

# Tâches à saveur sémantiques

---

[felipe@iro.umontreal.ca](mailto:felipe@iro.umontreal.ca)

**RALI**

Dept. Informatique et Recherche Opérationnelle  
Université de **Montréal**



V0.5

Last compiled: 22 novembre 2018



# Plan

Sémantique

Question-réponse

ARC

RACE

WDW

# Plan

Sémantique

Question-réponse

ARC

RACE

WDW

# Crédit

- ▶ la plupart de ces ressources (et d'autres) sont listées ici :  
<http://www.cs.cmu.edu/~mfaruqui/suite.html>

# Rubenstein and Goodenough

[http://www.cs.cmu.edu/~mfaruqui/word-sim/  
EN-RG-65.txt](http://www.cs.cmu.edu/~mfaruqui/word-sim/EN-RG-65.txt)

1	gem	jewel	3.94
2	midday	noon	3.94
3	automobile	car	3.92
4	cemetery	graveyard	3.88
5	cushion	pillow	3.84
6	...		
7	autograph	shore	0.06
8	fruit	furnace	0.05
9	noon	string	0.04
10	rooster	voyage	0.04
11	cord	smile	0.02

# WordSimilarity-353

- ▶ <http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353/>

- ▶ **set1** 153 paires de mots : 13 jugements de **relatedness**
- ▶ **set2** 200 paires de mots : 16 jugements

```
1 % more set1.csv | sort -t\, -k3,3n+
2 tiger,tiger,10.00,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10
3 coast,shore,9.10,9,9.75,10,10,6,6,10,8,9.5,10,10,10,10,10,10,10
4 football,soccer,9.03,9,9.9,9,10,8,7,9.5,10,8,10,9,9,9,9,9,9
5 fuck,sex,9.44,10,9.75,10,10,8,9,9,10,9,10,8,10,10,10,10,10
6 journey,voyage,9.29,9,9.75,10,10,6,7,9.5,10,9.5,10,10,10,10,10,10
```

- ▶ split de <http://alfonseca.org/eng/research/wordsim353.html>

similarité 203 paires  
relatedness 252 paires



# À propos du split de *[Agirre et al., 2009]*

- ▶ 103 paires en commun dans les deux fichiers :

```
more wordsim_relatedness_goldstandard.txt \
  wordsim_similarity_goldstandard.txt | \
  sort | uniq -c | awk '$1 == 2 {print $0}' | \
  sort -k4,4gr | head -n 5+
```

1	2	cup	food	5,00
2	2	monk	oracle	5,00
3	2	journal	association	4,97
4	2	car	flight	4,94
5	2	street	children	4,94

# MEN

[http:](http://clic.cimec.unitn.it/~elia.bruni/MEN.html)

[//clic.cimec.unitn.it/~elia.bruni/MEN.html](http://clic.cimec.unitn.it/~elia.bruni/MEN.html)

- ▶ 3000 paires de mots avec un score (entre 0 et 50) indiquant leur *relatedness*

cat-n feline-j	48	gun-n stair-n	2
copper-n metal-n	48	military-j tomato-n	2
flower-n garden-n	48	angel-n gasoline-n	1
grass-n lawn-n	48	feather-n truck-n	1
pregnancy-n pregnant-j	48	festival-n whisker-n	1

- ▶ paires vérifiant des contraintes de fréquences dans différentes ressources et échantillonnées de manière à générer une tâche équilibrée
- ▶ évaluation faite par comparaison binaire à 50 autres paires sur Mechanical Turk
- ▶ **tâche** : trouver les paires les plus similaires (ranking)





# TOEFL Tests *[Freitag et al., 2005]*

- ▶ accessible ici :  
<http://www.cs.cmu.edu/~dayne/wbst-nanews.tar.gz>
- ▶ chaque instance de test est de la forme :

*target* :  $\overbrace{[\text{synonym and 3 distractors}]}$  *stimulus*

ex :

comprehend : [encompass career admire pardon]

