

理論言語学の方法論を
証拠に基づく医療 (EBM) との比較を通じて
見直す

黒田 航

杏林大学医学部

第9回 Ling x NLP, 2015-09-06 (日曜日)

国立情報学研究所

はじめに 1/2

- (理論) 言語学の現状については、ホントに色々、色々、色々と言いたい事がありますw
- けれども、今日は自分が言いたい事の核心に迫りたいと思います
- Part II で黒田 (2012): 言語学と言語処理の共生は可能か？の論点との結びつけ
- Part I で、その前提になる事態の確認と情報提供

はじめに 2/2

- 私が言いたい事の核心
 - (理論) 言語学の方法論的基盤が脆弱だという事
 - けれども、方法論的基盤の整備は、厳密科学の方法論を安易に真似る事では達成できない(だろう)事
- 理由
 - 言語は(医療にとっての疾患/患者と同じく)厳密な対照実験で調査できる対象ではない
 - 言語には、(厳密科学の対象にはない)個別性と歴史性がある
 - そのため、それ相応の研究法の確立が必要

本日の概要

- Part I

- (理論) 言語学の方法論的基盤の脆弱性の指摘
- 証拠に基づく医療 (EBM) の紹介
 - 動機と目的, 方法 (特に疫学との関係について)
- EBMの成立から学べそうな事

- 答えの出ていない思索: 言語学にとって統計とは何か?

- Part II

- 理論言語学の今後
 - 特にNLP=応用言語学だとした時に

横野さん,

招待, ありがとうございます

Part I

言語学者の典型的な研究の仕方

- 自分の興味をもった現象を調べる
 - 興味をもつ理由も、文献中での遭遇が多い
- 調べると言っても、独自データ収集は最小限
 - しかも、データの追加は自作例基盤
 - 何と云っても文献渉猟が主な手段
- 先行研究の問題点を指摘し、対案を出す
 - が、記述力と説明力のトレードオフを考慮していない事が大半

何がいけないのか？ 1/3

- サンプリングの偏りが顕著
 - データが自作例に依存しているので、信頼性に乏しい
 - 稀な事例が観察の対象にならない
- 観察が知名度バイアスや確証バイアスの影響を受ける
 - 先行研究に言及がない現象は研究の対象になりにくい

何がいけないのか？ 2/3

- これには、自分が言語学の世界に足を踏み入れた時 (= 修士課程に入って) から、 **ずっと違和感**があった
- 自分に受入れ難い理由
 - 将来は生物学者 (特に分類学者か生態学者) になろうと (高校入学までは) 思っていた
 - 幼少時に学研の動物や昆虫の図鑑を見るのが最大の楽しみ
 - 昆虫生態観察の鬼ファーストの『ファースト昆虫記』の愛読者だった

何がいけないのか？ 3/3

- 今の (理論) 言語学では観察の重要性がまったく理解されていないし、強調されない
 - 大学院ですらマトモな指導を受けない
 - 説明，説明，説明と騒ぐが，土台がぐらぐら
 - でもまあ，人文系にはありがち？
- 疑問
 - しかし，どうして??
- まず Part I で，この問題を掘り下げる

理論言語学の方 法論？

何が、どうおかしいのか



理論言語学の三つの難点

- 私見では、理論言語学には少なくとも三つの難点がある
 - 軽い=>重いの順序で述べます
- 三つ目がもっとも深刻で、変更の難しい難点

(理論) 言語学の難点 1/3

- 他の研究分野 (e.g., NLP, 認知科学) との**接続可能性 (特に互換性)** を気にしなすぎ
 - 互換性がありそうな時は喧伝に勤しむ一方で、非互換性がありそうな時は無視
 - 大学院での指導で、学生の希望する関連分野との連携志向を指導教員が挫く場合すらある

要点

- 大学院の指導に関して言うと
 - 指導教員がモノを知らず、分野横断的研究への関与を避けたがるが最大の問題
- あり体に言うと、
 - 言語学には自分野の事しか知らない井の中の蛙 (aka 専門○カ) が多過ぎです
 - こういう方々は当然、自然科学で成立する基礎研究と応用研究の良い意味での緊張関係を知らない

個人的見解

- 言語学の (下らない) 論文や研究書をあれこれ読む位なら、代わりに
 - NLP やプログラミングの勉強を
 - 心理実験法の勉強を
 - (行動) 経済学や生態学の勉強を
 - 数学の勉強を
- すべきでしょう
 - 人生は短い...

(理論) 言語学の難点 2/3

- 観察 (observation), 記述 (description), 説明 (explanation, account) の区別が不明確
 - これらに関して十分な指導がない (大学院ですら)
 - そのため、自分が何のために、何をしているのか、自分の得た結果が何を意味しているのか自覚している人が少数派

補足

- N. Chomsky の観察的妥当性，記述的妥当性，説明的妥当性に言及する人も（それなりに）いるが，
 - 自然科学の観察，記述，説明の関係がちゃんとわかっている保証はないし，
 - わかっているようにも全然見えない。

観察, 記述, 理論の関係

- 観察
 - 十分な量の, 十分に信頼できるデータ D を集める事
- 記述
 - データ D に 妥当な分類 (proper classification) C を与える事
- C は 妥当な表示 (proper representation) R と等価
- 説明
 - データ D の記述に与えられた C/R を 正当化する事
=
 - 特定の C/D が, 他の C'/D' より良いか理由を明らかにする事

観察, 記述, 理論の歪み 1/3

- 歪んだ観察
 - 確証バイアス (confirmation bias) が強く働き, 十分な量のデータが集められる事も, 十分に信頼できるデータが集められる事も稀
- 典型的症状
 - 調査で無作為化は前提にならない事が常態
- 記述であれ説明であれ, データの選り好みするのが常態化
- 理論 (的主張) T の反証が目的になる事は稀
 - T が真であれば起きるはずのない事象を探す事は滅多に試みられない
 - 立証は, 既存の理論 (的主張) T に適合する例を追加する事ばかり

歪んだ観察の例 1/3

- 科学者は、二つの物事の相関関係 — 雲の種まき（人工的に雨を降らせる技術）と雨，心臓発作とコレステロール，[何]でも[良い]— を確定しようと思ったら、二つが両方ある場合（コレステロール値が高く心臓発作を起した人の数）と片方だけの場合（コレステロール値が高く心臓発作を起さなかった人の数と，コレステロール値が低く心臓発作を起した人の数）と両方ない場合（コレステロール値が低く心臓発作を起さなかった人の数）をすべて考慮したデータを用いて計算する。二つの物事に真の相関関係がある可能性を正確に評価するには、以上の数値が全部必要になる。
 - Daniel Gilbert 『明日の幸せを科学する』, Loc.1818
 - [K.K.: 因果関係を確定したいなら、更に別のデータも必要になる]

歪んだ観察の例 2/3

- こうしたことはみんな、[至極]もっともなことだ—統計学者にとっては。しかし、[普通]の人が二つの物事に相関関係があるかどうか調べたいと思ったら、[普通]は起きたことについて調べたり、注目したり、考慮したり、思い出したりするだけで、起きなかったことについては、あれこれ考えないことが研究で示されている。どうやら人はこの誤りを長い間犯してきたらしい。四世紀近く前、哲学者であり科学者でもあったサー・フランシス・ベーコン [Sir Francis Bacon] は、心の誤り方について記し、**欠如を考慮しないことがもっとも深刻だ**と指摘した。[引用略]

- Daniel Gilbert 『明日の幸せを科学する』, Loc.1822

歪んだ観察の例 3/3

- ベーコンはこの論点を説明するのに、ローマの神殿を訪れた男の物語を使った [実は Cicero からの引用]. この訪問者が神々の力に感銘を受けるよう、ローマ人は、先[頃]船が難破したのに、信仰心のおかげで生き残ったとされる敬虔な船乗りたちの肖像画を見せた. **これが奇跡の証拠だと信じるように[迫られ]た訪問者は、抜け目なく尋ねた.** 「ですが、神々に誓いを[立て]ていたのに**死んだ船乗りたちの肖像画はどこですか**」. 科学的にも示されているが、われわれのような凡人は、行方不明の船乗りの絵など [滅多に]見たがらない.

