

Automatically _____ gap-fill exercise items

Torsten Zesch

Assistant Professor
Language Technology Lab
University of Duisburg-Essen



Biographical Facts

Computer science background

2006 – 2012

Technische Universität Darmstadt (PhD/PostDoc)

- Lexical semantics / Wikipedia

2012 – 2013

German Institute for International Educational Research (DIPF), Frankfurt
(Substitute Professor)

- Automatic scoring (PISA data)

Now

University of Duisburg-Essen, “Language Technology Lab” (Assistant Professor)



Photo: Stadt Essen

Duisburg

- 490.000 inhabitants
- the most important steel production site in Europe
- logistical centre of Germany, with the largest inland port in Europe

Essen

- 570.000 inhabitants
- the cultural and economical centre of the Rhine-Ruhr region as well as a hotspot of the service industries

In the heart of Europe



In the heart of Europe



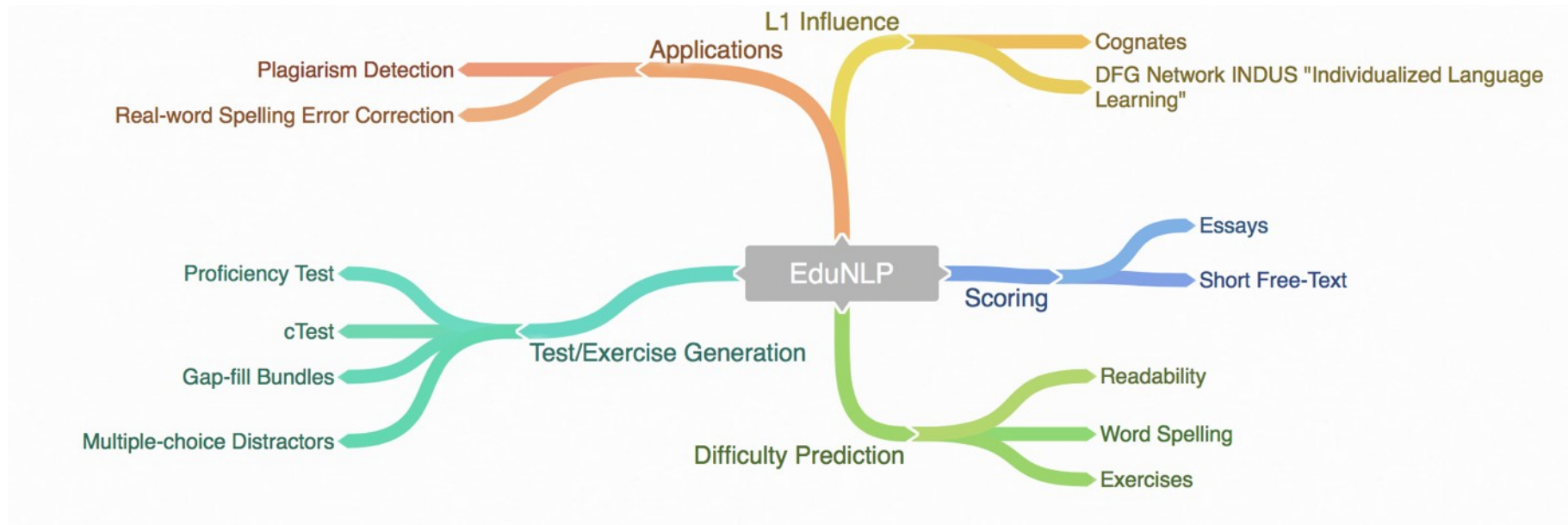
NLP for Education

Social Media Analysis

- Graduate school “User-centred Social Media”

Language Technology Infrastructure

- Reproducibility & Replicability



NLP4Edu (but not CALL) @ LTLab

Automatic scoring

- Essays
- Short free-text answers

Plagiarism detection

Spell checking

Language Proficiency Testing

- Lexical Recognition Tests (for Arabic)
- cTest difficulty prediction

Pred. Avg. Error Rate: 0.4

My friend and I bought the same book on Amazon , and it changed us , but in quite different ways . The bo ok
called ' Eating Animals ' b y Jonathan Safran Foer , was a n examination o f the grue some way pou ltry
and cat tie are slaug htered to pro duce cheap me at for pec ple like m e . A qua rter of t he way
thr ough I clo sed it , tuc ked it o n my sh elf - I kn ew that if I continued to read I would no longer be able
to enjoy roast chicken . While it opened my eyes to my failings and limits , Becca , who finished the book ,
became a vegetarian .

Learner Language

- Cognates
- L2 spelling errors
- Readability assessment

Czech	Human associations (German)	COP productions (German)
nudle	Nudel (15)	<u>nudel</u> , nadel, ode
švagr	Schwager (13)	sauger, <u>schwager</u> , berg
šlak	Schlag (12), Schlagsahne (3), schlagen (2)	stak
brýle	Brille (12), brüllen (4)	<u>brille</u> , brie

Exercise generation

- This talk □

Research network INDUS – “Individualized language learning”

- <https://sites.google.com/site/indusnetzwerk/home> (in German)

Automatically _____ gap-fill exercise items
generating

Is there a problem?

He lost his watch yesterday

Is there a problem?

He lost his _____ yesterday

Automatically _____ gap-fill exercise items

generating

checking

Checking – Ambiguity Problem

He lost his marbles yesterday

house
wife
car
job
hand
keys
glasses
wallet
soul
phone
confidence
hope
will
future
degree
friend
head
book
pen
rubber
hat
ring

Checking – Multiple Choice

He lost his _____ yesterday

- *go*
- *mind*
- *take*
- *him*
- *watch*
- *mind*
- *wife*
- *game*

Automatically _____ gap-fill exercise items

generating

checking

adapting

Adapting – Type of Knowledge

He lost his _____ yesterday

- *go*
- *mind*
- *take*
- *him*

He lost _____ mind yesterday

- *him*
- *his*
- *he*
- *her*

He lost his _____ yesterday

- *go*
- *mind*
- *take*
- *him*

His characteristic talk, with its keen _____ of detail and subtle power of inference held me amused and enthralled.