- ▶ 9887 instances de test pour les noms, 7398 pour les verbes

TOEFL Tests *[Freitag et al., 2005]*

force	violence	outburst encephalopathy sparta
force	power	plaintif vomiting deep
force	personnel	distress universe poe
force	strength	adrian character atlanta
force	effect	abuja flora share
buy	bribe	supply reinvent triumph
buy	corrupt	bar recount weight
buy	purchase	harp esteem forgive
record	commemorate	smack resort tremor
record	enter	ripple subpoena shirt
record	tape	drink pare pearl

# Word2Vec datasets (analogies)

<http://word2vec.googlecode.com/svn/trunk>

questions-words.txt 19544 équations (avec solution) :

<p>: capital-common-countries            Berlin Germany Islamabad Pakistan            Baghdad Iraq Oslo Norway</p> <p>: capital-world            Algiers Algeria Belmopan Belize            Abuja Nigeria Doha Qatar</p> <p>: currency            Armenia dram India rupee            Angola kwanza Denmark krone</p> <p>: city-in-state            Philadelphia Pennsylvania Atlanta Georgia            Chicago Illinois Stockton California</p> <p>: family            father mother husband wife            brother sister prince princess</p> <p>: gram1-adjective-to-adverb            cheerful cheerfully efficient efficiently            apparent apparently free freely</p> <p>: gram2-opposite            clear unclear impressive unimpressive            aware unaware efficient inefficient</p>	<p>: gram3-comparative            bright brighter simple simpler            big bigger bright brighter</p> <p>: gram4-superlative            bright brightest weird weirdest            big biggest dark darkest</p> <p>: gram5-present-participle            debug debugging code coding            dance dancing enhance enhancing</p> <p>: gram6-nationality-adjective            Australia Australian India Indian            Albania Albanian Sweden Swedish</p> <p>: gram7-past-tense            describing described playing played            dancing danced vanishing vanished</p> <p>: gram8-plural            bottle bottles man men            bird birds bottle bottles</p> <p>: gram9-plural-verbs            enhance enhances play plays            describe describes increase increases</p>
--	--

# Word2Vec datasets (analogies)

<http://word2vec.googlecode.com/svn/trunk>

questions-phrases.txt 3218 équations (avec solution) :

: newspapers

Boston Boston\_Globe Salt\_Lake Salt\_Lake\_Tribune  
Baltimore Baltimore\_Sun Cincinnati Cincinnati\_Enquirer

: ice\_hockey

Calgary Calgary\_Flames Ottawa Ottawa\_Senators  
Boston Boston\_Bruins Chicago Chicago\_Blackhawks

: basketball

Dallas Dallas\_Mavericks Boston Boston\_Celtics  
Boston Boston\_Celtics New\_York New\_York\_Knicks

: airlines

Finland Finnair Spain Spanair  
Austria Austrian\_Airlines Turkey Turkish\_Airlines

: people-companies

Samuel\_J.\_Palmisano IBM Jeff\_Bezos Amazon  
Eric\_Schmidt Google Larry\_Ellison Oracle

# À propos des datasets Word2Vec

<http://word2vec.googlecode.com/svn/trunk>

- ▶ 5031 analogies impliquant la relation **capitale\_of**, mais seulement 118 paires de pays / capitale ...
- ▶ 870 analogies impliquant la relation **plural-verbs**, mais seulement 30 paires de verbes ...
  - ▶ 27 avec la relation +s ; ex : **eat / eats**
  - ▶ et seulement 3 avec la relation +es : **go / goes**, **search / searches** et **vanish / vanishes**
  - ▶ **note** : [go: goes :: eat : eats] n'est pas une analogie formelle
- ▶ 1560 analogies impliquant la relation **past-tense**, mais seulement 40 paires de verbes ...
  - ▶ ex : **decreasing / decreased**, **knowing / knew**
  - ▶ dont la moitié (23) sont des verbes irréguliers ; ex : **sitted / sat** et **sleped / slept**

# MSR Sentence Completion dataset

<http://research.microsoft.com/en-us/projects/scc>

► 1040 questions comme celles-ci :

26 It may be that the solution of the one may \_\_\_\_\_ to be the solution of the other.

- a) prove
- b) learn
- c) choose
- d) forget
- e) succumb

27 We shall just be in time to have a little \_\_\_\_\_ with him.

- a) breakfast
- b) elegance
- c) garment
- d) basket
- e) dog

# MSR Sentence Completion dataset

<http://research.microsoft.com/en-us/projects/scc>

► 1040 questions comme celles-ci :