● Daniel Gilbert 『明日の幸せを科学する』, Loc.1818

疫学の問題設定 1

	ジドブジン	プラシーボ	条件差解消
-日和見感染 m	38	37	75
+日和見感染 n	1	7	8
総患者数 $N = m+n$	39	38	77
リスク m/N	0.026	0.184	0.104

- 早期HIV感染患者においてジドブジンと偽薬の投与による日和見感染リスクを比較した無作為化臨床試験/治験
- 『ロスマンの疫学: 科学的思考への誘い』, p. 84.のデータを改変

言語学の研究法の難点 1/2

- 素朴な例証
 - 主張 T (e.g., ドブシジンが日和見感染を抑制する) に該当する絶対数 (e.g., B3セルの事例の個数) が暗黙の主観的基準より多ければ, T の妥当性の十分な証拠になると考える
- 二重の欠陥
 - 該当しない例 (B2セルの事例数) と該当例の対比がない
 - そのため, 「リスク」に対応する指標が評価されず, 正例 positives と負例 negatives の区別がない
- 対照的環境 (B2セルのプラシーボ) との比較がない
 - そのため, 真の正例 false positives と偽の正例 false positives との区別がない

言語学の研究法の難点 2/2

- 試行

- Cicero/Bacon の逸話を疫学的に考えると、どうなるか？

疫学の問題設定 2a

	+信仰心	-信仰心	条件差解 消
助かった m	17	11	28
助からな かった n	13	9	22
$N = m+n$	30	20	50
効果率 m/N	0.567	0.550	0.560

架空のデータ 1/n

- B2: 信仰心があり助かった人の数
- B3: 信仰心があって助からなかった人の数
- C2: 信仰心がなくて助かった人の数
- C3: 信仰心がなくて助からなかった人の数

疫学の問題設定 2b

	+信仰心	-信仰心	条件差解 消
助かった m	25	17	42
助からな かった n	5	3	8
$N = m+n$	30	20	50
効果率 m/N	0.833	0.850	0.840

架空のデータ 2/n

- B2: 信仰心があり助かった人の数
- B3: 信仰心があって助からなかった人の数
- C2: 信仰心がなくて助かった人の数
- C3: 信仰心がなくて助からなかった人の数

疫学の問題設定 2c

	+信仰心	-信仰心	条件差解 消
助かった m	25	3	28
助からな かった n	5	17	22
$N = m+n$	30	20	50
効果率 m/N	0.833	0.150	0.560

架空のデータ 3/n

- B2: 信仰心があり助かった人の数
- B3: 信仰心があつて助からなかった人の数
- C2: 信仰心がなくて助かった人の数
- C3: 信仰心がなくて助からなかった人の数

要点

- 例 2a, b, c のうち, Cicero/Baconの逸話で信仰心に効果があると言える証拠になるのは, 例 2c のみ
- 例 2b では信仰心に効果があるとは言えない
 - 信仰ありの効果率が高くても, 信仰の有無の条件間で効果率の差がない

無効な立証の例1 1/3

- P: メタファーは (異なる) 概念領域間の写像である
 - $\forall x [\text{metaphor}(x) \Rightarrow \text{cross-domain-mapping}(x)]$
- を立証するのに、肯定例を追加し続けるのは、立証として無効
- 問われていない事
- Pの対偶: x が概念領域間写像でないなら, x はメタファーでない
 - $\forall x [\neg \text{cross-domain-mapping}(x) \Rightarrow \neg \text{metaphor}(x)]$
- 立証ですべき事
 - メタファーなのに領域間の写像が関与しない例がない事を確認する

無効な立証の例1 2/3

- 具体的には，次の S, T のいずれも真にならない事を立証
 - S: $\forall x [\neg \text{cross-domain-mapping}(x) \Rightarrow \text{metaphor}(x)]$
 - 領域間写像がないのにメタファーである
 - T: $\forall x [\text{cross-domain-mapping}(x) \Rightarrow \neg \text{metaphor}(x)]$
- 領域間写像なのにメタファーでない
- 注意点
 - 肯定証拠 (Pの実例) をいくら追加しても，説明 P の信頼度は上がらない
 - この立証を無作為化なしで実証するのは，無理

無効な立証の例1 3/3

- 参考
 - Pの裏: x がメタファーでないなら, x は概念領域間の写像でない?
 - $\forall x [\neg \text{metaphor}(x) \Rightarrow \neg \text{cross-domain-mapping}(x)]$
- Pの逆: x が概念領域間の写像なら, x はメタファー?
- $\forall x [\text{cross-domain-mapping}(x) \Rightarrow \text{metaphor}(x)]$

観察, 記述, 理論の歪み 2/3

- 歪んだ記述

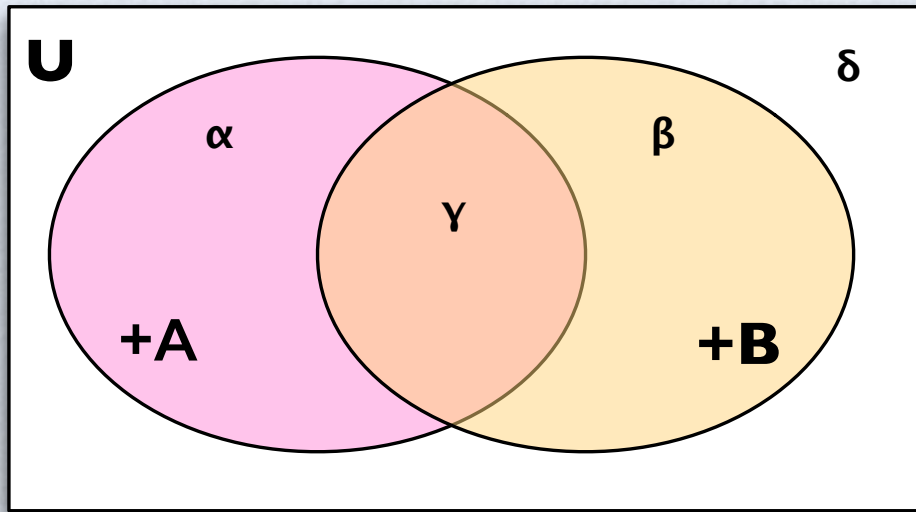
- 十分な根拠なく **排他分類** exclusive classification = hard clustering が想定される事が常態

- 典型的症状

- 血液型のA, B, AB, O分類のような現象を適切に表現できない

発表後の補足

Venn図で表示



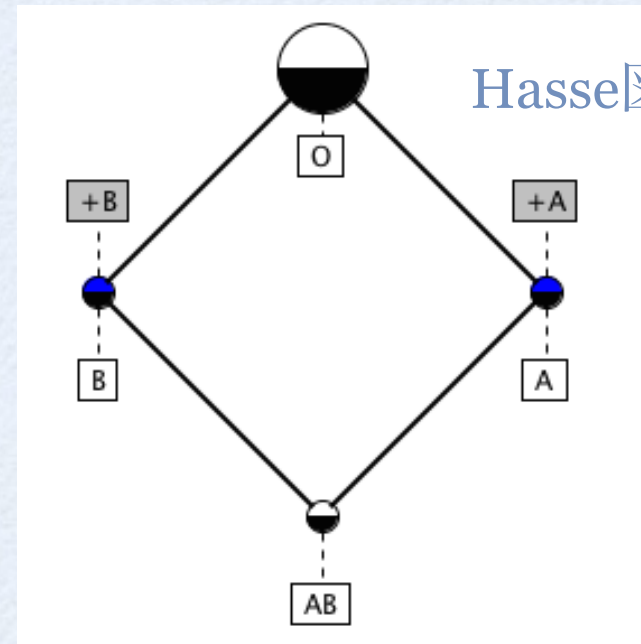
α : A型
 β : B型
 γ : AB型
 δ : O型

表で表示

	+A因子	-A因子
+B因子	AB型	B型
-B因子	A型	O型

- ABO血液型分類は、樹状構造にならない
- Venn図と表 (matrix) と Hasse図の表記はすべて等価

Hasse図で表示



歪んだ記述の例

- 統語構造が樹状構造 tree structure であるべき経験的理由は皆無
- 品詞分類が排他的になっているべき理由はない
- 要点
 - 排他分類が成立すると嬉しい事と、実際にそれが成立しているのは別の事

発表後の補足

- Formal Concept Analysis (FCA) という分類システムを使えば、素性/属性表現を使いつつ、排他分類の誤謬を回避できます
 - Ganter, B. and R. Wille (1999). *Formal Concept Analysis: Mathematical Foundations*. Berlin: Springer-Verlag (Translated by C. Franzke).
 - Ganter, B., G. Stumme, and R. Wille (eds.) (2005). *Formal Concept Analysis: Foundations and Applications*. Berlin/Heidelberg: Springer.
- FCA の無償ツールに Concept Explorer 1.3 があります
 - <http://conexp.sourceforge.net/>
 - 日本語を使いたい場合には、起動時に文字コードの指定が必要
 - `java -Dfile.encoding=utf-8 -jar conexp.jar`

観察, 記述, 理論の歪み 3/3

- 歪んだ説明

- 説明に付き纏うトレードオフ (= あちらを立てればこちらが立たず) を意識しないのが常態
- データ D に与えられた C/Rを正当化するために, 多くの場合, 十分性の提示しか行われない
- 必要性の提示がないので, 過剰般化 overgeneralizations が常態化

過剰般化の例

- 構文文法 Construction Grammar で言う “構文”
 - 構文の意味と語の意味の区別は，意味論の語用論の区別と相同
 - 一方があれば他方は無用なのではなく，2分化を説明できるモデルが必要
- 参照点構造
- Remarks
 - 過剰般化は機能言語学・認知言語学に目立つ
 - 過剰般化は生成言語学系で目立たないが，逆に過小般化=問題の矮小化が目立つ

過剰般化の由来

- 過剰般化の原因は、次のトレードオフを自覚しない事
 - A. 説明範囲の過度の拡大が、疑似正例を許容する
 - B. 説明精度の過度の増大が、疑似不例を許容する
- A 過大般化 (=過剰般化) の例
 - 万能薬 = “魔法の薬”, 万能理論
- B 過小般化の例
 - 適用範囲の限定され過ぎた理論

(理論) 言語学の難点2/3のまとめ

- 問い
 - 観察 (の妥当性), 記述 (の妥当性), 説明 (の妥当性)のうち, **どれが一番 (理論) 言語学で未発達か?**
- その答え
 - 観察 (の妥当性)
- 記述の妥当性, 説明の妥当性の難点は, 実際には観察の妥当性の欠如のオマケ
- 補足
 - 本点は非常に深刻な重要なのですが, 取り上げる問題ではありません
 - この問題を扱った教科書を書こうと頑張っています

