

意味とカテゴリー

—ロボットの概念獲得と自然言語の比較
に向けて—

九州大学大学院言語文化研究院
内田 諭



九州大学
KYUSHU UNIVERSITY

前回のまとめ

- 言葉の単位の認識が重要(二重分節性)
- 言葉の理解には「意図」の理解が重要
- 言葉は字義通りではなく、パラ言語的な要素に左右され、省略が多い
- 解決のための糸口(関連性理論、フレーム意味論)はありそうだが、「総合的なアプローチ」が必要(相互に影響しているため単独のユニットとしては成立しない)

【前回】**語用論**的な話

【今回】(認知)**意味論**的な話

目次

名詞

ロボットがことばの 카테고리を獲得するために...

- 古典的言語カテゴリー観について
- ことばのカテゴリーの諸相
- ロボットとカテゴリー(小まとめ)
- Tasks for ダイゴロー
- 「言語データ」からのカテゴリー抽出(試論)
- まとめ

※ウェブ公開版では画像等一部省略しています

- 古典的言語カテゴリー観について

古典的言語カテゴリー観

- チェックリスト意味論 (Fillmore 1975)
- man, woman, boy, girl

	human	male	adult
man	+	+	+
woman	+	-	+
boy	+	+	-
girl	+	-	-



adult

child



弁別素性

古典的言語カテゴリー観

【利点】(松本編2003を参考に作成)

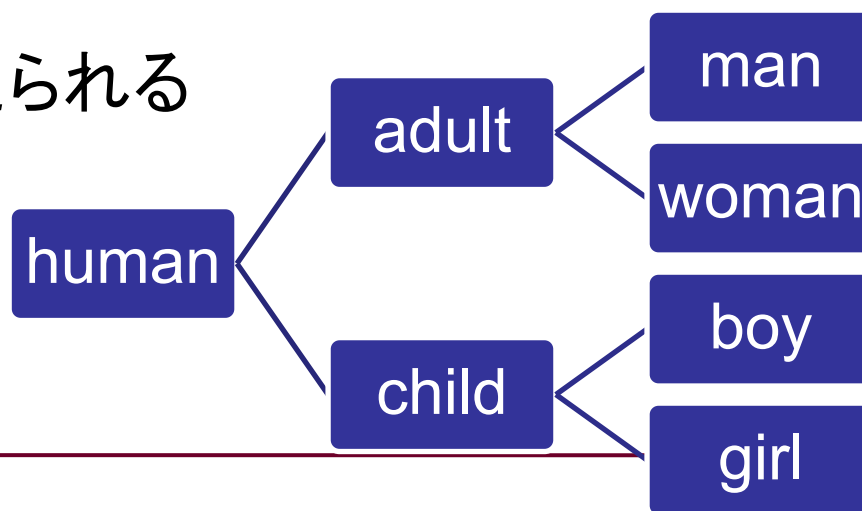
- **単語間**の関係がうまく捉えられる

→ man, womanは±maleという点で対義的

- 単語の分類がうまくできる(ように思える)

→ +humanなものと-humanなもので大きく生物は二分できる

- 単語の階層をうまく捉えられる



古典的言語カテゴリー観の問題点

【問題点】

- 2値以上の対立が扱いにくい

→animalの下のcat, dog, horseなどをどのように表現するか(±イヌ科、±ネコ科etc?)

- 意味素性が無限に存在する

→ある程度一般化は可能だが、類義語の素性を挙げると無限になり、モデル化が困難(例: big, largeの違いは?)

- 意味素性の境界値が曖昧

→boyはいつ(何歳で)manになるのか

古典的言語カテゴリーの問題点

- もっともcupらしいのは5？ (Labov 1973の議論)
- cup, glass, bowlを分別するのは「縦横の比率」(=連続的素性)(以下の定義はOALD onlineより)

【cup】 a small container shaped like a bowl, usually with a handle, used for drinking tea, coffee, etc.