26 It may be that the solution of the one may **prove** to be the solution of the other.

- a) prove
- b) learn
- c) choose
- d) forget
- e) succumb

27 We shall just be in time to have a little **breakfast** with him.

- a) breakfast
- b) elegance
- c) garment
- d) basket
- e) dog

# BLESS

<https://sites.google.com/site/geometricalmodels/shared-evaluation>

- ▶ 26 554 lignes comme celle-ci :

```
frog-n amphibian_reptile attri green-j
```

qui indique qu'une grenouille est un amphibien ou un reptile et qu'un de ses attributs est d'être verte (gasp).

- ▶ 200 concepts (reptiles, mammifères, oiseaux, fruits, insects, armes, etc.) , 5 relations (attribut, co-hyponyme, hyperonyme, meronyme, événement)

*On average, there are 14 attri, 18 coord, 19 event, 7 hyper and 15 mero relata per concept.*

- ▶ 14 400 concept-relation-relatum + 12 154 control tuples (random)
- ▶ une tâche suggérée



# BLESS/frog

<https://sites.google.com/site/geometricalmodels/shared-evaluation>

mero	blood-n bone-n eye-n foot-n gill-n head-n leg-n lung-n poison-n skin-n tongue-n wart-n
event	breathe-v catch-v chase-v croak-v die-v drink-v eat-v hop-v kill-v lay-v leap-v live-v sit-v swim-v
attri	amphibious-j aquatic-j brown-j colorful-j edible-j green-j little-j loud-j noisy-j old-j poisonous-j small-j ugly-j wild-j
coord	alligator-n bullfrog-n lizard-n snake-n toad-n turtle-n
hyper	amphibian-n animal-n beast-n chordate-n creature-n vertebrate-n

+ 50 relations aléatoires

frog-n amphibian\_reptile random-n doyen-n



# SemEval 2010

## Cause-Effect (1003)

He had chest pains and `<e1>headaches</e1>` from `<e2>mold</e2>` in the bedrooms.

## Component-Whole (941)

The `<e1>provinces</e1>` are divided into `<e2>counties</e2>` (shahrestan), and ...

## Entity-Destination (845)

The prosecutors carried `<e1>guns</e1>` into the `<e2>court room</e2>`.

## Product-Producer (717)

Next the `<e1>soldiers</e1>` put up a `<e2>wall</e2>` of stakes on the pile of dirt.

- + Entity-Origin (716), Member-Collection (690), Message-Topic (634), Content-Container (540), Instrument-Agency (504)

- ▶ **Tâche** : identifier la relation de 2717 paires de noms marquées dans des phrases

# SemEval 2012 (pilot) / 2013 (main task)

- ▶ sur une échelle de 0 à 5, noter la similarité de paires de phrases :

MT	When we are faced with a potential risk, it is important to apply the precautionary principle.
5	When we are faced with a potential risk, it is important to put into practice the precautionary principle.

MT	This rule can be applied equally to all recreational but dangerous drugs when no one but the person consuming the drug is likely to be affected.
4.8	This rule can, moreover, apply to all drugs <span style="color: lightblue;">récréationnelles</span> but dangerous provided that the consumer of drugs is the only run the risks.

MSR-VID	A man stares out a window.
4.8	A man looks out the window.

MSR-PAR	But at age 15, she had reached 260 pounds and a difficult decision : It was time to try surgery.
4.2	But at the age of 15, she weighed a whopping 117kg and came to a difficult decision : it was time to try surgery.

## SemEval 2014 / Task3

## ► Phrase2Word (TRAIN : 500, TEST : 500) — échelle [0, 4]

4	the structural and functional unit of all known living organisms	DNA	Lexicographic
4	traditionally the land on which a college or university and related institutional buildings are situate	campus	Lexicographic
4	watering hole	pub	Idiomatic
4	wear out one's welcome	exhaust	Idiomatic
4	whole ball of wax	total	Idiomatic
	:		
0	to make from scratch	breath	Idiomatic
0	what a vivid imagination	insipid	Newswire
0	workouts to strengthen abdominal muscles	fractions	Search

