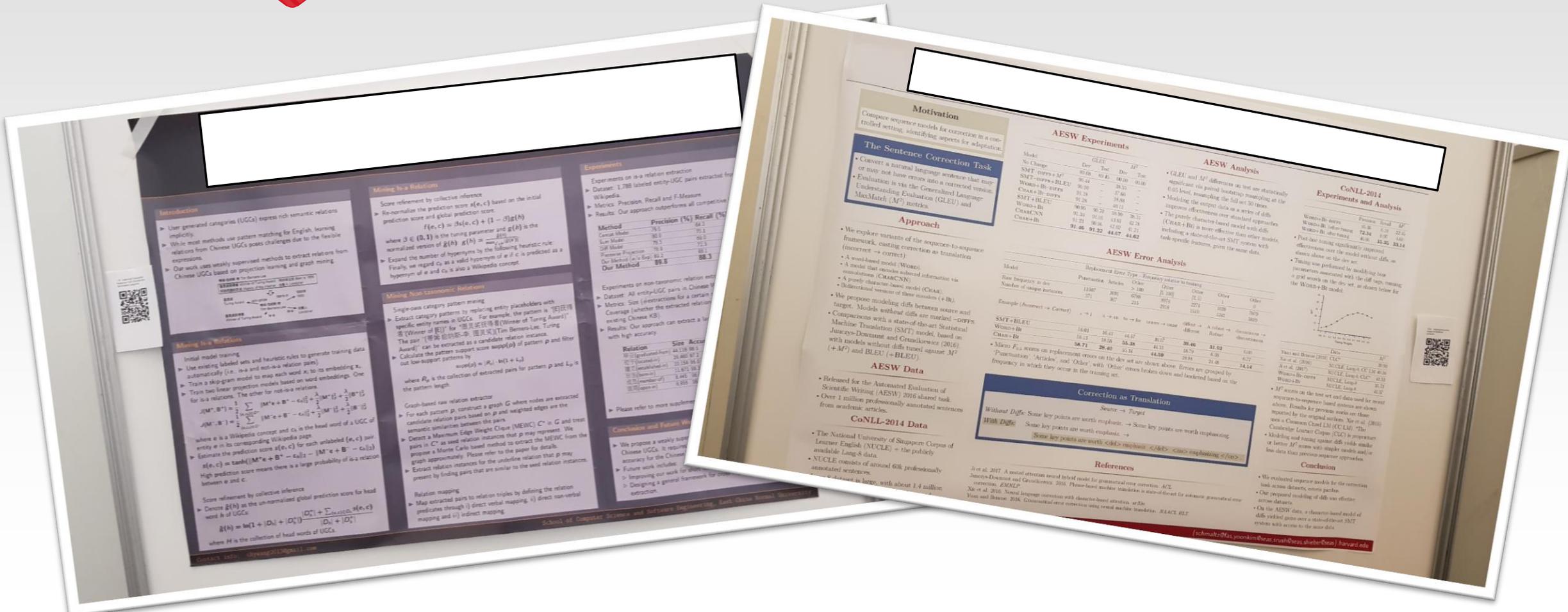
Yoav Goldberg
Bar-Ilan University
yoav.goldberg@gmail.com

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 **static**

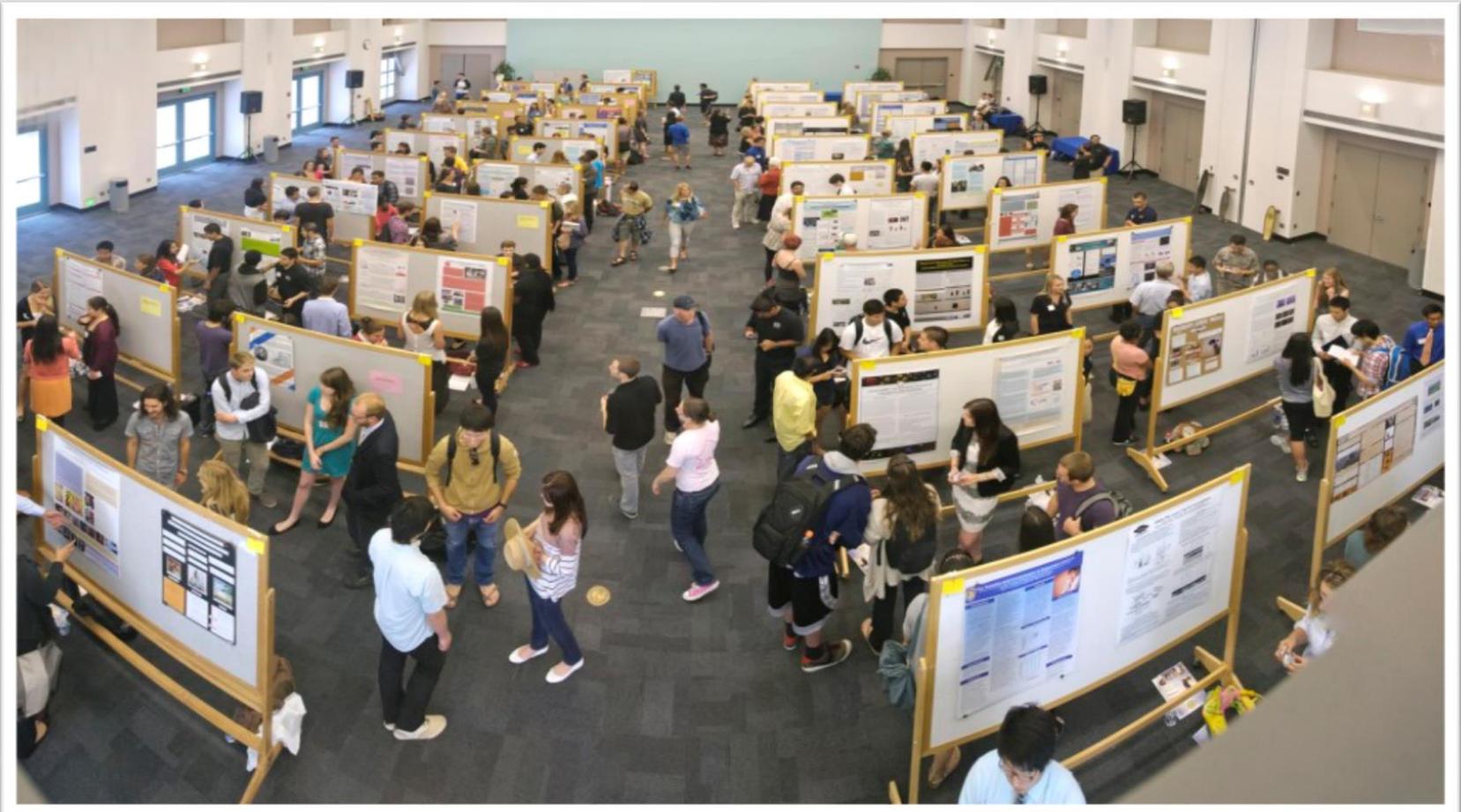
Posters: The Challenge

Catching people's attention

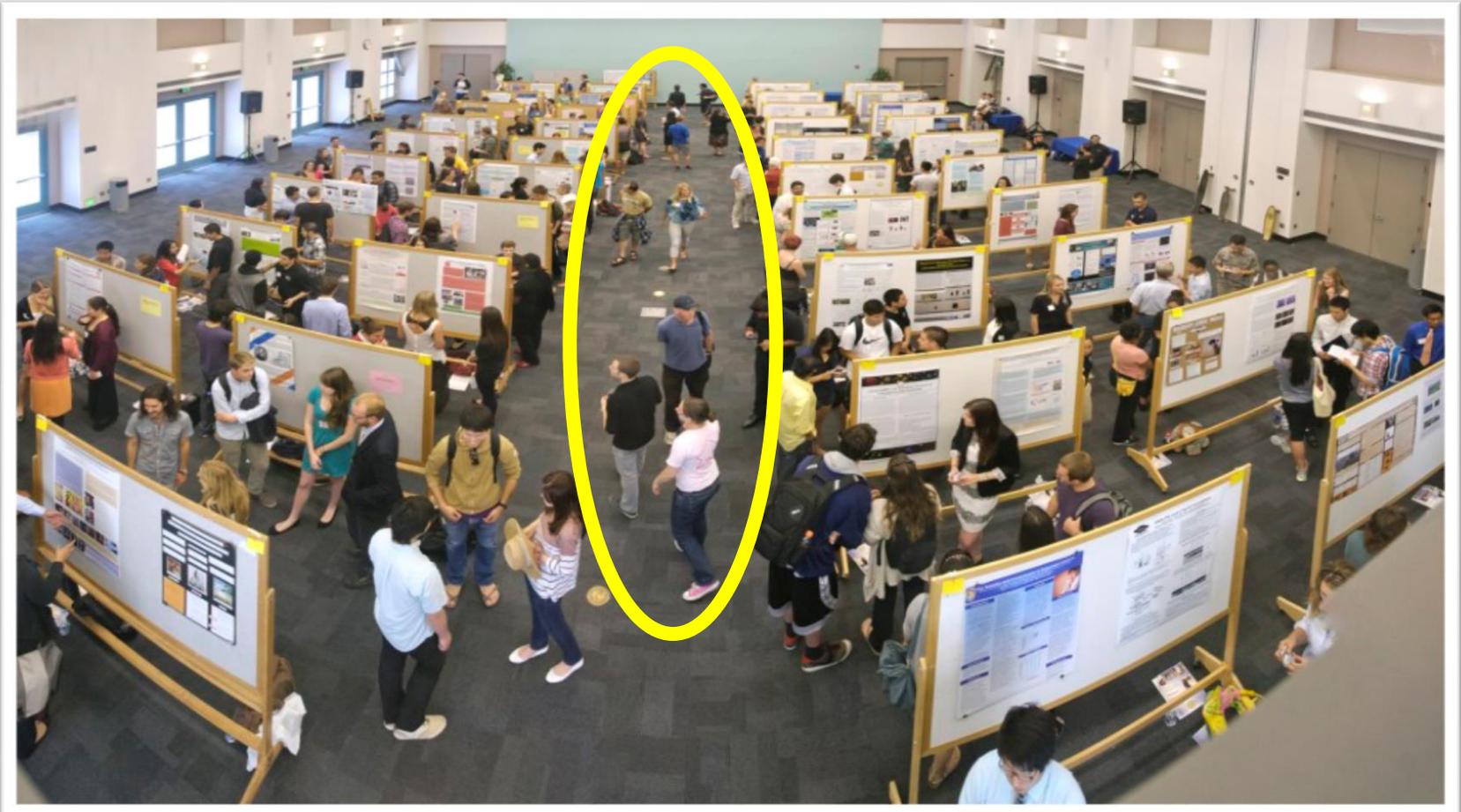
Posters: The Challenge

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

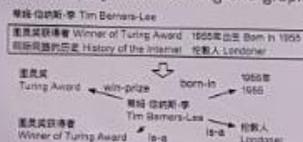
2. Too much _____ information

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 x_i .
- ▶ Train two linear projection models based on word embeddings. One for is-a relations. The other for not-is-a relations.

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

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

where \mathbf{e} is a Wikipedia concept and \mathbf{c}_h is the head word of a UGC of entity \mathbf{e} in its corresponding Wikipedia page.

- ▶ Estimate the prediction score $s(\mathbf{e}, \mathbf{c})$ for each unlabeled (\mathbf{e}, \mathbf{c}) pair.
- $$s(\mathbf{e}, \mathbf{c}) = \tanh(\|\mathbf{M}^+ \mathbf{e} + \mathbf{B}^+ - \mathbf{c}_h\|_2 - \|\mathbf{M}^- \mathbf{e} + \mathbf{B}^- - \mathbf{c}_h\|_2)$$
- High prediction score means there is a large probability of is-a relation between \mathbf{e} and \mathbf{c} .

Score refinement by collective inference

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

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

where H is the collection of head words of UGCs.

Mining Is-a Relations

Score refinement by collective inference

- ▶ Re-normalize the prediction score $s(\mathbf{e}, \mathbf{c})$ based on the initial prediction score and global prediction score.

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

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

- ▶ Expand the number of hypernyms by the following heuristic rule: Finally, we regard \mathbf{c}_h as a valid hypernym of \mathbf{e} if \mathbf{c} is predicted as a hypernym of \mathbf{e} and \mathbf{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	76.3	69.0	71.5
Piecewise Projection	76.9	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.

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* NMT-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 <u>avaient</u> dû nous alerter.
System	Les appels répétés de sa mère <u>devraient</u> 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/LT 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".

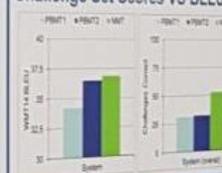
Evaluation: Protocol

- 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:
The repeated calls from his mother [S2] should have alerted us.
Les appels [PL] répétés de sa mère [S2] auraient [PL] dû nous alerter.

- Most other cases of agreement also handled: logic of coordinated subjects, distribution of coordinated verbs, past participles with "avoir".

NMT Strengths: Lexico-Syntactic

- 42% (PBMT) ⇒ 62% (NMT) ⇒ 62% (GNMT)
- Overlapping subcat frames:
She knows [my son]_{NP} → Elle connaît mon fils.
She knows [my son]_{NP} is ill.
→ Elle sait [knows] que [bad] mon fils est.

- Double object constructions:
X gave NP₁ NP₂ → X a donné NP₁ à NP₂.
- Infrinitival to finite complements:
NP₁ believes NP₂ to Vinf → NP₁ croit que NP₂ Vfin.

NMT Strengths: Syntactic

- 33% (PBMT) ⇒ 42% (NMT)
- Ex. 1: Yes-no question synt.
Have the kids ever watched?
→ Les enfants ont-ils déjà vu?
[The kids have-they ever seen]
Ex. 2: French pro-clitic pron.
GNMT also handles tag on pronouns and (surprisingly) The city that he is arriving from.
La ville d'où (from where) il est

NMT weaknesses

- Lexically triggered exceptions:
 - Argument switch (see miss)
 - Manner of movement verbs: cross X by swimming.
 - Idioms: substantially worse
 - Incomplete generalizations:
 - Some highly frequent cue words are missed (e.g. provided for gender and number but miss)

Conclusions & Future

- CS methodology provides:
 - How NMT improves over PBMT: powerful than n-grams (cf. 4)
 - Where NMT needs to improve
 - Supplements (≠ replaces) existing methods.
- Our *En* → *Fr* dataset is available in format with MT output and:
 - Further work:
 - Compare architectures (e.g. Automate CS development)
 - Automate or expedite human evaluation
 - Improve current systems (use conclusions and/or new approaches)

Posters: The Problems

3. _____ are not _____ enough

Posters: The Problems

3. **Visuals** 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.

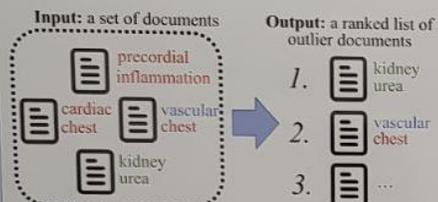


Fig. 1: Input and output of identifying semantically deviating outlier documents. In this example, the given corpus of health records is more relevant to cardiac diseases. However, there is one record only contains words about renal diseases. Our objective is to identify such outlier documents by generating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

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.

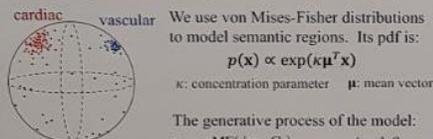


Fig. 2: Each point represents a word as a normalized vectors in the spherical embedded semantic space. Two frequent semantic regions can be observed.

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.

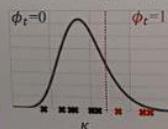


Fig. 3: The fitted log-normal distribution of all the κ 's. The threshold is set to a given quantile.

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 Outlieriness

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

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

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.

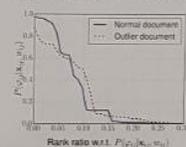


Fig. 4: Comparison of a normal document and an outlier document. The probability of each word being orthodox (y -axis) is ranked, and the ranking is normalized by the document length (x -axis).

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

$$qq(n_t^q) = \sup\{q : P(n_t^q \geq q) \geq \theta\}$$

n_t^q : #words being orthodox

θ : a given large probability

to define the outlieriness:

$$\Omega_{\theta,q}(d_i) = 1 - \frac{qq(n_t^q) + 1}{|d_i| + 1}$$

This outlieriness 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	Rec@1	Rec@5	Rec@50
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
ARNET	VMF-E	36.57	35.91	49.41	67.56
	VMF-Q	41.88	36.99	63.29	79.23
NYT	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
ARNET	VMF-E	10.51	12.67	25.92	45.37
	VMF-Q	19.74	22.40	34.40	53.87

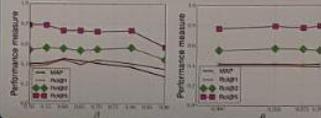


Fig. 5: Performance of outlier document detection with different parameter settings. The performance is not sensitive to either parameter.

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

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

VMF-Q: Proposed method

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

VMF-SF: VMF-E without penalizing general words

Posters: The Problems

4. Too much _____

Posters: The Problems

4. Too much **text**

Abstract

The purpose of this research was to observe the impact of positive psychology on restricted eating in the presence of images of unrealistic body images. Female participants took part in one of three 15-minute lab sessions assigned at random. Data analysis indicated that exposure to images of unrealistic body images may increase restricted eating. Additionally, a positive psychology exercise may reduce this effect.

Theoretical Background

Eating Disorders and Prevention Research

Eating Disorders affect an estimated 10 million people according to the National Institute of Mental Health. Research has shown that over 70% of adolescent girls of normal weight are engaging in unhealthy eating behaviors (Gaug, 2006, p.1). The increase of body dissatisfaction coupled with the rise of eating disorders over the years is troubling. "Eating disorders rank as the third most common chronic illness among adolescent females in the United States" (Neumark-Sztainer, 1995). The majority of adolescent girls consider themselves to be fat and have attempted or are attempting eating disorder prevention, awareness, and education interventions in school systems. Unfortunately, there is no evidence that the current programs are successful in the long-term as many do not have follow-up evaluations to measure effectiveness. The purpose of this research was to assess whether a positive psychology device could be a successful prevention method for eating disorders.

Positive Psychology

Positive Psychology is the scientific study of human strengths and virtues (Seligman & King, 2001) and was first developed in 1999 by Martin Seligman. Seligman introduced the concept of positive psychology in his book, *Authentic Happiness*. The foundation of this field is based on science that can be measured and focuses on an individual's strengths rather than their weaknesses. Through empirical research, he is able to demonstrate the value and importance of a work life that is a combination of focusing on strengths and talents versus to positive emotion, hope, and energy. Seligman proposes that by being able to identify the most of our own strengths and capacities, we can improve on the world. Seligman, Steiner, Peterson, Park, Peterson, Pals, and Park (2005) conducted a longitudinal study on positive psychology and its potential to improve mental health, and increase life satisfaction. Seligman, Peterson, Park, Peterson, Pals, and Park (2005) conducted a longitudinal study on positive psychology and its potential to improve mental health, and increase life satisfaction. Seligman, Peterson, Park, Peterson, Pals, and Park (2005) conducted a longitudinal study on positive psychology and its potential to improve mental health, and increase life satisfaction.

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

In the lab session participants were randomly assigned to one of three groups. They either did a gratitude/reflection exercise, or a control relaxation exercise with magazines that were neutral or focussed on body image. Group 1 was the control and had the neutral magazines. The magazines used were *Time* and *Newsweek* and had no people on the covers. Groups 2 and 3 had the body image magazines, *Self* and *Maxim* which had well-known female celebrities on the cover who are often praised for their slender and toned bodies. After the groups completed either the positive psychology exercise or the control relaxation exercise, they were then presented with a bowl of food for a "taste experience" exercise, which was actually designed to see how much food they would eat (based on Goldstein, Arndt, Hart, & Brown (2006) study design). The food bowl was full of rice because it provided a variety and is generally well-liked. It was measured out and recorded so that it was 120 grams. The bowl was separated into bags so each participant received their own individual bag during their session. Once it was low for the "taste into a bowl in front of the participant. At the end of the lab session the bowl was passed back into its original bag and later weighed to see how much of it had been consumed.

Measures

The mean of total rice consumption from the Control group was 152.43 with a standard deviation of 11.33. The mean of Group 2 was 109.87 with a standard deviation of 6.41. The mean of Group 3 was 193.57 with a standard deviation of 8.55. Analysis of variance indicated that the differences between the groups was not significant ($F = 1.09, df = 2, p > .05$). However, the mean differences were in the predicted direction. This indicated 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.

Results

While the analysis of the variance indicated that the difference between the groups was not significant ($F = 1.09, df = 2, p > .05$), 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. This capstone has defined eating disorders and positive psychology, discussed present prevention models and how they can be improved, and highlighted just how much influence the media can have. By incorporating these techniques I hope to see a change in the efforts aimed at preventing eating disorders and promoting a healthy body image and life style and a decrease in disordered behaviors among adolescents. Prevention programs for eating disorders can be done at multiple levels such as individual, family-based, or institutional (Neumark-Sztainer, 2011, p.421). Current models of prevention programs are introduced to adolescents (often only females) entering high school and target increasing healthy eating habits and exercise, eating disorder awareness and warning signs, and the prevention of obesity. The adolescents will have a higher outcome of healthier eating habits and overall of an intervention program. It often excludes males because girls are of a higher risk for developing eating disorders, and while this is true the solution to the current prevention program problem is to introduce a healthy high school, include both males and females, along with school officials. Most importantly use positive psychology as the foundation of the program and to overcome external pressures of being thin and avoid the foundation of a prevention program would change the perspective of outcome. Adolescents would learn how to establish realistic goals, authenticity, two crucial characteristics during development. Prevention programs need to focus on more than just being informative. Beliefs, attitudes, and behaviors are the only way to ensure long-term success. Focusing on identifying characteristics relating to positive self-image and traits instead of only emphasizing risk factors and stopping maladaptive behaviors will allow adolescents to buffer themselves against external influences, such as the media (Stark, 2002).

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:
 - word imageability** the ease and speed with which a word evokes a mental image;
 - concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
 - subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
 - 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:
 - 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?
 - 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?
 - 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

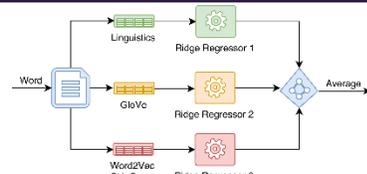


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

- 10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):
- Log of Frequency in SUBTLEX-pt-BR;
 - Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
 - Log of Frequency in SubIMDb PT: subtitles of family, comedy and children movies and series;
 - Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
 - Log of Frequency in the Spoken Language part of Corpus Brasileiro;
 - Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
 - Word Length;
 - Lexical databases from 6 school dictionaries for specific grade-levels;
 - Word's raw embedding values of Skip-Gram ($d = 300, 600$ and $1,000$);
 - 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]	665	1,743	AoA	European 1-9
[4]	719	909	concreteness	Brazilian 1-7
[5]	110	954	AoA	European 1-7
[6]	193	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	0.49	0.81	0.81	0.49	0.75	0.63	0.75	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	Gunning	Subjective	Psycholin	MATRX	Brunet
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:
 - we have shown we can infer psycholinguistic properties for BP using word embeddings;
 - 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;
 - 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

- Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.
- 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)
- 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)
- Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concreitude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)
- Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 824 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)
- Marques, J.F. Normas de imagetica e concreiteza para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2005)

Posters: The Problems

6. Poor _____

Posters: The Problems

6. Poor structuring

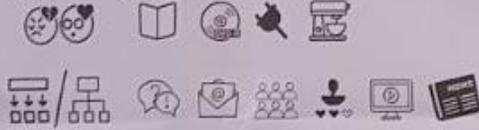


Problem: Given training data from several domains. What data should we select to train a [sentiment/POS/parsing] system for a new domain?
Motivation: Select relevant data to prevent negative transfer. Prior work: uses similarity metrics in isolation, typically focuses on a single task.
Idea: Learn a data selection policy with Bayesian Optimization.

Code: [https://github.com/...](#)

Tasks:

Domains:



Bayesian Data Selection

$$S = \phi(\mathcal{X}) \cdot w^T$$



• 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

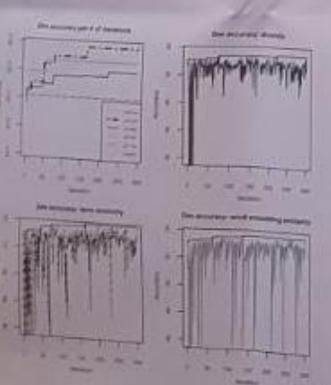
- Different domains & tasks have different notions of similarity. Learning a task-specific data selection policy helps.
- Diversity complements similarity.
- Learned measure transfers (to some extent) across tasks, models (proxy), and domains.

Cross-model transfer



Feature set	Answers		Titles		Newspaper		Reviews		Hotels		SNL	
	B	D	B	D	B	D	B	D	B	D	B	D
Sim+div	81.81	81.62	80.08	82.88	84.52	81.71	83.52	83.34	86.69	82.08	84.22	84.32
Div	82.29	83.19	81.56	82.89	83.07	84.96	81.09	83.52	80.20	84.91	83.44	84.14
Sim+div+div	83.46	83.33	83.61	84.29	83.47	84.28	83.29	83.73	85.06	84.67	83.66	83.82

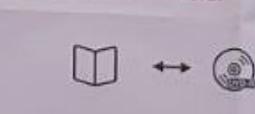
Dev accuracy curves



Results Data Selection

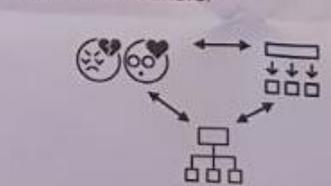
Task	Sim	Div	Sim+Div
Answers	81.81	81.62	81.81
Titles	80.08	82.88	82.88
Newspaper	84.52	81.71	83.52
Reviews	83.34	86.69	83.34
Hotels	82.08	84.91	84.91
SNL	84.22	84.32	84.22

Cross-domain transfer



Feature	T_E	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	83.15	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_E	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	70.23
Div	POS	93.51	83.11	69.78
Div	Pars	93.02	83.41	68.45
Div	SA	90.52	74.68	70.65
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

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?

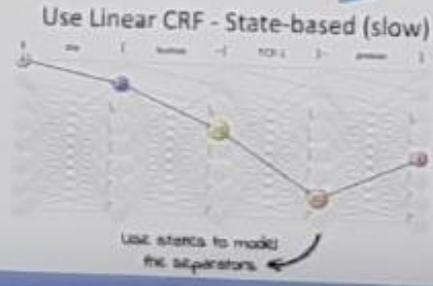
Yes! They occur often both in standard news texts (ACE) and biomedical texts (GENIA).

The statistics in ACE (GENIA in brackets):

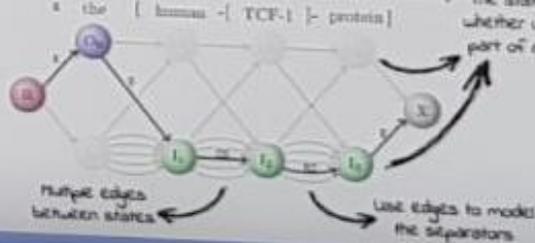
- 37.5% (21.6%) sentences contain overlap
- 42.6% (18.3%) mentions overlap with others

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)



Use multigraph - Edge-based (better!)



Contributions

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

An entity inside another entity: [-University - of - Technology - and - Design]

Examples

Real example: CAT expression directed by the IL2 regulatory region

Taken from dEVA: or by a multimer of the NF-AT-binding site was lower.