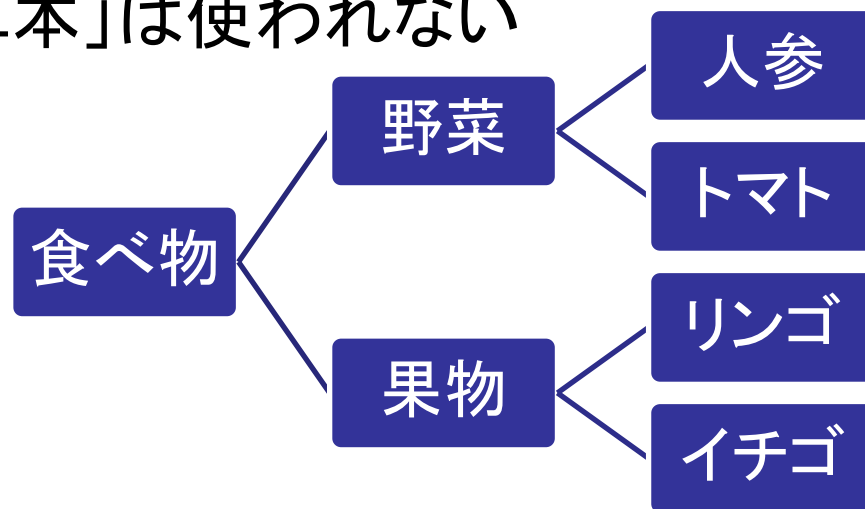
【bowl】 a deep round dish with a wide open top, used especially for holding food or liquid

【mug】 a tall cup for drinking from, usually with straight sides and a handle, used without a saucer

- ことばの 카테고리 の諸相

カテゴリーの階層性

- **hypernym** (上位語) と **hyponym** (下位語) の関係 (A is B)
- 上位語はしばしば抽象的
→ 「野菜」 (= **集合**) というものは存在しない
- 多用される上位語とそうではないものがある (参照点)
→ 「野菜」は頻出だが、「草本」は使われない



カテゴリーの階層性

- 部分語 (Meronym: A is a part of B) がある

→ 顔: 目、鼻、口、顎

→ 目: 瞳孔、虹彩、角膜 etc.

カテゴリー境界のあいまいさ

- カテゴリーを弁別する素性は離散的なもの(±humanなど)だけではない

→childとadultは**連続的**

→cupとglassを弁別する素性(縦横比率)も**連続的**

★「素性」をserialな値としてもってもよい

プロトタイプ性

- カテゴリーには「中心」がある
→「鳥」とは？
→もっとも「鳥らしい」ものは？

★確率的に表現できる？

人間との関わり形による分類

- 「おもちゃ」(機能的分類)
→(五感で捉えられる)形状的な分類ではない
- 「家具」(家族的類似)
→「家」を構成するもの
→場所に紐付いたカテゴリー？
- 「果物」 vs 「野菜」
- →「草木にできる」、「種子がある」、「甘い」、「生で食べる」、「食後に食べる」etc. (cf. トマト、スイカ)

言語相対性

- サピア・ウォーフの仮説
- 虹の色: 日本7色、アメリカ6色
- rain vs 小雨、霧雨、氷雨、時雨、五月雨、地雨 etc.

多義性

- 一つの単語が**複数の意味領域**を持つことがある
- →メタファー的拡張が多い(**具体的な**(物理的な)領域から**抽象的な**(概念的な)領域への拡張)
- →「**共通点**」(の写像)から意味が拡張される

【例】卵

- 卵を産む(字義通り)
- 役者の卵(メタファー拡張)

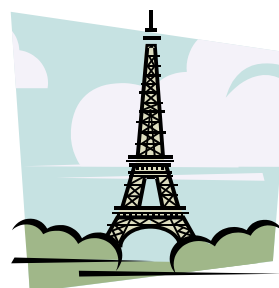
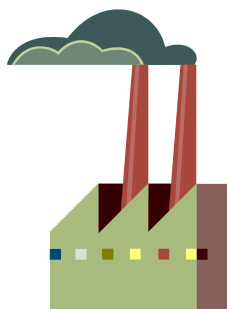
【共通点】「成長する前のもの」、「これから成長するもの」
(→知識?)

スキーマ性

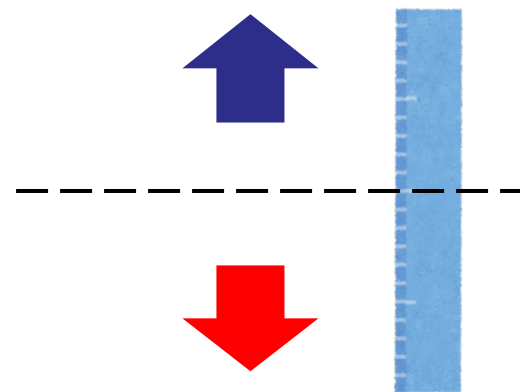
- 前置詞のon
- There is a pen on the desk.
- There is a painting on the wall.
- There is a spider on the ceiling.