(理論) 言語学の難点 3/3

- 研究者の**基本的心構え**が、**今だに直観(偏重)基盤** (overly) intuition-based で、**権威基盤** authority-based である
 - 私が見る限り、これが理論言語学の最大の難点
 - この状況を変えない限り、難点 1, 2 は改善不能
- だけど、待て！
- これは他の人文系の研究分野と同じなのでは？

以後の流れ

- 証拠に基づく医療 (evidence-based medicine: EBM) が成立した背景を概観し、
- 言語学が (医療科学や他の人文(科)学と同じく)
 - 権威主義になりがちなのはなぜか？
 - 直観偏重基盤に陥りがちなのはなぜか？
- の問いへの答えを見出し、
- 有効な対処法を模索する

EBMの成立背景



以下の考察の要点を先取り

- 言語学は証拠 (の量以上に) 質を気にするべき
 - 証拠の量も大して気にしないが、それ以上に証拠の質を気にしない
 - どんな証拠も同じ位の立証力を持っていると安易に想定
- これは
 - 世界はノイズに溢れている
 - 人の判断には間違いがつき物である
- 事を前提にすると、全然正しくない

なぜ EBM ? 1/2

- EBMが成立した背景を調べると、**医学は次の点で自然諸科学より、ずっと人文(科)学に近い事がわかる**
 - 医学は他の多くの自然科学(や社会科学)の研究と質的に異なっている
- 例
 - 対象となる現象が複雑
 - **交絡 (confounding) が常態**
 - 厳密な再試は実行不可能
 - **対象の歴史が影響する**
 - 人文科学と医療科学とは、**稀な現象を無視しない事を美德にする**
 - **稀な現象 (e.g., 疾患) は無視して構わない現象ではない**

なぜ EBM ? 2/2

- EBMが成立した背景を調べると、**医学は更に**次の点で自然諸科学より、**ずっと人文科学に近い**事がわかる
 - 権威主義の悪影響を受けやすい
 - 直観主義の悪影響を受けやすい
 - デマの影響を受けやすい
 - 利害関係の影響を受けやすい
- 問題設定
 - これを見本にして方針を立てられないのか？

因みに

- なぜ医学史・医療史に関心を持ったのか？
 - 現在の仕事が医学部で英語を教える事である
- それに加えて
 - 自分の父親が医師で、自分も医師になろうとして医学部を受験した
 - 2011/3/11 以降の放射能被害関連のデマに呆れた

権威主義の証拠

医療の権威主義の例

- [1] 脚気が vitamin D 不足で原因で起こる疾患だと認められるまでの紆余曲折
- [2] Ignaz Philipp Semmelweiss/Semmelweis Ignác Fülöp (1818-1815) の提唱した産褥熱の原因の受入れをめぐる紆余曲折
 - 以上の二つは板倉聖宣『新哲学入門』で紹介されています
- 1 の関連箇所を、(私にしては珍しく)長目に引用します

脚気 1/4 - 板倉聖宣 『新哲学入門』, pp. 82-84

- 私は、少し前に日本における脚気の研究史を調べて、『模倣の時代』上下二巻（仮説社）という大部の本をまとめましたが、そのとき「人間というものはなんと頑固なものか」と改めて思いしらされました。ふだん信用されている権威者たちがその判断を間違えると、一部の人たちが真理を発見しても、その発見は十年や二十年どころか、それこそその権威者たちが死んでしまうまで、三十年も認められないでいることがあるのです。

脚気 2/4 - 板倉聖宣 『新哲学入門』, pp. 82-84

- 脚気の研究の場合，ふつうの科学研究法に従う人なら，「麦飯が脚気に効く」という事実は，「一八八四（明治十七）年の大阪の陸軍での実験とその後の陸海軍の追試によって，一八八五年には確立された」ち断定してもよかったはずでした。ところが，その発見は，東大医学部の教授たちと陸軍の軍医本部の秀才たち，文豪としても知られる森林太郎（＝森鷗外）らによって，長いあいだ決定的に無視され続けたのです。
「そんなこと（麦飯が脚気に効くなんてこと）は現代医学の三大栄素説ではとうてい考えられない」というのがその理由でした。

脚気 3/4 - 板倉聖宣 『新哲学入門』, pp. 82-84

- 東大医学部の教授たちが「米ぬかに脚気予防効果がある」とはじめて認めたのは一九一七（大正六）年のことです。だから、東大医学部系統の人々はその事実を何と三十三年間も無視し続けたことになります。そのあいだに、日本陸軍の軍医本部は、日清戦争のとき戦線に麦を送ることを拒否して、戦死者に数倍する兵隊たちを脚気で死に至らしめ、日露戦争のときも戦線に玄米や麦を送ることを拒否して何十万という脚気患者を「製造」し、何万という脚気による死者を出すという悲惨な結果を招いたのです。

脚気 4/4 - 板倉聖宣 『新哲学入門』, pp. 82-84

- それなら、東大医学部の教授たちは、どうして最後には「米ぬかに脚気予防効果がある」と認めるようになったのでしょうか。それは、一九一二年に外国でビタミン学説が提唱されて、その研究が始まったからです。そこで、さすがの東大医学部の研究者たちも、米ぬかエキスの大量投与の研究を行って、それまでの実験結果とは違う結論を出したのです。もしも、外国でその種の研究が始まらなかったら、東大医学部の人々は「脚気は一種の栄養欠乏症だ」という事実を依然として認めなかったことでしょう。

産褥熱 1/2 - 板倉聖宣 『新哲学入門』, p. 85

- <産褥熱>というのは、お産のあと産婦が正体不明の熱を出して死ぬ病気です。ハンガリーの青年医師ゼンメルワイスの勤めたウィーンの産院でも、たくさんの産婦がその正体不明の病気で死にました。そのことを気にしたゼンメルワイスはその研究をすすめて、一八四七年、医者にその手指を消毒させることによって、産褥熱による死亡率を十分の一にも減らすことに成功したのでした。

産褥熱 2/2 - 板倉聖宣 『新哲学入門』, p. 85

- ところが、その意見は病院の採用するところとならず、彼は激しい反対にあつて病院をやめなければならなくなりました。彼は帰国してブダペスト大学の教授となつたのですが、その学説はいつまでも認められず、ゼンメルワイスは一八六五年、とうとう憤慨しながら四十七再の生涯を終えたのです。

真理の受入れと権威の関係

- 板倉『新哲学入門』（pp. 85-86）の主張
 - 真理は実験的に決められる前に、人々の党派的な思惑や感情などによって左右されてしまうことが少なくないのです。もっとも、その研究の成果を受入れることが既存の思想や権威の存続に大きな影響を与えずにすむ場合には、狭い社会だけの感情の対立があっても、その感情的な対立の外にいる人々に認めてもらうことができます。しかし、「その発見の受入れによって、既存の思想・権威を大きく変えざるを得なくなる」と考えられる場合には、多くの人々の抵抗は容易に克服できないでしょう。

発展 1/2

- 問い

- 権威主義はなぜ生じるのか？

- 私の答え

- 権威主義的な人は自分の判断力に自信がない人
 - 人は自分の判断に自信がない時，権威に縋る
- 対策は，多くを知るだけでなく，良くわかる，しかない

発展 2/2

- 問い

- 権威はゼロにできるか？

- 私の答え

- 無理

- 私でも、数学や物理学や化学の成果の大半は、権威の主張を受入れる

- その状態でできる事

- いつでも、どこでも 懐疑心を失わないでいる

- 絶対に確実だと言えるもの世の中に何もないと割り切る

直観主義の証拠

過度の直観基盤の弊害

- Archibold Cochrane (1909–1988)の逸話
 - 彼は問うた:
 - 心臓発作の患者が予後をどう過ごすのが最善か？病院か？自宅か？
 - 当時の医師の見解:
 - 病院に決まっているだろ！？！治験で確かめる？！ふざけた事をするな！！
 - 実際は逆だったが、これを納得させるのにトリックが必要だった
 - Tim Harford のTED 講演: *Trial, error and the Gold complex.*で紹介されている

直観の重要性

- 診断 diagnosis の特殊性を考えると、直観なしでは医療が成立しない
 - 医者の書いたエッセイには厄介な診断の報告が多数
 - Atul Gawande: *Complications*.
 - Lisa Sander: *Every Patient Tells a Story*.
 - Jerome Groopman: *How Doctors Think*.
- とはいえ、大量データ (big data) を基にした教師なし機械学習がヒトの直観を代替できる日が来ないとは言えない

証拠に基づく医療の概要



証拠の質を問う

- EBMの最大の動機は**権威主義と過度の直観主義からの脱却**
 - これは言語学を含めた人文学の目標と同じ！
- その目標を達成するため、EBMでは
 - 統計学的根拠に基づいて**証拠の質を分類し**、
 - その分類に基づいて**具体的処置の選択基準を明確にしている**
- 有用なアナロジー
 - 医療：人文科学 = 処置：説明理論/モデル

EBMの処置の推奨レベル

- 処置をするべきか否かが、証拠の質と量によって評価される

Grade	根拠の有無	採用指針
A	科学的根拠があり	実施を勧められる
B	科学的根拠がある とは言えないが	実施を勧められる
C1	科学的根拠がない けれども	実施が許容できる
C2	科学的根拠がない ので	実施が勧められない
D	無効性や有害性の 根拠があり	実施を禁じる

