

単語分散表現から類義語を考える

内田 諭

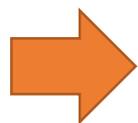
九州大学言語文化研究院



自己紹介

- 内田 諭 (うちだ さとる)
- 九州大学准教授
- 専門：コーパス言語学、フレーム意味論、英語教育学

フレーム意味論



言語資源

英語教育



自然言語処理

Commerce_buy

[Lexical Unit Index](#)

Definition:

These are words describing a basic commercial transaction involving a **Buyer** and a **Seller** exchanging **Money** and **Goods**, taking the perspective of the **Buyer**. The words vary individually in the patterns of frame element realization they allow. For example, the typical pattern for the verb BUY: **Buyer** buys **Goods** from **Seller** for **Money**.

Abby bought a car from Robin for \$5,000.

FEs:

Core:

The **Buyer** [Byr] wants the **Goods** and offers **Money** to a **Seller** in exchange for them.
Jess BOUGHT a coat.

Lee BOUGHT a textbook from Abby.

Goods [Gds]

The FE **Goods** is anything (including labor or time, for example) which is exchanged for **Money** in a transaction.
Only one winner PURCHASED the paintings

CVLA: CEFR-based Vocabulary Level Analyzer (ver. 2.0)

[Legend]

A1: example, A2: example, B1: example, B2: example, C1: example, C2: example, NA content words: example, NA others: example
#You can sort the table by clicking the table header.

ID	Sentence	Words	Verbs	AvrDiff
1	Writing is the act of recording language on a visual medium using a set of symbols .	16	3	2.0
2	The symbols must be known to others , so that the text may be read .	14	4	1.6
3	A text may also use other visual systems , such as illustrations and decorations .	13	1	2.22
4	These are not called writing , but may help the message work .	11	3	1.33
5	Usually , all educated people in a country use the same writing system to record the same language .	17	3	1.55
6	To be able to read and write is to be literate .	11	5	1.67

CEFR	ARI	VperSent	AvrDiff	BperA
A1	5.73	1.49	1.31	0.08
A2	7.03	1.82	1.41	0.12
B1	10.00	2.37	1.57	0.18
B2	12.33	2.88	1.71	0.26
Input	5.39	3.17	1.77	0.19
Estimated level	A1.2	C1	B2.2	B1.1

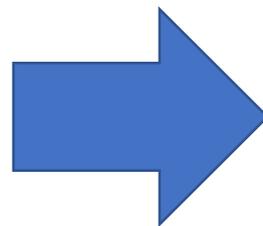
Mode: R, Estimated Text Level: B1.2

<https://framenet.icsi.berkeley.edu/fndrupal/>

<https://cvla.langedu.jp/>

類義語の分析方法

- 内省
- インフォーマント調査
- コーパス調査
- 計算 **【NEW!】**



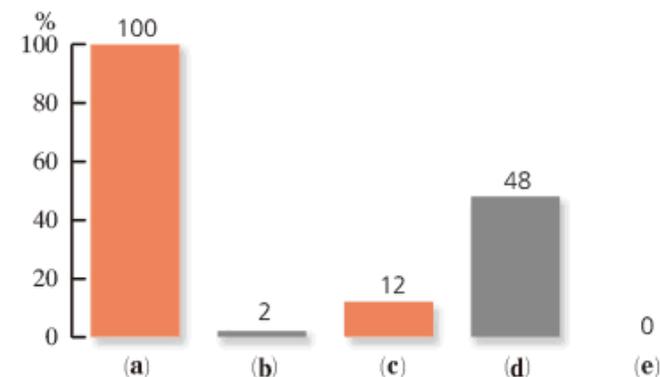
PLANET BOARD
76

「…高校の先生」を表すのに用いる前置詞は何か。

問題設定 「…高校の先生」という場合の teacher に続く前置詞に何をを用いるかを調査した。

次の (a) ~ (d) のどれを使いますか。(複数回答可)

- (a) He is a teacher **at** Asahi High School.
- (b) He is a teacher **of** Asahi High School.
- (c) He is a teacher **in** Asahi High School.
- (d) He is a teacher **from** Asahi High School.
- (e) どれも使わない



全員が (a) の at を用いるとしている。(d) の from を使う人は約半数いる一方、(b) の of と (c) の in の使用率は非常に低い。大多数が「『…高校に勤めている』という意味では at がふつう」と答えた。「from を用いるのは、その人を勤務校以外の場所でだれかに紹介する場合や、過去に…高校で働いていたが現在は別の学校で教師をしている場合、あるいは「…高校の出身である」と言いたい場合などである」とするコメントが目立った。teacher の代わりに student を用いた場合についても調査したが、同様の結果だった。

学習者への指針 ある人の職業について「…高校の先生」という場合は、a teacher at ... がふつう。of や in は使わない。from を用いる場合、「その人の(現在または過去における)所属先が…高校である」ことに重点が置かれる。

