

計算機による データに基づく言い換え

乾健太郎

東北大学 情報科学研究科

言い換えのあれこれ

A classification of paraphrases

Announcement

The classification and examples in this page are not continuously maintained (Last update was February 2007). If you need to get access to a newer version of classification, please refer to the one presented at the CBA workshop in December, 2010: [Typology of Paraphrases and Approaches to Compute Them](#).

はじめに

ひとくちに言い換えといっても、人間が実際に生成・認識している言い換えには様々な種類の現象がある。ここでは、そのうち、主に言語学的な知識に基づいてなされていると考えられる言い換え（語彙・構文的言い換え）の例を集め、(i)言い換えのスコップ、(ii)内容表現か機能表現か、(iii)必要な知識の種類の観点で分類・整理する。また、それぞれの言い換えを計算機上で実現するにあたっての課題を考察する。

なお、定義上は、「言い換え」=「同じ意味の言語表現」であるが、表現が異なるからには何らかの違いがあると考えられる。たとえば、能動文と受動文の対を考えてみよう。「誰が何をどうしたか」という命題部分の意味は同じと考えられるが、話者の視点やどの情報を強調するか、あるいはどの情報が新情報かといった情報構造の違いがある。また、[\[Inkpen and Hirst, 2006\]](#)で詳しく述べられているように、形式性や態度の違いが表現に現れる場合もある。したがって、今後は言い換えによって何が変わるのかという観点からも分類を進める予定である。

このページに掲載している言い換えの例は、今後の事例収集・整理において他の例に置き換えられたり、例番号を変更される可能性がある。引用なさる場合は、あらかじめご了承ください。

謝辞

ここで紹介した言い換えの例は、奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座（松本研究室）において2003年夏に実施された『言い換え百選』プロジェクト（メンバー：[乾健太郎氏](#)、[高橋哲朗氏](#)、[降幡建太郎氏](#)、[藤田篤](#)）の成果をベースに整理した。整理にあたっては、[佐藤理史氏](#)、[山本和英氏](#)にコメントを頂戴した。各氏に感謝の意を表す。

もくじ

節間の言い換え

[\[分裂文の言い換え\]](#) [\[文分割（連体節主節化）\]](#) [\[文分割（連用節・並列節の分割）\]](#) [\[接続表現の言い換え\]](#)

節内の言い換え

[\[否定表現の言い換え\]](#) [\[比較表現の言い換え\]](#) [\[態・使役の交替\]](#) [\[自他の動詞交替\]](#) [\[授受動詞の交替\]](#) [\[壁塗り／場所格交替\]](#) [\[湧出動詞の交替\]](#) [\[相互格の交替\]](#) [\[補文構文と格要素の交替\]](#) [\[可能動詞の言い換え\]](#) [\[授受の構文の言い換え\]](#) [\[修飾要素の交替（係り先の変更）\]](#) [\[数量詞の遊離\]](#)

文法カテゴリを変える言い換え

[\[名詞と動詞（機能動詞構文の言い換え）\]](#) [\[動詞と形容詞\]](#) [\[名詞と形容詞\]](#) [\[ナ形容詞とイ形容詞\]](#)

主辞交替

[\[名詞句\]](#) [\[句と節\]](#) [\[格と副詞句\]](#)

内容語の複合表現の言い換え

[\[複合名詞の分解・構成\]](#) [\[「AのB」⇔連体節\]](#) [\[複合動詞の分解・構成\]](#) [\[名詞接尾辞の着脱\]](#)

機能語／モダリティの言い換え

[\[機能語相当表現\]](#) [\[取り立て助詞の移動\]](#) [\[助詞による特徴づけの削除\]](#) [\[伝達のモダリティ\]](#) [\[敬語表現の言い換え\]](#) [\[文体の変換\]](#)

内容語句の言い換え

[\[名詞の言い換え\]](#) [\[動詞の言い換え\]](#) [\[形容詞の言い換え\]](#) [\[副詞の言い換え\]](#)

慣用表現の言い換え

[\[慣用句\]](#) [\[表記の揺れ／略語\]](#) [\[換喩\]](#)

日本語になくて他の言語にある言い換え

[\[いろいろとりまぜて\]](#)

（藤田篤，～2012） <http://paraphrasing.org/paraphrase.html>

節間の言い換え[±]

2つ以上の節にまたがる言い換えをここでは「節間の言い換え」と呼んでいる。主題が変化する場合は、それに適した名詞述語表現が必要になる。また、節間の修辭的關係を表す接続詞をあらためて選択しなければならない。このように、節間の言い換えでは、節間の順序や關係が変化するため、結束性の評価が必要になる。

分裂文の言い換え[±]

出典・参考文献：

[砂川, 1995] [Dras, 1999a] [吉見ら, 2000b]

英語の例：

- (1) a. **It was John who** wore his best suit to the dance last night.
b. **John** wore his best suit to the dance last night.
- (2) a. **What** Andrew wants most **is** to find a nice cover for his book.
b. Andrew wants most to find a nice cover for his book.