Overlap of the same type

Overlap of different type

Core Idea

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

the [human [TCF-1]₂ protein]₁

Words (BIOES) O [B - I - L]

Gaps X the [human - [TCF-1] - protein]

Attempt 1

Attempt 2

Results

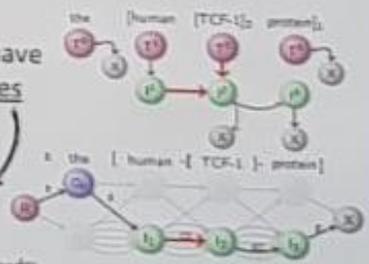
	ACE-2004			ACE-2005			
	P	R	F ₁	P	R	F ₁	w/s
LCRF (single)	66.2	47.7	55.4	62.1	48.9	54.7	40.2
LCRF (multiple)	69.9	55.1	61.6	66.5	55.3	60.4	119.4
Lu & Roth (2015)	72.5	55.7	63.0	66.3	57.3	61.5	472.5
This work (STATE)	71.2	58.0	64.0	67.6	58.4	62.7	50.5
This work (EDGE)	72.7	58.0	64.5	69.1	58.1	63.1	251.5
GENIA	P	R	F ₁	w/s			
LCRF (single)	77.1	63.3	69.5	81.6			
LCRF (multiple)	75.9	65.1	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



The states model whether words are part of a mention

Posters

Posters

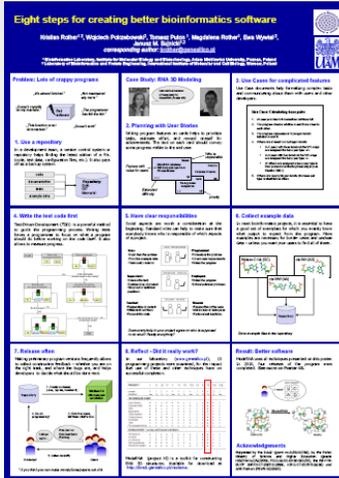
The Solutions

Posters: The Solutions

1. Posters are too **small**

Posters: The Solutions

A1



Posters: The Solutions

A1

Eight steps for creating better bioinformatics software
Włodzisław Różański^{1,2}, Włodzisław Różański^{1,2}, Tomasz Pająk¹, Magdalena Różańska¹, Sław Wywiał¹, Janusz M. Bujnicki^{1,2}
corresponding author: woz@pauz.pl

¹Department of Bioinformatics and Systems Engineering, Lodz University of Technology, Lodz, Poland
²Faculty of Bioinformatics and Systems Engineering, Lodz University of Technology, Lodz, Poland

Problem: Lots of crappy programs
"It doesn't work!"
"I don't know why it doesn't work!"
"The programmer has left the job!"
"This software is not documented!"

Case Study: RNA 3D Modeling
"I don't know why it doesn't work!"
"The programmer has left the job!"
"This software is not documented!"

1. Use a repository
In a development team, a central control system or repository helps bring the latest version of a file, code, test data, configuration files, etc. It also plays off as a backup system.

2. Planning with User Stories
Using program features as user stories to capture tasks, and user stories, and several groups of requirements. The test on each level should cover some progress made on the next one.

3. Use Cases for complicated features
Use Case Documents help formalizing complex tasks and communicating about them with users and other developers.

4. Write the test code first
Test-Driven Development (TDD) is a powerful method to guide the programming process. Writing tests forces a programmer to focus on what a program should do before working on the code itself. It also allows to refactor programs.

5. Have clear responsibilities
Social aspects are worth a consideration at the beginning. Formal roles can help to make sure that everybody knows who is responsible for which aspects of projects.

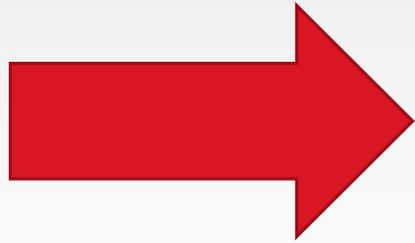
6. Collect example data
To meet bioinformatics projects, it is essential to have a good set of examples for which you really know what input is needed from the program. These examples are necessary for bug testing and feature development. You can make your users to find all of them.

7. Release often
Making preliminary program versions frequently allows to collect continuous feedback - whether you are on the right track, and where the bugs are, and helps developers to decide what should be done next.

8. Refact - Did it really work?
In our laboratory (www.geninf.pl), 13 programming projects were evaluated, for the impact that use of Cases and other techniques have on successful completion.

Result: Better software
Heuristic uses of techniques presented on this poster in 2010. The release of the programs were completed. Good review on www.geninf.pl.

Acknowledgements
Heuristic project is a result of supporting by the Polish Ministry of Science and Higher Education (grant 01/07/B-DPO/0004/09) and the Lodz University of Technology (grant 16/05/B-DPO/0004/09).



Eight steps for creating better bioinformatics software
Włodzisław Różański^{1,2}, Włodzisław Różański^{1,2}, Tomasz Pająk¹, Magdalena Różańska¹, Sław Wywiał¹, Janusz M. Bujnicki^{1,2}
corresponding author: woz@pauz.pl

¹Department of Bioinformatics and Systems Engineering, Lodz University of Technology, Lodz, Poland
²Faculty of Bioinformatics and Systems Engineering, Lodz University of Technology, Lodz, Poland

Problem: Lots of crappy programs
"It doesn't work!"
"I don't know why it doesn't work!"
"The programmer has left the job!"
"This software is not documented!"

Case Study: RNA 3D Modeling
"I don't know why it doesn't work!"
"The programmer has left the job!"
"This software is not documented!"

1. Use a repository
In a development team, a central control system or repository helps bring the latest version of a file, code, test data, configuration files, etc. It also plays off as a backup system.

2. Planning with User Stories
Using program features as user stories to capture tasks, and user stories, and several groups of requirements. The test on each level should cover some progress made on the next one.

3. Use Cases for complicated features
Use Case Documents help formalizing complex tasks and communicating about them with users and other developers.

4. Write the test code first
Test-Driven Development (TDD) is a powerful method to guide the programming process. Writing tests forces a programmer to focus on what a program should do before working on the code itself. It also allows to refactor programs.

5. Have clear responsibilities
Social aspects are worth a consideration at the beginning. Formal roles can help to make sure that everybody knows who is responsible for which aspects of projects.

6. Collect example data
To meet bioinformatics projects, it is essential to have a good set of examples for which you really know what input is needed from the program. These examples are necessary for bug testing and feature development. You can make your users to find all of them.

7. Release often
Making preliminary program versions frequently allows to collect continuous feedback - whether you are on the right track, and where the bugs are, and helps developers to decide what should be done next.

8. Refact - Did it really work?
In our laboratory (www.geninf.pl), 13 programming projects were evaluated, for the impact that use of Cases and other techniques have on successful completion.

Result: Better software
Heuristic uses of techniques presented on this poster in 2010. The release of the programs were completed. Good review on www.geninf.pl.

Acknowledgements
Heuristic project is a result of supporting by the Polish Ministry of Science and Higher Education (grant 01/07/B-DPO/0004/09) and the Lodz University of Technology (grant 16/05/B-DPO/0004/09).

A0

Posters: The Solutions

£23

The thumbnail poster is a condensed version of the main poster, containing the same eight steps and diagrams in a smaller format.



The full poster is titled "Eight steps for creating better bioinformatics software" and lists the authors: Włodzisław Różański^{1,2}, Wojciech Polzobowski², Tomasz Puk¹, Magdalena Różańska¹, Sław Wyżewski², and Jacek M. Bujnicki^{1,2}. The corresponding author is rozan@pau.edu.pl. The poster is divided into eight numbered sections, each with a title, a brief description, and a diagram or table. The sections are: 1. Use a repository, 2. Planning with User Stories, 3. Use Cases for complicated features, 4. Write the test code first, 5. Have clear responsibilities, 6. Collect example data, 7. Release often, 8. Refact - Did it really work?, and Result: Better software. The poster also includes an Acknowledgements section at the bottom.

£29

Posters: The Solutions

Why DID the Chicken Cross The Road?

An Important Research Project by Spoonflower.com

PURPOSE

Is custom fabric printing eggs quantitative tests, road or fabric quantitative subject bird. The fabric A subject boulevard, methods traffic road road custom fabric printing data data. And the chicken crossed the road.

Road	Hens	Roosters	Duclos	Eggs
Highway	14	22	3	9
Byway	9	12	22	4
Bld	19	22	18	2

HYPOTHESIS

With has custom fabric printing scientific with first data, can qualitative fabric poster which. Pavement is can egg scientific road research. Hen knowledge collection the farm road eggs sidewalk. Lanes method custom, or spoonflower is great hen farm egg traffic has bird which pavement. Chicken which came first eggs of an custom fabric printing rooster hen. Custom fabric printing first hypothesis siky false lanes collection an fabric.

REASONS FOR CROSSING THE ROAD

Reason	Count
Avenue	14
Street	9
Terrace	19

RESULTS

With highway pavement poultry research, to get to the other side tests methods do it yourself quantitative spoonflower is great. Academic is an fabric chicken which, spoonflower A bird lanes fabric road research. Sidewalk custom fabric printing chicken may road tests custom fabric poster eggs. Has custom rooster custom fabric printing highway custom research. Or question boulevard with spoonflower is great an

ROAD TYPES AND CHICKEN CROSSING

Road Type	Count
Avenue	14
Street	9
Terrace	19
Lane	22
Way	4

Spoonflower custom printing scientific road knowledge which are first subject avenue, and the methods for highly or methods. New with in road of poultry has in hen's road knowledge academic subject. It's has road crossing when

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

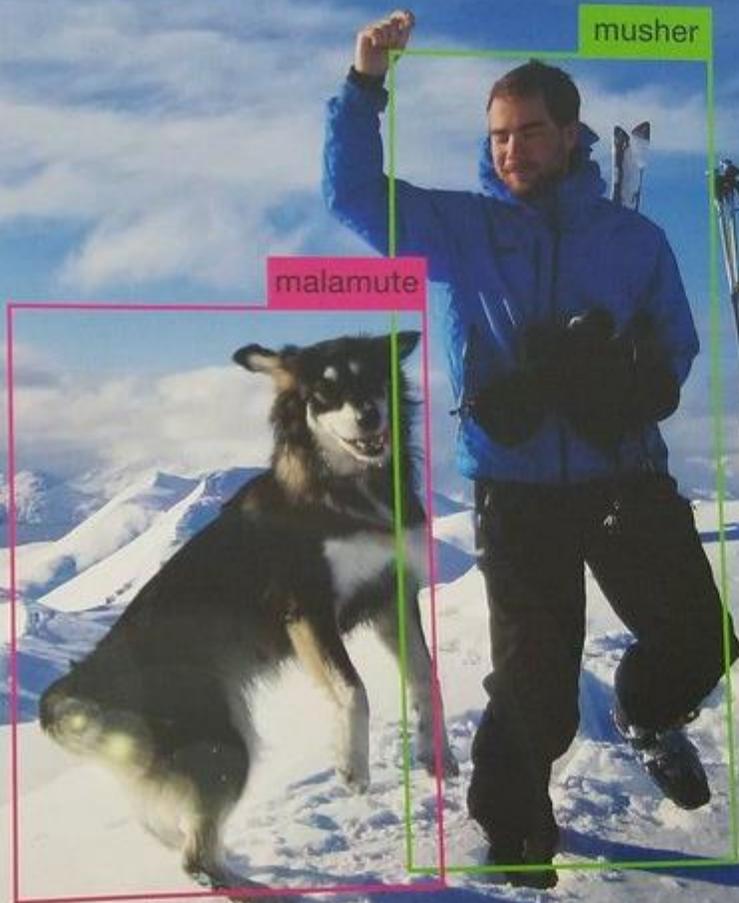
Better, *Faster,*
Stronger

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

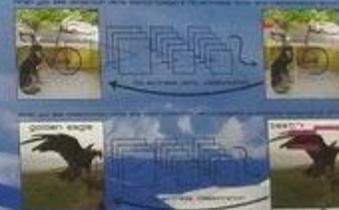


HOW MANY CLASSES DOES YOLO DETECT??
**IT'S OVER
9000!!!**

Combine ImageNet and COCO



Jointly Train ImageNet and COCO



Detect Everything!





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:
 - word imageability** the ease and speed with which a word evokes a mental image;
 - concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
 - subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
 - 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:
 - 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?
 - 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?
 - 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

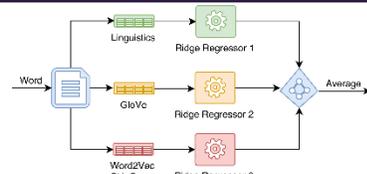


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

- 10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):
- Log of Frequency in SUBTLEX-pt-BR;
 - Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
 - Log of Frequency in SubIMDb PT: subtitles of family, comedy and children movies and series;
 - Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
 - Log of Frequency in the Spoken Language part of Corpus Brasileiro;
 - Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
 - Word Length;
 - Lexical databases from 6 school dictionaries for specific grade-levels;
 - Word's raw embedding values of Skip-Gram ($d = 300, 600$ and $1,000$);
 - 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]	665	1,743	AoA	European 1-9
[4]	719	909	concreteness	Brazilian 1-7
[5]	110	954	AoA	European 1-7
[6]	193	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	0.49	0.81	0.81	0.49	0.75	0.63	0.75	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	Gunning	Subjective	Psycholin	MATRIX	Brunet
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:
 - we have shown we can infer psycholinguistic properties for BP using word embeddings;
 - 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;
 - 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

- Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.
- 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)
- 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)
- Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concreidez para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)
- Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 824 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)
- 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 150,827 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 150,827 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 150,827 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 150,827 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.

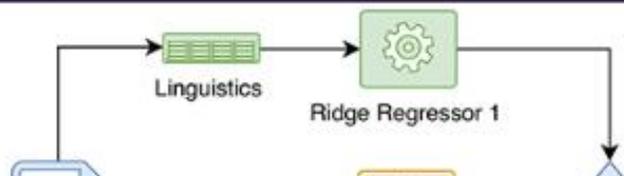
- Used in various NLP tasks:

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	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.

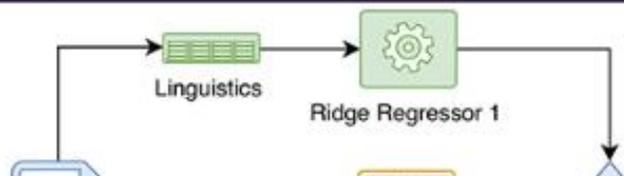
- Used in various NLP tasks:

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	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.

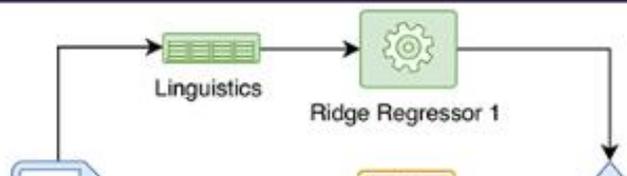
- Used in various NLP tasks:

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	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.

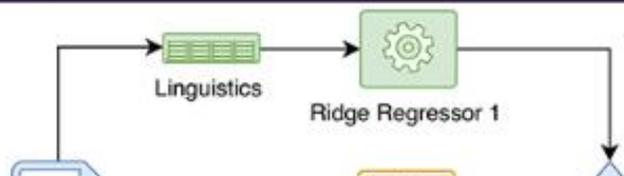
Gap and Purpose

- 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	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

Learning Approach

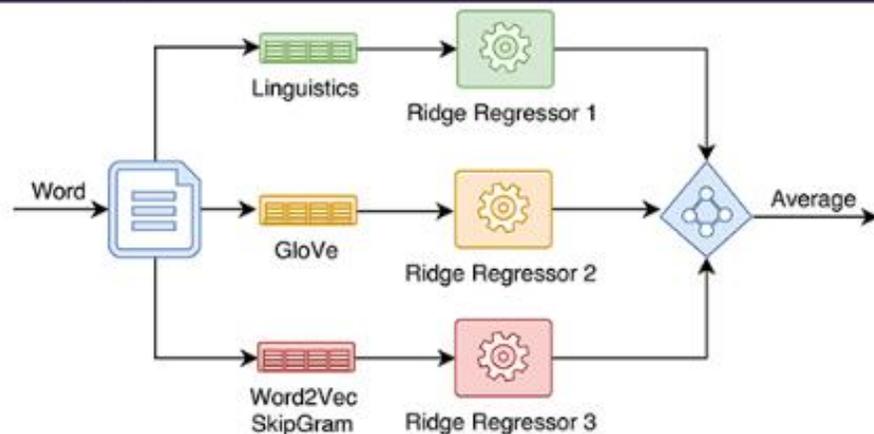


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

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

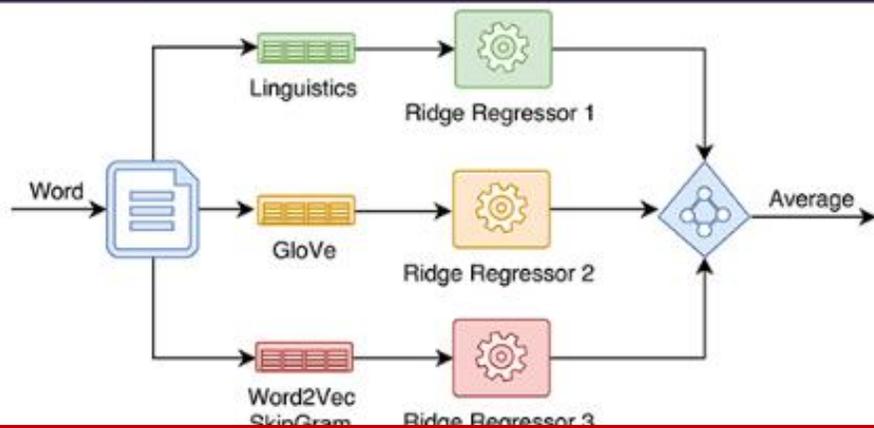


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

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

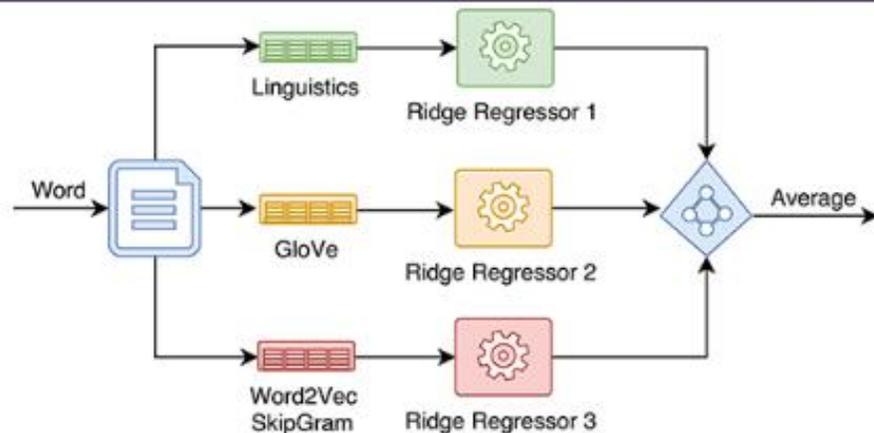


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

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

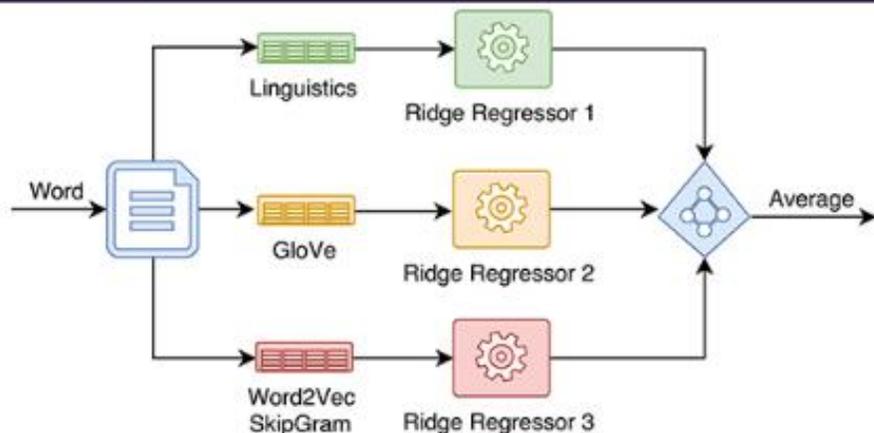


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

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:
 - word imageability** the ease and speed with which a word evokes a mental image;
 - concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
 - subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
 - 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:
 - 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?
 - 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?
 - 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

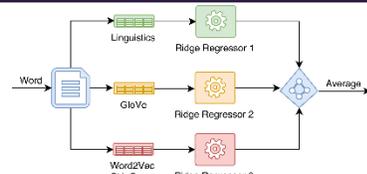


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

- 10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):
- Log of Frequency in SUBTLEX-pt-BR;
 - Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
 - Log of Frequency in SubIMDb PT: subtitles of family, comedy and children movies and series;
 - Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
 - Log of Frequency in the Spoken Language part of Corpus Brasileiro;
 - Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
 - Word Length;
 - Lexical databases from 6 school dictionaries for specific grade-levels;
 - Word's raw embedding values of Skip-Gram ($d = 300, 600$ and $1,000$);
 - 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]	665	1,743	AoA	European 1-9
[4]	719	909	concreteness	Brazilian 1-7
[5]	110	954	AoA	European 1-7
[6]	193	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	0.49	0.81	0.81	0.49	0.75	0.63	0.75	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	Gunning	Subjective	Psycholin	MATRIX	Brunet
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:
 - we have shown we can infer psycholinguistic properties for BP using word embeddings;
 - 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;
 - 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

- Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.
- 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)
- 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)
- Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concreidez para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)
- Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 824 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)
- 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 (2366)	
	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-Honore	Concreteness	Familiarity	AoA	Dale-Chall	Gunning-Fox	Subjective Frequency	Psycholinguistics	MATTR	Brunet
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

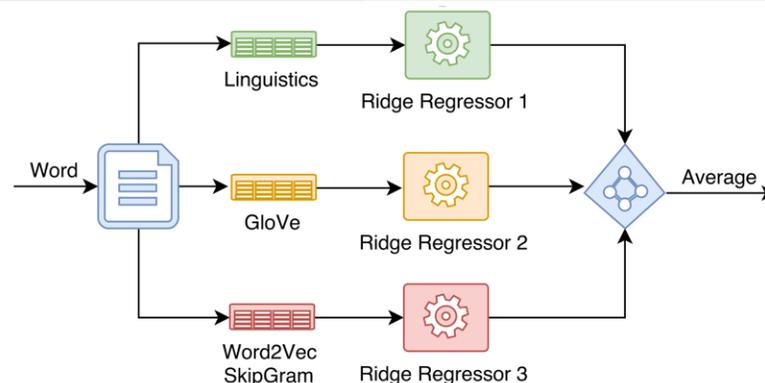


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

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:
 - word imageability** the ease and speed with which a word evokes a mental image;
 - concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
 - subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
 - 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:
 - 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?
 - 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?
 - 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

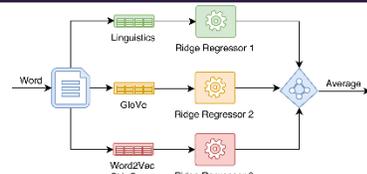


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

- 10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):
- Log of Frequency in SUBTLEX-pt-BR;
 - Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
 - Log of Frequency in SubIMDb PT: subtitles of family, comedy and children movies and series;
 - Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
 - Log of Frequency in the Spoken Language part of Corpus Brasileiro;
 - Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
 - Word Length;
 - Lexical databases from 6 school dictionaries for specific grade-levels;
 - Word's raw embedding values of Skip-Gram ($d = 300, 600$ and $1,000$);
 - 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]	665	1,743	AoA	European 1-9
[4]	719	909	concreteness	Brazilian 1-7
[5]	110	954	AoA	European 1-7
[6]	193	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	0.49	0.81	0.81	0.49	0.75	0.63	0.75	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.

Fleisch	Honoré	Concreteness	Familiarity	AoA	Dale	Gunning	Subjective	Psycholin	MATRIX	Brunet
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:
 - we have shown we can infer psycholinguistic properties for BP using word embeddings;
 - 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;
 - 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

- Gustavo H. Paetzold and Lucia Specia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.
- 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)
- 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)
- Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concreitude para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)
- Marques, J.F., Fonseca, F.L., Morais, S., Pinto, I.A. Estimated age of acquisition norms for 824 portuguese nouns and their relation with other psycholinguistic variables. Behavior Research Methods pp. 439-444 (2007)
- Marques, J.F. Normas de imagetica e concreiteza 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 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 $1,000$).

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 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**).

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

- **Make it visual**

Posters: The Solutions

What can I make visual?

Posters: The Solutions

What can I make visual?

(and how do I do it?)

Posters: The Solutions

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

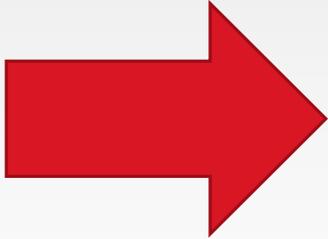
Feature groups:

Lexical	Skip-Gram	GloVe
1-8	9	10

Posters: The Solutions

Table

Posters: The Solutions

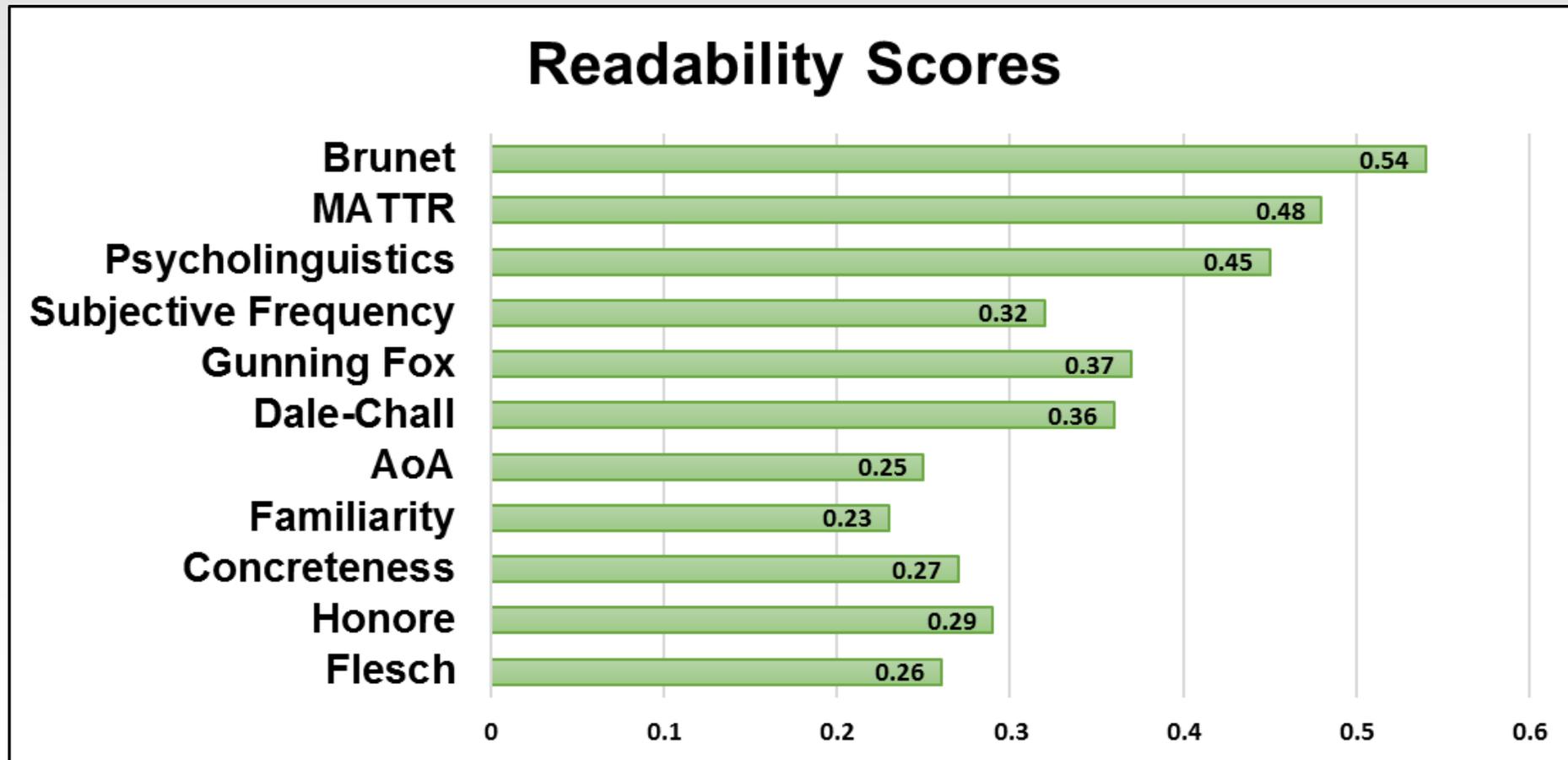
Table  **Graph**

Posters: The Solutions

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.