# SemEval 2014 / Task3

► Sentence2Phrase (TRAIN : 500, TEST : 500) — échelle [0, 4]

4	We completely agree and there's no daylight between us on the issue.	complete agreement	Idiomatic
4	do you now where i can watch free older movies online without download ?	streaming vintage movies for free	CQA

► Paragraph2Sentence (TRAIN : 500, TEST : 500) — échelle [0, 4]

4	so i bought an item a couple weeks ago on eBay and i still havent even been notified that my item has been shipped yet.. what do i do ?	What should I do when my eBay purchase still hasn't shipped after two weeks ?	CQA
---	---	---	-----

# Plan

Sémantique

Question-réponse

ARC

RACE

WDW

# Quelques jeux de test pour le QR (QA)

- ▶ SQUAD
- ▶ MS Marco
- ▶ NewsQA
- ▶ MC Test
- ▶ Cloze-style
- ▶ RACE
- ▶ CNN/Daily
- ▶ WDW (What did what)
- ▶ AI2 Elementary school Science questions
- ▶ triviaQA
- ▶ WikiHop



# ARC (AI2 Reasoning Challenge)

*[Clark et al., 2018]*

- ▶ 7787 questions à choix multiples en sciences naturelles :  
Easy set : 5197 questions  
Challenge set : 2590 questions qui résistent à des baselines basés sur la RI ou la co-occurrence
- ▶ questions pour étudiants de 8 à 13 ans

1 *Which property of a mineral can be determined just by looking at it?*  
(A) luster (B) mass (C) weight (D) hardness

**note** : il est dit qu'il n'y a pas sur le web de phrase comme  
luster can be determined by looking at





# What is saying Google about this ?



luster can be determined by looking at



All

Images

News

Shopping

Videos

More

Settings

Tools

About 2,370,000 results (0.53 seconds)

## What is Luster? - Definition, Types & Examples - Video & Lesson ...

<https://study.com/academy/lesson/what-is-luster-definition-types-examples.html> ▾

It is one of the properties mineralogists look at when trying to determine the identity ... The greasy type of luster can be found in minerals that look like they were ...

### People also ask

How do you find the luster of a rock? ▾

What is the luster of metals? ▾

How do you test for Luster? ▾

What is shiny luster called? ▾

*Feedback*

# ARC (AI2 Reasoning Challenge)

*[Clark et al., 2018]*

- ▶ 7787 questions à choix multiples en sciences naturelles :  
Easy set : 5197 questions  
Challenge set : 2590 questions qui résistent à des baselines basés sur la RI ou la co-occurrence
- ▶ questions pour étudiants de 8 à 13 ans

- 1 *Which property of a mineral can be determined just by looking at it?*  
(A) luster (B) mass (C) weight (D) hardness
- 2 *A student riding a bicycle observes that it moves faster on a smooth road than on a rough road. This happens because the smooth road has*  
(A) less gravity (B) more gravity (C) less friction  
(D) more friction



# ARC distribué avec

- ▶ 14M de phrases reliées (couvre 95% des connaissances requises pour les questions)
- ▶ 3 implémentations de réseaux profonds (top performers sur d'autres tâches de QA)
- ▶ <http://data.allenai.org/arc>



# Performances

Solver	Test Scores	
	Challenge Set	Easy Set
IR (dataset defn)	(1.02) <sup>†</sup>	(74.48) <sup>†</sup>
PMI (dataset defn)	(2.03) <sup>†</sup>	(77.82) <sup>†</sup>
IR (using ARC Corpus)	20.26	62.55
TupleInference	23.83	60.81
DecompAttn <sup>‡</sup>	24.34	58.27
Guess-all (“random”)	25.02	25.02
DGEM-OpenIE <sup>‡</sup>	26.41	57.45
BiDAF <sup>‡</sup>	26.54	50.11
TableLP	26.97	36.15
DGEM	27.11	58.97

<sup>†</sup>These solvers were used to define the dataset, affecting scores.

<sup>‡</sup>Code available at <https://github.com/allenai/arc-solvers>

Table 6: Performance of the different baseline systems. Scores are reported as percentages on the test sets. For up-to-date results, see the ARC leaderboard at <http://data.allenai.org/arc>.