日本語の例：

- (3) a. 今週当選した**のは**、奈良県の男性**でした**.
b. 今週**は**、奈良県の男性**が**当選**しました**.
- (4) a. ジョン**が**ダンスのために一番良いスーツを着ていた **のは**昨夜だ.
b. 昨夜ジョン**は**ダンスのために一番良いスーツを着ていた.

(藤田篤, ~2012) <http://paraphrasing.org/paraphrase.html>

節内の言い換え[±]

主題交替や格交替など、操作の対象が節内で閉じている言い換えである。変換パターンのバリエーションはそれほど多くなく、人手で書き尽くせる程度のように見えるが、変換パターンによっては適用の可否が語に依存するため、その判断に必要な語彙知識をいかにして発見・構築するかが課題となる。また、視点のような対人関係的意味や主題／陳述構造のような文脈レベルの意味が変化するため、これを捉えるモデルを形式化する必要もある。

否定表現の言い換え[±]

出典・参考文献：

[[林ら, 1991](#)] [[宮島ら, 1995a](#)] [[近藤ら, 2001](#)] [[飯田ら, 2001](#)] [[徳永, 2002](#)]

日本語の例：

- (14) a. 返信**しない**と、申込みは取り消され**ます**.
 b. 返信**する**と、申込みは取り消され**ません**.
- (15) a. 日本料理は一度**しか**食べたことが**ない**.
 b. 日本料理は一度食べたことが**あるだけだ**.

文法カテゴリを変える言い換え[±]

語の文法カテゴリの変化が中心となっている言い換えである。文法カテゴリの差異に起因する語の文法的な違いをどのような言語デバイスを用いて担うかが、言い換えを正確に生成するための鍵となる。とくにニュアンスの違いやコロケーションの問題のため、語と語に派生関係があっても、実際の文脈では言い換えには使えない場合もある。

名詞と動詞（機能動詞構文の言い換え）[±]

出典・参考文献：

[Mel'cuk and Polgère, 1987] [奥, 1990] [村木, 1991] [Iordanskaja et al., 1991] [森田, 1994]
[Dras, 1999a] [大泉ら, 2003] [鍛冶ら, 2003a] [降幡ら, 2004] [鍛冶ら, 2004a] [大竹, 2005]
[藤田ら, 2006b]

英語の例：

- (50) a. Employment **showed a sharp decrease** in October.
b. Employment **decreased sharply** in October.
- (51) a. Steven **made an attempt** to stop playing.
b. Steven **attempted** to stop playing.

日本語の例：

- (52) a. 住民の熱心な**要請を受け**、工事を中止した。
b. 住民に熱心に**要請され**、工事を中止した。
- (53) a. これは市場の**活性化にむけた**規制緩和策だ。
b. これは市場を**活性化する**規制緩和策だ。
- (54) a. 日焼けのため肌が**赤みを帯びている**。
b. 日焼けのため肌が**赤くなっている**。

(藤田篤, ~2012)

<http://paraphrasing.org/paraphrase.html>

多様な言い換え表現を
どうやって収集・管理するか？

パラレルコーパスからの言い換えの獲得

統計的機械翻訳と同じ問題

ただし、良質の大規模パラレルコーパスの入手は翻訳ほど容易でない

夜 11時 30分 まで 開いている レストラン があります
(今村01)

レストラン が 夜の 11時 半 まで 営業して います

夜11時30分 ⇔ 夜の11時半

(レストラン)が開いている ⇔ (レストラン)が営業している

(パラレルでない)