質問

- EBMの処置の推奨レベル A, B, C1, C2, Dに準拠するならば、言語学の説明はどのグレードに該当するか？
- その理由は？

EBMの証拠レベル/階層

強い



弱い

Level	内容	例	無作為化	対照化
1a	無作為化ありの比較治験の メタ分析		+	+
1b	少なくとも1つの無作為化 ありの比較治験		+	+
2a	無作為化なしの 同時並行対照コホート研究	prospective/ concurrent cohort study	-	+
2b	無作為化なしの過去との対 照コホート研究	retrospective/ historical cohort study	-	+
3	症例対照研究	retrospective study	-	+
4	対照群のない実験	処置前後の比較などの 前後比較	-	-
5	症例報告やケースシリーズ		-	-
6	専門家個人の見解	専門委員会報告	-	-

皆さんに質問

- EBMの証拠レベルの階層 1a, 1b, 2a, 2b, 3, 4, 5, 6
に準拠するなら、言語学の説明はどのレベルに
該当するか？
- その理由は？

私の解釈では

- 個々の言語学者の自作例の容認度判定は Level 6
かせいぜい Level 5
- コーパス言語学の調査結果も Level 3止まり
- 心理言語学の実験も多くは Level 3止まり

Seven alternatives to EBM という風刺論文

Basis for decisions	訳	Marker	Measuring device	Unit of measurement
Evidence	証拠	Randomized controlled trial	Meta-analysis	Odds ratio
Eminence	威光	Radiance of white hair	Luminometer	Optical density
Vehemence	声の大きさ	Level of stridency	Audiometer	Decibels
Eloquence (or elegance)	雄弁/優雅さ	Smoothness of tongue or nap of suit	Teflometer	Adhesion score
Providence	信仰	Level of religious fervor	Sextant to measure angle of genuflaection	International units of piety
Diffidence	臆病さ	Level of gloom	Nihilometer	Sighs
Nervousness	神経質さ	Litigation phobia level	Every conceivable test	Bank balance
Confidence	自信	Bravado	Sweat test	No sweat

Seven alternatives to EBM 1/2

- Isaacs と Fitzgerald (1999) のEBM のパロディー論文より:
 - **Eminence based medicine**— The more senior the colleague, the less importance he or she placed on the need for anything as mundane as evidence. Experience, it seems, is worth any amount of evidence. These colleagues have a touching faith in clinical experience, which has been defined as “making the same mistakes with increasing confidence over an impressive number of years.” The eminent physician’s white hair and balding pate are called **the “halo” effect**.
 - **Vehemence based medicine**— The substitution of volume for evidence is an effective technique for brow beating your more timorous colleagues and for convincing relatives of your ability.

Seven alternatives to EBM 2/3

- Isaacs と Fitzgerald (1999) のEBM のパロディー論文より:
 - **Eloquence based medicine**—The year round suntan, carnation in the button hole, silk tie, Armani suit, and tongue should all be equally smooth. Sartorial elegance and verbal eloquence are powerful substitutes for evidence.
 - **Providence based medicine**— If the caring practitioner has no idea of what to do next, the decision may be best left in the hands of the Almighty. Too many clinicians, unfortunately, are unable to resist giving God a hand with the decision making.

Seven alternatives to EBM 3/3

- Isaacs と Fitzgerald (1999) のEBM のパロディー論文より:
 - **Diffidence based medicine**— Some doctors see a problem and look for an answer. Others merely see a problem. The diffident doctor may do nothing from a sense of despair. This, of course, may be better than doing.
 - **Nervousness based medicine**— Fear of litigation is a powerful stimulus to overinvestigation and overtreatment. In an atmosphere of litigation phobia, the only bad test is the test you didn't think of ordering.
 - **Confidence based medicine**— This is restricted to surgeons (table).

皆さんに質問

Classification	元構造	先構造
Eminence-based		
Vehemence-based		
Eloquence (or elegance)-based		
Providence-based		
Diffidence-based		
Nervousness-based		
Confidence-based		

- EBMのパロディーの7例が比喩写像だとして、それぞれがどんな元構造と先構造の対応をもっているかを明示せよ

質問の解答例

Classification	元構造	先構造
Eminence-based	地位/知名度の高さ	権威主義
Vehemence-based	声の大きさ	自己主張の過剰
Eloquence (or elegance)-based	雄弁さや優雅さ	詭弁
Providence-based	神よりの啓示	直観主義
Diffidence-based	臆病さ	虚無主義
Nervousness-based	心配性	事なかれ主義
Confidence-based	自信	技能/判断への自惚れ

- 原典では“Confidence-based は外科医 surgeon に限る”と言っているが、言語学で自分の容認性判断に絶対の自信を持っている人の言動は、これに相当するかも

EBMの成立から 学べる事

証拠に基づく言語学 Evidence-
based Linguistics を望みつつ



EBMから学べる事 1/3

- 1. 無理に厳密科学のモノ真似をしなくて良いと割り切れる
 - 物理学のモノ真似をしたがる生成言語学の方法論は最善と言えない
 - どんな言語も歴史を反映した個性をもっている
 - それを分離するには入念な作業が必要

EBMから学べる事 2/3

- 2. 言語学の証拠の質を問う
 - 個々人の容認性判断だけで良いのか？
 - どんなに著名な人の容認性判断を寄せ集めても信頼性のある証拠にならない
 - EBMの証拠の階層を模範にすれば，無作為化された比較実験のメタ分析が一番信頼できる証拠
 - Level 6の証拠をいくら増やしても，単にノイズを増やすだけの結果に終わっている可能性も真剣に考慮すべき

EBMから学べる事 3/3

- 注意して欲しい事

- (理論)言語学の研究成果がすべて無価値という事ではない
- そうではなくて、証拠の質を問題にしない事で
 - 価値のある成果と価値のない成果が混在しており、
 - 今のところ、両者を区別できない
- という事
 - 方法論の基盤が脆弱なので、そうなるのは必至

EBL の成立には何が必要か？ 1/4

- 問い

- 証拠に基づいた医療 EBM に倣って、証拠に基づいた言語学 Evidence-Based Linguistics: EBL を構想すると、それには何が必要なのか？
- EBL は証拠に基づいた人文学 Evidence-based Humanities: EBH の下位部門

EBL の成立には何が必要か？ 2/4

- まず、絶対に必要なのは
 - 可能な限り記述や説明を測定可能なデータと結びつける
= 正誤判定の不能な記述や説明を拒絶する事
- という態度を貫く事
 - 証拠の質の向上には、これが不可欠

EBL の成立には何が必要か？ 3/4

- これが意味する事

- 特定の表示 R へのコミットメントは必要最小限にすべき

- 理由

- R が妥当かどうか？ (例えば統語構造が樹状構造かどうか？)
は、それを支持する証拠 (=肯定証拠) があっても、他のすべての可能性を排除する証拠がない限り、確立したものだとは言えない
- 統語構造は木構造じゃなくてラティスじゃないの？と言われたら、
どうする？

厄介な例 1/3

- 重複形態素1
 - 認知言語学会
 - ??認知言語学学会
- 重複形態素1
 - a. 認知言語学会長
 - b. 認知言語学学会長
 - c. 認知言語学会会長
 - d. ??認知言語学学会会長
- 重複形態素2
 - a. 国内外
 - b. 国内国外
- 重複形態素3
 - a. 日本国内外
 - b. ?*日本国国内外
 - c. 日本国内国外
 - d. ?日本国内日国外
- 一般にカッコ入れの逆理 (bracketing paradox) の例はどれも樹状構造による記述の妥当性に挑戦
- これは形態論に限った話?
 - だと良いですけど、その保証は全然ない

厄介な例 2/3

- パロディーの認識 1
 - a. {i. ダイエット; ii. 外国語学習; iii. 相場; vi. 婚活; v. 脆弱性対策; vi. 物件選び} に王道なし
 - b. 学問に {i. 近道; ii. 横道; iii. 国境} なし
 - c. 学問に抜け道あり
- パロディーの認識 2
 - a. 触らぬ {i. ブログ; ii. 姑; iii. クレーマー} に崇りなし
 - b. 下らぬ株に崇りなし
 - c. 触れる神に崇りあり

厄介な例 3/3

- RTの違い
 - a. その絵は壁にかかっていた
 - b. ?その男は壁にかかっていた
 - c. ?その絵は医者にかかっていた
 - d. その男は医者にかかっていた
- b, c は (容認不能でないにせよ) 解釈に時間がかかる

補足

- これらの例は pattern lattice モデル (PLM) で記述できると思います
 - PLMの詳細は Kuroda (2009), 黒田 (2011) をご参照下さい
- が, 何せ一般に受容されてはいません

EBL の成立には何が必要か？ 4/4

- EBL の成立に具体的に必要なのは，少なくとも次の事
 - 言語データの正例 positives と負例 negatives の関係論の確立
 - そのための，容認性判断 acceptability judgment の(統計的)実態の解明
 - 表示の理論は後回しで良いし，そうすべき
- 注意
 - 負例は，文脈さえ用意すれば存在しないと言う事は，問題の解決にならない

現状

- 現状のデータ駆動の方法論 (e.g, コーパス言語学) では, 正例と負例がうまく区別できない
 - コーパス中には (定義によって) 負例はない
 - (負例の) 作例には過剰生成力をもつ生成システムが必要だが, コーパス駆動の生成システムはあるとは聞かない
- 間接否定証拠 indirect negative evidence で十分?
- それが利用できていると言う主張 (Stefanowitsch 2006, 2008; Clark and Lappin 2009; Yang to appear) もあるけれど, 期待頻度と実頻度の落差=間接否定証拠性が大きい事が, 非容認性の必要十分条件であるとは私には思えない
 - 条件として十分なだけで, 必要ではないのか? 必要だが, 十分とは言えないのか? すら判っていない

