

Is China Winning the *Trade War*?



<https://www.newyorker.com/news/our-columnists/why-china-is-winning-the-trade-war>

Is China Winning the Trade War?



<https://www.newyorker.com/news/our-columnists/why-china-is-winning-the-trade-war>

Distributed Representation of Chinese Collocation

Bo Xia and Endong Xun

Introduction

- What is collocation?
- Why collocation is meaningful for us?
- How to represent a collocation?

What is collocation?

- English collocation

What is collocation?

- English collocation



Unit 60 The future



Vocabulary

New words

Activity

Word list

Practice

Section

W

• Er

intend to do sth.

意图做某事

hurry towards sb.

向某人匆匆走去

a relation of yours

您的一个亲戚（双重所有格）

the moment = as soon as

——就——



Wh



Unit 4

Wildlife Protection



- decrease to
- die out
- in peace
- burst into laughter
- appreciate doing sth
- before long
- for sure
- in relief
- according to
- have an effect on
- in danger of
- protect A from B
- long before
- come into being
- so that
- pay attention to
- in such a short time
- succeed in doing sth

What is collocation?

- English collocation
- Chinese collocation

Chinese collocation

✓ 露一手

✓ 【Annotation】

to make an exhibition of one's forte, speciality, ability or skill

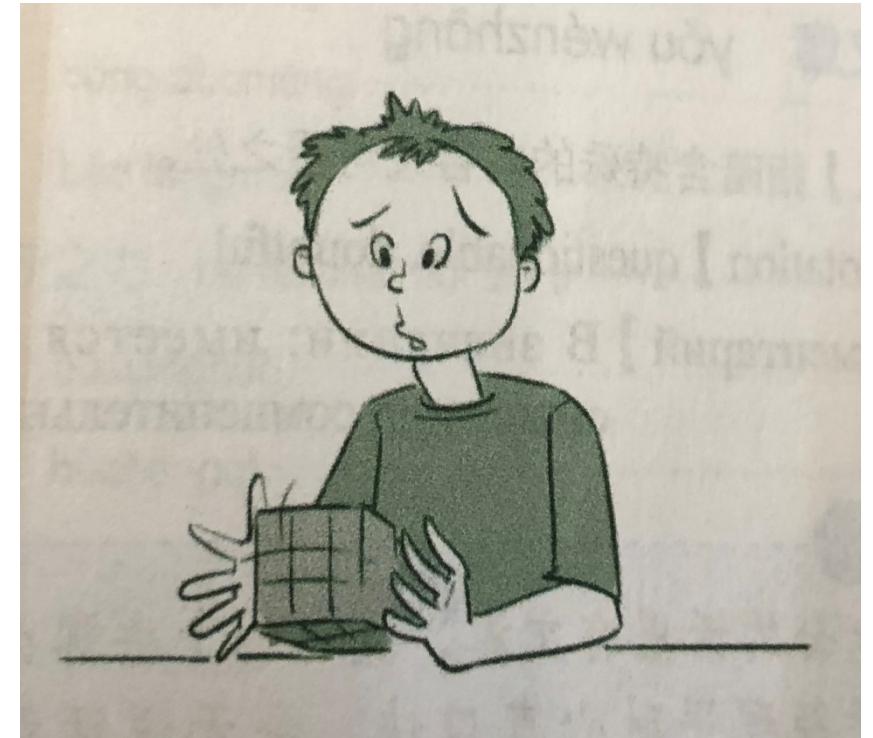


Chinese collocation

✓ 伤脑筋

✓ 【Annotation】

Knotty, troublesome, causing sb.
enough headache

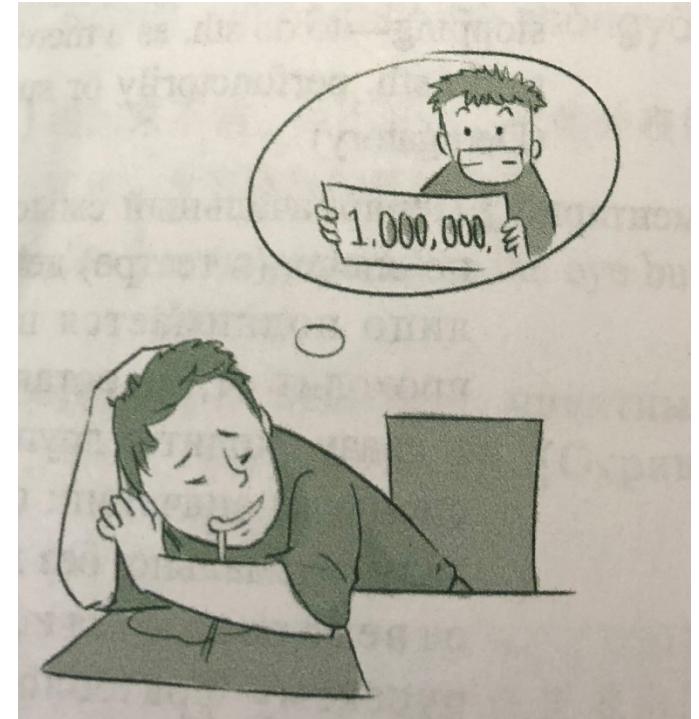


Chinese collocation

✓ 做白日梦

✓ 【Annotation】

To build castles in the air, to
daydream(Ironic)