生コーパスからの言い換えの獲得

Distributional Similarity

出現文脈の分布が似ている語は意味も似ている

局所文脈に出現する単語を数える

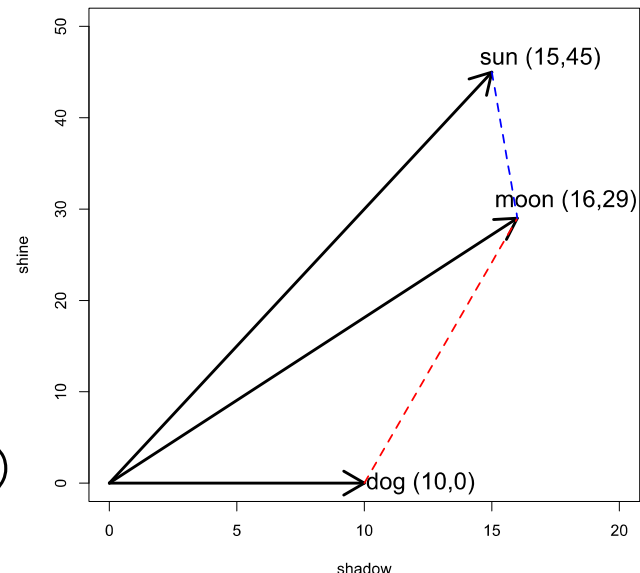
he curtains open and the moon shining in on the barely
ars and the cold , close moon " . And neither of the w
rough the night with the moon shining so brightly , it
made in the light of the moon . It all boils down , wr
surely under a crescent moon , thrilled by ice-white
sun , the seasons of the moon ? Home , alone , Jay pla
m is dazzling snow , the moon has risen full and cold
un and the temple of the moon , driving out of the hug
in the dark and now the moon rises , full and amber a
bird on the shape of the moon over the trees in front
But I could n't see the moon or the stars , only the
rning , with a sliver of moon hanging among the stars
they love the sun , the moon and the stars . None of
the light of an enormous moon . The plash of flowing w
man 's first step on the moon ; various exhibits , aer
the inevitable piece of moon rock . Housing The Airsh
oud obscured part of the moon . The Allied guns behind



単語間の共起頻度を表す行列

	planet	night	full	shadow	shine	crescent
moon	10	22	43	16	29	12
sun	14	10	4	15	45	0
dog	0	4	2	10	0	0

似ている



(Baroni EAACL2012 tutorial)のスライドより抜粋

次元削減の方法さまざま
(LSI(SVD), pLSI, LDA, NN, etc.)

語彙統計語パターンの言い換えの獲得も同様

(研究報告多数)

パターン×関係インスタンス行列

各行ベクトル
がパターンの
意味を表す
(パターン意味
ベクトル)

	カ フ カ	変 身	カ ミ ユ	異 邦 人	...	...	不 注 意	交 通 事 故	タ バ コ	病 気	...	...
Xが執筆したY		43		24	...			0		0	...	
Xの作品Y		34		22	...			0		0	...	
...		...		...	...			...		...	...	
Xが起こすY		0		0	...			23		68	...	
Xが原因のY		0		0	...			32		57	...	
...		...		...	...			...		...	...	

パターンの
クラスタ
(意味関係)

著者・著作関係

因果関係

パターン抽出

行列構築

類似度計算

クラスタリング

大規模なデータ
に対して高速・
高効率に動作

大規模コーパス
(60億文, 600GB)

powered by:

GRID
DATA FARM



パターン(x)	エンティティ(y)	エンティティ(z)
xを>抑制する>y	甲狀腺癌	イソジン
xを>抑制する>y	カビ	洗濯
xを>抑制する>y	津波	防潮堤
xを>撃退する>y	がん細胞	NK細胞
xを>撃退する>y	ジャイアン	どらえもん
xを>撃退する>y	悪者	正義の味方
xを>抑制する & yで>抑制する	癌質の吸収	食物繊維
xを>抑制する & yで>抑制する	甲狀腺癌	インジン
yにある>y	東京	府中
xにある>y	中央線沿線	府中
xにはある & yがある	府中	伊勢丹

関係パターンの
言い換え

SVD等の次元削減でパターンの
意味をよく表現
するベクトルを
獲得

(高瀬, 岡崎ら 2014)

パターン×関係インスタンス行列

(高瀬, 岡崎ら 2014)