スキーマ性

tall+ chimney, tree, building, man, tower, etc.



high+ speed, quality, level, temperature, rate, etc.



ロボットとカテゴリー(小まとめ)

【課題】

- 類似性を捉えられるか
- 階層(上位、下位、部分)を捉えられるか
- 典型性を捉えられるか
- 機能性を捉えられるか
- 多義性を捉えられるか
- スキーマを抽出できるか

【あり方】

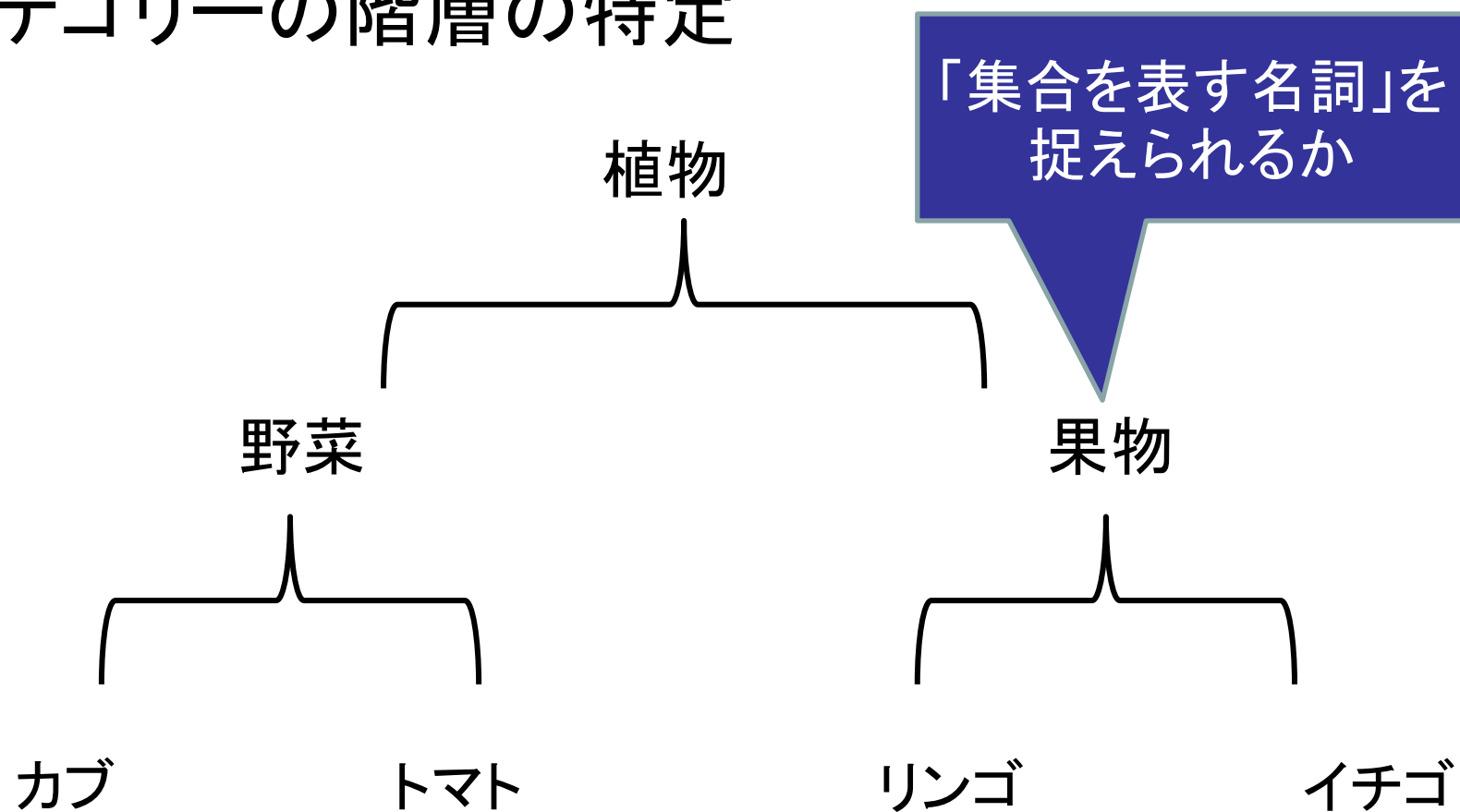
- カテゴリー境界はFuzzyであってよい
- カテゴリー性は連続的でよい
- カテゴリーラベルは相対的でよい(無理に人間に合わせる必要はない)