容認性判断の実態の解明

- 私見では、**容認性判定/判断** acceptability judgment **の統計的実態**をしっかりと理解する事が言語学の研究の今後の進展のために必須
 - 言語刺激 S への、十分に多い数の人々の反応が一般に一様なら、何も問題はない
 - だが、そんなことは経験から言ってあり得ない
 - 自分の研究の小規模の調査からも、NICT時代(2009-2010)のアノテーション監督経験から言っても、

調査例1 1/6

- F1

- a. 太郎が校庭を走る
- b. 稲妻が北の空を走る
- c. 戦慄が永田町を走る
- d. 汗が太郎の額を走る

- F2

- a. 校庭を太郎が走る
- b. 北の空を稲妻が走る
- c. 永田町を戦慄が走る
- d. 太郎の額を汗が走る

- G1

- a. 太郎が校庭に走る
- b. 稲妻が北の空に走る
- c. 戦慄が永田町に走る
- d. 汗が太郎の額に走る

- G2

- a. 校庭に太郎が走る
- b. 北の空に稲妻が走る
- c. 永田町に戦慄が走る
- d. 太郎の額に汗が走る

調査例1 2/6

- 実験
 - 16人の評定者に次の4段階の規準で F1, F2, G1, G2 の12文を評定してもらった。
- 評定基準
 - i. まったく違和感を感じない
 - ii. 軽く違和感を感じるが、言おうとしていることは簡単にわかる
 - iii. 強く違和感を感じるが、言いたいことがまったくわからないわけではない
 - iv. 何を言っているかわからないか、明らかに異常なことを言っていると思う
- i, ii, iii, iv を 3点, 2点, 1点, 0点として数値評価

発表後の補足

- 調査の難点

- 1. 提示順序をランダム化していないので、信頼性のある結果が得られている実験条件とは言えない
- 2. N=16は多いとは言えない
 - 更に自分も入っている
- 3. 被験者が全員、言語学研究室の院生、ないしはw研究者なので、実は偏りがある
- 1 はかなり致命的な難点

調査例1 3/6

Expression	Ind	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	av.	stdev
太郎が 校庭を 走る	F1a	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	0.00
永田町に 戦慄が 走る	G2c	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.00	2.88	0.50
北の空に 稲妻が 走る	G2b	2.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.00	2.81	0.54
校庭を 太郎が 走る	F2a	3.00	3.00	3.00	3.00	3.00	3.00	3.00	2.00	2.00	3.00	2.00	3.00	2.00	2.00	3.00	2.00	2.63	0.50
北の空を 稲妻が 走る	F2b	3.00	3.00	3.00	3.00	3.00	3.00	2.00	3.00	2.00	2.00	1.00	3.00	2.00	2.00	2.00	3.00	2.50	0.63
稲妻が 北の空を 走る	F1b	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.00	2.00	2.00	2.00	2.00	1.00	1.00	3.00	2.38	0.81
戦慄が 永田町に 走る	G1c	3.00	2.00	3.00	3.00	3.00	1.00	3.00	1.00	2.00	2.00	1.00	3.00	2.00	1.00	3.00	1.00	2.13	0.89
永田町を 戦慄が 走る	F2c	3.00	2.00	3.00	1.00	1.00	1.00	2.00	3.00	2.00	3.00	3.00	2.00	1.00	2.00	2.00	2.00	2.06	0.77
戦慄が 永田町を 走る	F1c	3.00	3.00	3.00	1.00	2.00	2.00	3.00	3.00	1.00	2.00	1.00	1.00	1.00	1.00	2.00	3.00	2.00	0.89
太郎の額を 汗が 走る	F2d	3.00	3.00	3.00	2.00	3.00	2.00	2.00	2.00	2.00	1.00	1.00	3.00	1.00	1.00	1.00	2.00	2.00	0.82
稲妻が 北の空に 走る	G1b	3.00	3.00	3.00	3.00	3.00	2.00	3.00	1.00	2.00	0.00	2.00	1.00	3.00	1.00	1.00	1.00	2.00	1.03
太郎の額に 汗が 走る	G2d	1.00	3.00	3.00	2.00	3.00	2.00	3.00	3.00	3.00	1.00	2.00	2.00	1.00	2.00	0.00	1.00	2.00	0.97
汗が 太郎の額を 走る	F1d	3.00	3.00	2.00	1.00	3.00	3.00	2.00	1.00	2.00	1.00	2.00	2.00	0.00	1.00	0.00	2.00	1.75	1.00
汗が 太郎の額に 走る	G1d	3.00	2.00	2.00	2.00	3.00	1.00	2.00	1.00	2.00	1.00	1.00	2.00	1.00	1.00	1.00	3.00	1.75	0.77
太郎が 校庭に 走る	G1a	3.00	1.00	0.00	1.00	1.00	3.00	2.00	1.00	1.00	0.00	0.00	0.00	1.00	3.00	3.00	3.00	1.44	1.21
校庭に 太郎が 走る	G2a	2.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	2.00	1.00	0.00	1.00	3.00	1.00	1.06	0.77
av.		2.75	2.56	2.50	2.19	2.56	2.25	2.50	2.13	2.00	1.69	1.81	2.13	1.63	1.75	1.94	2.00		
stdev		0.58	0.73	1.03	0.91	0.81	0.86	0.63	0.96	0.73	1.14	0.91	0.96	1.02	0.86	1.12	0.89		

表 1: $E \times R$

r は $av.$ と $stdev$ でソート

調査例1 4/6

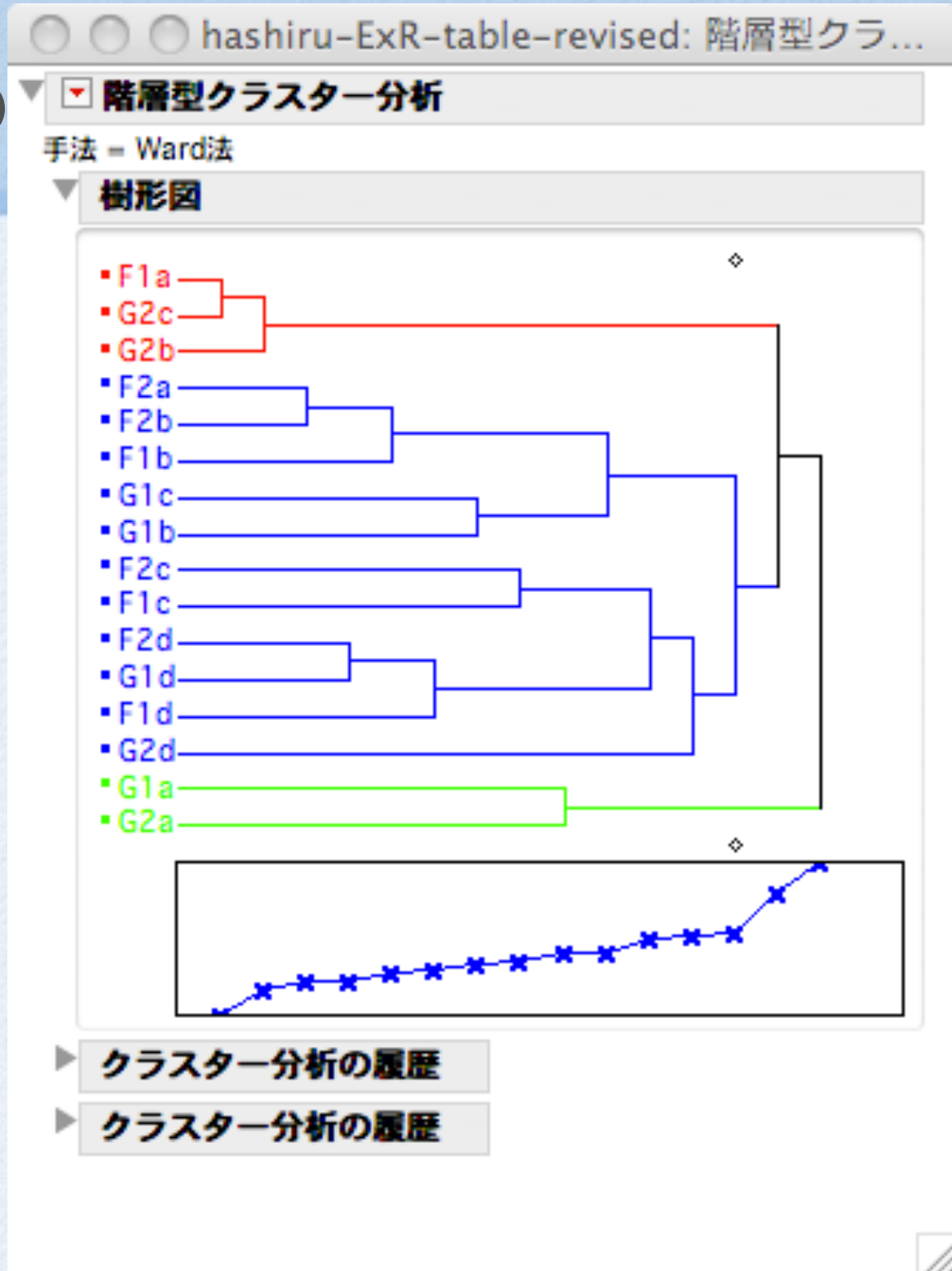
Rat	F1a	G2c	G2b	F2a	F2b	F1b	G1c	F2c	F1c	F2d	G2d	G1b	G1d	F1d	G1a	G2a	av	stdev
r3	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	2.00	2.00	0.00	0.00	2.50	1.03
r5	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.00	2.00	3.00	3.00	3.00	3.00	3.00	1.00	1.00	2.56	0.81
r4	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.00	1.00	2.00	2.00	3.00	2.00	1.00	1.00	1.00	2.19	0.91
r2	3.00	3.00	3.00	3.00	3.00	3.00	2.00	2.00	3.00	3.00	3.00	3.00	2.00	3.00	1.00	1.00	2.56	0.73
r6	3.00	3.00	3.00	3.00	3.00	3.00	1.00	1.00	2.00	2.00	2.00	2.00	1.00	3.00	3.00	1.00	2.25	0.86
r12	3.00	3.00	3.00	3.00	3.00	2.00	3.00	2.00	1.00	3.00	2.00	1.00	2.00	2.00	0.00	1.00	2.13	0.96
r7	3.00	3.00	3.00	3.00	2.00	3.00	3.00	2.00	3.00	2.00	3.00	3.00	2.00	2.00	2.00	1.00	2.50	0.63
r10	3.00	3.00	3.00	3.00	2.00	2.00	2.00	3.00	2.00	1.00	1.00	0.00	1.00	1.00	0.00	0.00	1.69	1.14
r15	3.00	3.00	3.00	3.00	2.00	1.00	3.00	2.00	2.00	1.00	0.00	1.00	1.00	0.00	3.00	3.00	1.94	1.12
r8	3.00	3.00	3.00	2.00	3.00	3.00	1.00	3.00	3.00	2.00	3.00	1.00	1.00	1.00	1.00	1.00	2.13	0.96
r13	3.00	3.00	3.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	1.00	3.00	1.00	0.00	1.00	0.00	1.63	1.02
r9	3.00	3.00	3.00	2.00	2.00	1.00	2.00	2.00	1.00	2.00	3.00	2.00	2.00	2.00	1.00	1.00	2.00	0.73
r14	3.00	3.00	3.00	2.00	2.00	1.00	1.00	2.00	1.00	1.00	2.00	1.00	1.00	1.00	3.00	1.00	1.75	0.86
r11	3.00	3.00	3.00	2.00	1.00	2.00	1.00	3.00	1.00	1.00	2.00	2.00	1.00	2.00	0.00	2.00	1.81	0.91
r1	3.00	3.00	2.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.00	3.00	3.00	3.00	3.00	2.00	2.75	0.58
r16	3.00	1.00	1.00	2.00	3.00	3.00	1.00	2.00	3.00	2.00	1.00	1.00	3.00	2.00	3.00	1.00	2.00	0.89
av	3.00	2.88	2.81	2.63	2.50	2.38	2.13	2.06	2.00	2.00	2.00	2.00	1.75	1.75	1.44	1.06	1.06	1.06
stdev	0.00	0.50	0.54	0.50	0.63	0.81	0.89	0.77	0.89	0.82	0.97	1.03	0.77	1.00	1.21	0.77	0.77	0.77