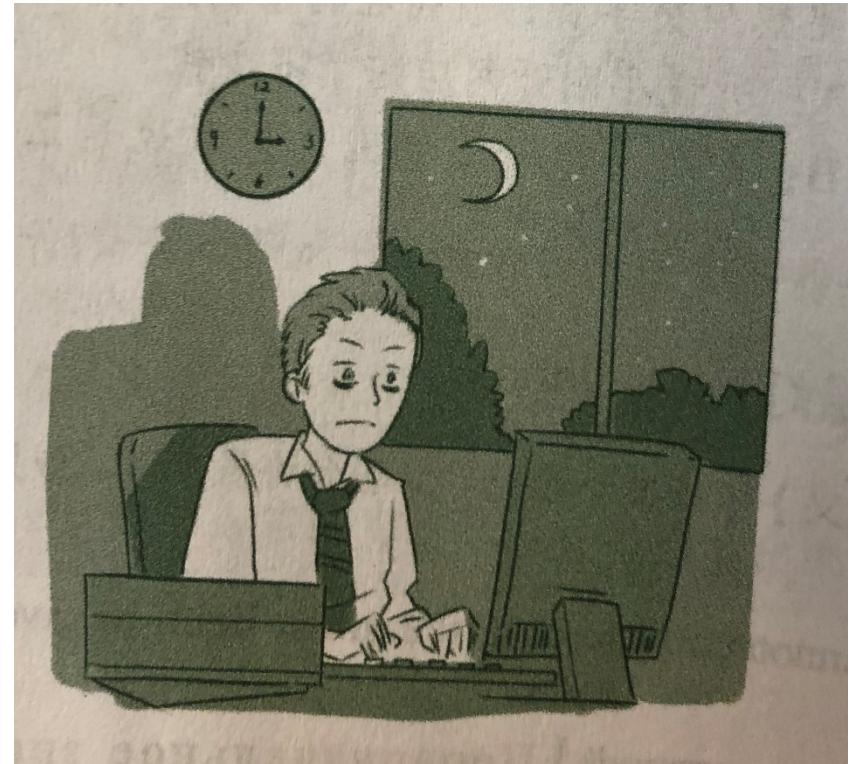


Chinese collocation

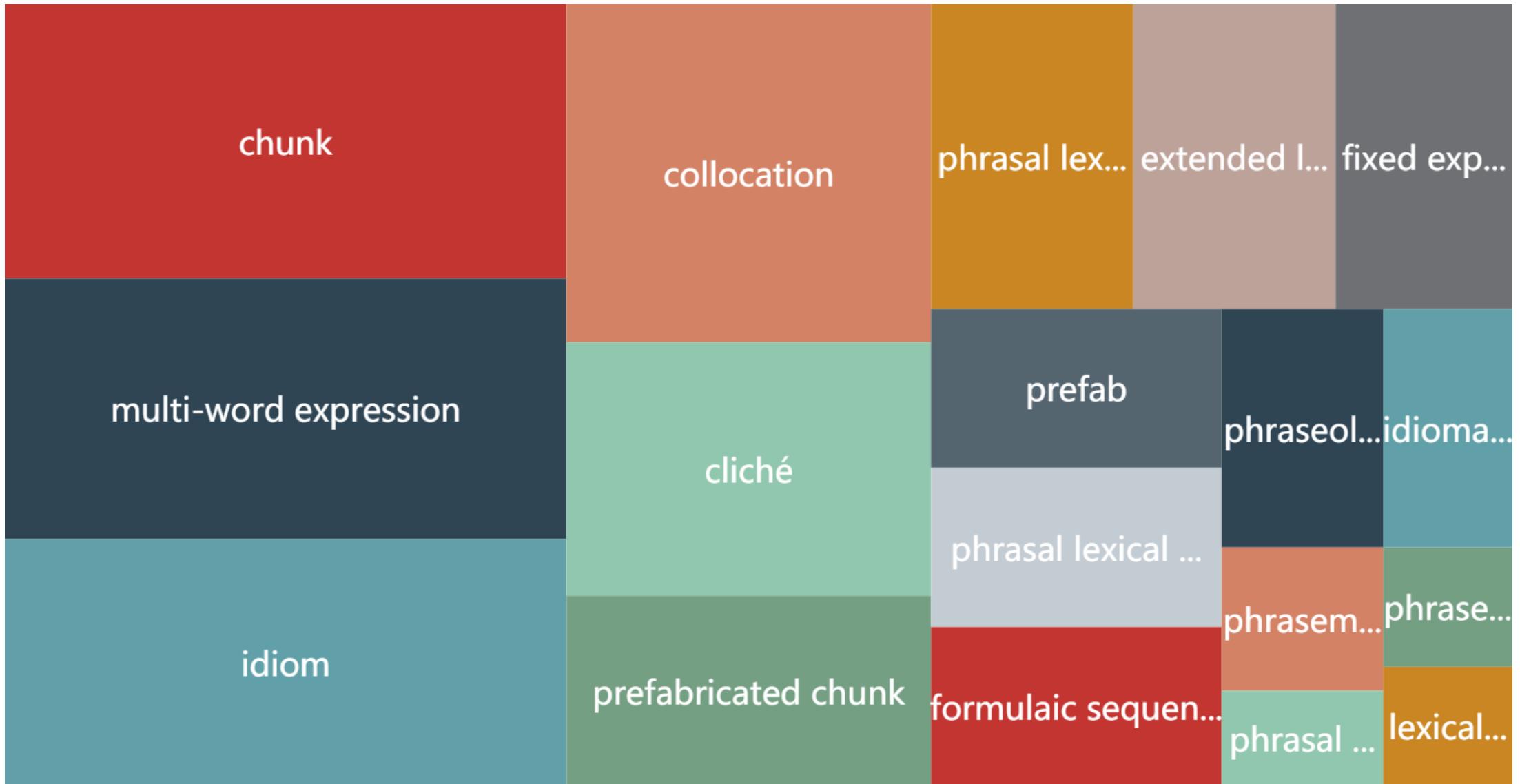
✓ 开夜车

✓ 【Annotation】

To drive at night – to work late into the night or