類義語の先行研究

言語分析、言語教育（日英）ともに
多くの研究がある

言語分析

- Fillmore, C. J., & Atkins, B. T. (1992). Toward a frame-based lexicon: The semantics of RISK and its neighbors. *Frames, fields and contrasts: New essays in semantic and lexical organization*, 75, 102.
- Gries, S. T., & Divjak, D. (2009). Behavioral profiles: a corpus-based approach to cognitive semantic analysis. *New directions in cognitive linguistics*, 57, 75.
- 杉本武 (2009). コーパスからみた類義語動詞:「ねじる」と「ひねる」. 文藝言語研究. 言語篇, 55, 109-122.
- 柴宝華, & 趙海城. (2017). コーパスに基づいた類義語分析:「見落とす」「見過ごす」「見逃す」を例に. 明星大学研究紀要人文学部, (53), 27-46.

言語教育

- Martin, M. (1984). Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal*, 68(2), 130-137.
- Laufer, B. (1990). Ease and difficulty in vocabulary learning: Some teaching implications. *Foreign Language Annals*, 23(2), 147-155.
- 岩佐靖夫. (2011). 日本語教育における類義語指導の一考察: 系統的な指導原理へ向けての提言. 尚美学園大学総合政策研究紀要, 20, 17-24.
- 大神智春. (2018). 類義語で形成されるコロケーションの理解と使い分け—中国語母語話者を対象として—. 日本語教育方法研究会誌, 24(2), 70-71.

言語資源

- **電子リソース**
- FrameNet
- WordNet
- VerbNetなど

今回はここにフォーカス

※近日中にフレーム意味論・FrameNetの入門書（研究社）が出ます！

- **紙媒体のリソース**
- Roget's International Thesaurus
- Oxford Dictionary and Thesaurus
- 日本語シソーラス類語検索辞典など

【本発表の目的】

FrameNetの類義語の体系を使って「計算」（BERT）から類義語を検証する
→類義語の「近さ」が、その種類によってどう異なるか？

FrameNetとは？

- 単語が喚起する言語フレームの辞書(<https://framenet.icsi.berkeley.edu/fndrupal/>)

例：Lee bought a textbook from Abby.

- bought (buy)はCommerce_buyフレームを喚起
- Commerce_buyフレームの上位フレームはGettingフレーム
- LeeはBuyer, a textbookはGoods要素でコアフレーム要素
- AbbyはSellerでこのフレームでは周縁的フレーム要素

→ 単語が喚起するフレーム、そのフレームの必須要素、フレーム要素の実現形と組み合わせ、**フレーム同士の関係**を記録

FrameNetにおける類義語の種類: フレーム間関係に基づいて

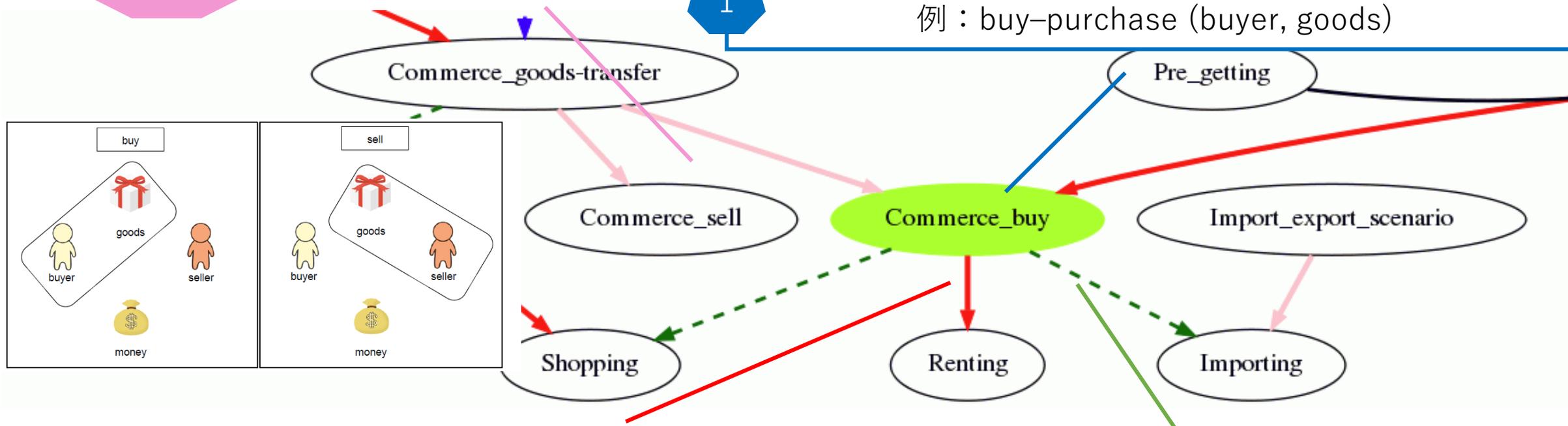
perspective_on (視点化) :
特定の事象を別視点から描写する
例: buy-sell

