

もし「主語」や「動詞」がなかったら

— 主語無し発話・動詞無し発話から探る会話と文法の関係性 —

吉川 正人 (慶應義塾大学)

1. はじめに: 会話と文法

「主語」や「動詞」という概念は自然言語の文法を考える際には極めて重要なものとされ、文法研究においては中心的な役割を担ってきた。しかし一方で、日常会話を典型とする話し言葉においては定型的な発話が多数であり、「主語」「動詞」「目的語」といった文法関係を用いた分析は適切ではないという指摘も少なからず見受けられる。

日常会話における文法の役割や位置付けについては、談話機能言語学 (Discourse-functional Linguistics)、あるいは近年では相互行為言語学 (Interactional Linguistics) と呼ばれる一連の研究によって盛んに研究が行われてきた。その中では、文法は旧来理論言語学で想定されていたような無味乾燥な規則の体系や認知的な演算システムとしてではなく、コミュニケーション上の問題解決を図るツールとして再解釈され (e.g., Du Bois, 1985), 「主語」「動詞」といった抽象度の高い文法概念よりも、「疑似分裂文 (pseudocleft sentences)」 (Hopper, 2001) や「目的格補文 (object complements)」 (Thompson, 2002) といった、特定の文法パターンや構文にフォーカスが当てられてきた。また例えば主語や目的語といった文の「基本要素」とされる要素が不足なく揃った「基本的」な文が日常会話に現れることがいかに稀であることを示す研究 (大野・サドラー・第十, 2020) や、書き言葉と話し言葉、新聞社説など、ジャンルやレジスター間でいかに文法的要素の頻度分布が異なるかを示す研究 (Iwasaki, 2015) も存在し、これらの研究は、書き言葉や作例をベースに構築されてきた抽象度の高い、画一的な文法観を疑う必要性を投げかけるものであると言える。

しかしながら、よく指摘されるように (e.g., Newmeyer, 2010), 頻度の偏りが大きかったとしても、「主語」「動詞」といった抽象度の高い文法概念を持ち出すことで有意義な記述のできる発話や、会話の参与者にとってもそのような概念を理解に利用していると考えた方が自然と思えるような「複雑」な発話も存在はしているはずであり、言語使用者たちがそのような文を産出・理解することが可能であるものまた事実である。以下の例文は全て話し言葉を収録した同一コーパス (Simpson, Briggs, Ovens, & Swales, 1999) の同一データ (mtg425jg004.cha) から取得したもののだが、確かに (1a) のように「主語」が存在しない発話や (1b) のように「主語」「動詞」「目的語」が認定できるものの、極めて定型性が高く、文法関係を用いた分析があまり意味をなさないような発話も会話データからは多数見受けられる。それでも、(1c) のように定型的な表現とは言い難い複雑な構造を持ち、会話の参与者にとっても文法関係で表わされるような文中の要素同士の対応関係を理解しなければ解釈が難しい発話も散見されるという事実は無視できない。

- (1) a. Sounds good.
- b. I like it.
- c. You can look at what I've already started to write up for the problem statement.

このように会話など話し言葉における文法関係の働きを考える際には、「必要／不必要」「重要／重要でない」のような二項対立で考えるべきではなく、問題の本質は、いつ、どのような状況で文法関係がものをいうような「複雑」な発話が用いられるか、そしてその背景にはどのような要因があるのか、ということであり、その点を詳らかにすることが、会話と文法の関係性を解き明かすためには必要不可欠である。そこで本稿では、そのための第一歩として、英語の文における「必須要素」と考えられる「主語」と「動詞」に着目し、これらのいずれか一方が欠けている、文として「不完全」あるいは「不適格」とされる発話に着目することで、会話における文法の役割の一端を解き明かすことを試みる。具体的には、英語の話し言葉を対象とした統語解析済みのコーパスを用い、「主語」あるいは「動詞」を含まない発話の分布を様々な視点から分析していく。

2. 分析

2.1 データ

分析には、CABank (MacWhinney & Wagner, 2010) に収録されている、アカデミックジャンルにおけるアメリカ英語の話し言葉コーパスである The Michigan Corpus of Academic Spoken English (MICASE: Simpson et al., 1999) を使

用した。MICASE はアカデミックジャンルに限定されているが、150 万語を越える比較的大きな規模 (1,848,364 語) の話し言葉コーパスである点、アカデミックジャンルの中でも、講義やディスカッションなどにとどまらず、オフィスアワーやキャンパスツアーなど、比較のカジュアルな会話データも含め幅広く収録されている点、そして何より、統語解析済み (syntactically-parsed) である点で、本稿の分析に最適であると判断した。