by candlelight, to put in extra time at night, to burn the midnight oil.



What is collocation?



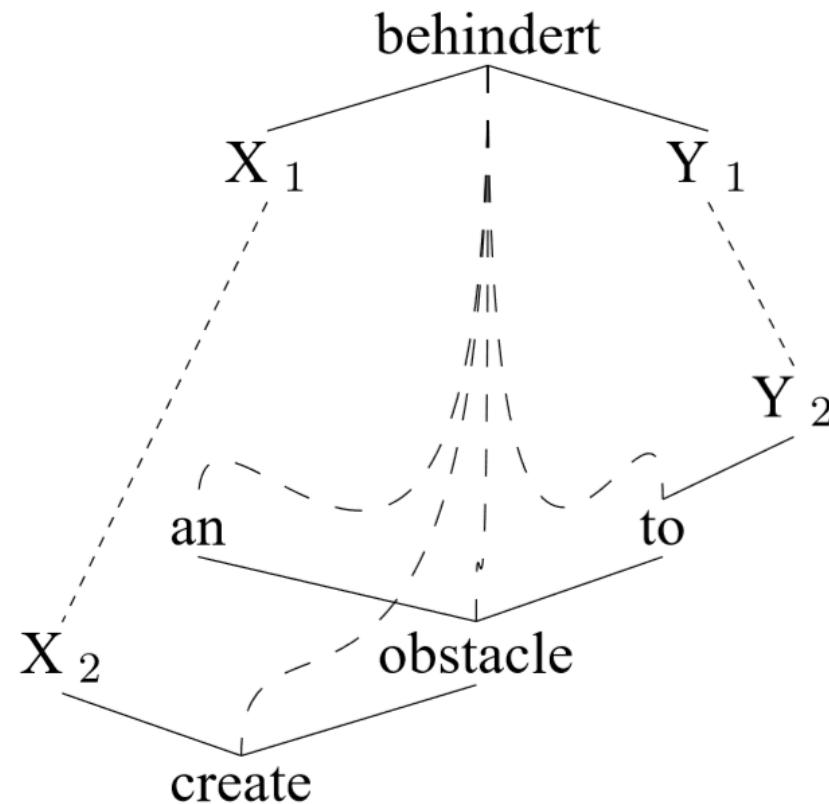
Why collocation is meaningful for us?