543件

X が -> 引き起こす -

X

Y

並び順 ▾

Pattern	X	Y	Freq	PMI
X が -> 引き起こす -> Y	密室性	犯罪	228	14.8331205257
X が -> 引き起こす -> Y	ケロロ小隊	騒動	114	14.7290329967
X が -> 引き起こす -> Y	深部[の]疲労	タンパク質[の]硬化	63	14.5569113848
X が -> 引き起こす -> Y	ホルモンバランス[の]崩れ	老化	40	14.298825183
X が -> 引き起こす -> Y	宇宙飛行[の]無重力状態	筋萎縮[の]現象	32	14.1260312992
X が -> 引き起こす -> Y	発達障害	二次障害	70	14.0405816694
X が -> 引き起こす -> Y	ユーザー	ウイルス感染	30	13.981354542
X が -> 引き起こす -> Y	人間模様	本格サスペンス	25	13.8260694086
X が -> 引き起こす -> Y				13.6444318891
X が -> 引き起こす -> Y				13.5984088051
X が -> 引き起こす -> Y				13.5579896451
X が -> 引き起こす -> Y				13.5579896451
X が -> 引き起こす -> Y				13.4829801644
X が -> 引き起こす -> Y				13.4588722281
X が -> 引き起こす -> Y	体温低下	エネルギー浪費	19	13.4588722281

- 日本語62億文から5300万事例を抽出
 - 57万種類のパターン候補 (頻度 $\geq 1,000$)
 - 3500万種類の関係インスタンス候補 (頻度 ≥ 10)
- 4コア × 12台 = 48CPU分散並列処理で1週間程度
- Apache Spark上の実装を公開予定

獲得された類義関係パターンの例

(高瀬, 岡崎ら 2014)

Xが->|Yを->|引き起こす

Pattern	Similar
Xや->ウイルスによる->Y	0.311595
Xが->考えられている->Y	0.273014
Xが-> Yを-> 引き起こすと	0.257079
Xが-> Yを-> 引き起こすという	0.228389
Xが-> Yを-> 引き起こします	0.224239
Xは-> Yを-> 引き起こします	0.218061
Xが->なり-> Yに-> なってしまいます	0.205171
Xは-> Yを-> 誘発します	0.204377
Xから->引き起こされる->Y	0.195493
Xが->なって->起こる->Y	0.190255
Xで-> Yに-> なっちゃうよ	0.185453
Xによって-> Yを-> 引き起こす	0.185016
Xが-> Yは-> 影響していると	0.182419
Xが->入って-> Yに-> なる	0.179039
Xは-> Yを-> 引き起こしました	0.159263
Xにより->引き起こされる->Y	0.149517