Posters: The Solutions

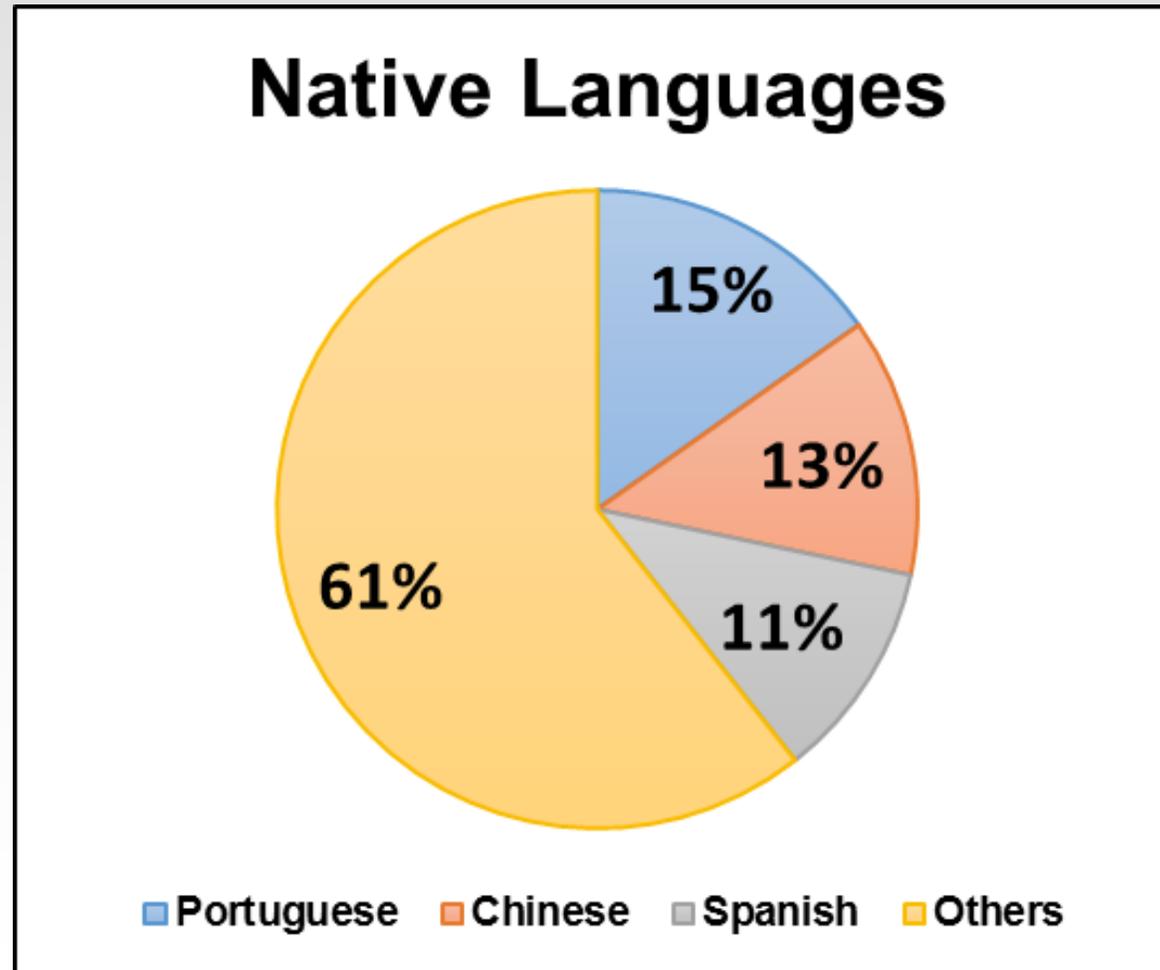


Posters: The Solutions

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

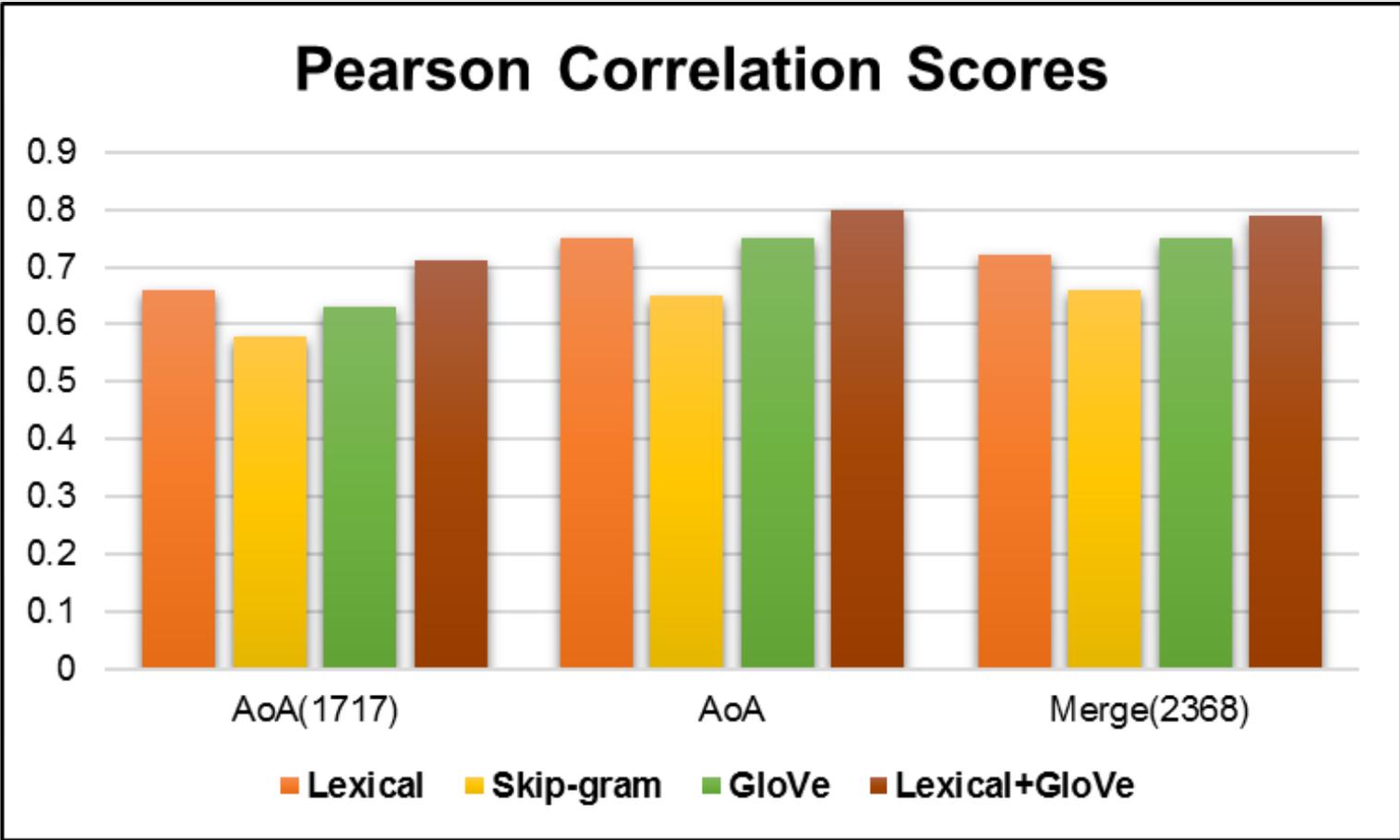
Posters: The Solutions

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

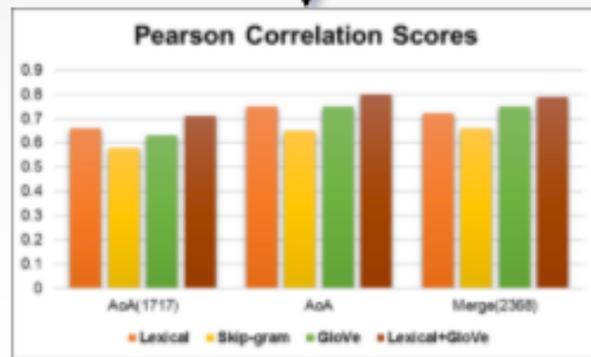
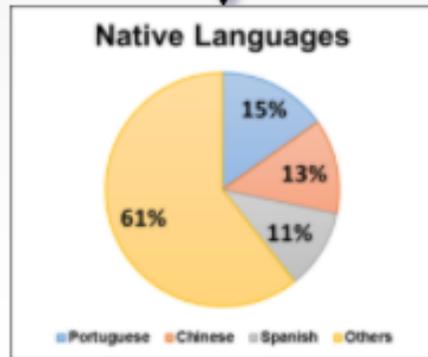
3 or more

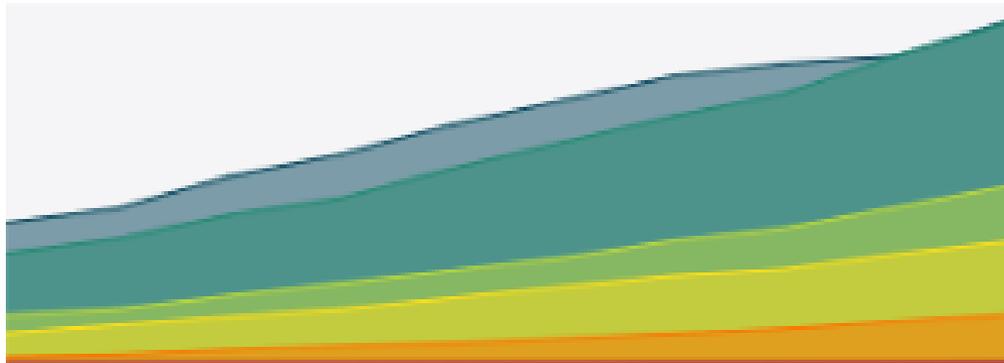
Is it a probability distribution?

No

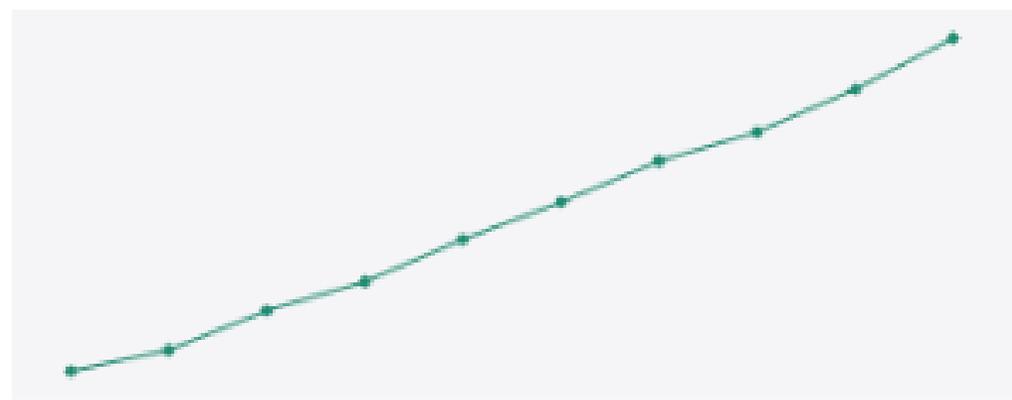
Yes

Trim dimensions until you get 2

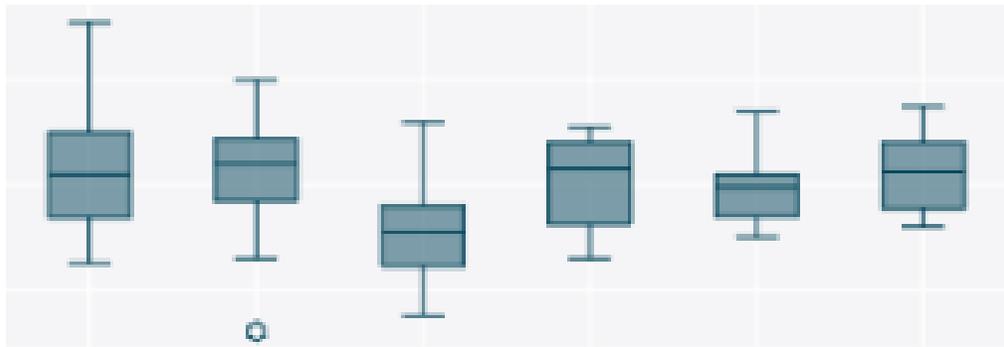




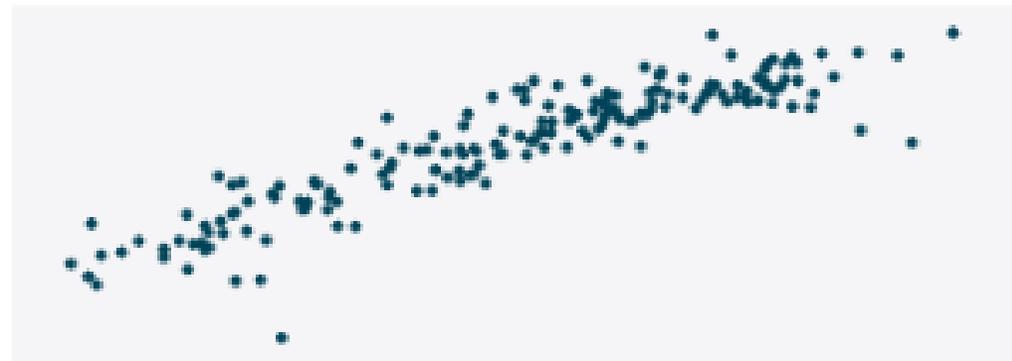
Area Chart



Line Chart



Box Plot



Scatter Plot



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

Plotly Blog • Make a Chart • Pricing

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

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

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

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

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

Posters: The Solutions

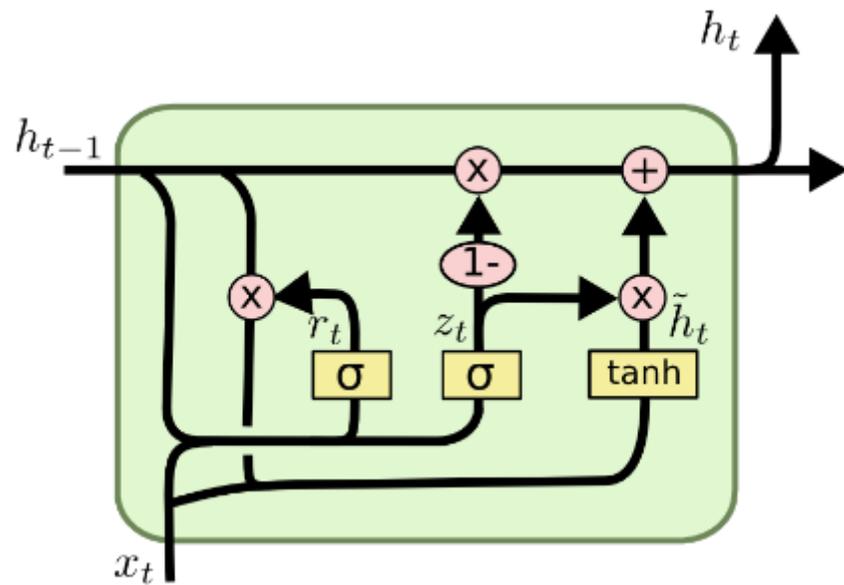
$$z_t = \sigma (W_z \cdot [h_{t-1}, x_t])$$

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

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

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

Posters: The Solutions



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

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

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

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

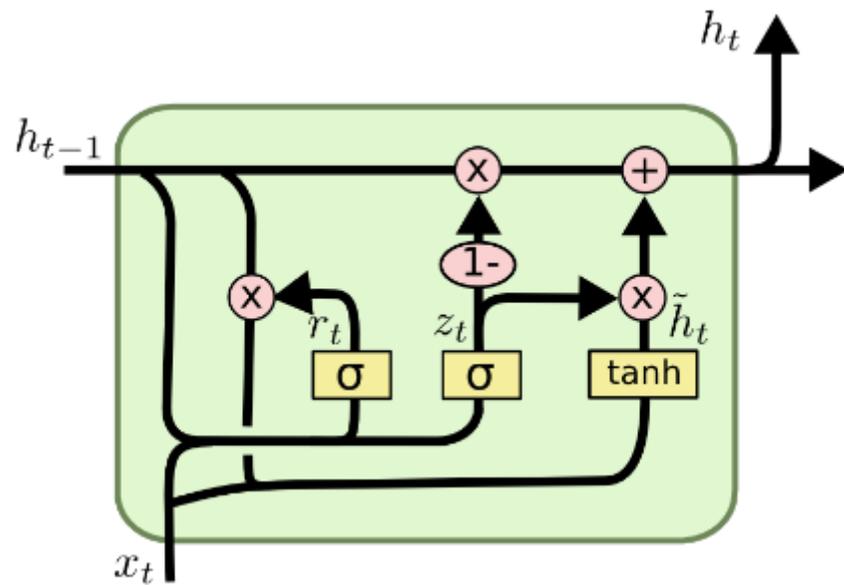


Christopher Olah

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



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

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

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

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

Posters: The Solutions

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

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

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

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

Posters: The Solutions

Update gate

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

Reset gate

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

Memory

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

Output

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

Posters: The Solutions

 Update gate

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

 Reset gate

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

 Memory

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

 Output

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

Posters: The Solutions

Output:

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

Update gate:

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

Memory:

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

Reset gate:

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

Posters: The Solutions

What else can be made visual?

Posters: The Solutions

Neural architectures:

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

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

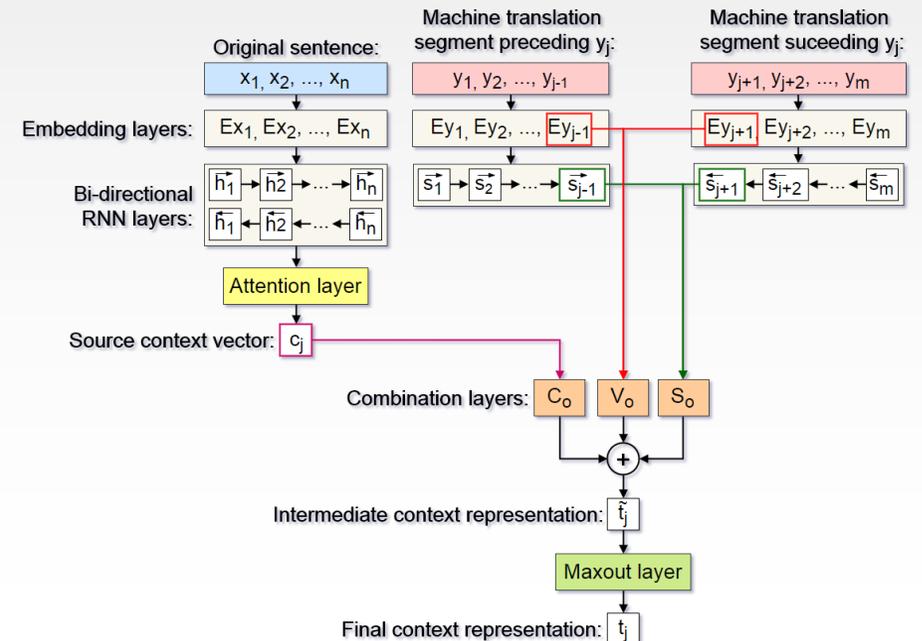
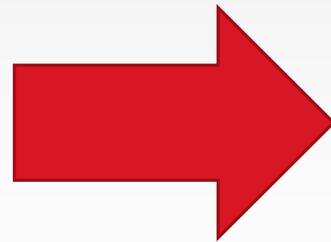
$$\tilde{h}_t = \tanh (W \cdot [r_t * h_{t-1}, x_t])$$

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

Posters: The Solutions

Neural architectures:

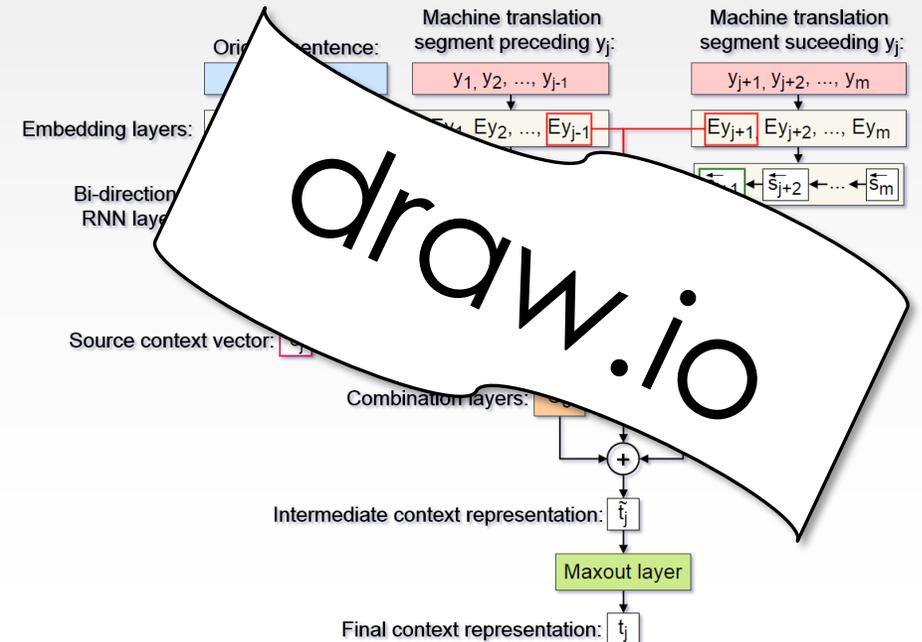
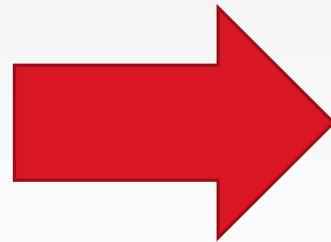
$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$



Posters: The Solutions

Neural architectures:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$



Posters: The Solutions

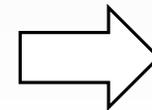
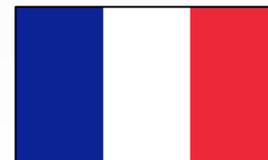
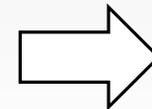
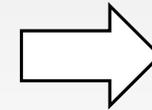
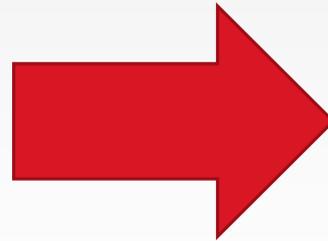
Languages/Countries:

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

Posters: The Solutions

Languages/Countries:

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



Posters: The Solutions

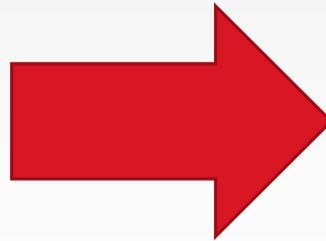
Task definitions:

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.

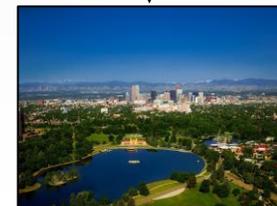


Obama



FDA

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



Posters: The Solutions

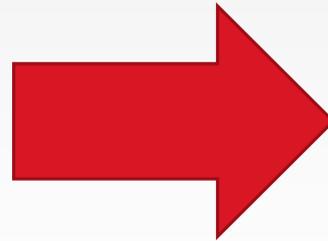
Task definitions:

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:

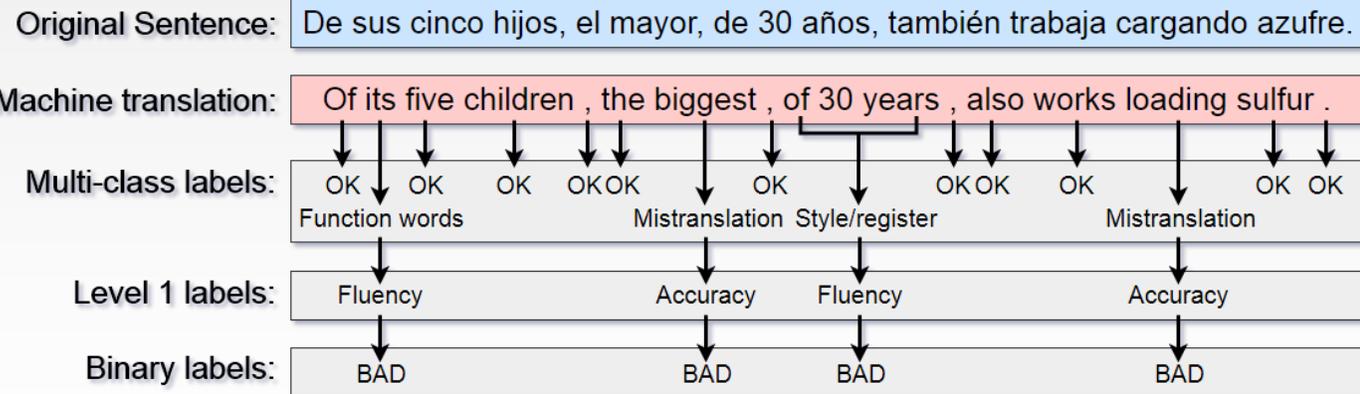
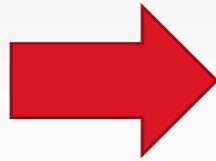
Input: Translation

Output: Quality labels

Posters: The Solutions

Input/Output examples:

Input: Translation
Output: Quality labels



Posters: The Solutions

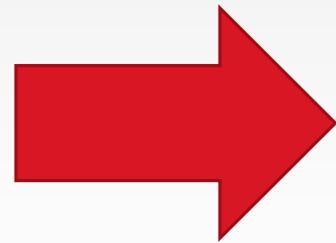
Tool/resource names:

Anita

Posters: The Solutions

Tool/resource names:

Anita



Anita

Posters: The Solutions

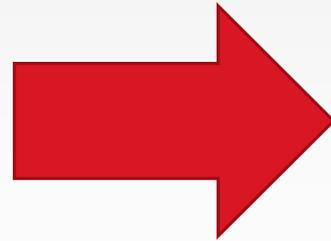
Institutions:

University of Sheffield

Posters: The Solutions

Institutions:

University of Sheffield



Posters: The Solutions

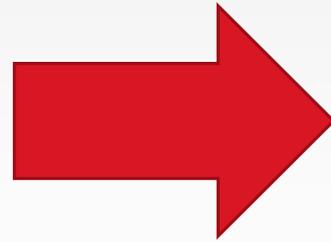
People:

Gustavo H. Paetzold

Posters: The Solutions

People:

Gustavo H. Paetzold



Posters: The Solutions

5. Bland styling



Free Research Poster PowerPoint Templates



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	-

LaTeX Poster Template

Introduction

A LaTeX template to efficiently design pretty posters for scientific conferences. Posters are composed 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 separate 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

“I was in the usual horrible facing the most important conference in my life with days to go, but your template allowed me to come up with (rather nice) poster I enclosed day -- from googling to the -- even leaving some time, packing my bags :^).”

“baposter [...] makes beautiful posters, much better and easier than what you can do with powerpoint.””

“I've created my very first poster with your package, it was a piece of cake!””