- Tasks for ダイゴロー

Tasks for ダイゴロー

- 部分語の抽出
- カテゴリーの典型性の抽出(典型度の計算)
- 「人間基準」を介在した分類(文房具など)
- スキーマの抽出

カテゴリーの階層の特定



多義性の獲得

【ガラス】

- ガラスの器
- ガラスの心

【脚】

- 人の脚
- 机の脚

【おもちゃ】

- おもちゃで遊ぶ
- おもちゃにされる

- 「言語データ」からのカテゴリー抽出（試論）

Word2Vecを用いた分析

【Word2Vecの概要】

Word Embeddingの手法の1つで、単語の「文脈」をベクトル化(cf. Mikolov et al., 2013)

→似た意味を持つ単語は似た文脈に出現する(分布仮説)

【特徴】

- ベクトルからコサイン類似度を計算し、「似た振る舞いをする単語」(文脈が似ている単語)を抽出可能。
→Sweden : Norway, Denmark, Finland, Switzerland etc.
- ベクトルを用いて四則演算が可能。
→King – Man + Woman = Queen

Word2Vecを用いた分析

【Word2Vecの設定】

- Pythonのgensimパッケージを使用
- モデルは日本語Wikipediaのデータから構築したものを使用 (ウェブから取得: 白ヤギコーポレーション: <https://shiroyagi.co.jp/news/> (2017年5月アクセス))
- イチゴ, カーテン, スイカ, テーブル, トマト, ペン, リンゴ, 椅子, 鉛筆, 家具, 果物, 机, 消しゴム, 食べ物, 定規, 文房具, 野菜, ソファ, クレヨン, ノートの20語が対象

Word2Vecを用いた分析

【Word2Vecの設定】

- Pythonのgensimパッケージを使用
- モデルは日本語Wikipediaのデータから構築したものを使用 (ウェブから取得: 白ヤギコーポレーション:
<https://shiroyagi.co.jp/news/> (2017年5月アクセス))
- イチゴ, カーテン, スイカ, テーブル, トマト, ペン, リンゴ, 椅子, 鉛筆, 家具, 果物, 机, 消しゴム, 食べ物, 定規, 文房具, 野菜, ソファ, クレヨン, ノートの20語が対象