- *instincts*
- *observance*
- *presumption*
- *implements*

Reducing Ambiguity



Horsmann, Tobias; Zesch, Torsten

Towards Automatic Scoring of Cloze Items by Selecting Low-Ambiguity Contexts.

In: 3rd workshop on NLP for computer-assisted language learning, Uppsala, Sweden, 2014.

Ambiguity Problem

He lost his marbles *yesterday*
house
wife
car
job
hand
keys
glasses
wallet
soul
phone
confidence
hope
will
future
degree
friend
head
book
pen
rubber
hat
ring

Try to find contexts that are
less ambiguous
(or even unambiguous)

Ambiguity Problem

house
wife
car
job-
hand
keys
glasses
wallet
soul
phone
He lost his bag of marbles *yesterday*
confidence
hope
will
future
degree
friend
head
book
pen
rubber
hat
ring

Finding Low-Ambiguity Contexts

Hearst Patterns

- "Z such as X and Y"
- *Already small amounts of _____ such as beer or wine can affect your ability to drive. → alcohol*

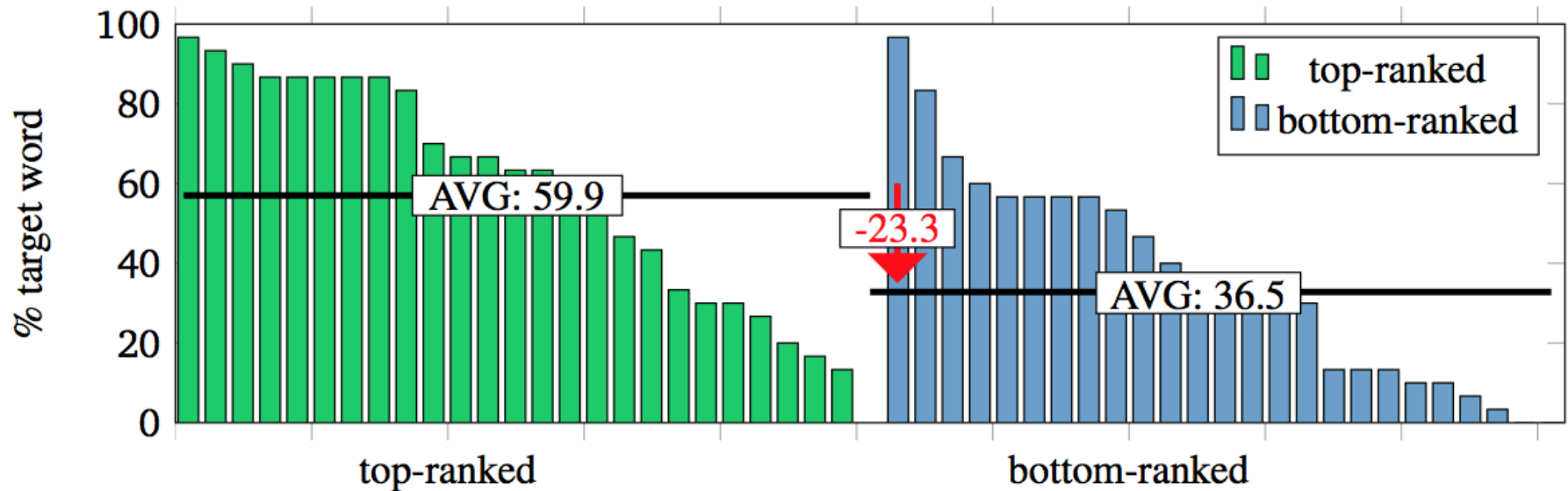
Collocations

- *Lessons learned from successful forest _____ prevention campaigns were presented. → fire*

Word Repetition

- *Go through the _____ and mark each affected slice. → slices*

Evaluation Results



- top-ranked are significantly less ambiguous
- still too many valid solutions

Finding unambiguous gaps might be impossible

Challenging Distractors



Zesch, Torsten; Melamud, Oren

Automatic Generation of Challenging Distractors Using Context-Sensitive Inference Rules.

In: Proceedings of the 9th Workshop on Innovative Use of NLP for Building Educational Applications at ACL, Baltimore, USA, 2014.

Tradeoff – Checking & Difficulty

Microsoft _____ Skype for 8.5b dollar.

Solution **acquires**

Too easy **drinks**

Challenging **gains**

Invalid **purchases**

acquire \square gain

?