Xを->取り除く->Y

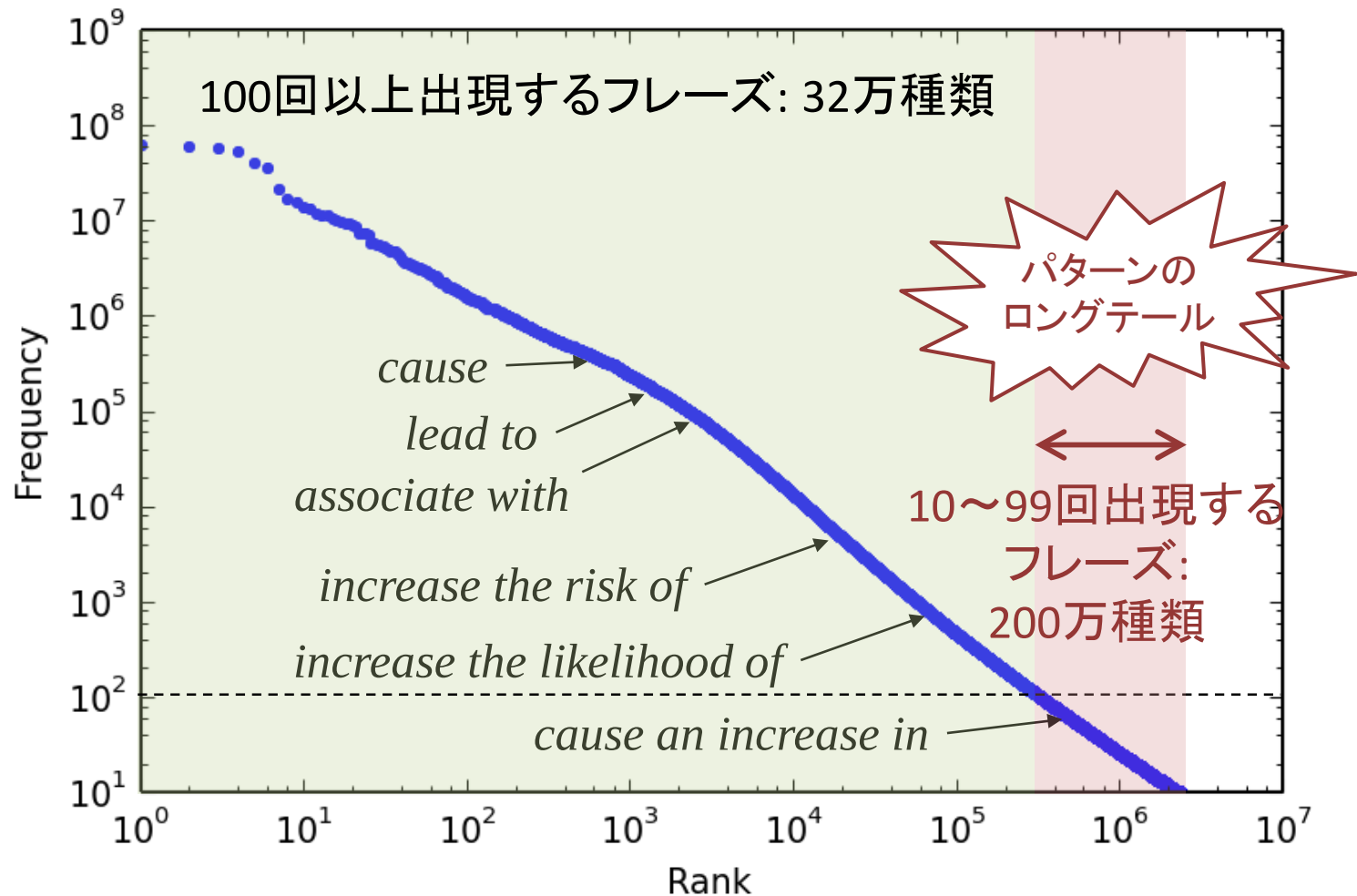
Pattern	Similar
Xを->除去する->Y	0.476163
Xを->取り除いてくれる->Y	0.449809
Xを->取り去る->Y	0.444928
Xを->除去してくれる->Y	0.386644
Xを->取り除く->ため[の]->Y	0.368256
Xを->取り除くような->Y	0.365391
Xを->とりのぞく->Y	0.358086
Xを->無害化する->Y	0.341775
Xを->無毒化する->Y	0.330335
Xを->剥離する->Y	0.328211
Xを->排出させる->Y	0.32533
Xを->排泄させる->Y	0.317296
Xを->分解除去する->Y	0.309232
Xから->逃れる->ためには-> Yを->続けるしか-> ありません	0.307414
Xを->解毒する->Y	0.302393
Xを->和らげる->ため[の]->Y	0.29903

2つの問題

頻度の低い関係表現^{ロングテール}パターンの
意味の学習が難しい

文脈の類似性だけでは
類義と反義の区別が難しい

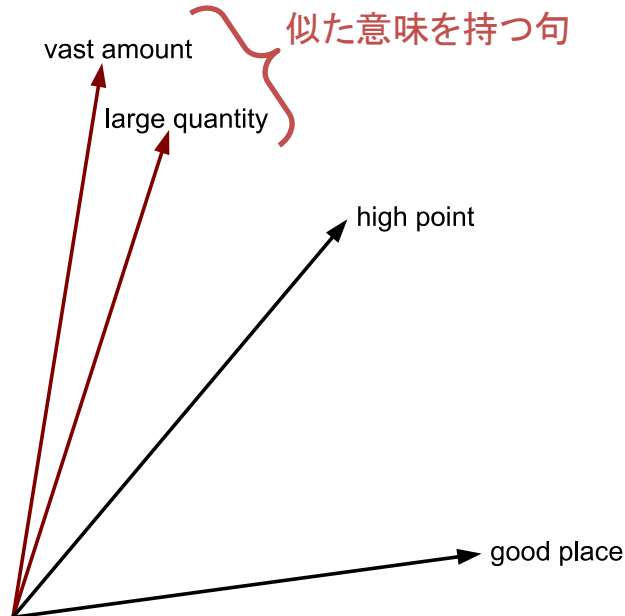
パターンのロングテールをどう学習するか？



ukWaCコーパス中出现する名詞句・動詞句の出現頻度とその順位

Compositional Distributional Semantics (Distributed)