- Important role of language phenomenon in syntactic and semantic,
- But need to use it well.
- e.g.
- **look up**
- **bus stop**

How to represent a collection?

- Lexicon
- Table
- Triple
- Rules, Features or labels
- e.g.
- < ‘make’ , ‘decision’ >
- < ‘make’ , ‘decision’ , ‘verb–object’ >

Example of a typical syntactic MWE configuration



A conceptual lexicon

- Development of a lexicon
- of the encoding of MWEs
- and their treatments.

Slot	Values
mwe_type	<i>Fixed; collocation</i>
Lexical_variance	<i>Boolean (yes, no)</i>
Variants	<i>string</i>
Is_actually	<i>Boolean (yes, no)</i>
maintains_meaning	<i>String</i>
has_meta_meaning	<i>String</i>

Example of encoding schema for nouns & verbs

Basic Notions

- A Neural Probabilistic Language Model
- Collocation
- NNLM + Collocation = ?

Basic Notions

- A Neural Probabilistic Language Model
- Collocation
- NNLM + C



Basic Notions

- A Neural Probabilistic Language Model
- Collocation
- NNLM + Collocation =
- Distributed Representation of Collocation

Basic Notions

“

By collocation,
we mean a directed association of **head word** and **dependency word**
that constructed by **fully random combination**
in a sentence.

”

Model Architecture

- ✓ The objective of Collocation prediction
- ✓ Continuous Skip-gram Model
- ✓ A combination of ontology and hierarchical softmax