- ▶ 1 pour une unique bonne réponse
- ▶ 1/k pour k réponses dont la bonne

# RACE [*Lai et al., 2017*]

- ▶ 28k passages, 100k questions posées par des enseignants de l'anglais langue seconde à des étudiants Chinois âgés de 12 à 18 ans
- ▶ deux niveaux : RACE-M (12-15 ans), RACE-H (15-18)
- ▶ vocabulaire restreint (apprenants de l'anglais)
- ▶ dataset : <http://www.cs.cmu.edu/~glail/data/race/>
- ▶ code : [https://github.com/qizhex/RACE\\_AR\\_baselines](https://github.com/qizhex/RACE_AR_baselines)
- ▶ voir aussi : MCTest [*Richardson et al., 2013*] (histoires fictives)



**Passage:**

In a small village in England about 150 years ago, a mail coach was standing on the street. It didn't come to that village often. People had to pay a lot to get a letter. The person who sent the letter didn't have to pay the postage, while the receiver had to. "Here's a letter for Miss Alice Brown," said the mailman.

"I'm Alice Brown," a girl of about 18 said in a low voice.

Alice looked at the envelope for a minute, and then handed it back to the mailman.

"I'm sorry I can't take it, I don't have enough money to pay it", she said.

A gentleman standing around were very sorry for her. Then he came up and paid the postage for her.

When the gentleman gave the letter to her, she said with a smile, " Thank you very much, This letter is from Tom. I'm going to marry him. He went to London to look for work. I've waited a long time for this letter, but now I don't need it, there is nothing in it."

"Really? How do you know that?" the gentleman said in surprise.

"He told me that he would put some signs on the envelope. Look, sir, this cross in the corner means that he is well and this circle means he has found work. That's good news."

The gentleman was Sir Rowland Hill. He didn't forgot Alice and her letter.

"The postage to be paid by the receiver has to be changed," he said to himself and had a good plan.

"The postage has to be much lower, what about a penny? And the person who sends the letter pays the postage. He has to buy a stamp and put it on the envelope," he said . The government accepted his plan. Then the first stamp was put out in 1840. It was called the "Penny Black". It had a picture of the Queen on it.

**Questions:**

- |   |   |
|---|---|
| <p>1): The first postage stamp was made ...<br/>A. in England B. in America C. by Alice D. in 1910</p> <p>2): The girl handed the letter back to the mailman because ...<br/>A. she didn't know whose letter it was<br/>B. she had no money to pay the postage<br/>C. she received the letter but she didn't want to open it<br/>D. she had already known what was written in the letter</p> <p>3): We can know from Alice's words that ...<br/>A. Tom had told her what the signs meant before leaving<br/>B. Alice was clever and could guess the meaning of the signs<br/>C. Alice had put the signs on the envelope herself<br/>D. Tom had put the signs as Alice had told him to</p> | <p>4): The idea of using stamps was thought of by ...<br/>A. the government<br/>B. Sir Rowland Hill<br/>C. Alice Brown<br/>D. Tom</p> <p>5): From the passage we know the high postage made ...<br/>A. people never send each other letters<br/>B. lovers almost lose every touch with each other<br/>C. people try their best to avoid paying it<br/>D. receivers refuse to pay the coming letters</p> <p><b>Answer:</b> ADABC</p> |
|---|---|

Table 1: Sample reading comprehension problems from our dataset.

# Performances

	RACE-M	RACE-H	RACE
Random	24.6	25.0	24.9
Sliding Window	37.3	30.4	32.2
Stanford AR	44.2	43.0	43.3
GA	43.7	44.2	44.1
Turkers	85.1	69.4	73.3
Ceiling Performance	95.4	94.2	94.5

Table 5: Accuracy of models and human on the each d previous publications. DM denotes Daily Mail and WD

SW tf-idf de  $\text{conc}(q, a)$  et d'un span candidat

AR attentive reader (neuronal)

*[Chen et al., 2016]*

GA gated-attention reader (neuronal)

*[Dhingra et al., 2017]*

CPerf % de questions avec réponses atteignables (échantillon de 500 questions)