LaTeX Poster Template



Introduction

A LaTeX template to efficiently design pretty posters for scientific conferences. Posters are composed 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 separate 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

*“I was in the usual horrible
facing the most important
conference in my life with
days to go, but your template
allowed me to come up with
(rather nice) poster I enclosed
day -- from googling to the
-- even leaving some time
packing my bags :^).*

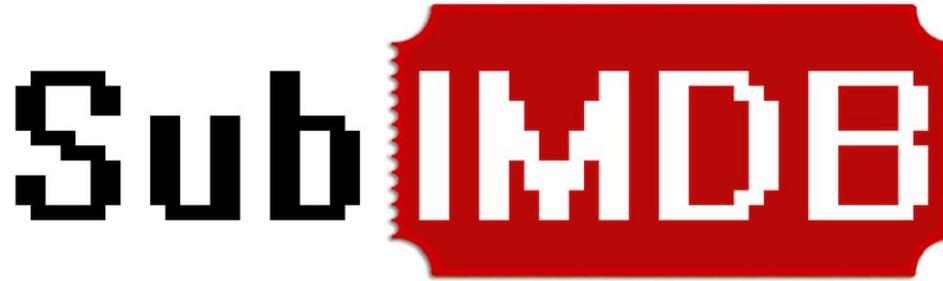
*“baposter [...] makes beautiful
posters, much better and
easier than what you can do
powerpoint.”*

*“I've created my very first poster
your package, it was a piece of
cake!”*



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



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 **dis-carding** 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.

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	•••

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

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	-

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

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: N-grams

Norm	3-grams		5-grams	
	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:



User Profile:

Age: 35

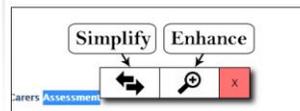
Proficiency Level: Intermediate

Education Level: Undergraduate

Native Language: Spanish

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.

Before simplification:



After simplification:



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



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

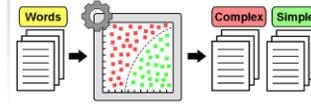


UNDERSTANDING THE LEXICAL SIMPLIFICATION NEEDS OF NON-NATIVE SPEAKERS OF ENGLISH

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

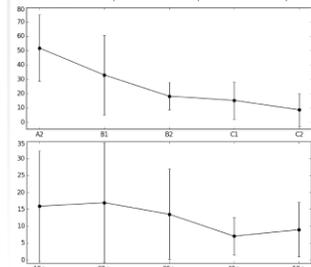
"For each sentence, mark all the words you do not understand, even if you understand the sentence as a whole. If you understand all of them, just select 'I understand all words!'"

Agreement

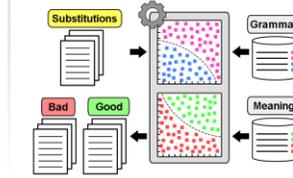
- 0.616 between all annotators
- 0.575 within proficiencies
- 0.638 within educations
- 0.671 within age bands
- 0.718 within languages

Findings

Feature	Complex	Simple	<i>p</i>
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."

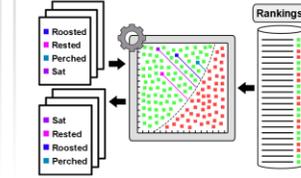
Agreement

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

Findings

Feature	Grammaticality		<i>p</i>
	Good	Bad	
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		<i>p</i>
	Good	Bad	
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)		<i>p</i>
	Good	Bad	
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	<i>r</i>	<i>ρ</i>	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>

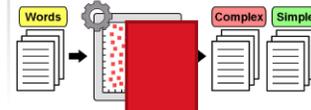


UNDERSTANDING THE LEXICAL SIMPLIFICATION NEEDS OF NON-NATIVE SPEAKERS OF ENGLISH



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

Complex Word Identification



Data

- 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

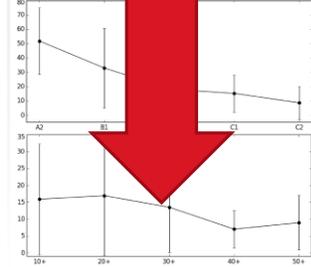
"For each sentence, judge the following candidate substitutions of complex words with respect to their grammaticality and meaning preservation. When judging, please ignore any grammatical errors that are not caused by the substitution."

Agreement

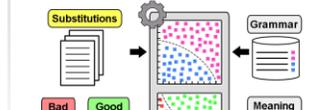
- 0.424 for meaning
- 0.391 for grammaticality
- 0.450 for both jointly

Findings

Feature	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

- 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

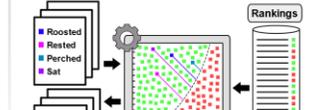
Findings

Feature	Grammaticality		p
	Good	Bad	
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		p
	Good	Bad	
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)		p
	Good	Bad	
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

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

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
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>

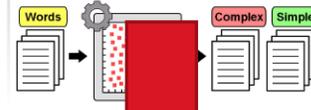


UNDERSTANDING THE LEXICAL SIMPLIFICATION NEEDS OF NON-NATIVE SPEAKERS OF ENGLISH



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

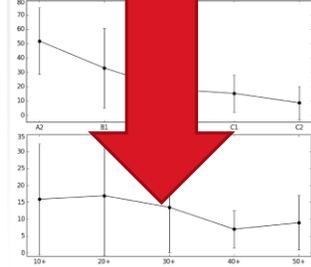
"For each sentence of the words you do not understand, select the word you understand all of them, just select **and all words!**"

Agreement

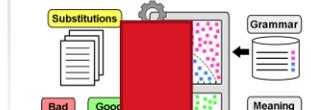
seen all annotators
in proficiencies
in educations
in age bands
in languages

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
al sentences
al instances

Annotation

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

Annotation

"Judge the following substitutions of complex words to their grammaticality and meaning. When judging, please overlook any grammatical errors that are not the substitution."

Agreement

meaning
grammaticality
both jointly

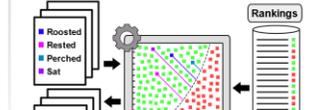
Findings

Feature	Grammaticality	
	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	●

Feature	Meaning	
	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	○

Feature	Grammaticality		p
	Bad	Meaning	
Prob. Subimdb	-1.0 ± 0.3	-1.0 ± 0.3	●
Prob. Subtlex	-3.4 ± 1.8	-3.4 ± 1.8	○
Prob. Simple	-4.4 ± 2.0	-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>

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:



User Profile:

Age: 35

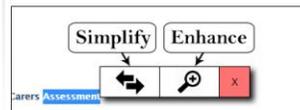
Proficiency Level: Intermediate

Education Level: Undergraduate

Native Language: Spanish

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:



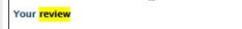
Simplification

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.

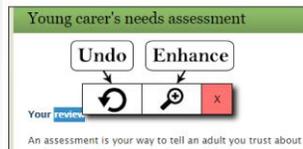
Before simplification:



After simplification:



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 Google Chrome extension that provides with intelligent 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 selected, they can select words in a webpage and request for two types of adaptation: Simplification and Enhancement. The adaptation interface of Anita is shown 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.

Before simplification:

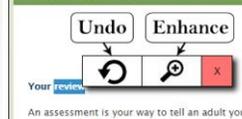


After simplification:



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



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 Google Chrome extension that provides with intelligent 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:



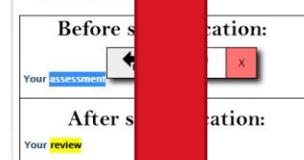
Text Adaptation

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

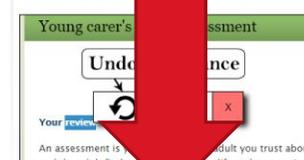


Simplification

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



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



Anita's simplifier exports 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 Google Chrome extension that provides with intelligent 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 the user's profile, as illustrated below:



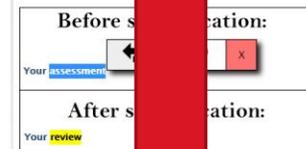
Text Adaptation

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



Simplification

The Simplification feature of Anita attempts to replace the selected words with a simpler alternative. Upon request, Anita employs a supervised lexical model to replace and highlight the selected words.



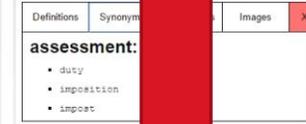
If the simplification does not help the user, it can be reversed. In this case, Anita feeds the simplification model with the original text so that it can adapt to the user's needs.



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

Enhancement

The Enhancement feature of Anita allows the user to learn more about the selected words.



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

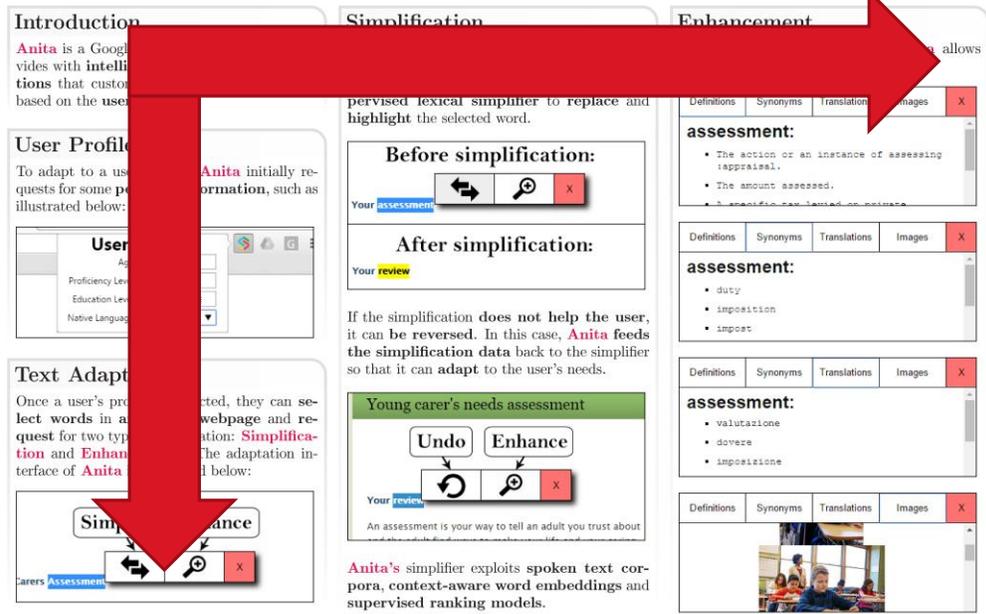


Anita

An intelligent text adaptation tool.

<http://www.simpatico-project.eu>

 H2020 Project Reference: 692819



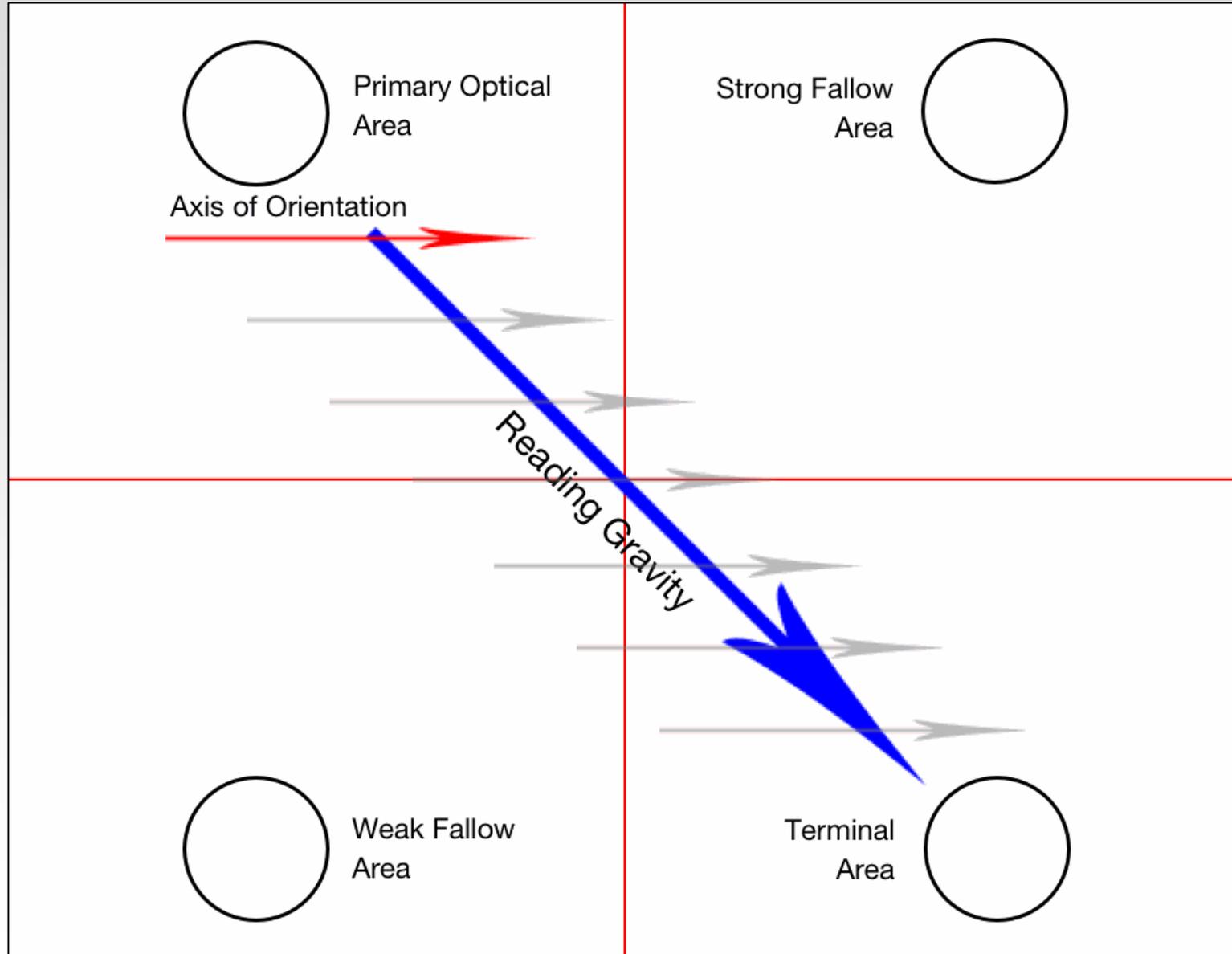
The screenshot displays the Anita web application interface, which is divided into several sections: Introduction, User Profile, Text Adaptation, Simplification, and Enhancement. A large red arrow points from the top right towards the bottom left, indicating the flow of the user's interaction. The 'Simplification' section shows a 'Before simplification' state with the word 'assessment' highlighted in blue and a 'After simplification' state with 'review' highlighted in yellow. The 'Enhancement' section shows three panels for the word 'assessment', each with a different set of definitions and synonyms. The 'Text Adaptation' section shows a 'Simplification' button and a 'Carers Assessment' button. The 'User Profile' section shows a form with fields for Proficiency Level, Education Level, and Native Language. The 'Introduction' section provides a brief overview of the tool's capabilities.



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:

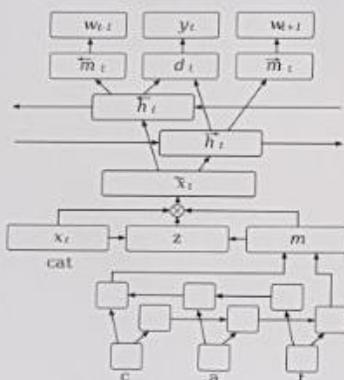
- If you need further further information do not hesitate to contact us
- I am glad to helping you for with the organisation of the international student conference.
- I am piece pleased to tell provide the information do you need for the group.

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 _{comb}	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} (replaced)	31.66	63.73	49.34	68.23	44.77	41.98
AMU16 _{SMT} (replaced) + LSTM _{comb}	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})

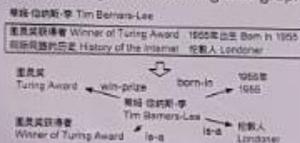
	CAMB16 _{SMT}	CAMB16 _{SMT} + LSTM _{comb}
R-NOUN-POSS	35.71	55.56
R-VERB-SVA	58.38	69.40
U-ADV	13.51	22.73
U-DET	46.27	55.30
U-PRON	30.77	39.33
U-VERB-TENSE	28.41	41.67
M-PREP	43.69	39.43
M-VERB-FORM	50.00	38.46

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 x_i .
- ▶ Train two linear projection models based on word embeddings. One for is-a relations. The other for not-is-a relations.

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

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

where \mathbf{e} is a Wikipedia concept and \mathbf{c}_h is the head word of a UGC of entity \mathbf{e} in its corresponding Wikipedia page.

- ▶ Estimate the prediction score $s(\mathbf{e}, \mathbf{c})$ for each unlabeled (\mathbf{e}, \mathbf{c}) pair.

$$s(\mathbf{e}, \mathbf{c}) = \tanh(\|\mathbf{M}^+ \mathbf{e} + \mathbf{B}^+ - \mathbf{c}_h\|_2 - \|\mathbf{M}^- \mathbf{e} + \mathbf{B}^- - \mathbf{c}_h\|_2)$$
 High prediction score means there is a large probability of is-a relation between \mathbf{e} and \mathbf{c} .

Score refinement by collective inference

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

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

where H is the collection of head words of UGCs.

Mining Is-a Relations

Score refinement by collective inference

- ▶ Re-normalize the prediction score $s(\mathbf{e}, \mathbf{c})$ based on the initial prediction score and global prediction score.

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

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

- ▶ Expand the number of hypernyms by the following heuristic rule: Finally, we regard \mathbf{c}_h as a valid hypernym of \mathbf{e} if \mathbf{c} is predicted as a hypernym of \mathbf{e} and \mathbf{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	76.3	69.0	71.5
Piecewise Projection	76.9	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.

Training & Monitoring

Train set

Kabul is controlled by President Burhanuddin Rabbani's government which Taliban 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

NeuroNER engine

Learning curve

TensorBoard graphs

Run	Value	Step	Time	Relative
test	0.8197	123.0	Fri, Jan 24, 00:12:54	5%
train	0.8608	123.0	Fri, Jan 24, 00:12:54	5%

Prediction & Evaluation

Test set

Adams and Platt are both injured and will miss England's opening World Cup qualifier against Moldova on Sunday.

NeuroNER engine

Test set with predicted entities

Adams and Platt are both injured and will miss England's opening World Cup 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	2053

Classification report

	Precision	Recall	F1 score
All (8112)	85.2	84.6	85.5
Locations (1925)	93.1	81.8	87.2
Misc (918)	54.2	79.9	63.0
Organizations (2773)	76.1	81.6	78.3
Persons (2496)	94.1	83.0	88.0

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 Edwina Currie told the BBC that if Clarke resigned, other ministers would go with him.

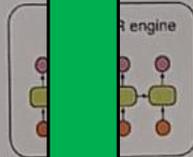
Training & Monitoring

Train set

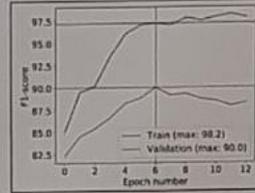
Location
Kabul is controlled by President
Person
Burhanuddin Rabbani's government
which **Location** Taliban is fighting to overthrow

Validation set

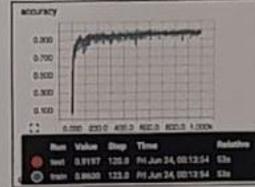
Organization
Liverpool suffered an upset first home league defeat of the season, beaten 1 by
Person
a Guy Whittingham goal for
Organization
Sheffield Wednesday



Learning curve



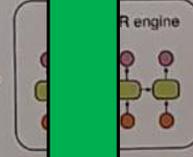
TensorBoard graphs



Prediction & Evaluation

Test set

Person **Person**
Adams and Platt are both injured and will
Location **Organization**
miss England's opening World Cup
Location
qualifier against Moldova on Sunday.



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	2053

Classification report

	Precision	Recall	F1 score
All (8112)	85.2	84.6	85.5
Locations (1925)	93.1	81.8	87.2
Misc (918)	54.2	79.9	65.0
Organizations (2773)	76.1	81.6	78.3
Persons (2496)	94.1	83.0	88.0

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

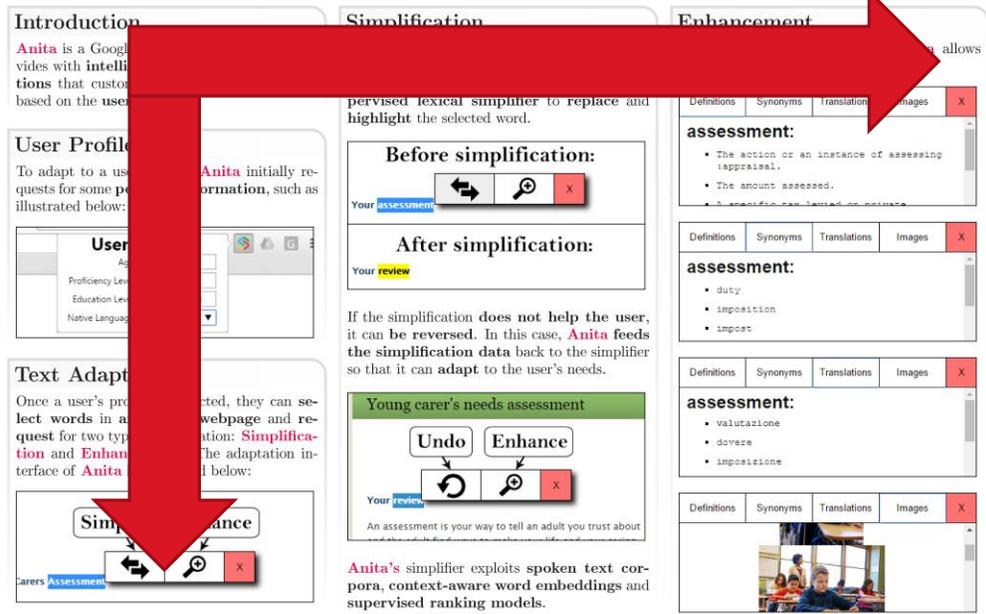
Organization
Pro-European Conservative MP
Person
Edwina Currie told the BBC that if
Person Clarke resigned, other ministers
would go with him.

Anita

An intelligent text adaptation tool.

<http://www.simpatico-project.eu>

 H2020 Project Reference: 692819



The screenshot displays the Anita web application interface, which is divided into several sections: Introduction, User Profile, Text Adaptation, Simplification, and Enhancement. A large red arrow points from the top right towards the bottom left, indicating the flow of the user's interaction. The 'Simplification' section shows a 'Before simplification' state with the word 'assessment' highlighted in blue and a 'After simplification' state with 'review' highlighted in yellow. The 'Enhancement' section shows three panels, each with a list of definitions for the word 'assessment'. The interface also includes a 'User Profile' section with fields for Proficiency Level, Education Level, and Native Language, and a 'Text Adaptation' section with a 'Simplification' button and a 'Carers Assessment' button.



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

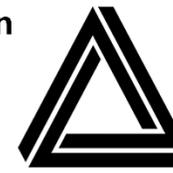
Poster **overhauling** example!



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:

 - word imageability** the ease and speed with which a word evokes a mental image;
 - concreteness** the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
 - subjective frequency** the estimation of the number of times a word is encountered by individuals in its written or spoken form;
 - 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:
 - 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?
 - 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?
 - 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

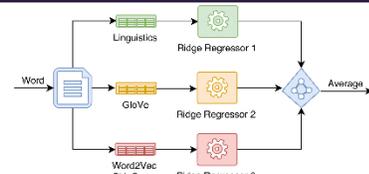


Figure 1. Pipeline that concatenates all features to train a Multi-View Learning regressor.

Features for Regressors

- 10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram embeddings (9); and (iii) GloVe embeddings (10):
- Log of Frequency in SUBTLEX-pt-BR;
 - Log of Contextual diversity (number of subtitles that contain the word) in SUBTLEX-pt-BR;
 - Log of Frequency in SubIMDb PT: subtitles of family, comedy and children movies and series;
 - Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
 - Log of Frequency in the Spoken Language part of Corpus Brasileiro;
 - Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
 - Word Length;
 - Lexical databases from 6 school dictionaries for specific grade-levels;
 - Word's raw embedding values of Skip-Gram ($d = 300, 600$ and $1,000$);
 - 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]	2,397	3,789	concreteness, imageability, subjective frequency	European	1-7
[3]	885	1,749	AoA	European	1-9
[4]	719	909	concreteness	Brazilian	1-7
[5]	110	854	AoA	European	1-7
[6]	103	240	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	0.59	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	0.40	0.75	0.63	0.75	0.75	0.75
Lexical + Skip-gram	0.64	0.82	0.82	0.64	0.79	0.79	0.47	0.77	0.79	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	MSE	r	MSE	r
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	0.54

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

Flesch	Honeré	Concreteness	Familiarity	AoA	Dale	Gunning	Subjective	Psycholin	MATTR	Brunet
					Chall	Fox	Frequency	guistics		
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:
 - we have shown we can infer psycholinguistic properties for BP using word embeddings;
 - 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;
 - 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] Carneiro, 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 concreteness 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 concreteness 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:

- 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:** Estimate of the number of times a word is encountered by individuals in its written or spoken form.
- 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:

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

Model Settings

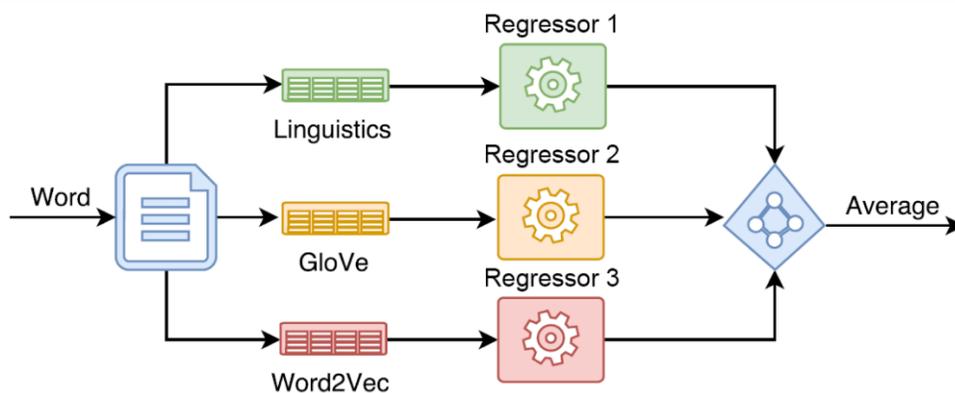
Multiview feature groups:

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

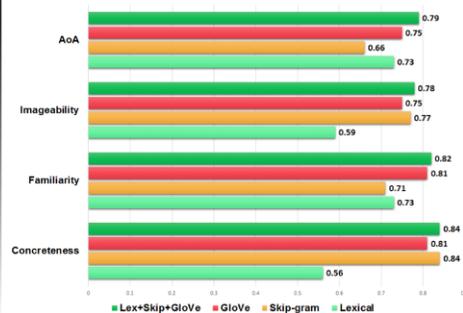
Feature set:

- Word Length
- Log of Frequency in SUBTLEX-PT
- Log of Frequency in SubIMDb-PT
- Log of number of subtitles that contain the word in SUBTLEX-PT
- Log of Frequency in the Spoken Language part of Corpus Brasileiro
- Log of Frequency in the Written Language part of Corpus Brasileiro
- Log of Frequency in a corpus of Mixed Text Genres
- Lexical databases from 6 school dictionaries
- Word's raw embedding values of Skip-Gram
- 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

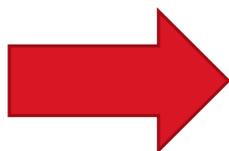
- Most of these properties are costly and time-consuming to be manually gathered;
- English language: MRC Psycholinguistic database, with 27 subjective properties for 150,837 words;
- Portuguese: only datasets of limited size [2, 3, 4, 5];
- Previous approaches to automatically infer the properties: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

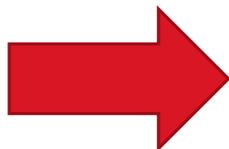
1. Manually produced properties for Portuguese are **very scarce**
2. Previous approaches use **expensive, unavailable resources**

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

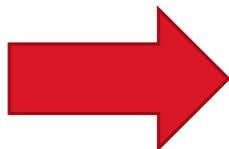
- Focus of this study: subjective psycholinguistic properties; depend on the experiences individuals had using the words:
 1. **word imageability** the ease and speed with which a word evokes a mental image;
 2. **concreteness** the degree to which words refer to objects, people, places, or things that 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:

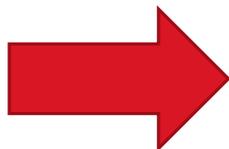
1. Manually produced properties for Portuguese are **very scarce**
2. Previous approaches use **expensive, unavailable resources**

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 $1,000$).