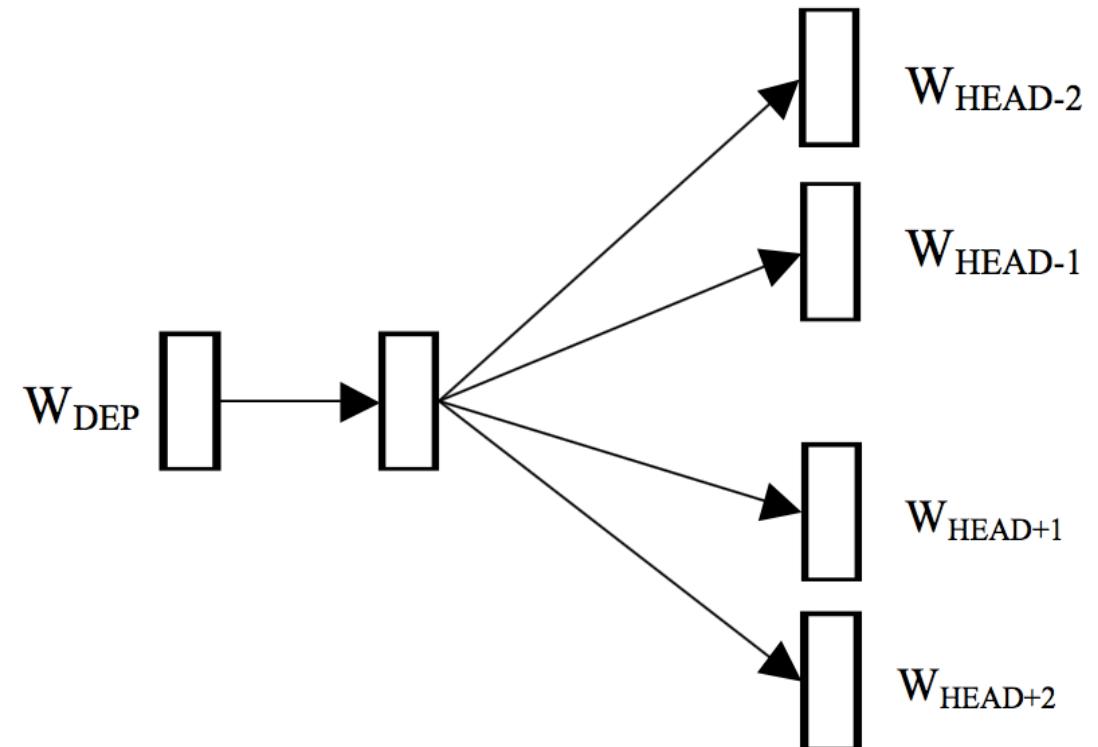
The objective of Collocation prediction

- $P(\text{Context}(\text{HEAD}) \mid \text{DEP})$



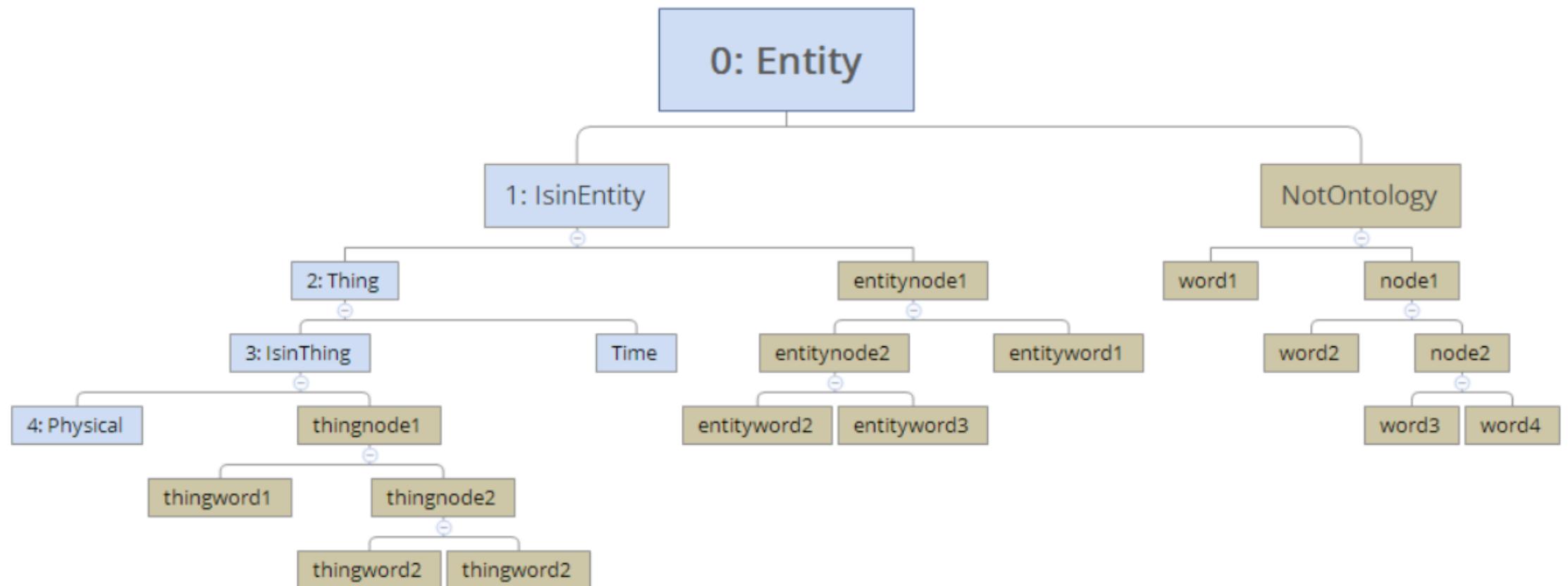
Continuous Skip-gram Model

- consider each DEP word representation as a bag of HEAD word contexts

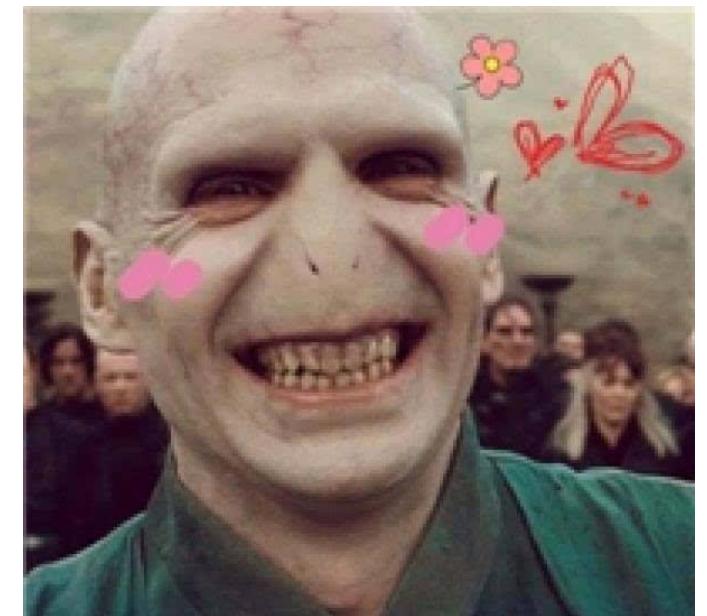


The DEP word predicts surrounding words given the HEAD word of current collocation.

A combination of ontology and hierarchical softmax



Experiment



Experimental setup

- 100,000,000 sentence, with 3,377,634,987 Chinese tokens of roughly 100,906 words.
- Averaged sentence length of test set is 14.79 by word

Neighbors of one collocation

主张	proposal	后空翻	backward somersault
诉诸武力	to appeal to arms	接特卡切夫	Tkachev (gymnasts)
一中一台	Taiwanization	直体	straight
国统	Ruling of KMT	屈体	pike
台独	Taiwanization	9.725	9.725

Measuring collocation intensity

因素				9.643	3.969	4.534	4.969	7.45	
关键		9.416	9.431	3.957	4.637	5.009			7.354
个	4.39	9.313	9.561	3.889	4.679		7.256		7.185
一	5.729	4.495	9.438	9.682	3.919		5.098	7.319	7.191
的	5.699	4.448	9.49	9.632		4.631	4.999	7.359	7.32
自杀	5.699	4.448	9.694		3.951	4.656	4.983	7.245	7.297
阻止	5.707	4.395		9.77	3.934	4.567	4.903	7.33	
是	5.695		9.439	9.663	3.944	4.688	5.003		
时间		3.904	9.503	9.574	3.899	4.644			

Verb-verb phrase recognition

- 他/r 不/d 愿意/v 离婚/v , /w 因为/c 这样/r 以后/f 他/r 也/d 不/d 可能/v 幸福/a 。/w