単語のベクトルから句や文のベクトルを構成的に計算したい
近年、研究報告多数



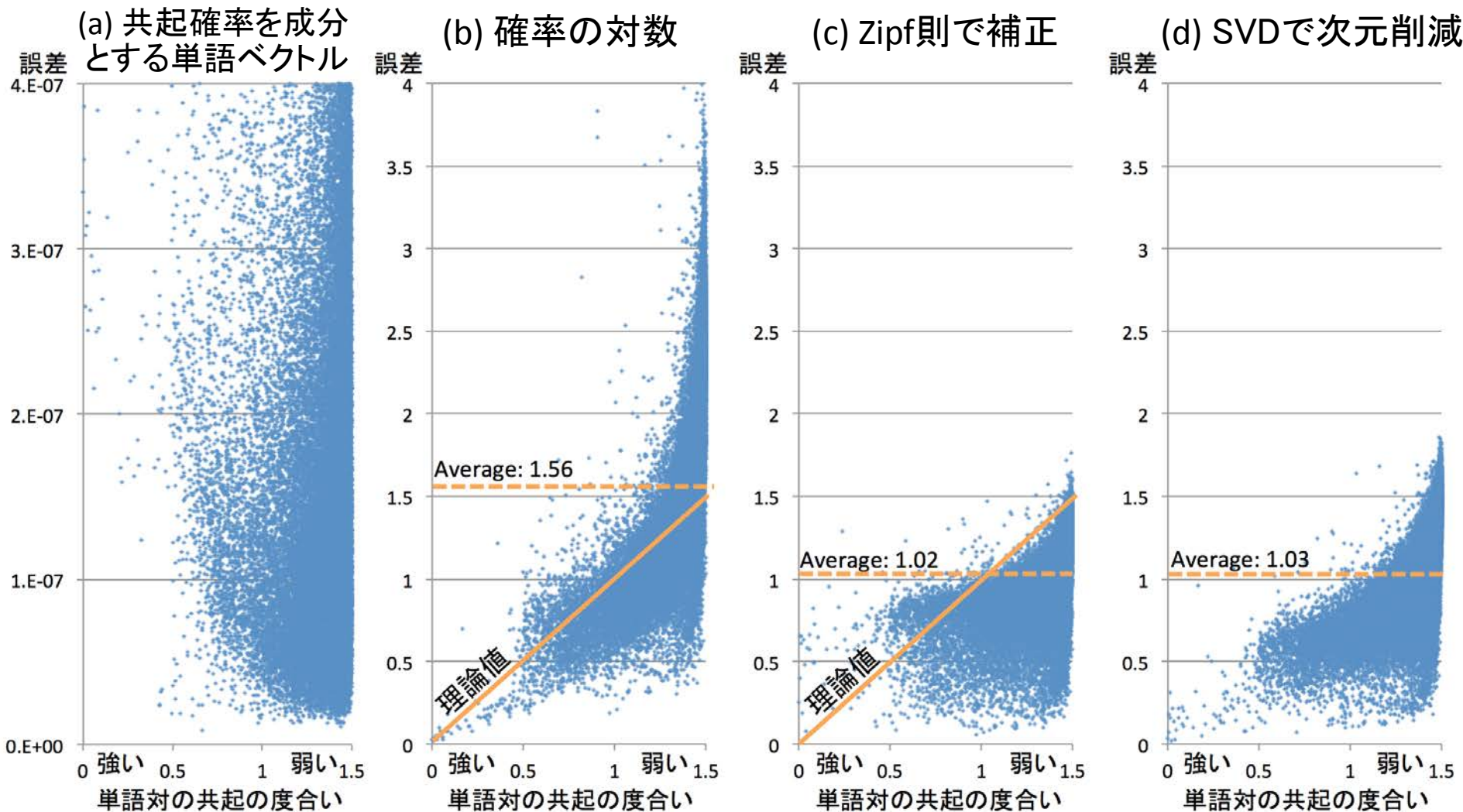
	planet	night	space	color	blood	brown
red	15.3	3.7	2.2	24.3	19.1	20.2
moon	24.3	15.2	20.1	3.0	1.2	0.5
red+moon	39.6	18.9	22.3	27.3	20.3	20.7
red⊕moon	371.8	56.2	44.2	72.9	22.9	10.1
red(moon)	24.6	19.3	12.4	22.6	23.9	7.1