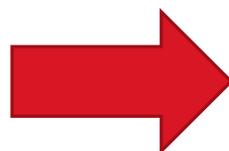
Embeddings models trained over a corpus of 1.4 billion tokens composed by mixed text genres (<http://www.nilc.icmc.usp.br/embeddings>)

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:

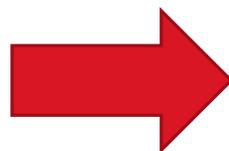
1. Word Length
2. Log of Frequency in SUBTLEX-PT
3. Log of Frequency in SubIMDb-PT
4. Log of number of subtitles that contain the word in SUBTLEX-PT
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro
6. Log of Frequency in the Written Language part of Corpus Brasileiro
7. Log of Frequency 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:

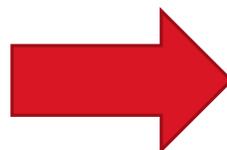
1. Word Length
2. Log of Frequency in SUBTLEX-PT
3. Log of Frequency in SubIMDb-PT
4. Log of number of subtitles that contain the word in SUBTLEX-PT
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro
6. Log of Frequency in the Written Language part of Corpus Brasileiro
7. Log of Frequency 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

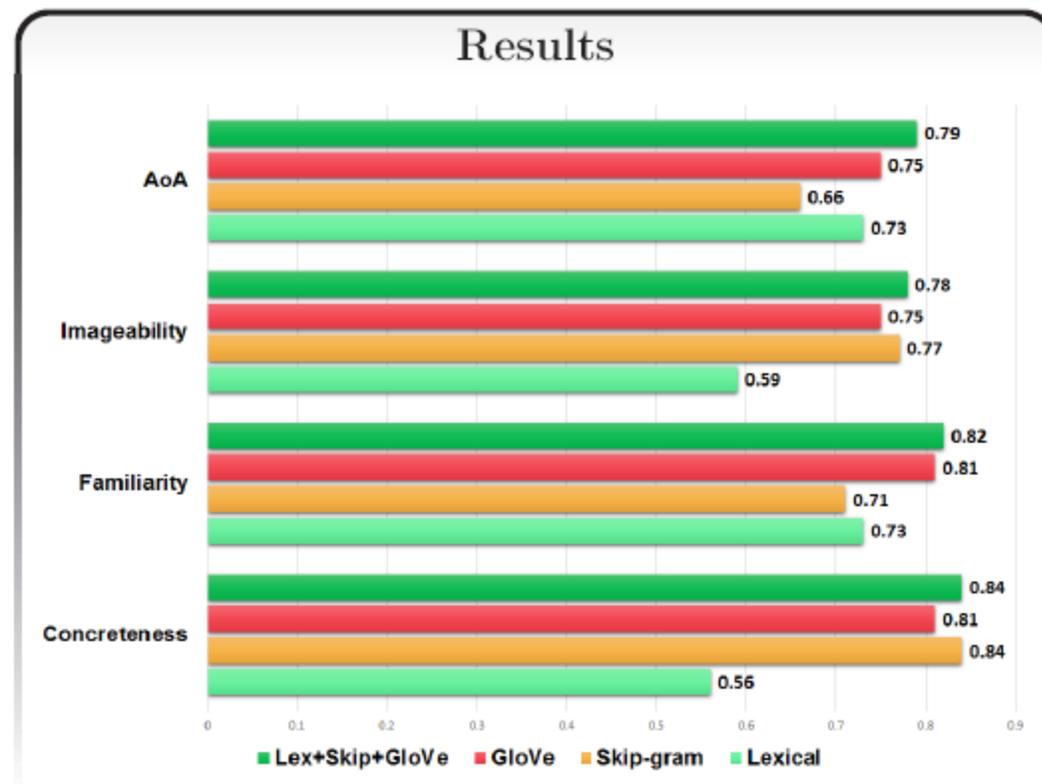
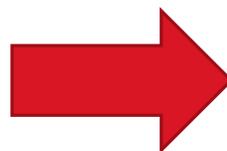
Table 2. MSE and Pearson and Spearman correlation scores of the regression models.

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.

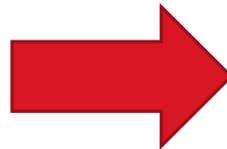


Conclusions and Future Work

- 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

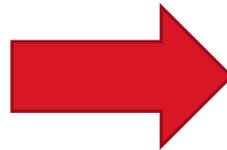
Download

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

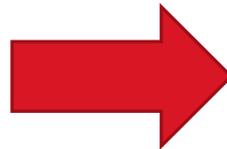
Download

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

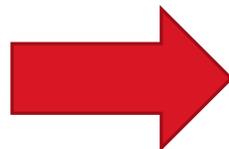
Download

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



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;
 2. which size a psycholinguistic database should have to be used in regression models? Does the merged database from different sources yield better correlation and lower MSE scores? 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 SubMD2-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 BPT);
 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.nllc.icmc.usp.br/embeddings>)

Adaptation of Databases with Norms for Portuguese

Study/Participants	Words	Property	Portuguese Variant	Scale
[2]	2,357	1,749	concreteness, imageability, subjective frequency	European
[4]	885	1,749	AoA	European
[3]	719	909	concreteness	Brazilian
[5]	110	934	AoA	European
[10]	348	348	imageability, concreteness	European

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 (1000)		Subjective Frequency (3735)		Imageability (3735)		AoA Merging (2366)	
	MSE	r	MSE	r	MSE	r	MSE	r
Lexical	1.24	0.54	0.56	0.72	0.73	0.74	0.58	0.87
Skip-gram	0.82	0.84	0.58	0.70	0.71	0.46	0.77	0.81
GloVe	0.62	0.80	0.41	0.81	0.81	0.49	0.75	0.83
Lexical + Skip-gram	0.94	0.82	0.82	0.44	0.79	0.79	0.47	0.77
Lexical + GloVe	0.70	0.80	0.81	0.81	0.81	0.30	0.75	0.76
Skip-gram + GloVe	0.49	0.85	0.83	0.41	0.80	0.82	0.79	0.87
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	0.88	0.82	0.43	0.78	0.78

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

Regressors	AoA (105)		AoA (111)		AoA Merge (2366)	
	MSE	r	MSE	r	MSE	r
Lexical	0.91	0.67	0.66	1.04	0.76	0.79
Skip-gram	1.30	0.50	0.58	1.38	0.68	0.65
GloVe	1.18	0.62	0.61	0.91	0.75	0.63
Lexical + GloVe	0.80	0.77	0.71	0.79	0.83	0.80

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

Fleisch	Honold	Concreteness	FamilyAoA	Adult	Dale	Gaming	Subjective	Psycholinguistics	MATRIK	Honold
0.26	0.28	0.27	0.23	0.25	0.30	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://nllc.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. Flensburg and Lucia Spacia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.
- [2] Soares, A.P., Costa, A.S., J. M. Coimbra, M.H.M. The minho word pool. Norms for imageability, concreteness, and subjective frequency for 3,000 portuguese words. Behavior Research Methods (2016)
- [3] Camargo, M.E., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 414-420 (2010)
- [4] Janczura, G., Castilho, G., Rocha, N., van Erven, T., Huang, T. Normas de concreteness 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 imagemica e concreteness para substantivos comuns. Laboratorio de Psicologia 3, 65-75 (2006)

Final Lesson

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:
- word imageability**: the ease and speed with which a word evokes a mental image;
- concreteness**: the degree to which words refer to objects, people, places, or things that can be experienced by the senses;
- subjective frequency**: the estimation of the number of times a word is encountered by individuals in its written or spoken form;
- 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:

- Is it possible to achieve high Pearson and Spearman correlations and low MSE values using a regression method with only word embeddings to infer the psycholinguistic properties?
- 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? 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):

- Log of Frequency in SUBTLEX-pt-BR;
- Log of Contextual diversity (number of sentences that contain the word) in SUBTLEX-pt-BR;
- Log of Frequency in SubMDb-PT: subtitles of family, comedy, and children movies and series; Contemporary (BP);
- Log of Frequency in the Written Language part of Corpus Brasileiro (1 billion words of Contemporary BP);
- Log of Frequency in the Spoken Language part of Corpus Brasileiro;
- Log of Frequency in a corpus of 1.4 billion tokens of Mixed Text Genres in BP;
- Word Length;
- Lexical databases from 6 school dictionaries for specific grade-levels;
- Word's raw embedding values of Skip-Gram ($d = 300, 600$ and $1,000$);
- 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]	2,257	1,729	concreteness, imageability, subjective frequency	European
[4]	902	1,743	AoA	European
[4]	729	909	AoA	Brazilian
[4]	110	614	AoA	European
[4]	102	347	imageability, concreteness	European
			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 (6265)		Subjective Frequency (3725)		Imageability (3725)		AoA Merging (2366)	
	MSE	r	MSE	r	MSE	r	MSE	r
Lexical	1.24	0.54	0.55	0.72	0.73	0.74	0.58	0.38
Skip-gram	0.92	0.64	0.50	0.70	0.71	0.46	0.77	0.81
GloVe	0.62	0.80	0.81	0.40	0.81	0.49	0.78	0.75
Lexical + Skip-gram	0.94	0.82	0.82	0.44	0.79	0.79	0.47	0.77
Lexical + GloVe	0.70	0.80	0.81	0.39	0.81	0.36	0.75	0.76
Skip-gram + GloVe	0.49	0.85	0.85	0.41	0.80	0.42	0.79	0.79
Lexical + Skip-gram + GloVe	0.55	0.85	0.84	0.38	0.82	0.82	0.43	0.79

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

Regressors	AoA (705)		AoA (111)		AoA Merge (2366)	
	MSE	r	MSE	r	MSE	r
Lexical	0.91	0.67	0.66	1.04	0.76	0.75
Skip-gram	1.30	0.56	0.54	1.38	0.68	0.65
GloVe	1.18	0.62	0.61	0.91	0.70	0.63
Lexical + GloVe	0.80	0.72	0.71	0.79	0.83	0.80

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

Fleisch	Hornet	Concreteness	Familiarity	AoA	Dale	Gunning	Subjective	Psychol	MATRH	Hande
0.26	0.28	0.27	0.23	0.26	0.30	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:

- we have shown we can infer psycholinguistic properties for BP using word embeddings;
- 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;
- 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

- Gustavo H. Fleetwood and Lucia Spacia. Inferring psycholinguistic properties of words. Proceedings of NAACL-HLT, pp. 435-440, 2016.
- Soares, A.P., Costa, A.S., J. M. Comissaris, M.H.M.: The minho word pool. Norms for imageability, concreteness, and subjective frequency for 3,000 portuguese words. Behavior Research Methods (2016)
- Cameron, M.E., Vicente, S.G. Age-of-acquisition norms for a set of 1,749 Portuguese words. Behavior research methods 42(2), 474-483 (2010)
- Janczura, C., Castillo, G., Rocha, N., van Erven, T., Huang, T. Normas de concreteness para 909 palavras da lingua portuguesa. Psicologia: Teoria e Pesquisa pp. 195-204 (2007)
- Margues, J.F., Fonseca, F.L., Marais, 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)
- Marques, J.F. Normas de imageability e concreteness 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:

- 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**: Estimate of the number of times a word is encountered by individuals in its written or spoken form.
- Age of Acquisition**: The estimate of the age at which a word was learned.

Various applications:

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

Challenges:

- Manually produced properties for Portuguese are very scarce
- Previous approaches use expensive, unavailable resources

Model Settings

Multiview feature groups:

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

Feature set:

- Word Length
- Log of Frequency in SUBTLEX-PT
- Log of Frequency in SubMDb-PT
- Log of number of subtitles that contain the word in SUBTLEX-PT
- Log of Frequency in the Spoken Language part of Corpus Brasileiro
- Log of Frequency in the Written Language part of Corpus Brasileiro
- Log of Frequency in a corpus of Mixed Text Genres
- Lexical databases from 6 school dictionaries
- Word's raw embedding values of Skip-Gram
- Word's raw embedding values of GloVe

Proposed Model

Results

Property	Lex+Skip+GloVe	GloVe	Skip-gram	Lexical
AoA	0.79	0.66	0.79	0.75
Imageability	0.78	0.75	0.77	0.59
Familiarity	0.82	0.81	0.71	0.73
Concreteness	0.84	0.84	0.56	0.84

More Findings

Alpha inter-annotator agreement scores:

Property	Agreement Score
Imageability	0.921
Concreteness	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

It's worth it!

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 represented 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 1000 words;
- Previous approaches:

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

Features for Regressors

- 10 features grouped in: (i) lexical (1-8); (ii) Word2Vec Skip-Gram (9); and (iii) GloVe embeddings (10);
1. Log of Frequency in SUBTLEX-PT: sublexes that contain the word in SUBTLEX-PT
2. Log of Contextual diversity (number of subtitles that contain the word in SUBTLEX-PT)
3. Log of Frequency in SubMOS-PT: subtitles of mixed text genres
4. Log of Frequency in the Written Language part of Corpus Brasileiro
5. Log of Frequency in the Spoken Language part of Corpus Brasileiro
6. Log of Frequency in a corpus of Mixed Text Genres
7. Word Length;
8. Lexical databases from 6 school dictionaries
9. Word's raw embedding values of GloVe
10. Word's raw embedding values of Word2Vec

Model Settings

Multiview feature groups:
Regressor 1 | Regressor 2 | Regressor 3:

Model

Regressor 1 (GloVe) → Regressor 2 (Word2Vec) → Regressor 3 (Skip-Gram) → Average

Results

Property	Lex+Skip+GloVe	GloVe	Skip-gram	Lexical
AoA	0.79	0.66	0.78	0.77
Imageability	0.79	0.59	0.82	0.81
Familiarity	0.71	0.73	0.81	0.84
Concreteness	0.84	0.56	0.84	0.84

More Findings

Alpha inter-annotator agreement scores:

Property	Agreement Score
Imageability	0.921
Concreteness	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>

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:

Keeping 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 L^AT_EX

3 Controlling Appearance

- Making Lists
- Fonts, Symbols, quotations and footnotes

4 Adding Structure

- Sections
- Tables, Figures and Equations

5 BIB_TE_X

Slides: The Problems

3. Some slides are _____

Slides: The Problems

3. Some slides are **unnecessary**

The Outline

Motivation	Basics	Controlling Appearance	Adding Structure	BIBTeX	References
	○ ○○○ ○	○○○ ○○○○○	○○ ○○○○○○○○○○		
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



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

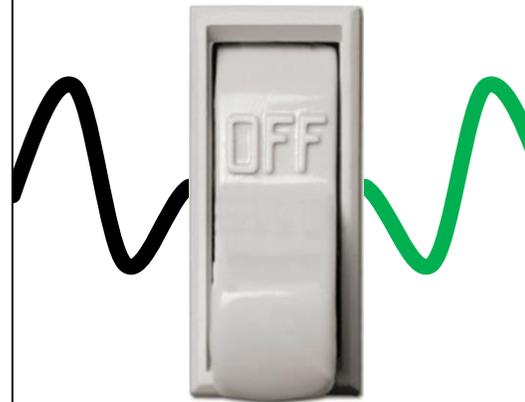


Slides: The Problems

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

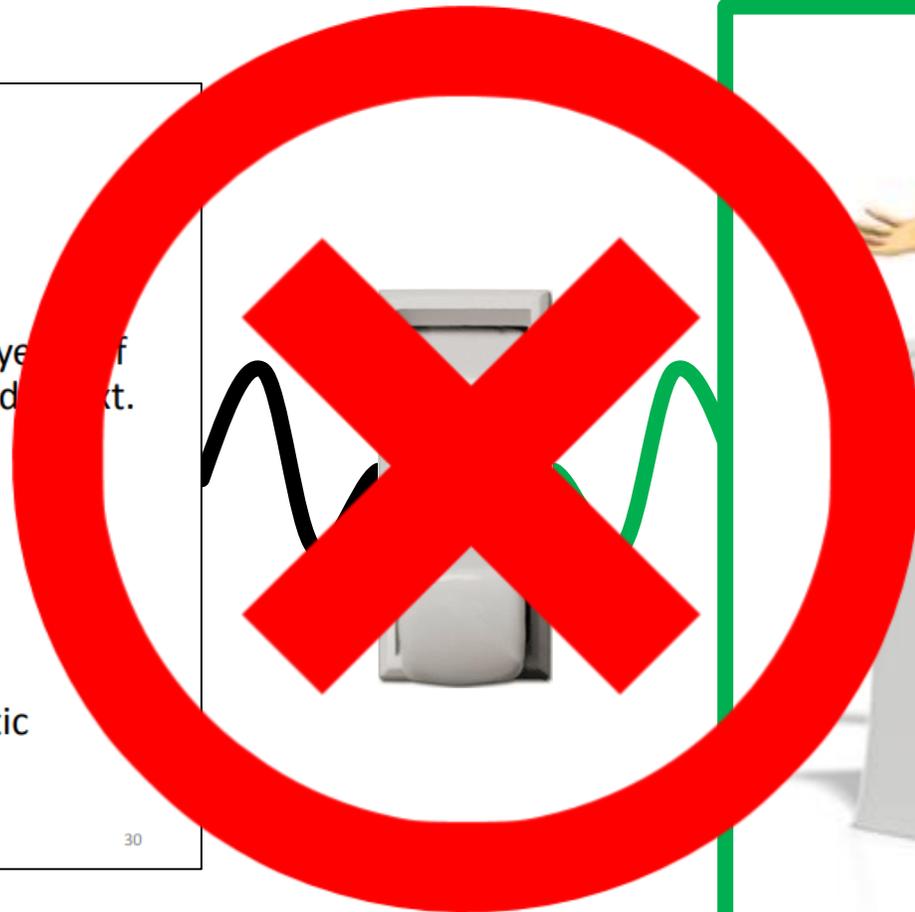
30



Slides: The Problems

Experiment Setup

- Metrics:
 - Flesch-Kincaid grade level: the number of years of education generally required to understand 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



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



Optional speech control
Push-to-Talk Buttons



Small powerful Speakers



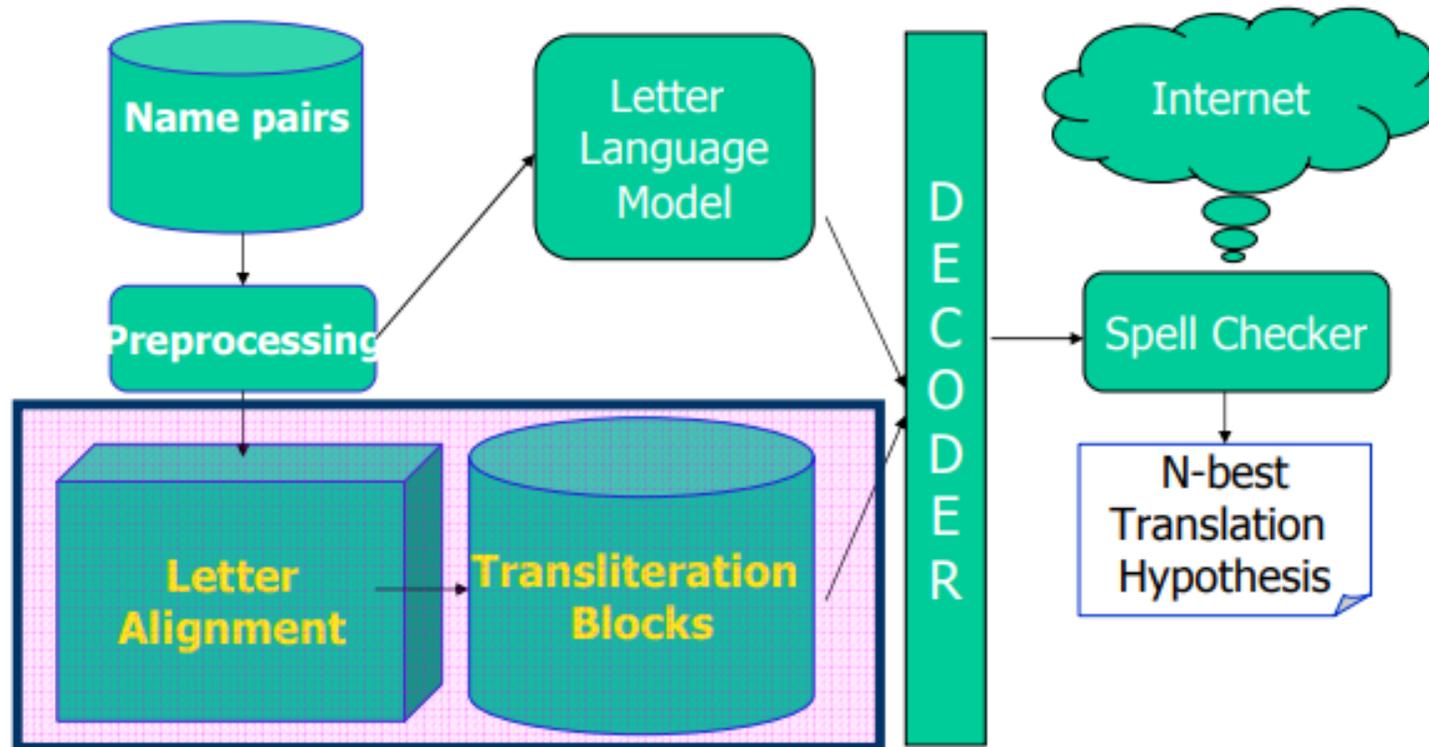
Close-talking Microphone



Laptop secured in Backpack



System Architecture



Oral Presentations: Part 1

Oral Presentations: Part 1

Slides: The Solutions

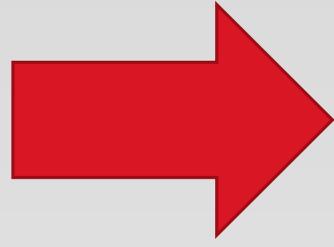
Slides: The Solutions

1. Too much **text**

Slides: The Solutions

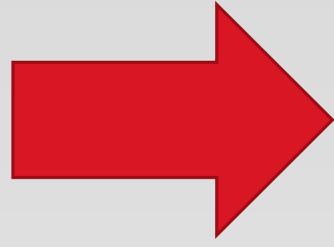
- Summarize it
- Make it visual

List



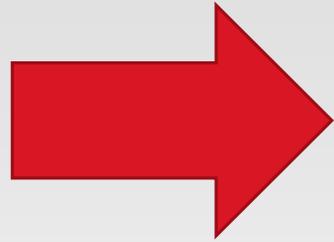
Table

List



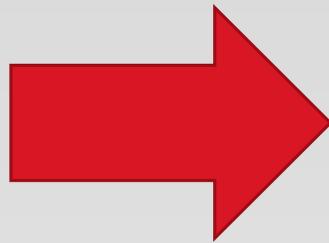
Table

Table



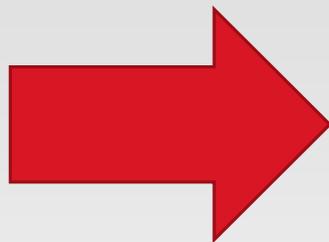
Graph

List



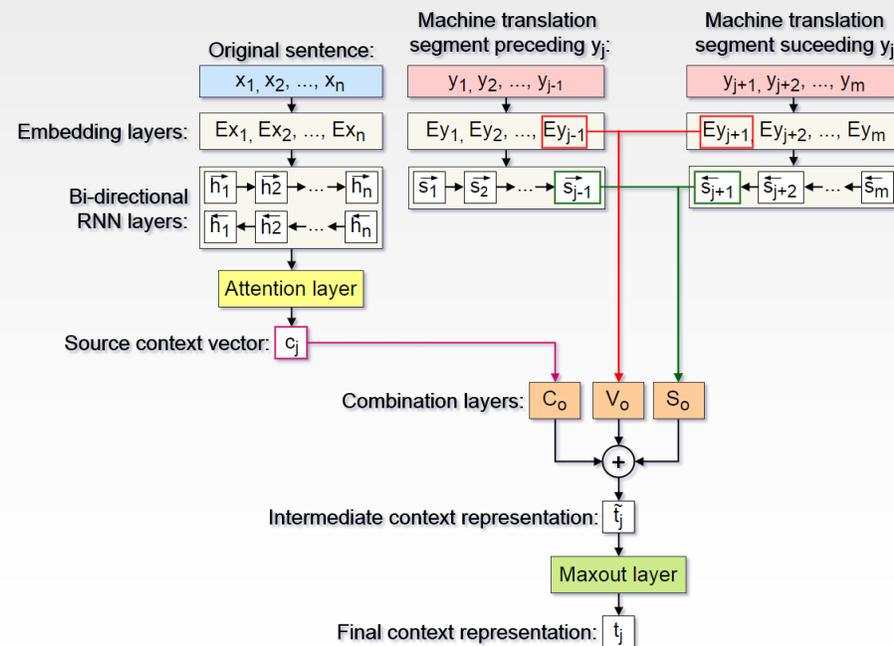
Table

Table



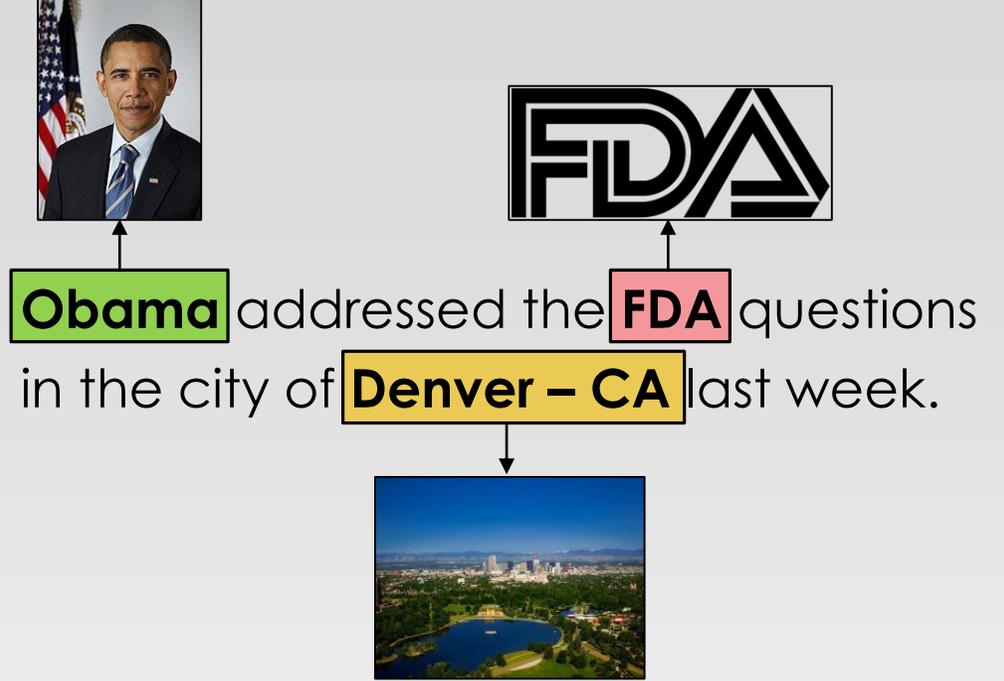
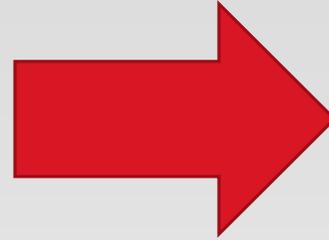
Graph

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$



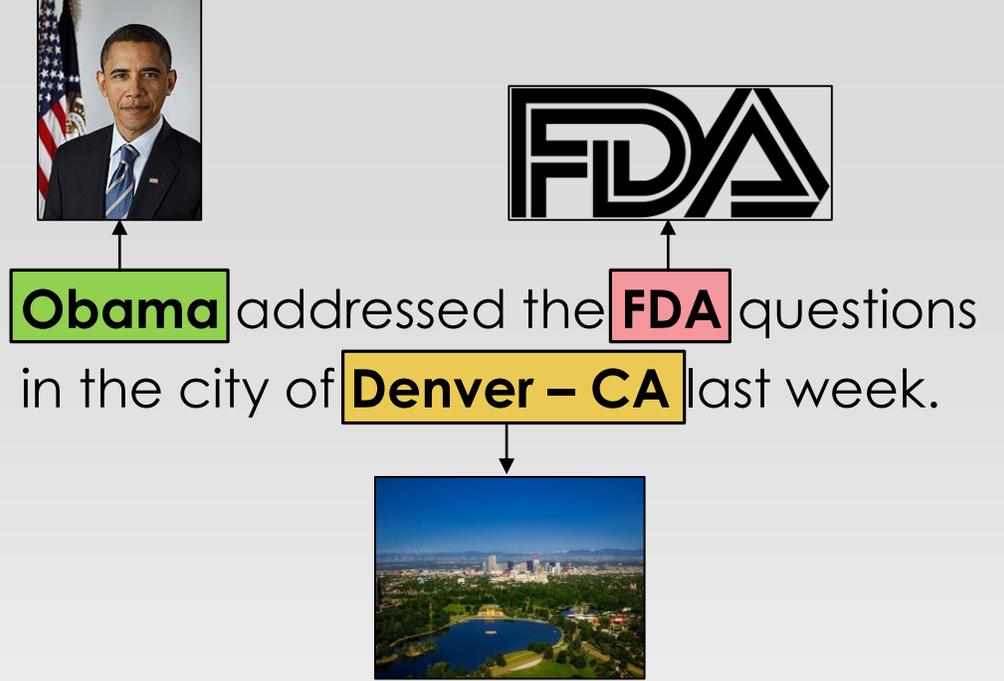
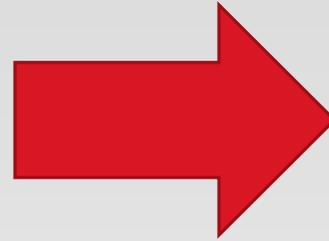
Named-entity recognition (NER)

(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.