Word2Vecを用いた分析

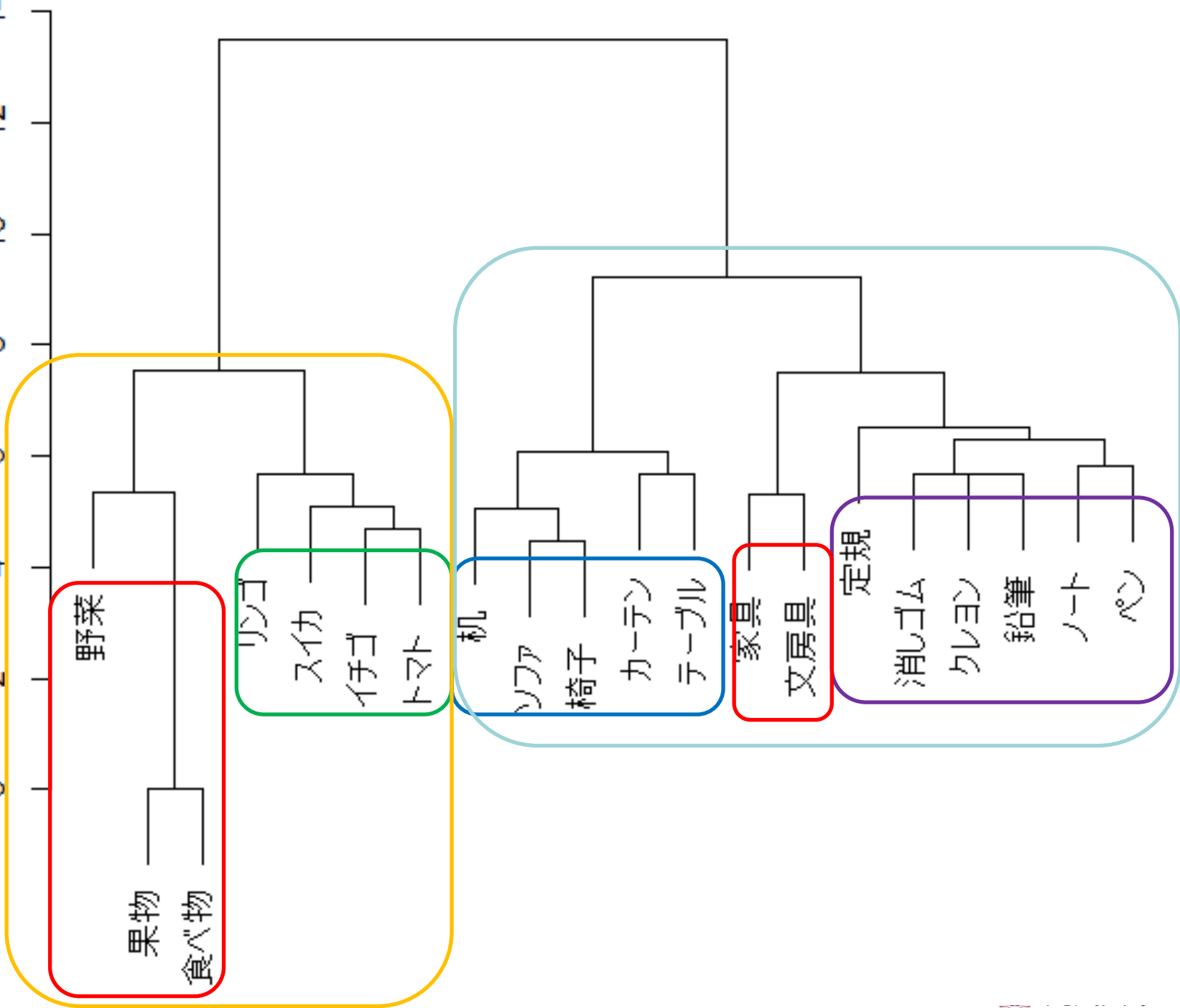
- コサイン類似度の上位20語をそれぞれ抽出
- 共起関係から距離行列に変換(ユークリッド距離)
- Rのhclust関数でクラスタ分析(method="ward.D")

単語	順位	sym	cos
イチゴ	1	キャベツ	0.93951
イチゴ	2	トマト	0.93427
イチゴ	3	メロン	0.933073
イチゴ	4	ニンジン	0.931125
イチゴ	5	パイナップル	0.929049
イチゴ	6	ダイコン	0.927189
イチゴ	7	レタス	0.926877
イチゴ	8	果物	0.926793
イチゴ	9	キュウリ	0.921576
イチゴ	10	マンゴー	0.920143
イチゴ	11	カボチャ	0.913331
イチゴ	12	ハウレンソウ	0.910531
			0.904093
イチゴ		ギ	0.903051
カーテン			0.900136
クレヨン			0.897209
スイカ			0.897111
ソファ			0.896829
テーブル			0.896184
トマト			0.896091
ノート			

	イチゴ	カーテン	クレヨン	スイカ	ソファ	テーブル	トマト	ノート		
イチゴ	20	0	0	8	0	0	9	0	0	
カーテン	0	20	0	0	8	4	0	0	0	
クレヨン	0	0	20	0	0	0	0	2	0	
スイカ	8	0	0	20	0	0	7	0	0	イモ
ソファ	0	8	0	0	20	5	0	0	0	
テーブル	0	4	0	0	5	20	0	0	0	
トマト	9	0	0	7	0	0	20	0	0	ノボ
ノート	0	0	2	0	0	0	0	20	0	ノ

Height

0 2 4 6 8 10 12 14



まとめ

【カテゴリーのキーワード】

- 階層性、典型性、機能性、多義性、スキーマ

【カテゴリーの実態】

- カテゴリー境界はFuzzy
- カテゴリー素性は連続的
- カテゴリーラベルは相対的

→ 言語は**離散的**だが、カテゴリー(素性)は**連続的**

【言語データ】

- 大規模言語データから「カテゴリー」を学習できる可能性

参考文献

- Fillmore, C. J. (1975, September). An alternative to checklist theories of meaning. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 1, pp. 123-131).
- Labov, W. (1973). The Boundaries of words and their meanings. In *New ways of analyzing variation in English*. C. J. Bailey and R. W. Shuy (eds). Georgetown University Press. (pp. 140-173).
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).
- 松本曜(編)(2003).『認知意味論』大修館書店.

• Thank you for your attention!