- < 愿意, 离婚 >
- 以后 2.392401
- 离婚 0.986100
- 因为 0.710415
- 这样 0.651951
- 他 0.310175
- 不 0.273814

Verb-verb phrase recognition

- 昨晚/t， /w 股权/n 分置/v 改革/vn 试点/vn 工作/vn 宣布/v 启动/v
- < 宣布, 启动 >
- 启动 5.668315
- 试点 3.171463
- 改革 0.476158
- 工作 0.409356
- 分置 0.313457

Verb-verb phrase recognition

- 她/r 心里/s 拒绝/v 他/r 的/u 来到/v , /w 因为/p 他/r 会/v 挤/v 走/v 现在/t 的/u 父亲/n 。/w
- < 拒绝, 来到 >
- 心里 2.190928
- 来到 0.870547
- 因为 0.569144
- 她 0.312641
- 他 0.255559
- 的 0.224716

Verb-verb phrase recognition

collocation-pretreatment	maximal verbs	if bi-directional	nCorrect	nTotal	Accuracy
TRUE	3	TRUE	129	161	0.80124
TRUE	4	TRUE	78	109	0.71559
TRUE	5	TRUE	38	59	0.64406
TRUE	3	FALSE	127	161	0.78881
TRUE	4	FALSE	76	109	0.69724
TRUE	5	FALSE	36	59	0.61016
FALSE	3	TRUE	143	185	0.77297
FALSE	4	TRUE	89	127	0.70078
FALSE	5	TRUE	42	70	0.60000
FALSE	3	FALSE	156	185	0.84324
FALSE	4	FALSE	103	127	0.81102
FALSE	5	FALSE	54	70	0.77142

Conclusion and future work

- How it works in English Collocation ?
- Does triple collocation feature can be learned?

Distributed Representation of Chinese Collocation

- Bo Xia and Endong Xun
- *Beijing language and culture university*
- https://github.com/xiabo0816/Chinese_MWEs_Representation

Distributed Representation of Chinese Collocation

Bo Xia and Endong Xun
Beijing language and culture university