(Baroni EACL2012 tutorial) のスライドより抜粋

理論的にも経験的にも性質の理解が進みつつある

例えば、Additive Compositionality: (red + moon)/2 $\rightarrow$ red_moon

単語ベクトルから構成した句ベクトルの予測誤差の比較 (Tian+ in prep)



2つの問題

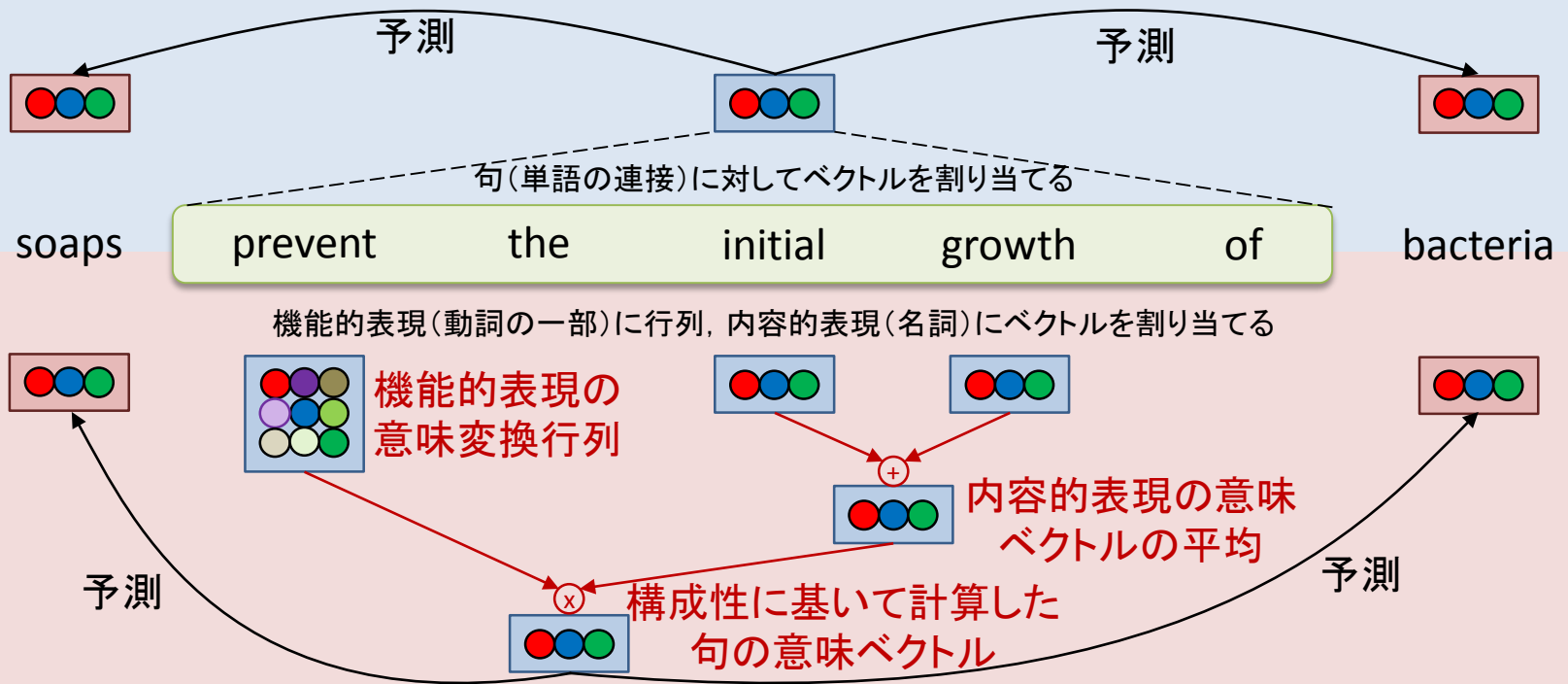
頻度の低い関係表現^{ロングテール}パターンの
意味の学習が難しい

文脈の類似性だけでは
類義と反義の区別が難しい

機能的表現の意味の学習

従来手法

疎データ問題により, 句の意味ベクトルの質が低下する
学習時に存在しなかった句の意味ベクトルを計算できない



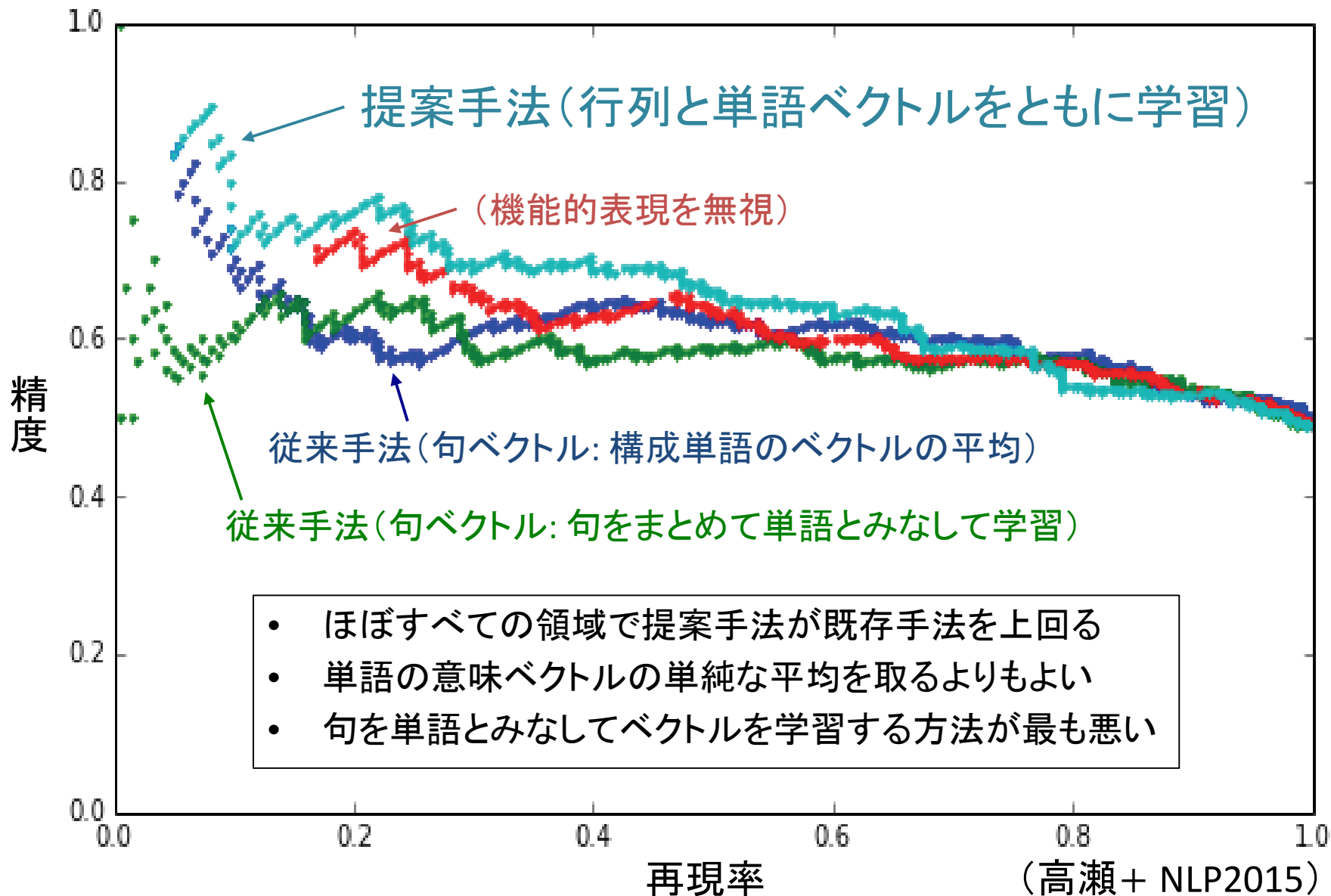
提案手法

動詞による意味の変性をモデル化できる (promote, prevent など)
学習時に存在しなかった句の意味を構成的に計算できる

skip gram (Mikolov + 2013) の拡張 (高瀬 + NLP2015)

述部の類似性判定のデータセットで

(Zeichner+, 12)



データに基づく言い換え

- 大規模コーパスから言い換え獲得
 - 2000年前後以降、活発に研究
 - Distributional Semantics研究の発展
 - 語や句の意味の分散表現、構成性
 - 規模・精度ともに一定レベルまで成熟
 - ソリューション現場で使えるツール化が課題
- 認識問題には応用事例多数
 - 情報抽出、QA、翻訳
- 生成問題への応用はまだこれから
- 論理推論、機械学習との融合？