students **acquire** knowledge



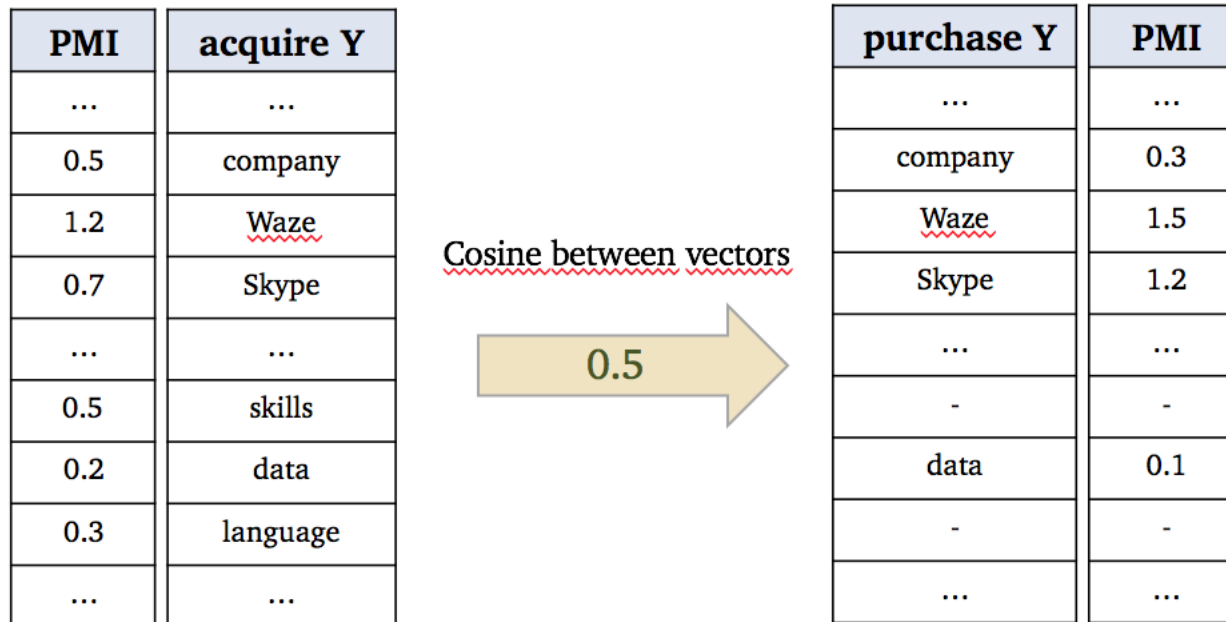
students **gain** knowledge

Microsoft **acquires** Skype



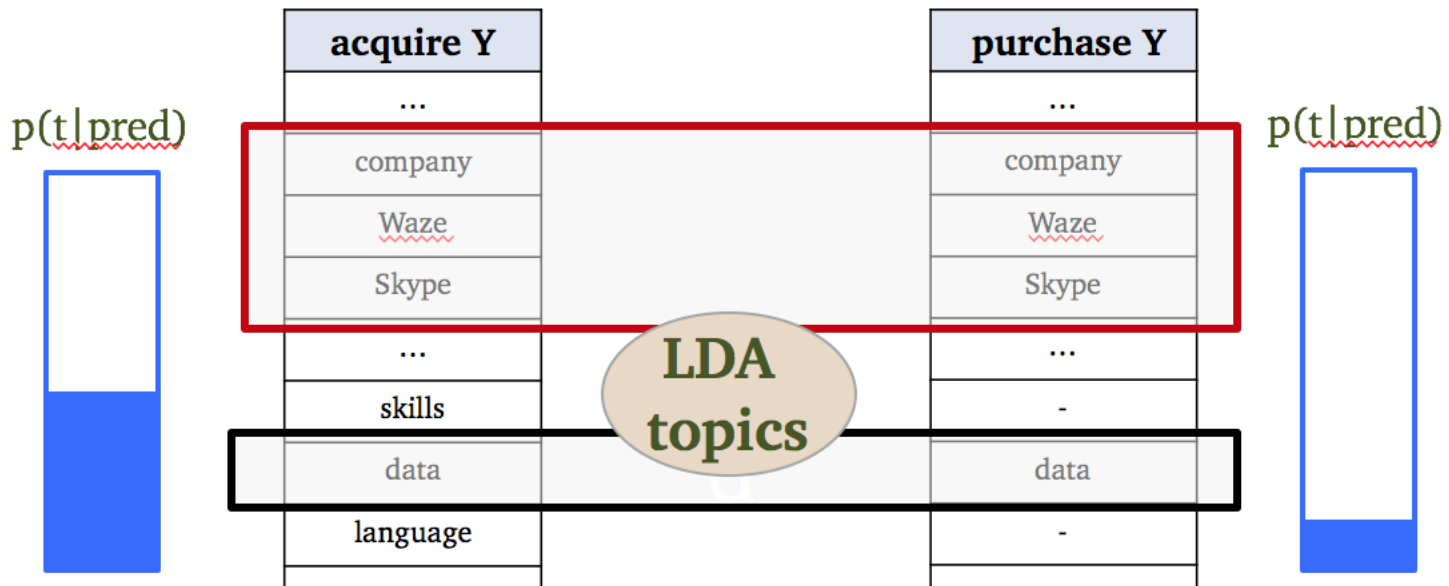
Microsoft **gains** Skype

Context insensitive rules (DIRT-like)



Context-sensitive Rules

Learn inference rules per sense &
Apply according to senses induced by context



Melamud Oren, Jonathan Berant, Ido Dagan, Jacob Goldberger, Idan Szpektor.

A two level model for context sensitive inference rules.

In: Proceedings of ACL 2013, Best paper runner-up

Finding Challenging Gap-fillers

Microsoft acquires Skype for 8.5b dollar.

Context-sensitive Gap-fillers

buy
have
lose
obtain
own
possess
purchase
retain
sell

Context-insensitive Gap-fillers

attain
gain
have
lack
lose
obtain
purchase
recover
sell

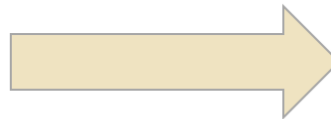
Finding Challenging Gap-fillers

Microsoft acquires Skype for 8.5b dollar.