# WDW [Onishi et al., 2016]

- ▶ 200k questions à trou (cloze) à choix multiples construites sur le Gigaword newswire corpus
- ▶ processus pour sélectionner des paires  $(q, p, c)$ 
  - 1 choisir  $q$  la première phrase d'une nouvelle
  - 2 considérer ses entités nommées de  $q$ , une sera retirée
  - 3 identifier une autre nouvelle  $p$  "proche" via Lucene (question :  $q$ )
  - 4 identifier les entités nommées présentes dans  $p$  comme réponse potentielle ( $c$ )
- ▶ plus filtrage, retrait de questions trop faciles
- ▶ corpus, leaderbord,  
[https://tticnlp.github.io/who\\_did\\_what/](https://tticnlp.github.io/who_did_what/)





## WDW [*Onishi et al., 2016*]

**passage** : Britain's decision on Thursday to drop extradition proceedings against Gen. Augusto Pinochet and allow him to return to Chile is understandably frustrating ... Jack Straw, the home secretary, said the 84-year-old former dictator's ability to understand the charges against him and to direct his defense had been seriously impaired by a series of strokes. ... Chile's president-elect, Ricardo Lagos, has wisely pledged to let justice run its course. But the outgoing government of President Eduardo Frei is pushing a constitutional reform that would allow Pinochet to step down from the Senate and retain parliamentary immunity from prosecution. ...

**question** : Sources close to the presidential palace said that Fujimori declined at the last moment to leave the country and instead he will send a high level delegation to the ceremony , at which Chilean President Eduardo Frei will pass the mandate to **XXX**.

**choix** : (1) Augusto Pinochet (2) Jack Straw (3) Ricardo Lagos



# WDW [Onishi et al., 2016]

System	WDW	CNN
Word overlap	0.47	–
Sliding window	0.48	–
Distance	0.46	–
Sliding window + Distance	0.51	–
Semantic features	0.52	–
Attentive Reader	0.53	0.63 <sup>I</sup>
Attentive Reader (relaxed train)	0.55	
Stanford Reader	0.64	0.73 <sup>II</sup>
Stanford Reader (relaxed train)	0.65	
AS Reader	0.57	0.70 <sup>III</sup>
AS Reader (relaxed train)	0.59	
GA Reader	0.57	0.74 <sup>IV</sup>
GA Reader (relaxed train)	0.60	
Human Performance	84/100	0.75+ <sup>II</sup>

**Table 4:** System performance on test set. Human performance was computed by two annotators on a sample of 100 questions. Result marked *I* is from (Hermann et al., 2015), results marked

*II* are from (Chen et al., 2016) Tâches à saveur sémantiques



**Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Paşca, M., and Soroa, A. (2009).**

A study on similarity and relatedness using distributional and wordnet-based approaches.

In *Proceedings of Human Language Technologies : The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, pages 19–27.



**Chen, D., Bolton, J., and Manning, C. D. (2016).**

A thorough examination of the cnn/daily mail reading comprehension task.

In *54th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, pages 2358–2367.



**Clark, P., Cowhey, I., Etzioni, O., Khot, T., Sabharwal, A., Schoenick, C., and Tafjord, O. (2018).**

Think you have solved question answering ? try arc, the AI2 reasoning challenge.

*CoRR.*



**Dhingra, B., Liu, H., Yang, Z., Cohen, W., and Salakhutdinov, R. (2017).**

Gated-attention readers for text comprehension.

*In 55th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, pages 1832–1846.

Association for Computational Linguistics.



**Freitag, D., Blume, M., Byrnes, J., Chow, E., Kapadia, S., Rohwer, R., and Wang, Z. (2005).**

New experiments in distributional representations of synonymy.

*In Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 25–32.



**Lai, G., Xie, Q., Liu, H., Yang, Y., and Hovy, E. (2017).**

Race : Large-scale reading comprehension dataset from examinations.

*In Conference on Empirical Methods in Natural Language Processing*, pages 785–794. Association for Computational Linguistics.



**Onishi, T., Wang, H., Bansal, M., Gimpel, K., and McAllester, D. A. (2016).**

Who did what : A large-scale person-centered cloze dataset.  
*CoRR*.



**Richardson, M., Burges, C. J., and Renshaw, E. (2013).**

Mctest : A challenge dataset for the open-domain machine comprehension of text.  
In *Conference on Empirical Methods in Natural Language Processing*, pages 193–203.