3 or 4

※数字は類義語の関連度の高さの直感的な順位

1

同一フレームを喚起: 共通するコアフレーム要素を持つ
例: buy-purchase (buyer, goods)



2 inheritance (継承) : コアフレーム要素をすべて引き継ぐ
例: buy-rent (buyer[Lessee], goods要素が共通)
※反意語(increase-decreaseなど) を含む

3 or 4 use (使用) : コアフレーム要素の一部として
上位フレームのコア要素を使用
例: buy-import

実験の内容

- 同一フレームを喚起する語、inheritanceの関係の語、perspective_onの関係の語、Useの関係の語をそれぞれ5ペアずつ取り上げ（文例、動詞に限定）、OALD(<https://www.oxfordlearnersdictionaries.com/>)から用例を抽出(1,035件)
- 見出し語を[MASK]した例文をBERT(bert-base-uncased)を使ってTOP5のcandidateをリスト
- Target（見出し語）およびその類義語が含まれているかを検証

例：give-offer（useの関係）

I was asked to give evidence at the trial. → I was asked to [MASK] evidence at the trial.

give(0.497) , provide(0.279), present (0.089) , testify (0.013) , witness (0.009)

→見出し語はtop 1, 類義語はNA

実験結果1：見出し語の予測確率

	NA	Top 1	Top 2	Top 3	Top 4	Top 5	総計	
Inheritance	93	133	43	24	16	15	324	71.3
perspective_on	103	158	34	19	13	11	338	69.5
same	31	35	15	12	4	6	103	69.9
use	72	130	35	18	7	8	270	73.3
総計	299	456	127	73	40	40	1035	

- Targetとなった単語（見出し語）を正しく予測できた確率
- どの関係でも約70%の確率でTop 5に入る（Top 1で約42%）

対象語：claim-confirm (inheritance), consider-judge (inheritance), get-buy (inheritance), say-reveal (inheritance), vanish-evacuate (inheritance), buy-sell (perspective_on), charge-pay (perspective_on), employ-work (perspective_on), give-receive (perspective_on), succeed-fail (perspective_on), build-construct (same), complete-finish (same), cure-heal (same), increase-decrease (same), supply-provide (same), break-injure (use), cook-grill (use), destroy-shred (use), give-offer (use), say-promise (use)

実験結果2：類義語がTop 5に入る割合

	NA	Top 1	Top 2	Top 3	Top 4	Top 5	総計		
Inheritance	295	7	4	8	7	3	324		9.0
perspective_on	305	7	10	4	6	6	338		9.8
same	70	8	9	7	5	4	103		32.0
use	232	9	5	10	10	4	270		14.1
総計	902	31	28	29	28	17	1035		

- 同一フレームを喚起する場合に目立って類義語が多く含まれる

類義語が逆転する割合	件数	総数	割合
Inheritance	6	324	1.9
perspective_on	7	338	2.1
same	11	103	10.7
use	8	270	3.0

- 同一フレームで類義語が逆転する場合が多い
- 何が起きているか、**言語学的に興味深い**

類義語が逆転した具体例

target	syn	relation	example	candidate 1	candidate 2	candidate 3	candidate 4	candidate 5
charge	pay	perspective_on	He only charged me half price.	SVOO			charged	pays
construct	build	same	They constructed a shelter out of fallen branches.	[MAKE] O out of ~			dug	constructed
decrease	increase	same	The price of wheat has decreased by 5 per cent.	[CHANGE] by [RATE]			dropped	decreased
give	offer	use	They were all thirsty so I gave them a drink.	SVOO			handed	got
offer	give	use	They decided to offer the job to Jo.	SVO to O			hand	sell
sell	buy	perspective_on	I sold James my car for £800.	S [BUY-SELL] O for [MONEY]			offered	loaned
supply	provide	same	The company supplied schools with sports equipment.	S [PROVIDE] O with [ITEM]			supports	provided

- 構文的に中立なもの（他のものに入れ替え可能で汎用性があるもの）が立ち浮かんでくる
- **構文(construction)の発見**に寄与する可能性がある

まとめと課題

【まとめ】

- 「計算」による類義語の発見は言語学的には新しいもの
- その詳細はまだまだ未検討のところが多い
- FrameNetを類義語のリソースとして使ってBERTで実験したところ、「同一フレーム」を喚起する単語で、「見出し語と類義語」の逆転現象が観察された（→構文の発見につながる現象？）

【課題】

- 予備調査に留まる（約1,000例）
→言語学者だけだと大規模化が難しい
- Attentionの可視化等からキーとなるheadや構文的な振る舞いを観察したい
→言語学者だけだと技術やアイデアが限られる

NLPと言語学の協働・共創が必要！
どんなテーマでもよいのでご関心があればご連絡ください