Context-sensitive Gap-fillers

buy
have
lose
obtain
own
possess
purchase
retain
sell

Remove
invalid distractors



Context-insensitive Gap-fillers

attain
gain
have
lack
lose
obtain
purchase
recover
sell

Evaluation Dataset

Dataset properties

- Sentences with main verb
- Medium-sized (5-12 tokens)
- Understandable by general audience

Examples

- "His petition charged mental cruelty"
- "The ball floated downstream"
- "The plan does not cover doctor bills"

Difficulty of Resulting Items

Randomly split 52 learners in two groups

- Group 1: our approach
 - Group 2: randomly selected distractors (with comparable corpus frequency)
-
- 7 control items (same for both groups)
 - 13 test items

Error rates		
	Group 1	Group 2
Control Items	0.24 ± 0.12	0.20 ± 0.12
Test Items	0.18 ± 0.17	0.18 ± 0.15

Conclusion

Works well for clearly separated verb senses

- For example: “He played basketball” □ “compose”, “tune”

Problems with fine-grained distinctions

- For example: “The ball floated downstream” □ “glide”, “travel”

Contrary to hypothesis, generated distractors not more difficult

- Subjects with high language proficiency □ re-test with beginners
- Error rate does not directly measure difficulty / learning, as subjects might be confused but still solve it in the end □ measure response time

Generating difficult distractors is a hard problem

Bundled Gap-fill



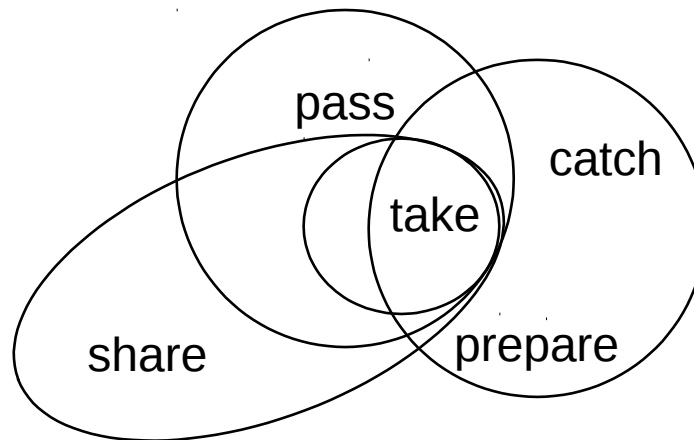
Wojatzki, Michael; Melamud, Oren; Zesch, Torsten

Bundled Gap Filling: A New Paradigm for Unambiguous Cloze Exercises

In: Proceedings of the Building Educational Applications Workshop at NAACL, ACL, San Diego, USA, 2016.

Bundled Gap-fill

The students have to ____ the test.
Their cook will ____ three salmons.
All passengers should ____ their seats.
Both authors ____ credit for this.



Exercise Type Overview

Type	Example	Ambiguity
gap-fill	The students have to ____ the test.	high
multiple-choice gap-fill	The students have to ____ the test. a) take b) fold c) entertain d) fry	low to moderate depending on selected distractors
bundled gap-fill	The students have to ____ the test. Their cook will ____ three salmons. All passengers should ____ their seats. Both authors ____ credit for this.	low to moderate depending on selected gaps

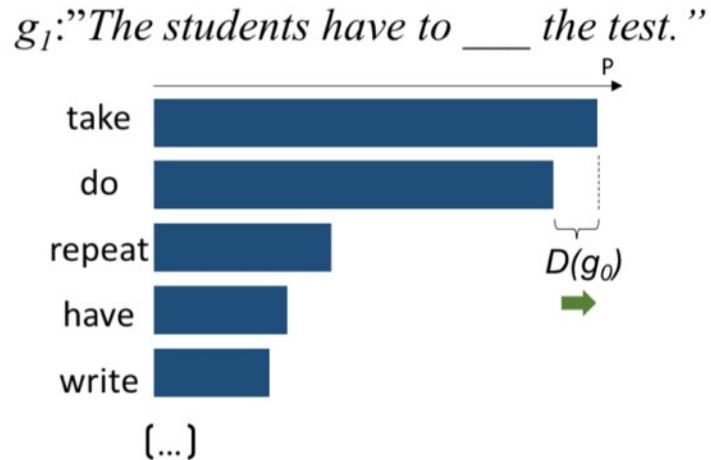
Exercise Type Overview

Type	Example	Ambiguity	Automatability
gap-fill	The students have to ____ the test.	high	high
multiple-choice gap-fill	The students have to ____ the test. a) take b) fold c) entertain d) fry	low to moderate depending on selected distractors	low as it's hard to control both ambiguity and difficulty
bundled gap-fill	The students have to ____ the test. Their cook will ____ three salmons. All passengers should ____ their seats. Both authors ____ credit for this.	low to moderate depending on selected gaps	high using the scheme proposed in this paper

Exercise Type Overview

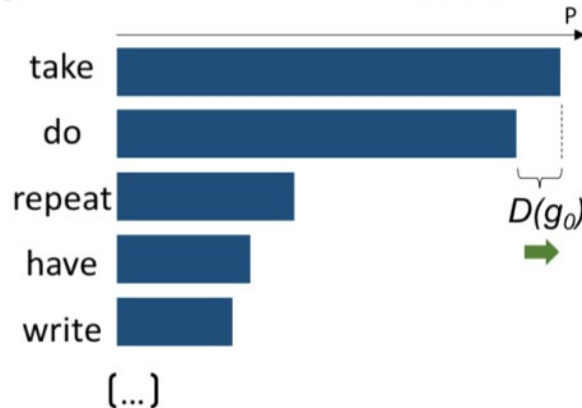
Type	Example	Ambiguity	Automatability	Nature of Task
gap-fill	The students have to ____ the test.	high	high	production
multiple-choice gap-fill	The students have to ____ the test. a) take b) fold c) entertain d) fry	low to moderate depending on selected distractors	low as it's hard to control both ambiguity and difficulty	recognition
bundled gap-fill	The students have to ____ the test. Their cook will ____ three salmons. All passengers should ____ their seats. Both authors ____ credit for this.	low to moderate depending on selected gaps	high using the scheme proposed in this paper	production