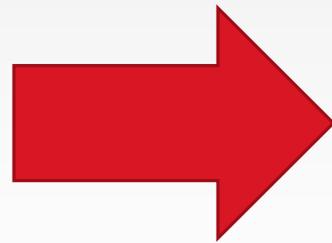


Named-entity recognition (NER)

(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.



Anita



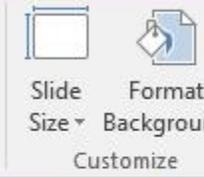
Anita

Slides: The Solutions

2. Bland styling



Variants



Enable Content Updates from Office.com...

- Browse for Themes...
- Save Current Theme...



Communicating and sharing your work effectively with colleagues, supervisors and the general public often requires the preparation of a suitable presentation tailored to that audience. These templates make it easy to create such a presentation, and the resulting set of slides is available for LaTeX Beamer – perfect for sharing before or after your lecture, seminar or talk.



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

汪曦之
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, NITTTT, Bhopal, India
1 anushrivastava1999@gmail.com; 2 marizvi@nittrapt.ac.in

May 28, 2015

Krylov Subspace Methods in Model Order Reduction

Mohammad Umar Rehman

PhD Candidate, EE Department, IIT Delhi
umar.ee.11td@gmail.com

March 8, 2016

FACULTY OF ECONOMICS AND ADMINISTRATION
Bhargh University

Presentation Title

Presentation Subtitle

Author's Name



Startup Keynote Template
By slidefusion



WATERCOLOR
Keynote

Watercolor Keynote Template
By Jumsoft



Martin Keynote Business Template
By SimpleSmart



Sella Keynote Template
By Levato



ELFA
Powerpoint Presentation Template

ELFA - Keynote Template
By Inspirasign



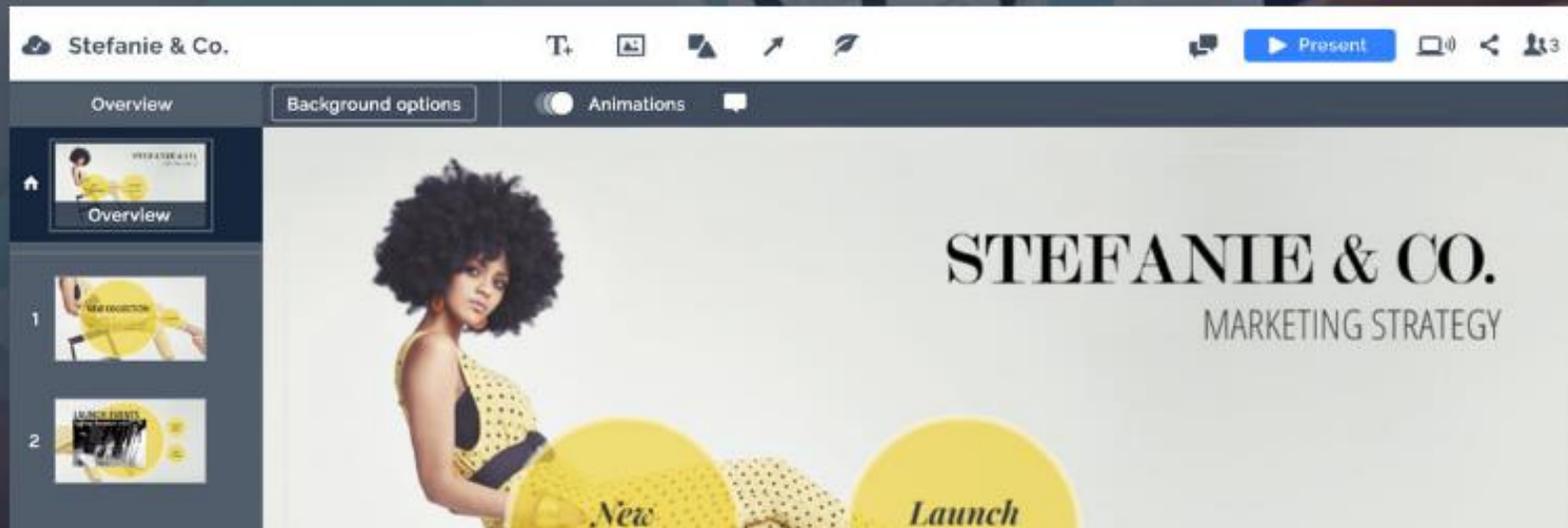
SIMPLECO: Keynote presentation ...
By TITO



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



The screenshot shows the Prezi Next interface. At the top, the title bar reads "Stefanie & Co." and includes icons for text, images, shapes, lines, and arrows, along with a "Present" button and a user count of 3. Below the title bar, there are tabs for "Overview", "Background options", and "Animations". The main content area displays a slide with a woman in a yellow polka-dot dress holding a large yellow circle. The text on the slide reads "STEFANIE & CO. MARKETING STRATEGY" and "New Launch". A sidebar on the left shows a list of slides, with the first slide selected and labeled "Overview".

▶ 1. Overview

Prezi's one-of-a-kind open canvas lets you organize and view your presentations in a way that's unique to your business.

▶ 2. Smart structures

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

Go easy on the transitions 😊



The screenshot shows a Prezi presentation titled "Stefanie & Co." with a navigation menu on the left and a main content area. The main content area features a woman with a large afro hairstyle, a yellow polka-dot dress, and two large yellow circles labeled "New" and "Launch". The text "STEFANIE & CO. MARKETING STRATEGY" is displayed in the background. The navigation menu includes "Overview", "Background options", and "Animations". The right side of the screen shows a list of sections: "1. Overview" and "2. Smart structures".

Stefanie & Co.

Overview Background options Animations

Overview

1

2

STEFANIE & CO.
MARKETING STRATEGY

New Launch

1. Overview
Prezi's one-of-a-kind open c
organize and view your pres

2. Smart structures

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

IPython

Interactive Computing

[Install](#) · [Documentation](#) · [Project](#) · [Jupyter](#) · [News](#) · [Cite](#)

The Jupyter Notebook

(Formerly known as IPython Notebook)

The IPython Notebook, now known as the Jupyter Notebook, is an interactive computing environment. In this environment, you can combine code execution, rich text, and images. For more details on the Jupyter Notebook, please see the [Jupyter](#) website.

Slides: The Solutions

3. Some slides are **unnecessary**

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:

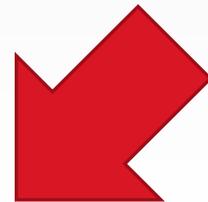
1. Introduction
2. Motivation/Challenges
3. Approach
4. Experimentation setup
5. Results
6. Output examples/comparison
7. Main conclusions
8. Page numbers (without total pages)

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



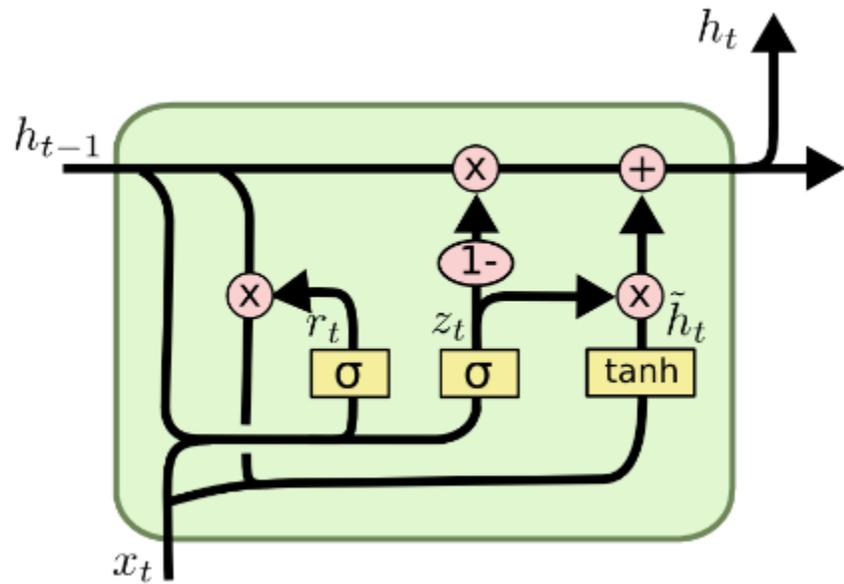
Slides: The Solutions

4. Too much **information** per slide

Slides: The Solutions

Stepification

Gated Recurrent Units



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

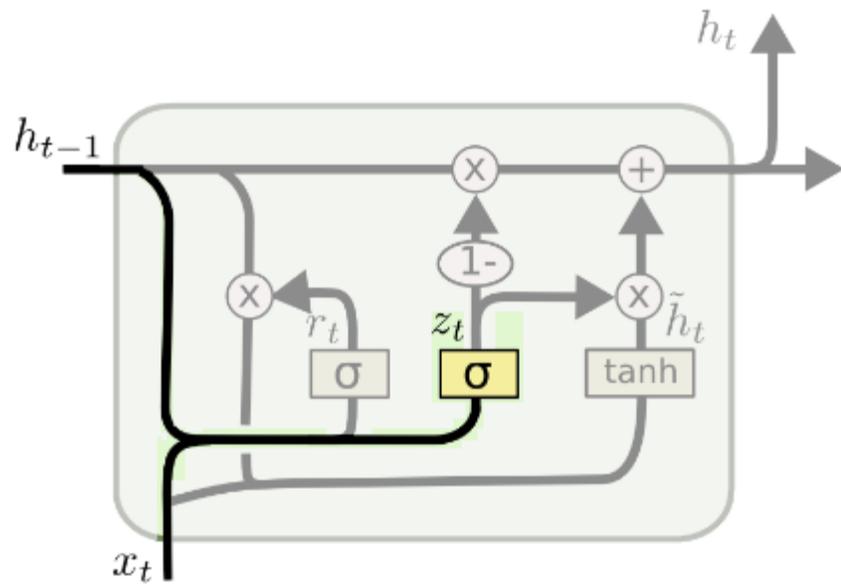
$$r_t = \sigma (W_r \cdot [h_{t-1}, x_t])$$

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

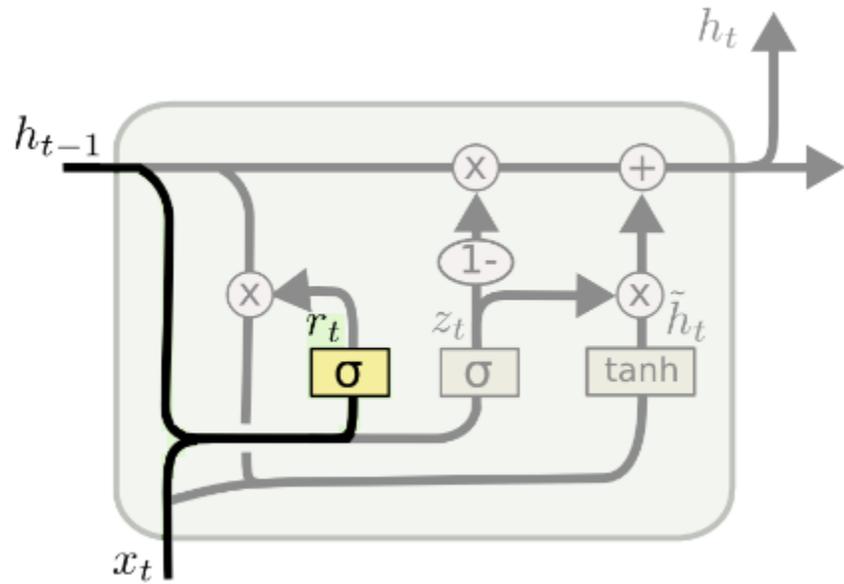
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

The Update Gate

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

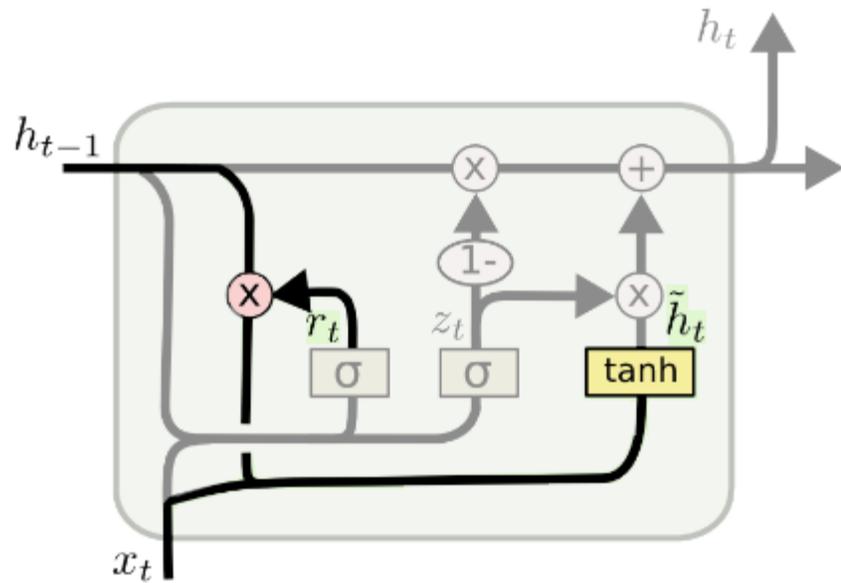


The Reset Gate



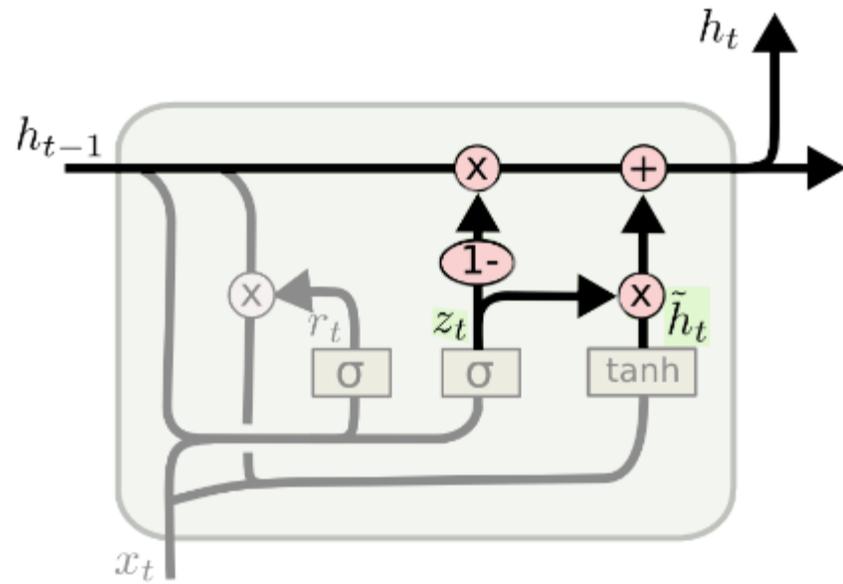
$$r_t = \sigma (W_r \cdot [h_{t-1}, x_t])$$

The Memory



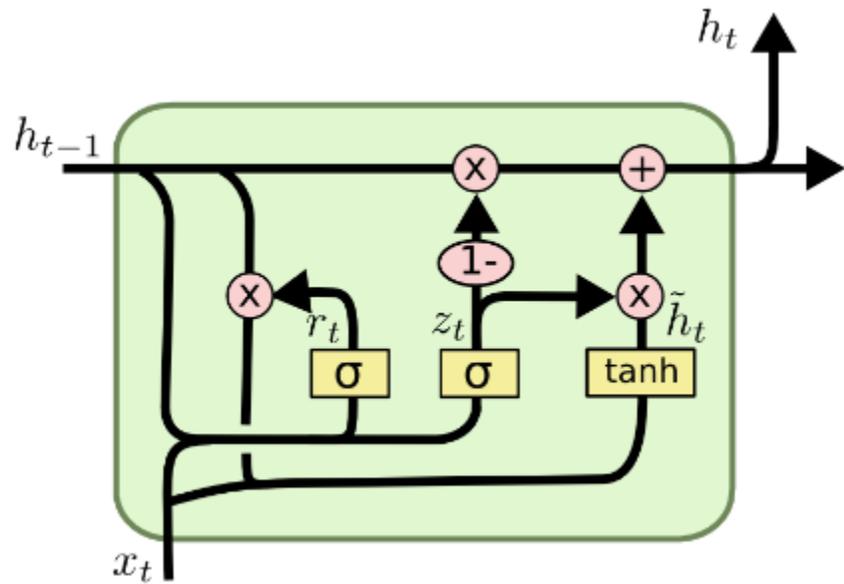
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

The Output



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

Gated Recurrent Units



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

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

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

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

Gated Recurrent Units

Output:

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

Update gate:

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

Memory:

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

Reset gate:

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

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:

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

The Memory

Output:

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

Update gate:

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

Memory:

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

The Reset Gate

Output:

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

Update gate:

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

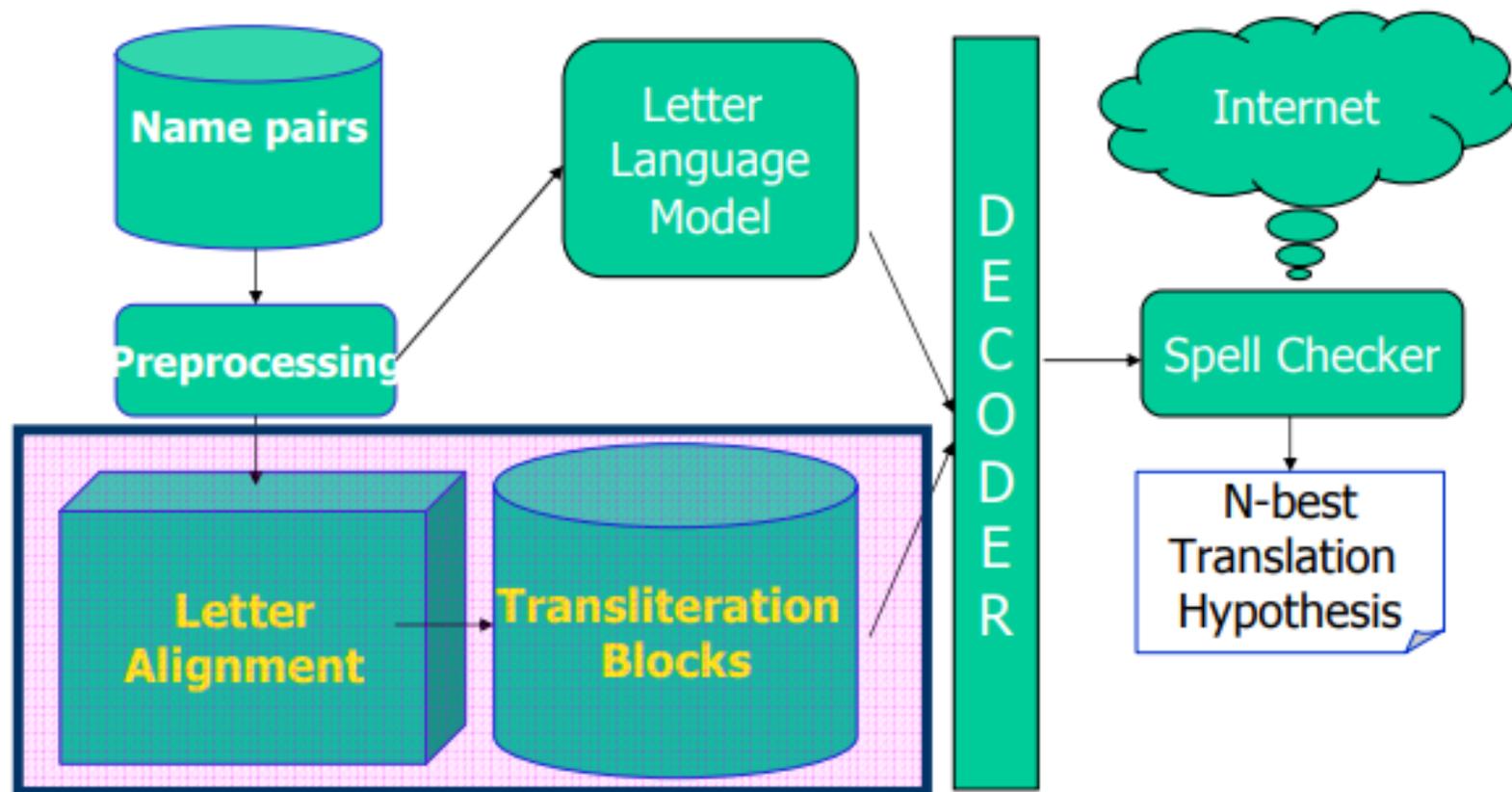
Memory:

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

Reset gate:

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

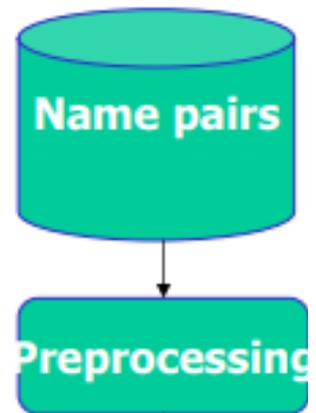
System Architecture



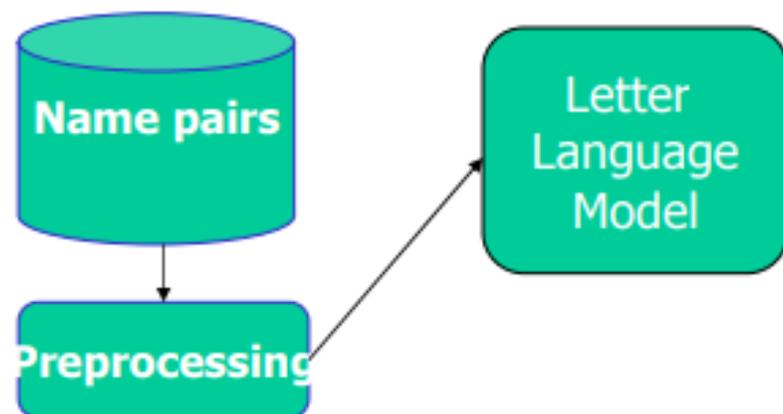
System Architecture



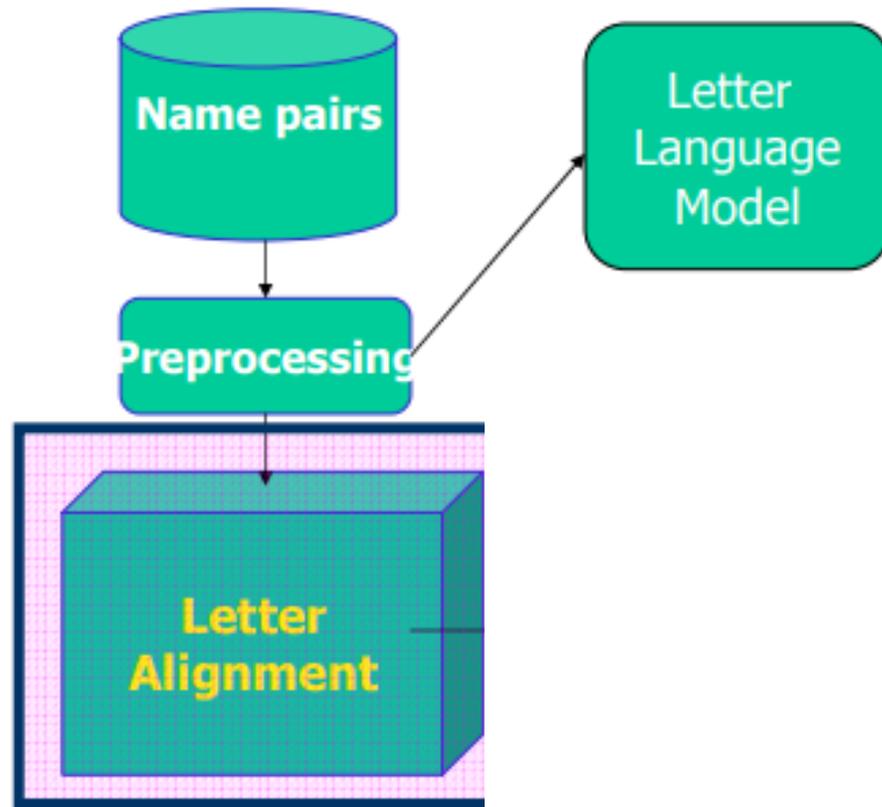
System Architecture



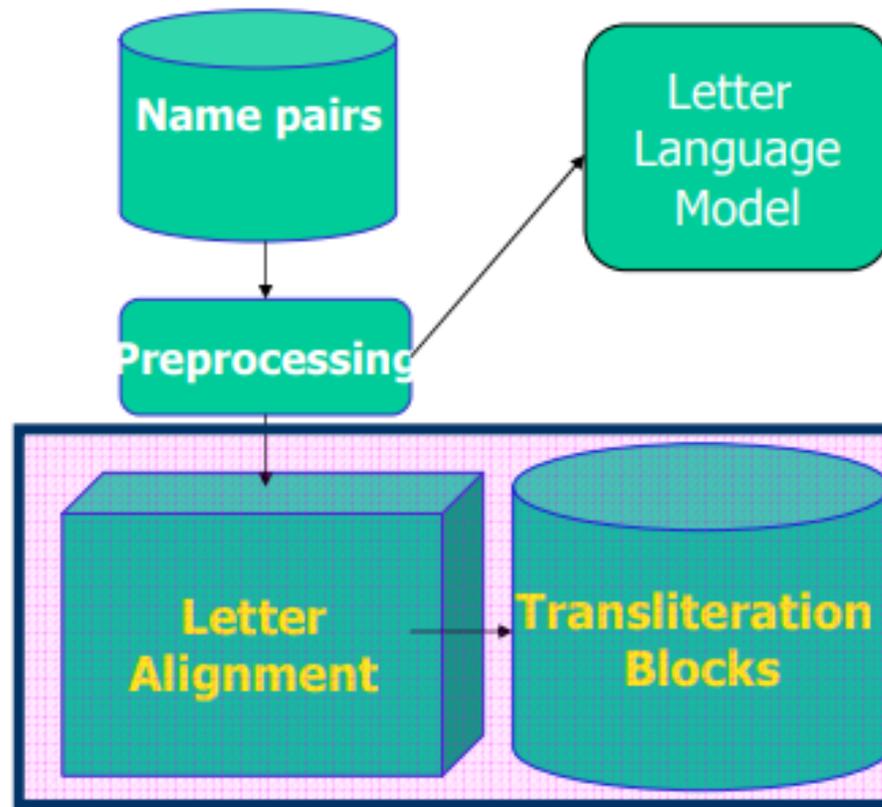
System Architecture



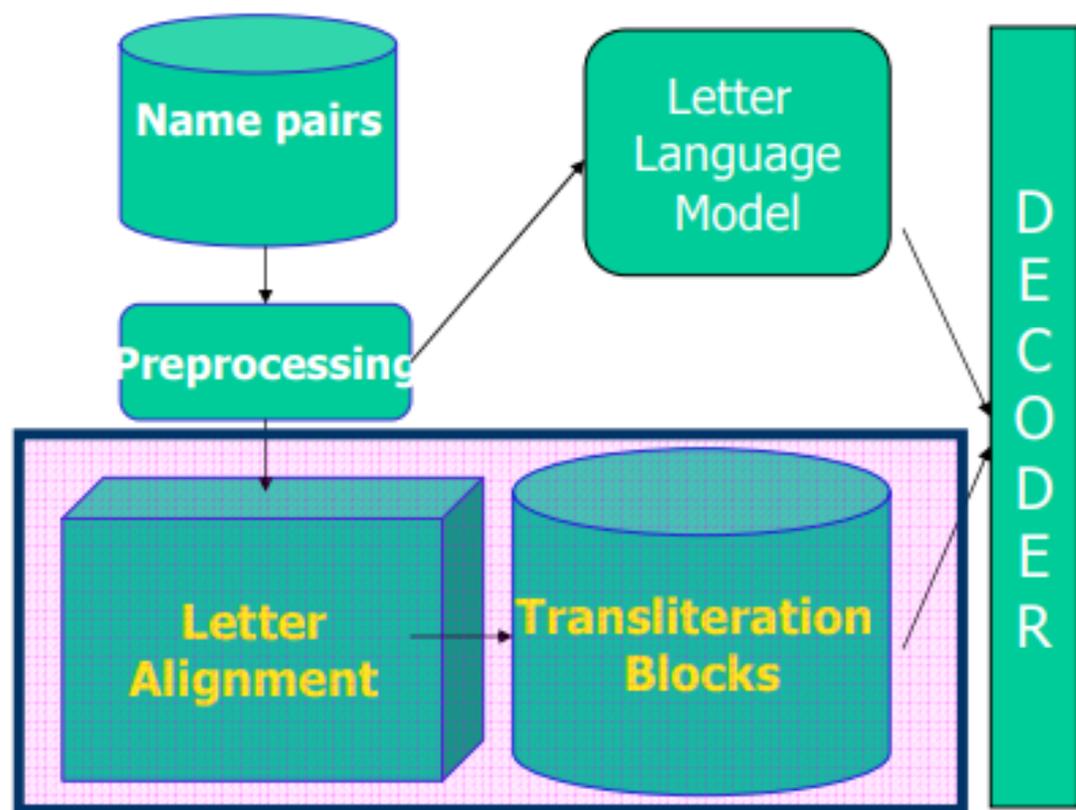
System Architecture



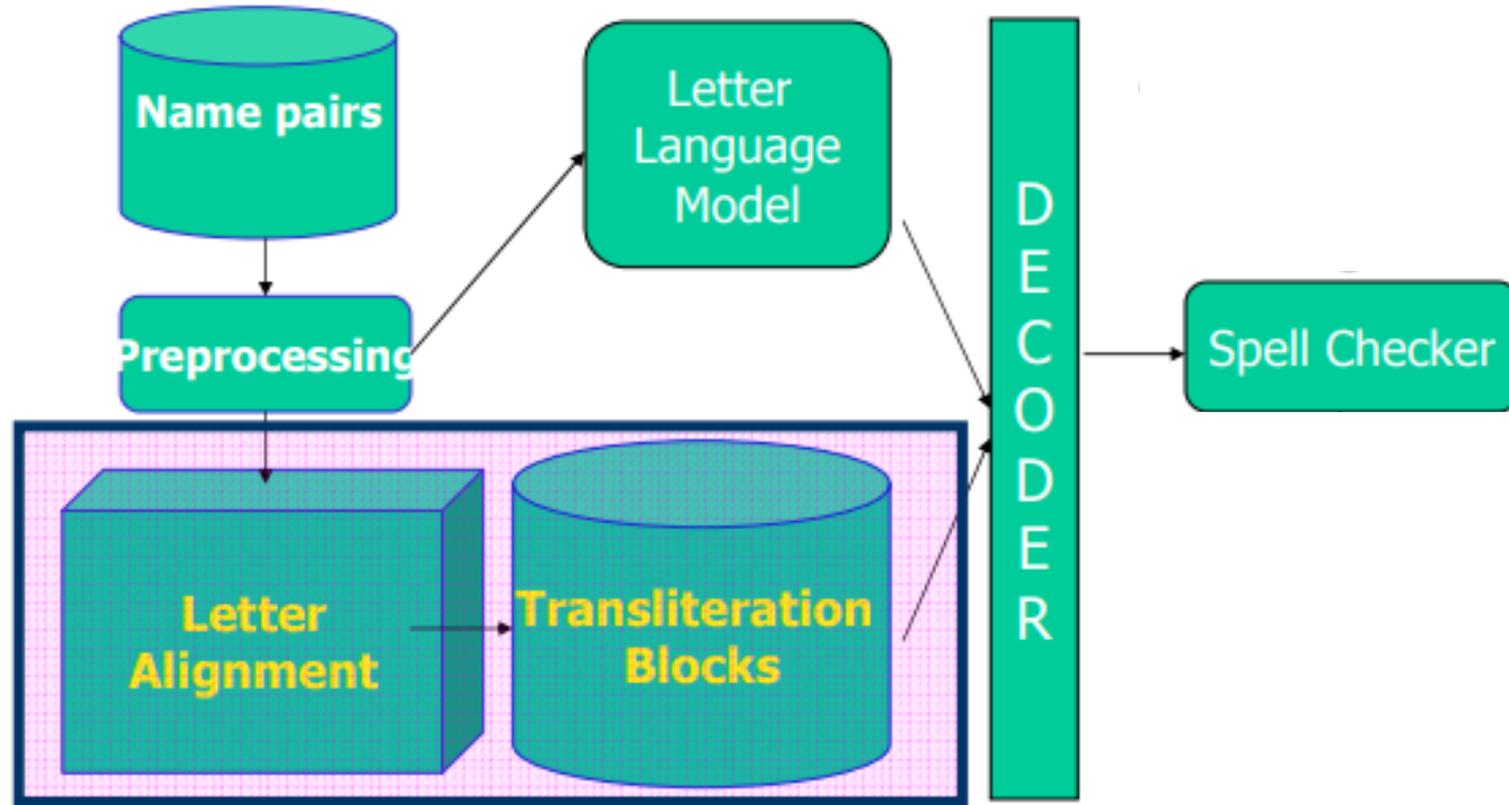
System Architecture



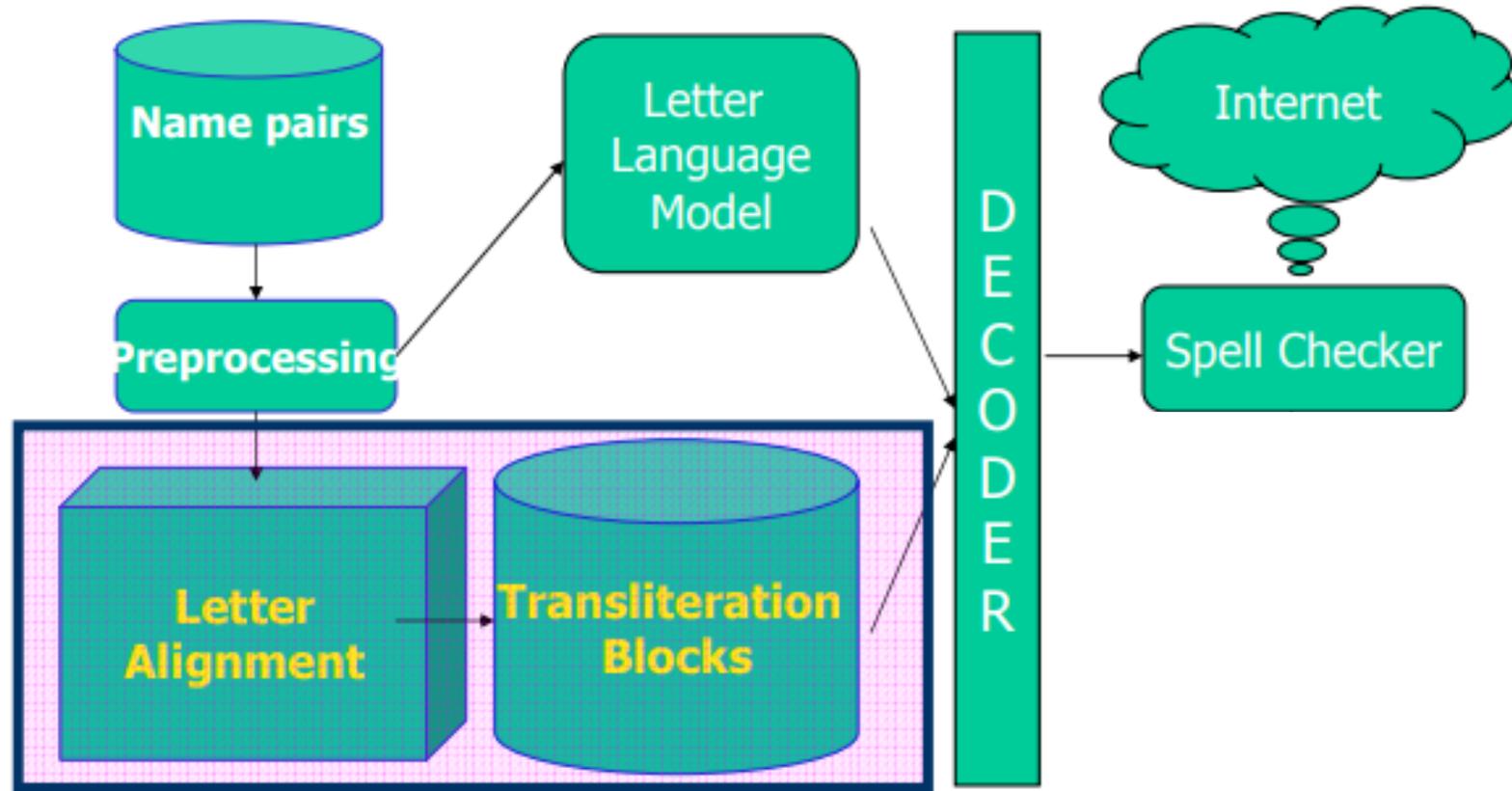
System Architecture



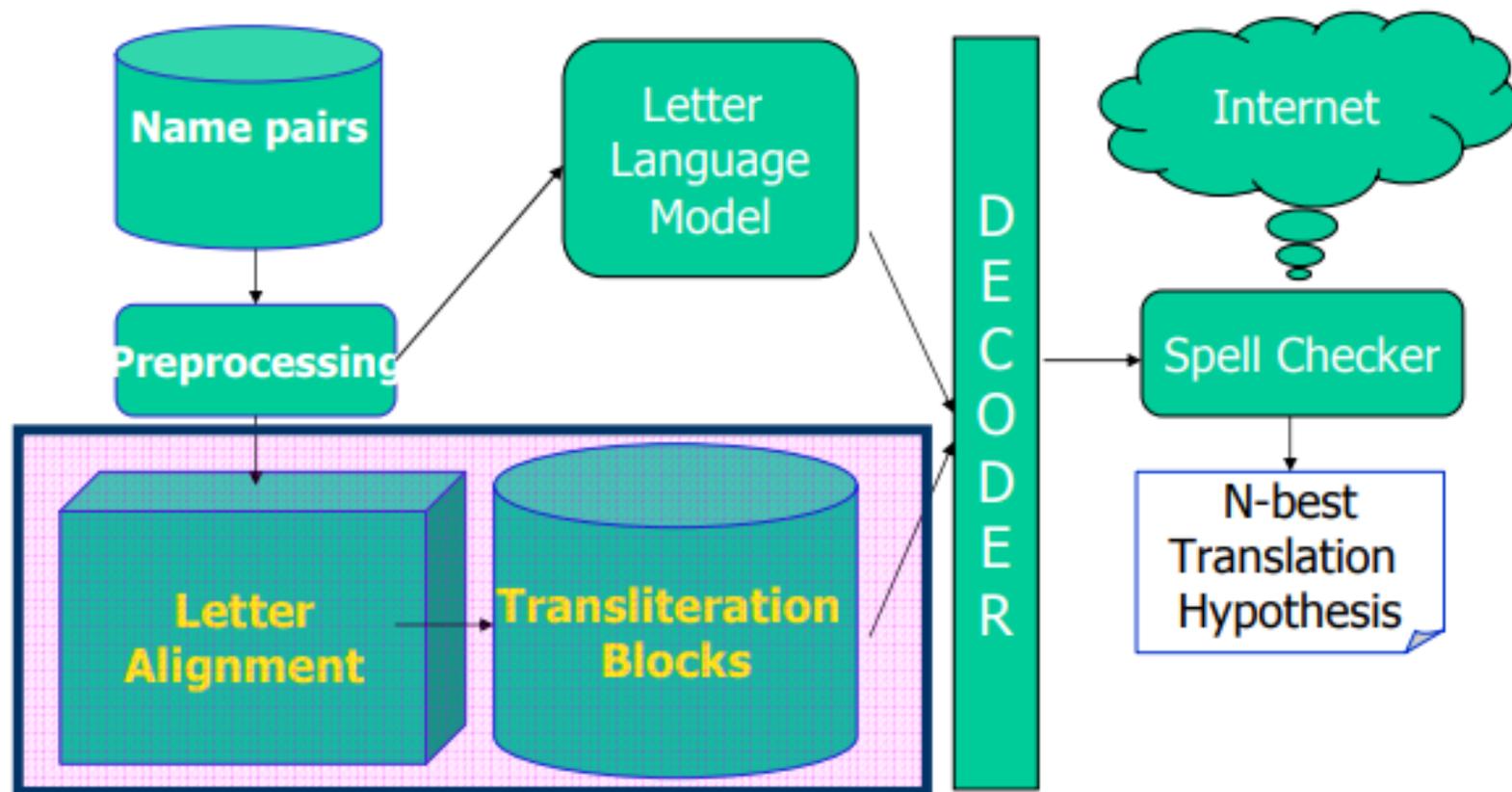
System Architecture



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

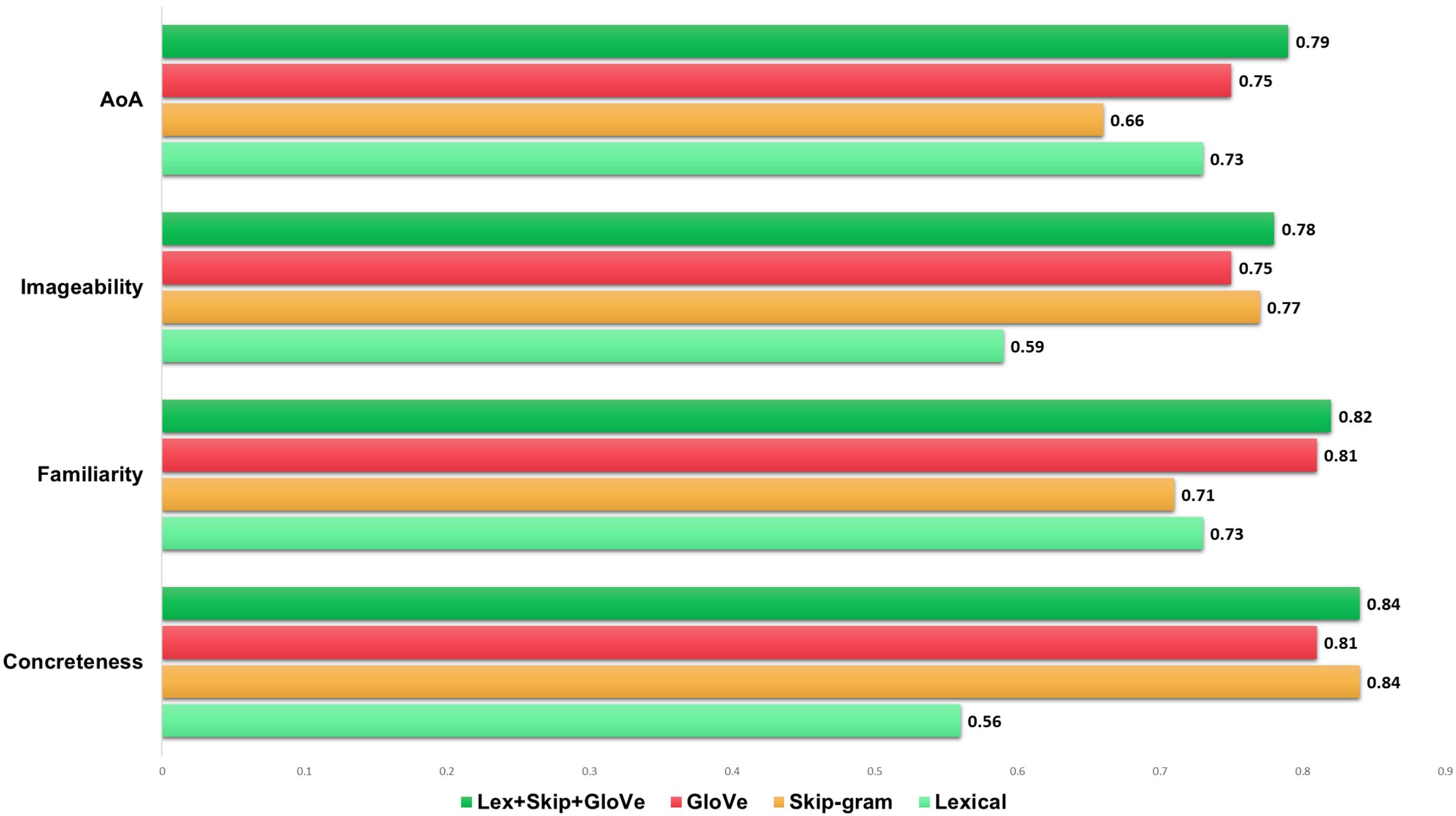
- Metrics:
 - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
 - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
 - Training set: 754 sentences
 - Unseen test set: 100 sentences
 - 70% Wikipedia

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

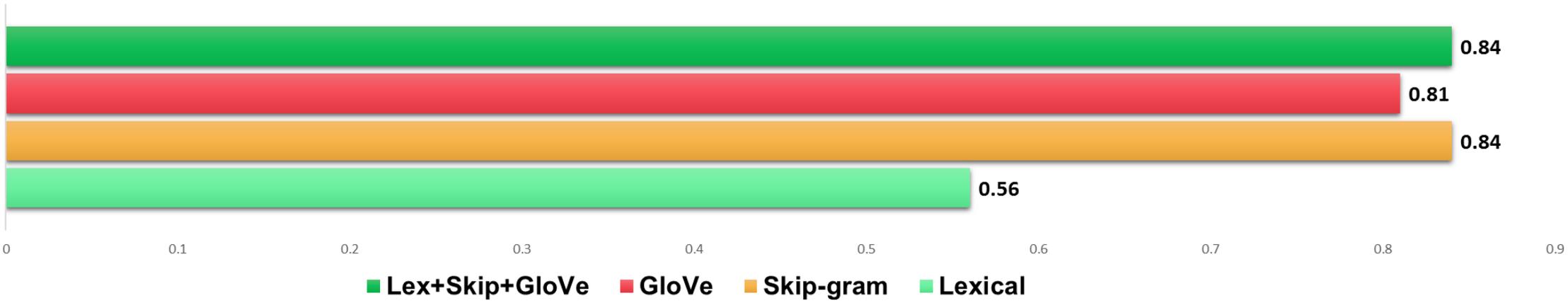
Experiment Setup

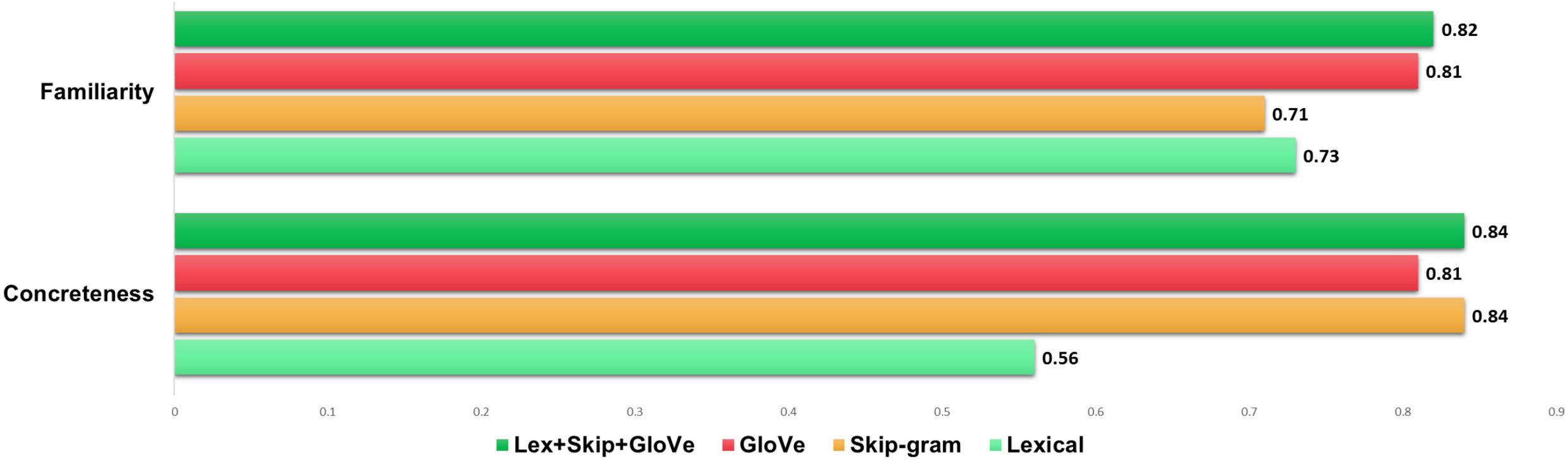
- Metrics:
 - Flesch-Kincaid grade level: the number of years of education generally required to understand a text.
 - ROUGE-n: n-gram co-occurrence between hypothesis and reference
- Data
 - Training set: 754 sentences
 - Unseen test set: 100 sentences
 - 70% Wikipedia, 25% NY Times, 5% synthetic

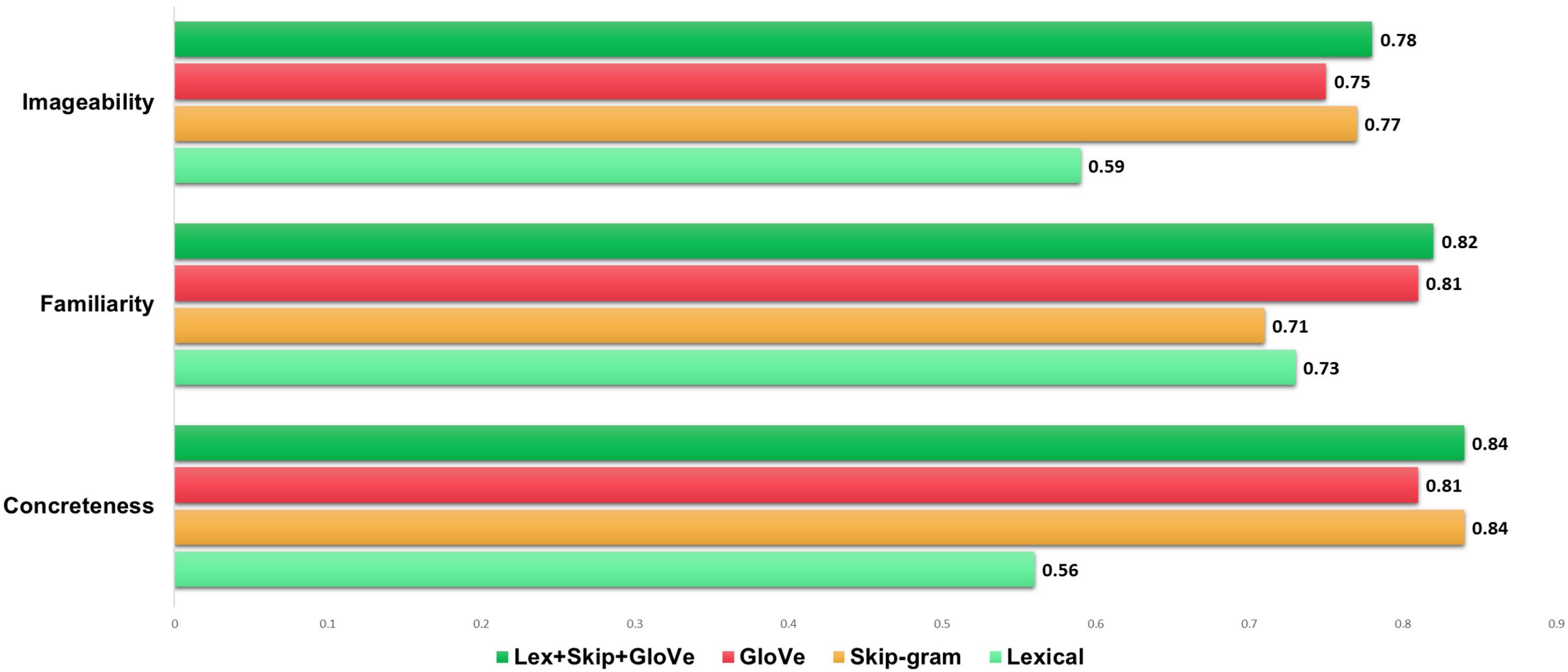


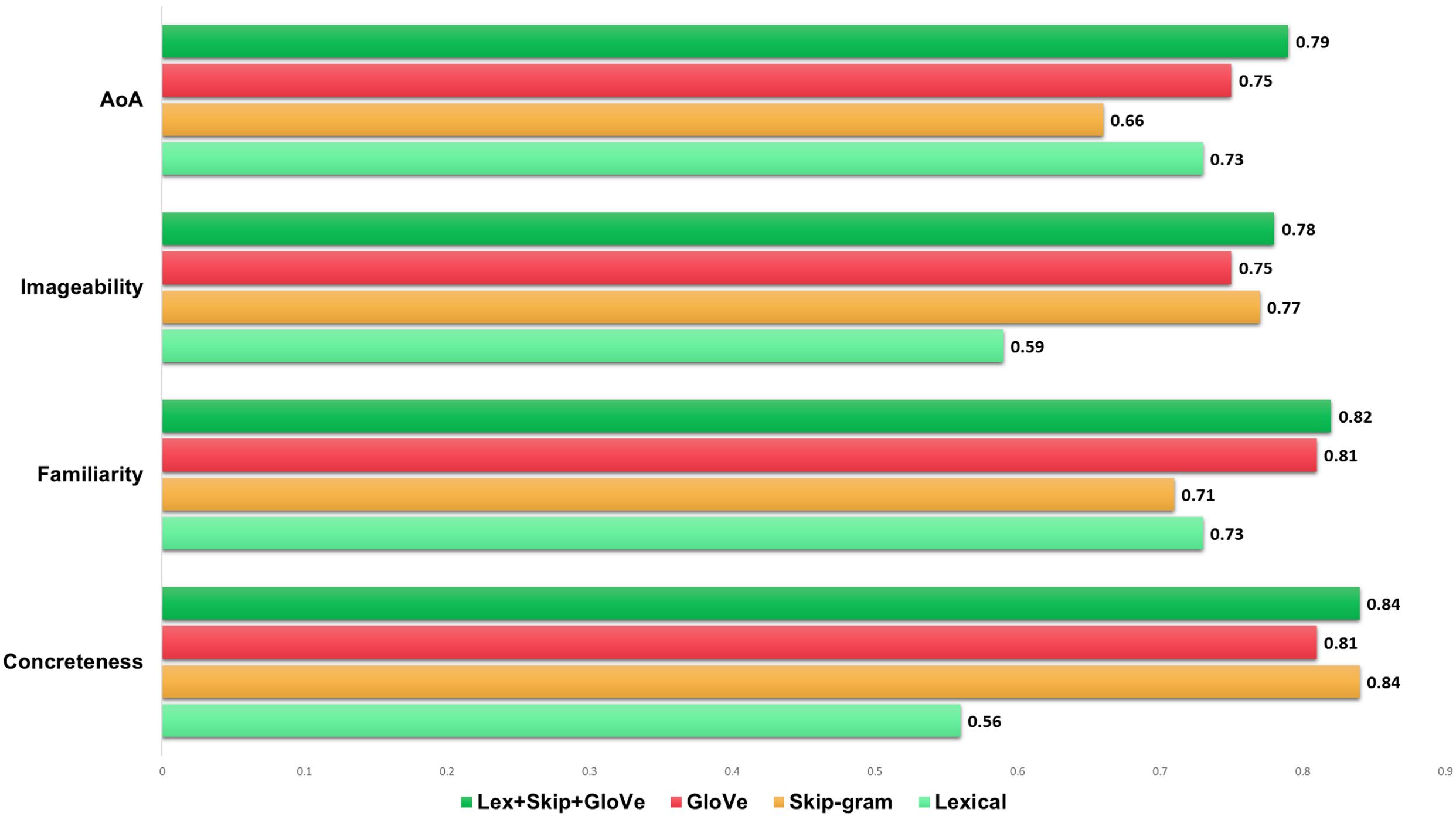
■ Lex+Skip+GloVe ■ GloVe ■ Skip-gram ■ Lexical