表 2: $R \times E$

F は av. と stdev でソート

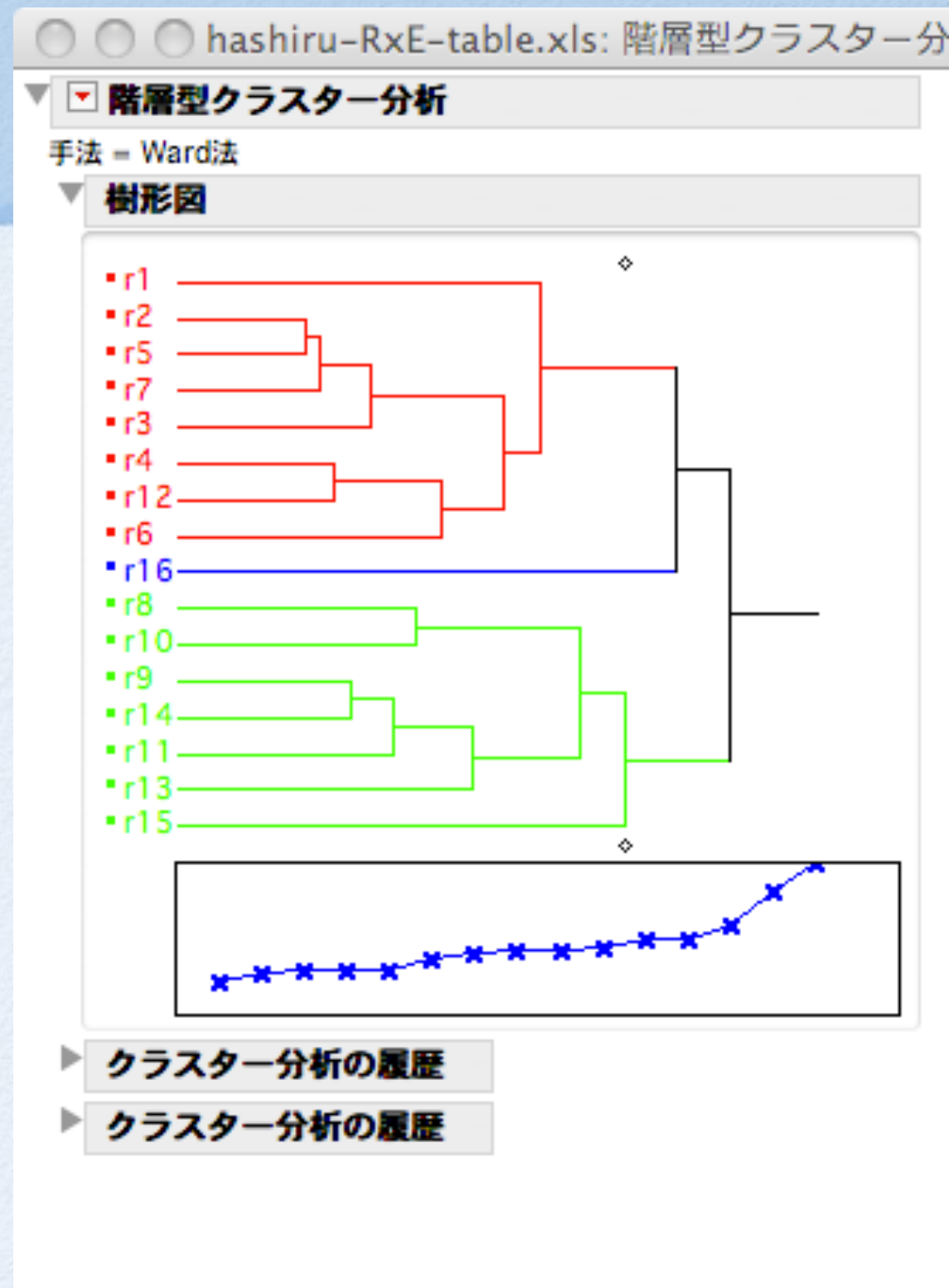
調査例1 5/6

- 表1 ExR の数値データを，E について階層クラスター解析した結果
 - クラスタ-1: {F1a, G2c, G2b}
 - クラスタ-2: {F2a, F2b, F1b, G1c, G1b}
 - クラスタ-3: {F2c, F1c, F2d, G1d, F1d, G2d}
 - クラスタ-4: {G1a, G2a}



調査例1 6/6

- 表1 E x R の数値データを, R について階層クラスタ解析した結果
 - クラスタ1: {r1, r2, r5, ..., r6}
 - クラスタ2: {r16}
 - クラスタ3: {r8, r10, r9, ..., r13, r15}



要点 1/4

- 容認度を構成するのは**程度の差**で，連続分布
 - 容認される vs されないの二値対立は
 - 容認度高い \Leftrightarrow 容認度低い
 - の分布の両極(性)に過ぎない
 - 反応に三つのクラスターがある事からわかる
 - ただ，容認度分布は非線型 (で，おそらくシグモイダル)
 - クラスター2が更に2分岐する事からわかる

要点 2/4

- 講演者からの挑戦 1

- 一般にどんな表現=刺激に，どんな容認度分布が対応するか？
という問い答える事が，言語学の究極の目的の一つ
 - ですよ？

- 刺激集合 E と反応者集合 R があり，

- $r_i \in R$ を固定した時に，系列 $e_1, e_2, \dots (e_i \in E)$ の評定値は？
- これは，ベイズ推定を使えば原理的に答えられる問い

要点 3/4

- 容認性の評価には幾つかの戦略がある
 - R のクラスター分析の結果が、3グループに分れている事からわかる
 - グループ 2 = {r16} の反応が一貫しているか怪しい事もわかる

要点 4/4

- 講演者からの挑戦 2
 - 容認度判定の戦略の記述 = 一般に誰がいつ、どんな言語表現をどう容認するか?を説明する事も、言語学の重要な課題の一つ
 - ですよね?
- 刺激集合 E と反応者集合 R があり,
 - $e_i \in E$ を固定した時に、系列 $r_1, r_2, \dots (r_i \in R)$ のタイプ値は?
- これは、ベイズ推定を使えば原理的に答えられる問い