Approach – Probability of Gap Fillers

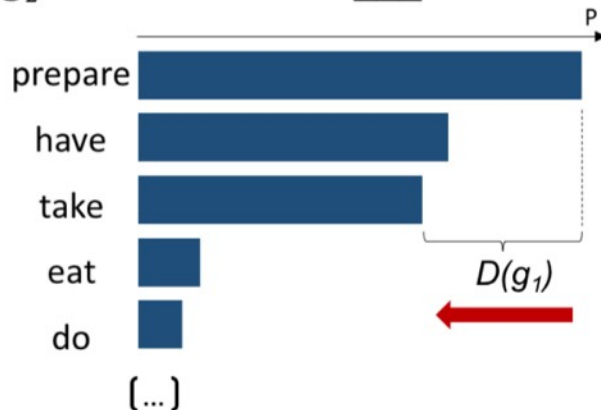


Approach – Probability of Gap Fillers

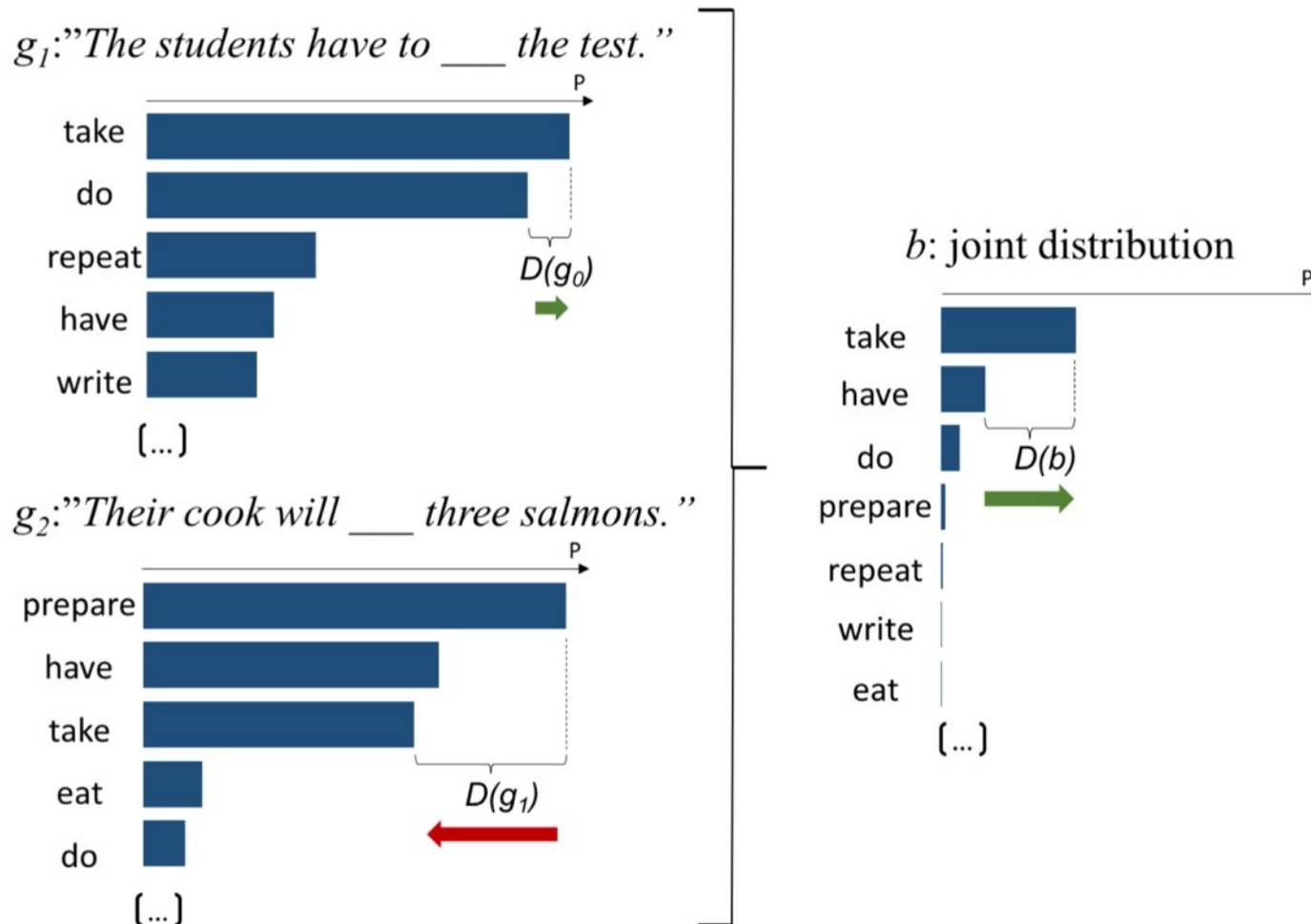
g_1 : "The students have to ____ the test."



g_2 : "Their cook will ____ three salmons."



Approach – Probability of Gap Fillers



Target words

- controlled for word frequency and word classes

Target Words	Frequency Class	Word Class
new	5	Adjectives
best	6	
full	7	
final	8	
people	5	Nouns
language	6	
information	7	
room	8	
make	5	Verbs
want	6	
add	7	
give	8	

Table 1: List of target words

Generate Bundles

Target words

- controlled for word frequency and word classes

Sentence base

- single gap selected at random
- GUM corpus (Zeldes, 2016)

Background Corpus

- ukWaC English (Baroni et al., 2009)

Probability distribution of gap-filler words

- KenLM toolkit (Heafield et al., 2013)
- FASTSUBS to efficiently compute vectors (Yuret., 2012)

Short Wrap-Up

Approach to generate bundled gap-fill exercises

The students have to ____ the test.
Their cook will ____ three salmons.
All passengers should ____ their seats.
Both authors ____ credit for this.