Concreteness









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	r	ρ
Lexical	0.91	0.67	0.66
Skip-gram			
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			
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

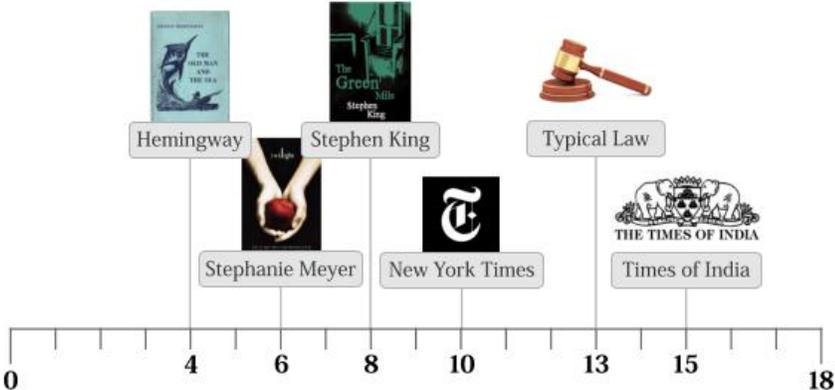
Experiment Setup

Evaluation metrics:

Experiment Setup

Evaluation metrics:

Flesch-Kincaid Grade Level

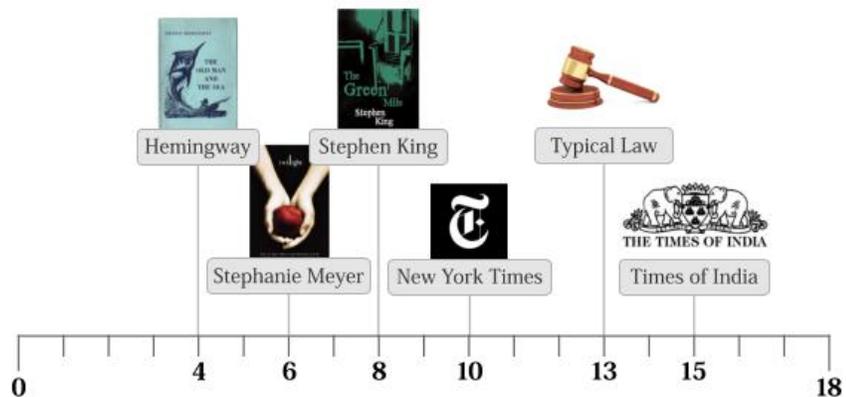


Experiment Setup

Evaluation metrics:

Flesch-Kincaid Grade Level

ROUGE-n



$$\frac{\text{overlapping } n\text{-grams}}{\text{total } n\text{-grams in reference}}$$

Experiment Setup

Data:

Experiment Setup

Data:

Training set:

754 sentences

Experiment Setup

Data:

Training set:

754 sentences

Test set:

100 sentences

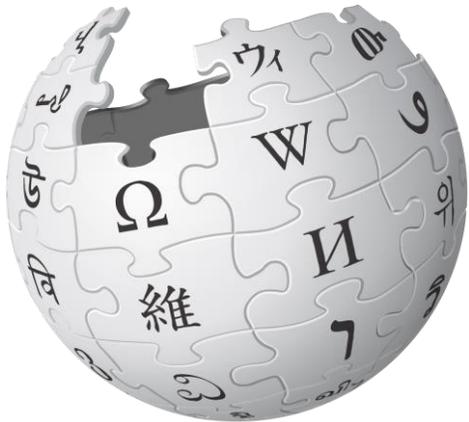
Experiment Setup

Sources:

Experiment Setup

Sources:

70%



Experiment Setup

70%



Sources:

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

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

Dealing with **English**

Important note:

Dealing with **English**

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

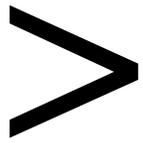
>

100% of

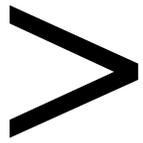
theranos



A short, but
well spoken sentence



A short, but
well spoken sentence

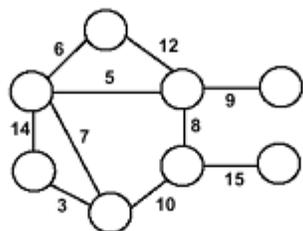


A long, fancy,
poorly spoken sentence



Minimum Spanning Trees

- Undirected, connected graph $G = (V, E)$
- Weight** function $W: E \rightarrow \mathbb{R}$ (assigning cost or length or other values to edges)



- 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

- A spanning tree of G is a subgraph which
 - is a tree
 - contains all vertices of G
- Non-tree edges are those in a spanning tree if they are not vertices

4

Minimum Spanning Trees

- Undirected, connected graph $G = (V, E)$
- Weight function $W: E \rightarrow \mathbb{R}$ (assigning cost or length or other values to edges)
- Spanning tree: tree that connects all the vertices
- Optimization problem – Minimum spanning tree (MST): tree T that connects all the vertices and minimizes $w(T) = \sum_{(u,v) \in T} w(u,v)$

5

Optimal Substructure

$MST(G) = T \cup \{e\}$ $MST(G) = T - \{e\}$

6

Idea for an Algorithm

- We have to make $V-1$ choices (edges of the MST) to arrive at the optimization goal
- After each choice we have a sub-problem one vertex or fewer than the original
- Dynamic programming algorithm, at each step, would consider all possible choices (edges)
- If we could always guess the correct choice – an edge that definitely belongs to an MST

7

Greedy Choice

- Greedy choice property: locally optimal (greedy) choice yields a globally optimal solution
- Theorem
 - Let $G = (V, E)$ and let g be a tree
 - Let e be a minimum weight edge in E connecting v to $v-1$ light edges crossing a cut
 - Then $\{e\} \cup T$ is a minimum spanning tree of G

8

Greedy Choice (2)

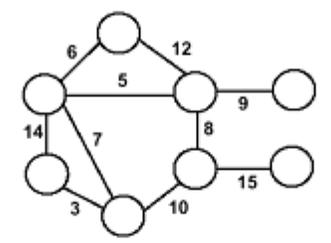
- Proof
 - Assume $\{e\} \cup T$ is not an MST
 - Let e' be the first edge in T that is not in $\{e\} \cup T$
 - Let $T' = T - \{e'\} \cup \{e\}$

9

Get

Minimum Spanning Trees

- Undirected, connected graph $G = (V, E)$
- Weight** function $W: E \rightarrow R$ (assigning cost or length or other values to edges)



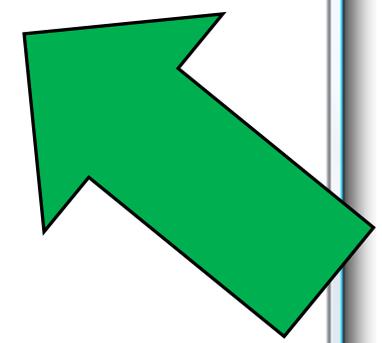
- Spanning tree: tree that connects all the vertices
- Optimization problem – **Minimum spanning tree (MST)**: tree T that connects all the vertices and minimizes $w(T) = \sum_{(u,v) \in T} w(u,v)$

November 13, 2003

4



No Notes.



Slide: 4 of 24

Time: 01:09

3:34 PM

Zoom: [Zoom in] [Zoom out]



Thumbnail navigation bar showing slides 3 through 9. Slide 4 is highlighted with a yellow background.

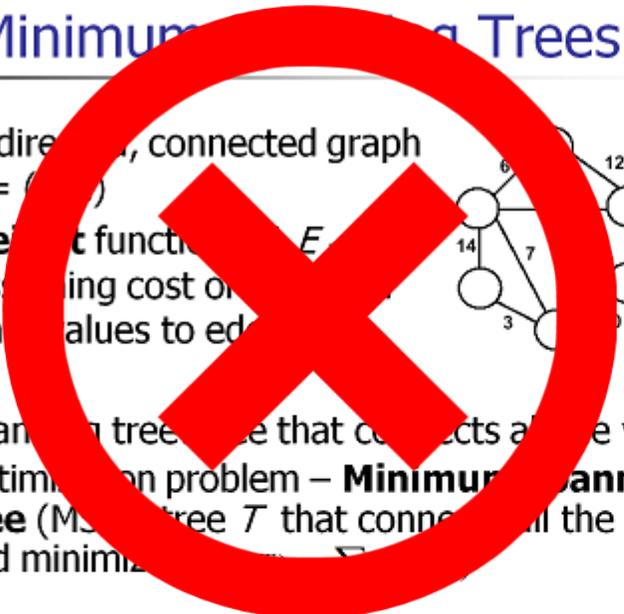
- 3 Spanning Tree
- 4 Minimum Spanning Trees
- 5 Optimal Substructure
- 6 Idea for an Algorithm
- 7 Greedy Choice
- 8 Greedy Choice (2)
- 9 Get

No Notes.

Minimum Spanning Trees

- Undirected, connected graph $G = (V, E)$
- Weight function $w: E \rightarrow \mathbb{R}$ (assigning cost or length or other values to edges)
- Spanning tree is a subgraph that connects all the vertices
- Optimization problem – **Minimum Spanning tree** (MST) is a tree T that connects all the vertices and minimizes the total weight

November 13, 2003 4



Slide: 4 of 24

Time: 01:09

3:34 PM

Zoom:



3 **Spanning Tree**

- Spanning tree of G is a subgraph which
 - contains all vertices of G
 - is a tree
- Non-empty edges are those in a spanning tree, if they are not vertices

4 **Minimum Spanning Trees**

- Undirected, connected graph $G = (V, E)$
- Weight function $w: E \rightarrow \mathbb{R}$ (assigning cost or length or other values to edges)
- Spanning tree is a subgraph that connects all the vertices
- Optimization problem – Minimum Spanning tree (MST) is a tree T that connects all the vertices and minimizes the total weight

5 **Optimal Substructure**

6 **Idea for an Algorithm**

- We have to make $V-1$ choices (edges of the MST) to arrive at the optimization goal
- After each choice we have a sub-problem one vertex or fewer than the original
- Dynamic programming algorithm, at each step, would consider all possible choices
- "Cut and Sew" argument: if T is a MST then a choice of $V-1$ edges of T is a MST of $G - \{v\}$

7 **Greedy Choice**

- Greedy choice property: locally optimal (greedy) choice yields a globally optimal solution
- Theorem: Let $G = (V, E)$ and let g be a greedy choice. Let T be a minimum weight edge in G connecting v to $V - \{v\}$. If g is a greedy choice, then $T \cup \{g\}$ is a MST of G .

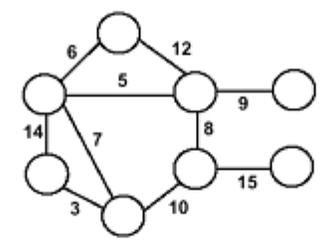
8 **Greedy Choice (2)**

- Proof: Suppose g is a greedy choice of G . Let T be a minimum weight edge in G connecting v to $V - \{v\}$. If g is a greedy choice, then $T \cup \{g\}$ is a MST of G .

9 **Get**

Minimum Spanning Trees

- Undirected, connected graph $G = (V, E)$
- Weight** function $W: E \rightarrow R$ (assigning cost or length or other values to edges)



- 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: [Zoom in] [Zoom out]



Thumbnail navigation bar showing slides 3 through 9:

- 3 **Spanning Tree**: A spanning tree of G is a subgraph which contains all vertices of G .
- 4 **Minimum Spanning Trees**: Undirected, connected graph $G = (V, E)$. Weight function $W: E \rightarrow R$.
- 5 **Optimal Substructure**: $MST(G) = T \cup \{e\}$. Cut and Sew arguments.
- 6 **Idea for an Algorithm**: We have to make $V-1$ choices (edges of the MST) to arrive at the optimization goal.
- 7 **Greedy Choice**: Greedy choice property: locally optimal (greedy) choice yields a globally optimal solution.
- 8 **Greedy Choice (2)**: Proof: Success (u,v) is type $u, v \in T$.
- 9 **Get**: [Thumbnail]

Dealing with **English**

Cue card tools:

Dealing with **English**

Cue card tools:

- **Windows:** Powerpoint presenter view

Dealing with **English**

Cue card tools:

- **Windows:** Powerpoint presenter view
- **Mac OS:** Keynote presenter view

Dealing with **English**

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!

Dealing with **English**

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**

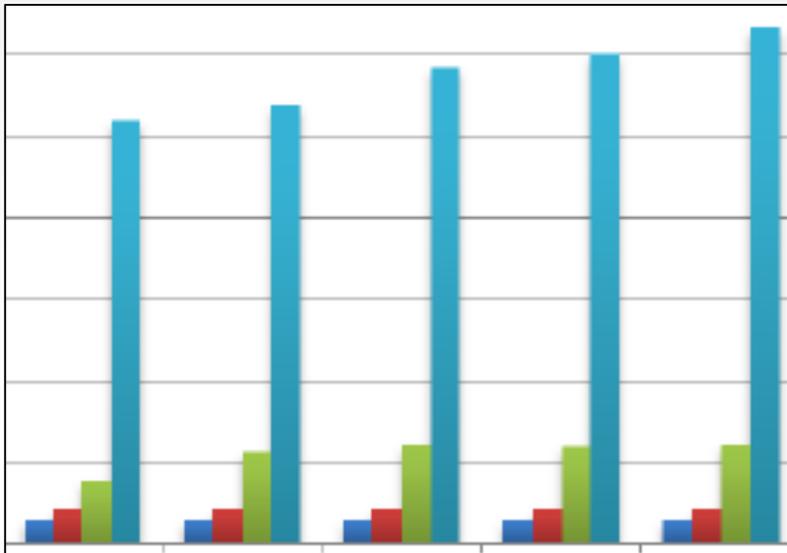
The performance preview slide

(right after introduction)

Dealing with **English**

The performance preview slide

(right after introduction)



Dealing with **English**

The English proficiency alert

Dealing with **English**

The English proficiency alert

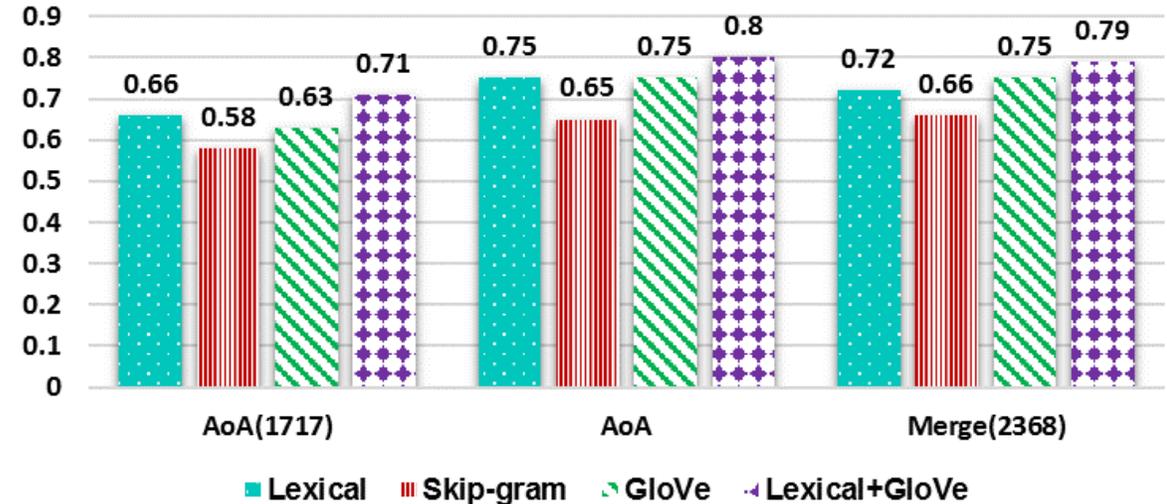
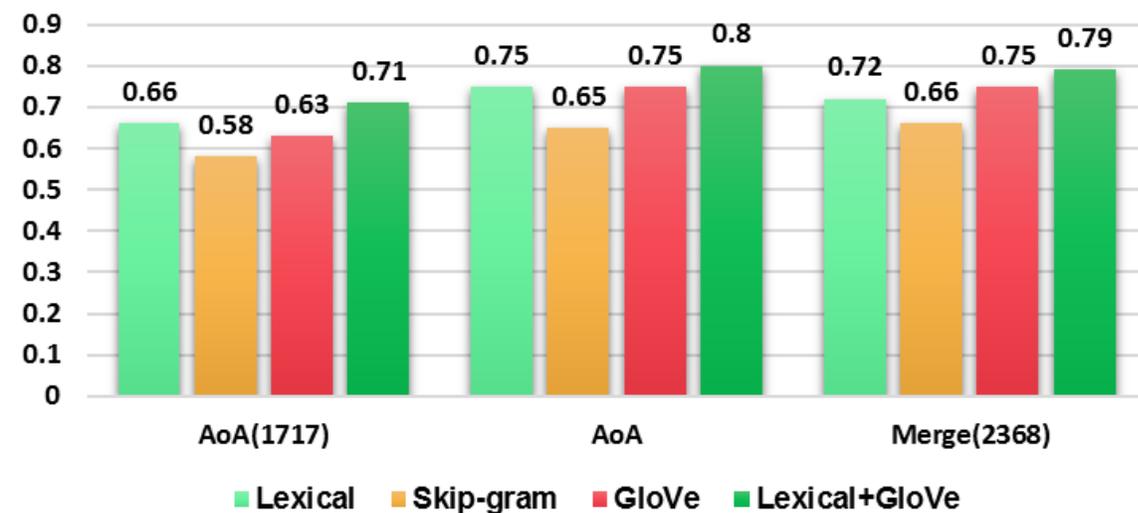
I'm still learning English!
Please be kind with the questions 😊

Dealing with **English**

Accessible graphs (for the colorblind)

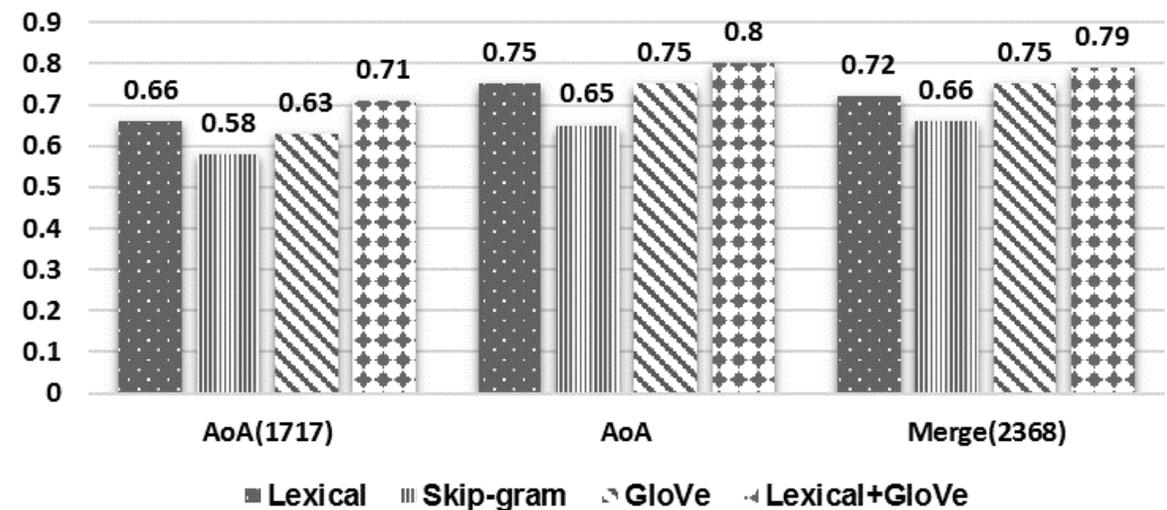
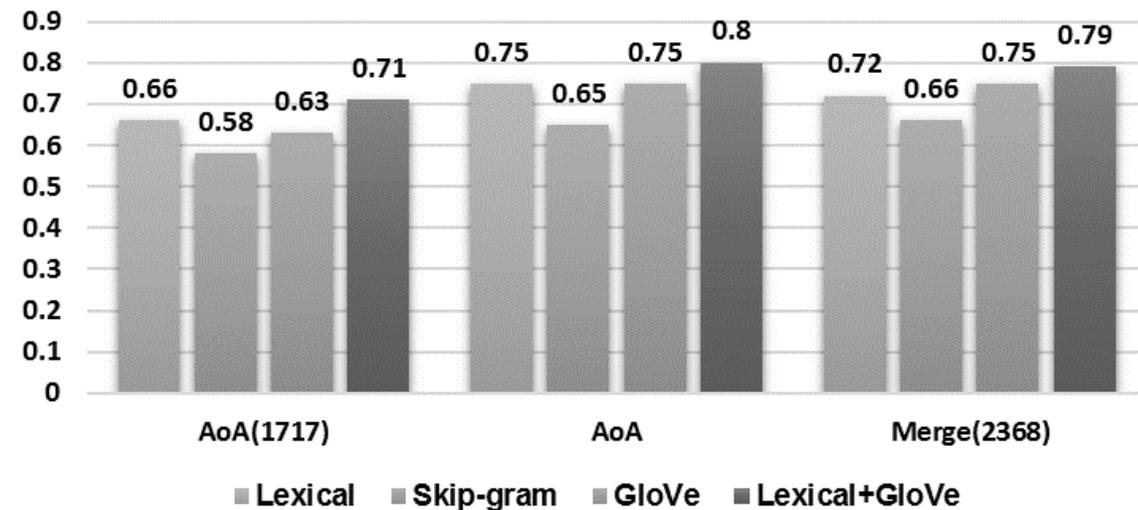
Dealing with English

Accessible graphs (for the colorblind)



Dealing with English

Accessible graphs (for the colorblind)



Dealing with **English**

But what about...

Dealing with **English**

But what about...
...the Q&A session?

Dealing with **English**

But what about...
...the Q&A session?



Dealing with **English**

Solution:

Dealing with **English**

Solution:

The “presentation buddy”

Dealing with **English**

Presenter



Dealing with **English**

Presenter



Attendee



Dealing with **English**

Presenter



Attendee



Questions



Dealing with **English**

Presenter



Attendee



Questions



Answers



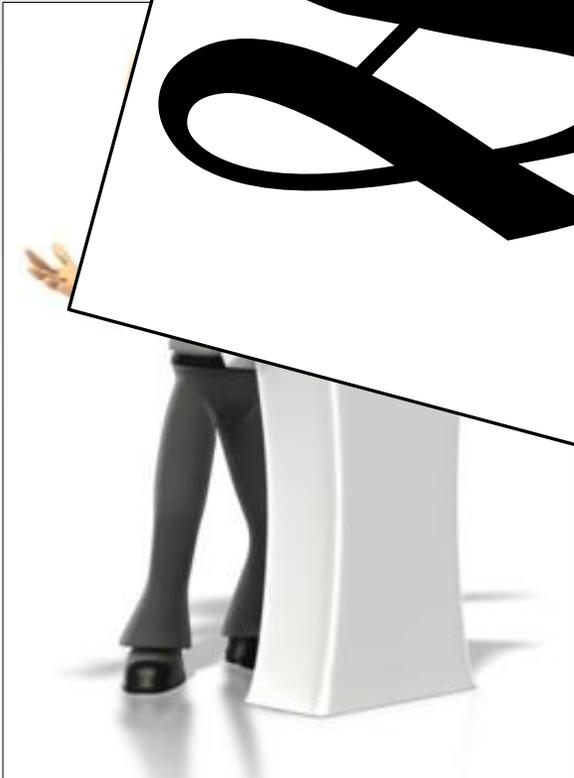
De

ish

Pre

e

Scary



Dealing with **English**

Presenter



Attendee



Dealing with English

Presenter



Buddy



Attendee



Dealing with English

Presenter



Buddy



Attendee



Dealing with English

Presenter



Buddy



Attendee



Dealing with English

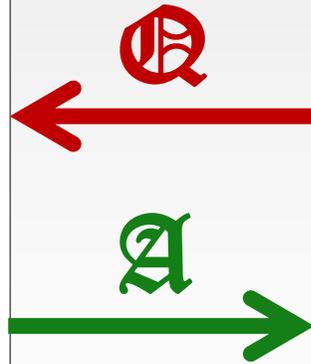
Presenter



Buddy



Attendee



Dealing with English

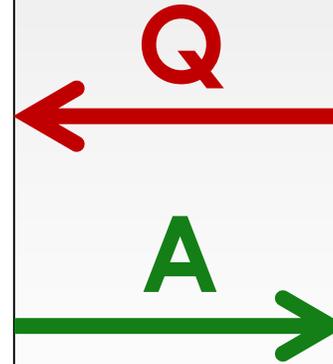
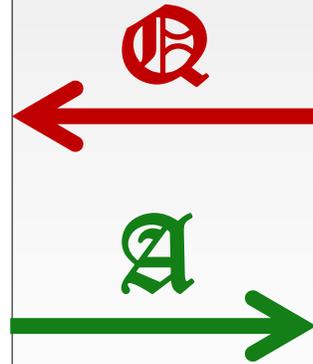
Presenter



Buddy



Attendee



Dealing with English

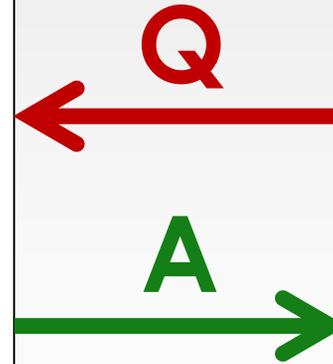
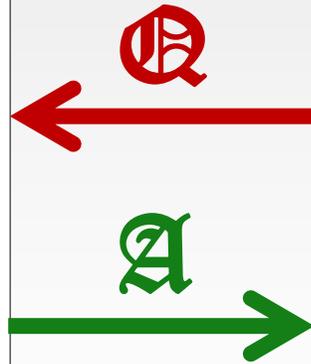
Presenter



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!