容認性判断の研究例2

- 斎藤 (2015) の研究
 - 無作為に作成した新造 NN複合語に対し, 5件法の容認度判定を実施
 - 刺激となる N1-N2 に触れた事がないという条件を作り出すのが調査の肝
- 結果を H, M, L とする
- H, M, L について, 要素の直積の 4字 XY 複合語 (X, Yは2文字名詞)を作成
- これらから i) X=Yの場合と ii) 実例のあるものを除いて実験刺激とする
- 目的
 - 容認度を決めているのは頻度か?
- 刺激の作成
 - BCCWJから高/中/低頻度域ごとに2文字名詞を無作為に抽出
- 結果
 - 頻度は N1-N2 の容認度を説明しない
 - 反応を説明するのは, N2のスキーマ性
 - X先生, X主義

発表後の補足

- 任意の人 p の容認度判定力の信頼性を格づけする方法
 - 1. 事前に、一般の人の容認度評定値と分散がわかっている表現集合 S を用意する
 - s_i の評定値の平均値を a_i , 分散を v_i とする
 - 2. 評定を受ける人 p は, S の個々の要素 s_i について, (i) s_i の容認度評定をし, かつ (ii) s_i の分散の大きさを予測してもらう
- p による s_i の評定値を a_p , 予測分散値を v_p とする
- 3. $|a_i - a_p|$ の差の総和 A_p と $|v_i - v_p|$ の差の総和 V_p を求める
- 4. A_p, V_p と高水準 H と低水準 L の二つの参照値との相対位置
 - $A_p < AH, H < A_p < L, L < A_p, V_p < H, H < V_p < L, L < V_p$
- によって,
 - **H**: 信頼の高い判断をする人, **M**: 信頼できる判断をする人, **L**: 信頼の低い判断をする人
- の三つに格づけできる

EBMを見習うと？

メタファー研究の例

- 概念メタファー研究に必要な事
 - (i) 修辞性の有無と (ii) 写像の関与の有無を要因として分離させる
- 具体的手順
 - 1. サンプルを取るデータ D を決める
 - 例えば $D = \text{BCCWJ}$ 全体
 - 2. D から無作為抽出してサンプル S を得る
 - 3. S の全表現について、比喩性の感じられる表現を網羅的に列挙する。結果を T とする。
 - この際、修辞性の理由は問わない
 - 4. S の全表現について、領域間写像の関与する表現を網羅的に列挙する。結果を U とする。
 - 5. T と U の関係を調べる。
 - 場合 A: $T = U$
 - 場合 B: T が U に真に包含される
 - 場合 C: T と U の積集合が空でない
 - CMTの主張「メタファーは比喩写像である」がどれくらい妥当かは、A, B, Cのどれが正しいかに拠る
- 重要な点
 - 手順 3 と手順 4 を独立に行う

言語学にとって統計とは何か
答えの出ていない個人的思索

疑問

- 言語学は直観なしで済ませるべきなのか？
 - 言語学が医療と同質なら
 - 自然科学と違って，得られるデータや証拠にノイズが多い
 - 対象の挙動が複雑過ぎて，わかる事が限定的
 - 交絡 confounding という特有な現象
- 私見では
 - 直観なしで済ませるべきではない
 - 個別性を無視できる訳ではないし，すべきでない
- とは言え...

統計を使う事の意味 1/3

- EBMの難点

- EBMに反対する人たち (Hickery and Roberts 2011) の言い分の一つが、患者/疾患の個別性を統計で処理するのは誤りだという指摘

- これが意味する事

- 統計が意味をもつのがどんな場合か自覚していないと、とんでもない事になる

統計を使う事の意味 2/3

- 医療の対象の二面性
 - 患者/疾病という対象は,
 - 一面では、平均化可能な、個別性のない存在であるが、
 - 他面では、平均化不可能な、個別な存在である
 - 統計が意味をもつのは前者の場合
 - 研究対象がこの意味での二面性を持つのは、人文学の研究でも同じ事

統計を使う事の意味 3/3

- 言語研究で統計が意味を持つ事は明らか
- ただし、理由には二つの質的に異なる可能性が考えられる
 - I. 言語という知識の実態が統計的 (表示的統計解釈)
 - II. 言語という知識の実態は膨大な事例記憶で、記憶というデータを処理する方法が統計 (処理的統計解釈)
- 選択
 - 統計基盤か？ 記憶基盤か？
- 私見
 - パロディの認識力の一般性を考えると、後者が正しい見こみの方が高い
 - NLPの処理系としては、(Big Dataとか言っていますから) I から II に以降中？

興味深い対比 1/2

- 疫学に携わる統計学者の重要な実績は、彼らが次のような難題に日々直面していることを考えると、いっそう感心させられる。
 - 最小限のデータ (10件を下回るケースから判断を下さなければならない)
 - 緊急性 (患者が死にかけている)
 - 不完全な情報 (聞き取り調査には「思い出せない」という回答も)
 - 信頼できない情報 (人は想像で語るときがある)
 - 原因を必ず見つけなければならない (この至上命題がさまざまな間違いにつながる)
 - 間違いがもたらす結果の重大さ (説明するまでもない)
 - カイザー・ファング『ヤバい統計学』, Loc. 1003より

興味深い対比 2/2

- これは、統計学者にとって当たり前の世界ではない。他の統計学者はもっと寛大な状況に恵まれている。
 - 豊富なデータ (文字どおり数百万件を分析する)
 - 時間的余裕 (結論は繰り返し検証され、精緻なものになっていく)
 - パターンだけに注目する (原因はほとんど気にしない)
 - リスクが小さい (誰かが死ぬわけではない)
 - カイザー・ファンク『ヤバい統計学』, Loc. 1003より

発展的問題

- 言語学は疫学者と普通の統計学者のどっちに近いのか？
 - 私には疫学者のように思える
 - 理由
 - 「豊富なデータ (文字どおり数百万件を分析する)」と言うけれど、それで十分なのか？
 - ヒトの言語が世界に誕生してから (外延として) 存在した言語表現の全体 W があると考えると、現代の言語データ C がどんなに膨大でも、 W の極く一部でしかないのは確実
 - C を基に W を推定するのは、サンプリングの偏りが大き過ぎて、無理
- 上の問いの答えが疫学者だとすると、それから敷衍できる事
 - 言語学は統計解析に帰着できる訳でない
 - 個人の言語は、(患者/疾病と同じく) 個別な歴史の上に成立する複雑系で
 - 対照実験の反復が無条件に想定可能でない
 - 無作為化が集団知識として言語を表わしても、それは個人の言語を表わさない
 - ただ、最初の問いに対する答えは研究者によって違って構わない
 - 折衷主義を受入れる方が無難

Part I のまとめ

- 言語学の方法論の難点を三つ指摘
- その解決モデルとしてEBMの取り組みを紹介
- EBL の成立に必要な条件の考察と実装例の紹介
- 統計と言語との関係に関する、答えの出ていない思索

Part II

言語学と言語処理 の関係を整理する

注意

以後の内容は，9/6の発表
で話していません



はじめに 1/2

- 私の言語観
 - 言語学は言語研究の基礎部門で、言語処理が応用部門
 - 物理学が基礎で工学が応用、生物化学が基礎で農学が応用、...
 - 理論の正しさを実証するのは、実用/応用である
 - 研究は確証バイアスの影響を受けるので、実用/応用より信頼性が低い

はじめに 2/2

- 私の不満
 - 応用研究と基礎研究の乖離が絶望的に大きい
 - 言語処理で生じた実用上の必要性から言語学が発展する事はあって良いし、あるべき
 - 科学史を見ても、基礎が先で応用が後という順序づけはしばしば誤り
- 以上は、黒田 (2012): 言語学と言語処理の共生は可能か？の論点
- 以後、これらに類例と論点を追加

言語学での応用の軽視

- 言語処理 (NLP) と言語学 (Ling) が必ずしも仲良くやれてない最大の理由は、**言語学の多数派が応用研究を軽視しているから**
 - 言語処理関係者が基礎を作り直さないといけないことが多い
- ただ、これは言語学に限った事ではなく、数学と応用数学の間にも高頻度に見られる事
- 例えば

統計学ですら応用の軽視 1/3

- 『異端の統計学ベイズ』, pp. 323-324 から長目に引用
 - イェジ・ネイマンがカリフォルニア大学バークレー校に気づいた頻度主義の要塞は、一九七〇年代を通してアメリカの統計センターだった。スタンフォード大学の大規模な統計学部もまた、マッカーシー旋風の[中]で忠誠の誓いに著名する[事]を拒んだチャールズ・スタインなどのカリフォルニア大学の教授たちに支えられた熱烈な頻度主義の牙城で、教授たちの研究室の扉には反ベイズの標識が掲げられていた
 - 一方ベイズ派は、伸び悩んでいた。立ち泳ぎをしながら、知らず[知らず]コンピュータの進歩が追いつくのを待っているような格好だった。

統計学ですら応用の軽視 2/3

- 使いやすく強力なコンピュータもソフトウェアもなかったのに、ベイズ派の多くが—[更]には反ベイズ派の人々も—統計を現実に応用する[事]はできないと考えて理論的数学に引きこもった。海軍研究事務所の問題に触発されて統計学の業績を上げる[事]が多かったハーバード・チャーノフは、入念な一般化を目指して邁進する理論家たちにすっかり[苛立]って、一九七四年にはスタンフォード大学からMITに移り、[更]にハーバード大学に転じた。「ついに、我々が一段と集中的にコンピュータと向き合わなければならない時代がやって[来]た。それはまた、もっとたくさんの応用研究をしなければならない時代でもある。……思うに、この先我々が目指すべき方向に関して[何]がしかの洞察を得たければ、理論に磨きをかける[事]に集中するのではなく、現実への応用との接点をもっと増やすべきだ」。