Remaining task:

- How well does it work?

User Study

- Near native speakers (N=35)
- Participants have to successively solve bundles with increasing sizes

- Near native speakers (N=35)
- Participants have to successively solve bundles with increasing sizes

The students have to ____ the test.



How many participants filled the solution word? □ Success rate

- Near native speakers (N=35)
- Participants have to successively solve bundles with increasing sizes

The students have to ____ the test.

Their cook will ____ three salmons.



How many correct now?

If bundles work, score should increase

- Near native speakers (N=35)
- Participants have to successively solve bundles with increasing sizes

The students have to ____ the test.

Their cook will ____ three salmons.

All passengers should ____ their seats.

Both authors ____ credit for this.



How many correct now?

Results

Average Success Rate	
Single Gap	.27

Results

	Average Success Rate
Single Gap	.27
Bundle ₂	.59

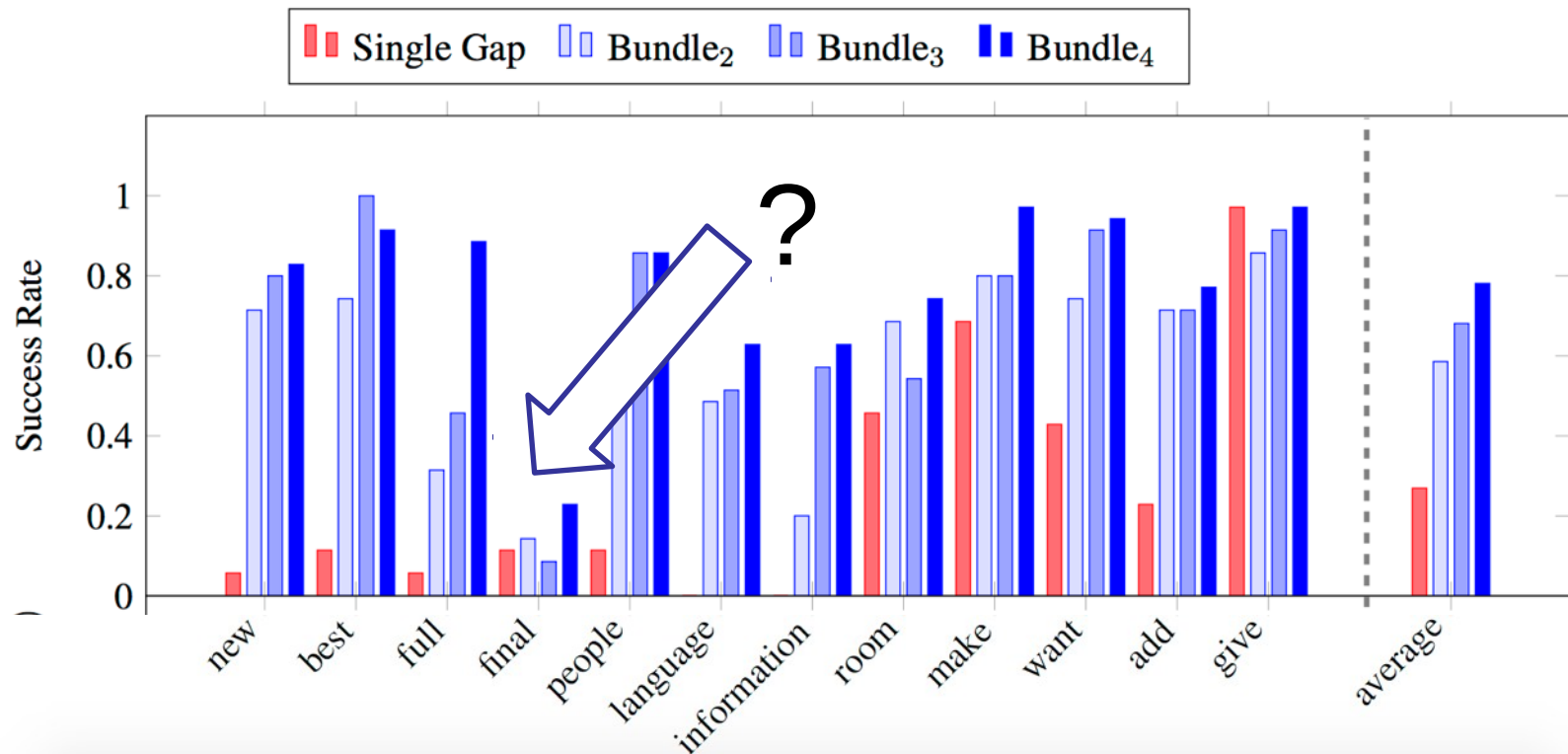
Results

	Average Success Rate
Single Gap	.27
Bundle ₂	.59
Bundle ₃	.68
Bundle ₄	.78

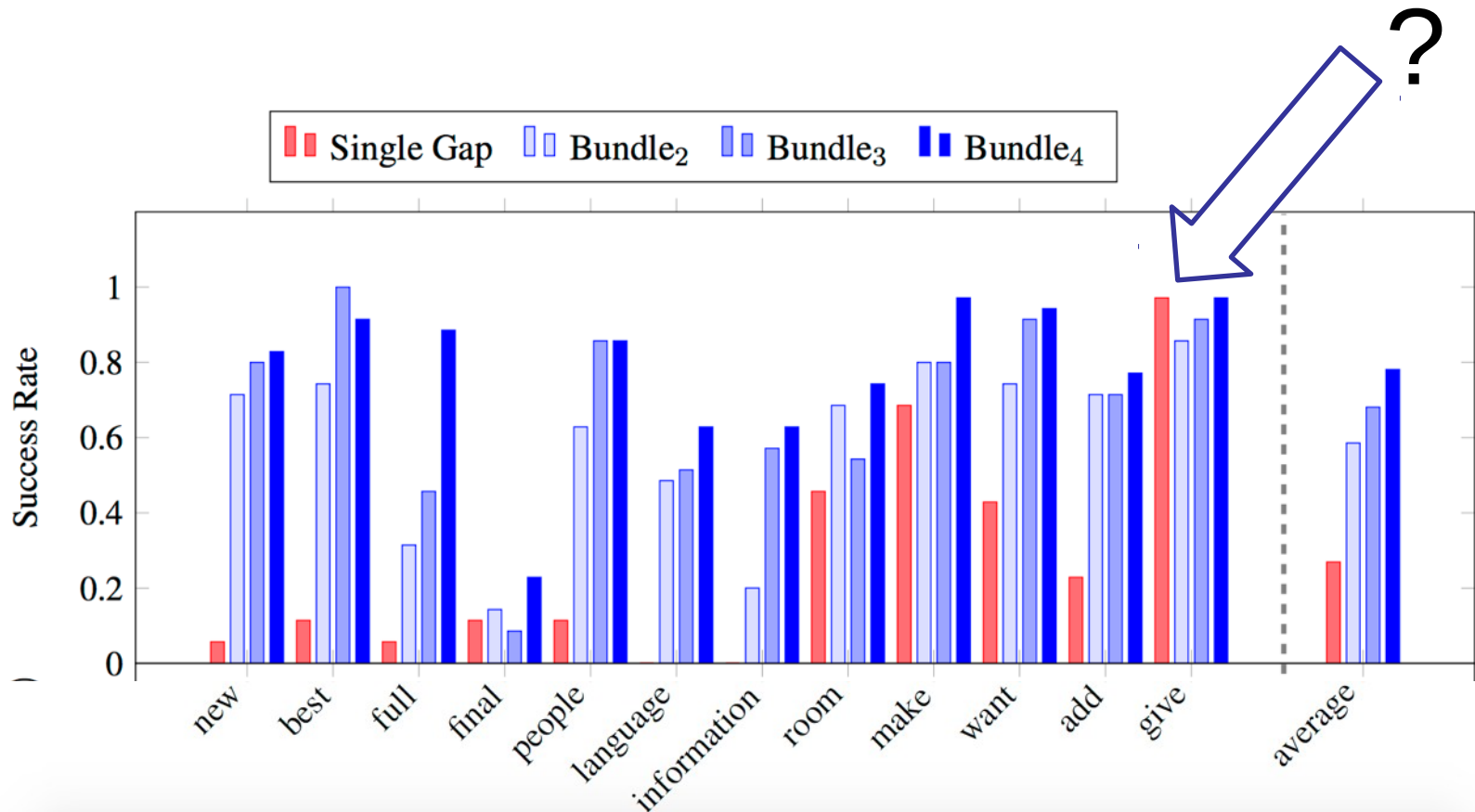
Results

	Average Success Rate	Average Disambiguation Measure
Single Gap	.27	-0.50
Bundle ₂	.59	4.00
Bundle ₃	.68	7.75
Bundle ₄	.78	11.06