統計学ですら応用の軽視 3/3

- この後に次のように続くが、これは本発表とは無関係
 - チャーノフはベイズ派ではなかったが、当時研究者として第一歩を踏み出したばかりだった統計学者のスーザン・[ホームズ]に、難しい問題に直面した時に構えを次のように説いている。「その問題について、まずベイズ流のやり方で考えてみる。すると正解が得られるから、[後]はそれが正しい[事]を、好みの方法で証明すればよろしい」

論点の補足

- 科学は基礎部門より先に応用部門で進歩する事が稀でない(と言うより, その方が多い)
 - 統計学(部)を言語学(科)に置き換えると身につまされる
- ただし, 私自身は
 - 「数学/言語学はコンピュータとアルゴリズムがありさえすれば, 数学/言語学はまったくの用済みになる」とは思わないし,
 - 言語学が応用統計学に還元可能だとは考えない

教条を捨てる事の大切さ 1/2

- 『異端の統計学ベイズ』, p.383
 - ベイズ派も頻度主義者も、次元の呪いには大いに悩まされた。大学人統計研究者の世界では、[相変]わらず大勢の人々が、そもそもコンピュータを駆使した分析に[溺]れて[良]いものかどうかを議論していた。この当時の統計学者はたいてい数学者で、古き[良]き手動式のブルンスビガや電動式のファシットなどの計算機と新たに登場した電子計算機をごっちゃにする者が多かった。[...]ベイズ派はロバート・シュレイファーやハワード・ライフアと[言]った先駆者たちのお[陰]で、ビジネススクールや理論経済学の世界でこそ幅を利かせていたが、大学の統計学部は[相]も[変]わらず、未知数がごまんとあるデータではなく、未知数がほとんどないデータセットに重点を置く頻度主義者たちの牙城だった。

教条を捨てる事の大切さ 2/2

- 『異端の統計学ベイズ』, p.383
 - その結果, 多くの統計学部が, プレート・テクトニクスやパルサーや進化生物学, 汚染や環境や経済や健康, 法律や社会科学などのデータを分析している物理学者や生物学者たちをただ傍観する[事]になった. [統計学者ではなく]エンジニアや計量経済学者やコンピュータ科学者や情報科学者たちはじきにマンネリに陥った統計学者たちには無縁な名声を手に入れ, 統計学部は孤立して守りに入り, もはや下り坂だと鼻であしらう声上がる始末だった.
 - 統計学の一流雑誌は[余]りに数学的でほとんど読める人がおらず, 現実とかけ離れ[過]ぎていて読もうとする人すらいないと[言]われた. そしてもっと若い連中は, コンピュータとアルゴリズムがありさえすれば, 数学はまったくの用済みになると考えているようだった.

追加する論点 1

- 折衷主義と教条主義からの離脱の勧め
- 私の研究と人生のモットー
 - 還元論には気をつけろ！
- 応用
 - 私は言語の特性が統計に帰着できないと考えるが、これがその理由の一つ

折衷主義の勧め 1/2

- 『異端の統計学ベイズ』, p.285
- [Federalist papersの著者推定で業績を上げたフレデリック] モステラーは、人の助言に耳を貸す人間だった。それに、サヴェッジやリンドレーやライフアやシュレイファーとは違って、熱狂的なベイズ派でもなかった。役に立つ技法を好む折衷派の問題解決者だったのだ。[更]に、信念の程度として確率[=ベイズ派の確率解釈]と相対頻度としての確率[=頻度派の確率解釈]は、どちらも正当だと考えていた。

折衷主義の勧め 2/2

- 『異端の統計学ベイズ』, p.285
 - モステラーの見たところ, 問題は「ハミルトンが五十二番目の論文を書いた」というような一回[切り]の出来事をサンプリング理論で処理するのは難しいという点にあった. ベイズ的な信念の度合い[としての確率解釈]の[方]が, 具体的に述べるのは難しくても, 応用範囲は広いはずだ.
 - しかもモステラーは, 矛盾を避けて教科書の例に逃げるよりも, 重要な社会問題に取り組む[事]を好んだ. 問題にリアリティーがある[方]が, 緊張感が増す. 本人曰く「安楽椅子に座ったままで」見つけた困難と, 現場や化学実験室で見つけた困難が似ている[事]は稀なのだ.

追加する論点 1

- 応用研究が開花する条件は,
 - 誰でも使えるデータと使いやすい解析ツール/ソフトウェアの普及

実用の敷居を下げる事の大切さ 1/3

- 『異端の統計学ベイズ』, pp.393-394
- [ベイズ派の普及に貢献した]リンドレーは一九六五年以降, 自分自身のコンピュータのプログラミングをしており, ベイズはコンピュータを使った計算にもってこいだと考えていた. 「公理とデータを入力しさえすれば, [後]はコンピュータを算術の法則に[従]わせるだけで[済]む」. リンドレーに[言]わせれば「ベイズのクランクを回す」のだ. ところがリンドレーの[下]で学んでいたエイドリアン・スミスには師には見えていないものが見えていた. ベイズを仕事場で役立つものにするには, より精練された理論ではなく計算の容易さが鍵になる. [...]

実用の敷居を下げる事の大切さ 2/3

- 『異端の統計学ベイズ』, pp.393-394
- スミスは多くの統計学部が守りの姿勢に入っているのには知らん顔で、まったく新たな方向に向けて攻撃を開始した。
[...]スミスはイタリア語を習い、デ・フィネッティ『確率論』と題する二巻本を英訳して世に問うた。こうして多くの英米の統計学者たちが、[初]めてデ・フィネッティの主観主義的アプローチを使えるようになった。[更]にスミスはフィルタを開発し、この実際的な計算の工夫が、後にベイズの計算を大いに楽にする[事]になった。

実用の敷居を下げる事の大切さ 3/3

- 『異端の統計学ベイズ』, pp.393-394
 - 次にスミスはリンドレー, ホセ・M・ベルナルドとモリス・デ[フ]ルートの三人と力を合わせて, ベイズ派のためにスペインのバレンシアで一連の国際会議を組織した. [...]
 - スミスは一九八四年にマニフェスト[...]を発表した. 「ベイズの手法が[更]に広く使われるかどうかは, 効率的な数値積分法の手順の有無によって決まる」. [...]
 - スミスはいかにも実際家らしく, 自分が所属してノッティンガム大学の学生たちに, 空間統計や疫学におけるベイズ統計の問題を解くのに必要なユーザーフレンドリーで効率的なソフトウェアを開発させる[事]にした.
- こうして開発されたのが, Markov Chain Monte Carlo法=MCMC法

論点の拡張 1/2

- 電子化された言語コーパスは今や誰にでも使えるデータになっている
- でも、言語研究で部外者にも使いやすい解析ツール/ソフトウェアは存在するのだろうか？
 - その前提になるのは特定の解析結果へのコミットメント

論点の拡張 2/2

- その前提になるのは特定の解析結果へのコミットメント
- ジレンマ発生!!
 - だが、特定の解析/表示へのコミットメントは必要最小限にすべき...
- これには私も答えが出せません

まとめ

- 言語学の方法論の難点を三つ指摘
- その解決の手本としてEBMの取り組みを紹介
- 言語学と言語処理の関係を整理する(私独自の)視点を紹介

ご静聴, ありがとうございます

Thank you for your attention.

Děkuji vám za pozornost.

Köszönöm a figyelmeted.

תודה על תשומת הלב שלך.

References 1/2

- 板倉 聖宣 (1999): 『新哲学入門』. 仮説社.
- 黒田, 航 (2011). 並列疑似エラー補正法に基づく「破格」な言語表現の(疑似)解釈: 「不自然処理」のための理論的枠組み. In 言語処理学会第17回年次大会発表論文集, 1139–1142.
- 黒田 航 (2012). 言語学と言語処理の共生は可能か? In 人工知能学会誌, Vol. 27, No. 3 の特集「ポスト経験主義の言語処理」, pp. 326–332. 2012.
- 斎藤 幹樹 (2015). 「容認性判断に関わる認知的要因: 認知文法的観点からの分析」. 未刊行修士論文, 京都大学.
- ギルバート, D. (2014): 『明日の幸福を科学する』. 早川書房. [Gilbert, D. (2010): Stumbling on Happiness. の邦訳]
- ファング, K. (2011): 『ヤバい統計学』. メディアハウス. [Fung (2010): Numbers Rule Your World}, 1st edition. McGraw-Hill Education, 2010の邦訳]
- ロスマン, K. J. (2004): 『ロスマンの疫学: 科学的思考への誘い』. 篠原出版新社. [Rothman (2002): Epidemiology: An Introduction の邦訳]

References 2/2

- Cochrane, Archibald and Max Blythe (1989). *One Man's Medicine: An Autobiography of Professor Archie Cochrane*. BMJ Publishing Group.
- Isaacs, D. and D. Fitzgerald: *Seven alternatives to evidence based medicine*. British Medical Journal 319: 1618, 1999.
- Kuroda, Kow (2009). Pattern lattice as a model for linguistic knowledge and performance. In *Proc. of the 23rd Pacific Asia Conference on Language, Information and Computation*, Vol.~1, pp. ,278–287.
- Stefanowitsh, Anatol (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory*, 2(1): 61--77.
- Stefanowitsch, Anatol (2008). Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics*, 19(3): 513–531.
- Yang, Charles (to appear). Negative knowledge from positive evidence. *Language*, ??(??):??.