Detailed Results



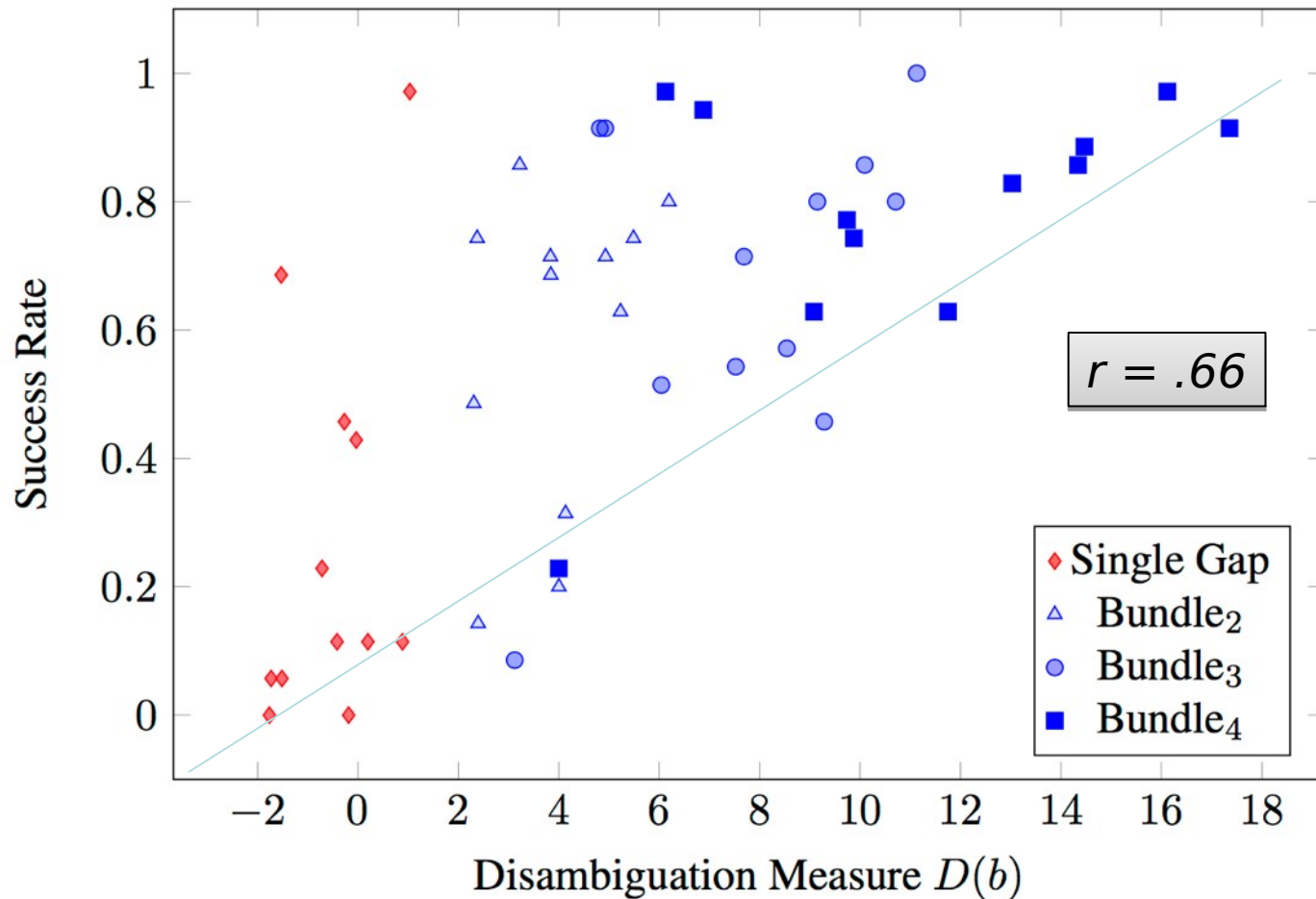
Detailed Results



Error Analysis – “give” bundle

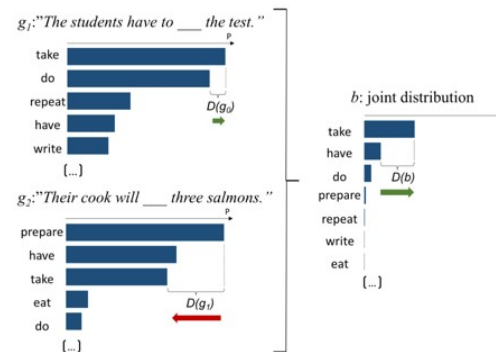
1. A string of islands and reefs along the eastern edge of the Group shelters the area from strong winds and ocean swells and humpback whales come to these protected waters to ____ birth.
2. Tape the bottom of the box securely so that it doesn't ____ way.
3. We take document security seriously , " she said , but refused to ____ any more details about how the papers came to be on a road.
4. Clap your hands or ____ it a gentle shove until it jumps up and walks away.

Detailed Results

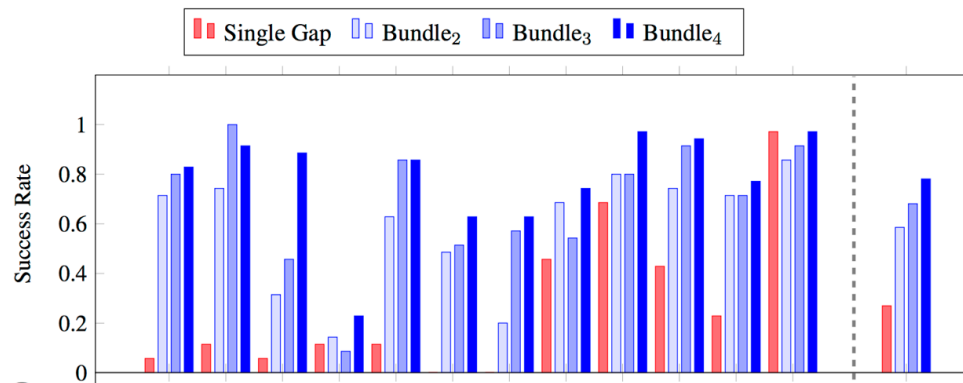


Summary – Bundled Gap-fill

Bundled gap fill tests can be effectively generated



Gap bundles decrease ambiguity



Wrap Up

Reducing ambiguity of cloze tests

- Works to some extent, but not suitable for automatic generation

house
wife
car
job
hand
keys
glasses
wallet
soul
phone
He lost his marbles *yesterday*
confidence
hope
will
future
degree
friend
head
book
pen
rubber
hat
ring

Wrap Up

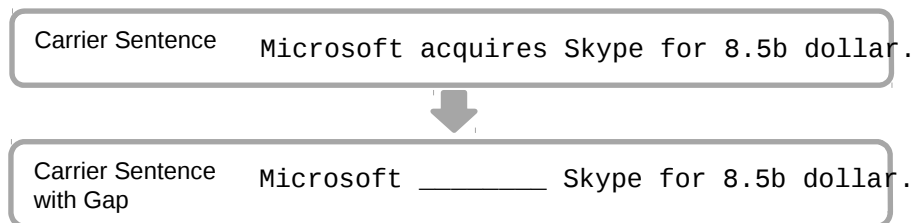
Reducing ambiguity of cloze tests

- Works to some extent, but not suitable for automatic generation

Generating challenging distractors

- Works to some extent, but not good enough for automatic generation

Fill-the-gap Example



Distractor Difficulty



Reducing ambiguity of cloze tests

- Works to some extent, but not suitable for automatic generation

Generating challenging distractors

- Works to some extent, but not good enough for automatic generation

New approach □ bundled gap-fill

- Can be reliably generated
- Many open questions ...

The students have to ____ the test.

Their cook will ____ three salmons.

All passengers should ____ their seats.

Both authors ____ credit for this.

Future Work (Bundles only)

- Adapt to other languages
- Improve gap probability model
- Relation to established proficiency tests
- Predict and adapt bundle difficulty

Thanks for your _____
Special _____ should be paid to this work
_____ is given to these exercises
_____ is a currency

It's all team work