2.2 方法

データは以下の要領で前処理を行った (以降の処理は全て筆者がスクリプト言語 Python [ver. 2.7.13] を用いて作成したスクリプトを使用)。MICASE には 152 の会話ファイルが存在するが、下記の前処理の結果、45 ファイルが分析対象として抽出された。総発話数は 124,676、総語数は 1,337,779 である。

- (2) a. 対象となるコーパスのデータを読み込み、発話を抽出した上で、
- b. 書き起こしに用いられている種々の記号 (e.g., .+, [/]), および特殊文字 (e.g., ↗, ≡, °) を全て削除し、¹⁾
- c. 文字表記を全て小文字で統一させた；
- d. 発話の抽出に際しては、話者情報を表すヘッダー (e.g., ”*S1:”) が付されているものを一つの「発話」としてカウントし、複数行に渡る発話は全て一つの発話として統合した；
- e. 統語解析結果が付与されている発話を分析対象とするため、一部統語解析結果が付与されていない発話を除外し、残った発話数が 1000 を超える会話ファイルのみを分析対象とした。

分析の手順は以下の通りである：

- (3) a. コーパスに付与された統語構造 (依存構造) のタグで主語を表す SUBJ, および埋め込み節の主語を表わす CSUBJ を含む発話が、当該コーパスにおける各会話ファイルの発話全体のうちどれほどであるかを、「主語率 (SUBJ_rate)」として算出、
- b. 文の「主辞」となっている動詞を表わす ROOT というタグを含む発話が当該コーパスにおける各会話ファイルの発話全体のうちどれほどであるかを、「動詞率 (ROOT_rate)」として算出した上で、
- c. 上記 2 指標が会話の性質や会話上の他の要因 (疑問文率、話者交代密度) とどのような関係にあるかを分析した

主語率、動詞率については、会話ファイル全体の割合とともに、30 発話のウィンドウ幅を設定し、それを 1 発話ずつスライドさせることで、30 発話単位の主語率・動詞率 (30 発話中主語のある発話、動詞のある発話は何発話あるか) を算出し、その変化の傾向も分析の対象とした。疑問文率と話者交代密度についても同様で、30 発話のウィンドウ幅で疑問文 (発話末のクエスチョンマーク “?” で判断) が何発話あるかを算出したものを疑問文率 (question_rate), 前発話と話者が変わっている発話は何発話あるかを算出したものを話者交代密度 (turn-taking_density) とした。

主語・動詞の無い発話の例を、大規模講義のデータ (le1185su066.cha) よりいくつか紹介する (S1 は教員)：

- (4) a. S1: other thoughts about this ?
- b. S25: have a higher demand than others .
- c. S5: rewarding A while hoping for B .

なお (4b) は SUBJ が無い発話として解析されていたもので、文脈を見ると、先行する発話を動詞以降から言い直しているものと思われる。他の 2 例は主語・動詞共に欠落した発話として解析されている。

2.3 結果

MICASE の各ファイルには、Highly monologic から Highly interactive までの 5 段階 (Highly monologic, Mostly monologic, Mixed, Mostly interactive, Highly interactive) で会話データのインタラクティブ性の評定が付与されている。この 5 段階評価を、Highly monologic を 1, Highly interactive を 5、つまり、インタラクティブ性が高い会話データほど数値が高くなるように数値に変換し、ファイル全体の主語率・動詞率との相関をピアソンの相関係数を用いて求めた。結果、主語率・動詞率共に、中程度の負の相関を示す結果となった (主語率: $r = -0.556$, $p = 7.109e-05$ / 動詞率: $r = -0.48$, $p = 0.001$)。データの一覧を表 1 に示す。インタラクティブ性が低いデータが少ないのは発話数 1000 以上のデータに限定しているためである。

疑問文率および話者交代密度と主語率・動詞率の関係については、30 発話毎に計算された主語率・動詞率が疑問文率・話者交代密度とどの程度相関しているかをピアソンの相関係数で求め、その相関係数と、ファイル全体の主語率・動詞率との対応関係を分析の対象とした。これはいわば、「相関の相関」を見ていることになる。2つのファイルを例に、これらの指標の時系列ごとの変化をグラフにプロットしたものを図 1, 2 に示す。両ファイルは発話数、主

¹⁾ CABank の “Special CA characters” ページ <https://ca.talkbank.org/codes.html> より抽出した。

表 1: 各会話ファイルの発話数 (#gU) と主語率・動詞率・インタラクティブ性 (int)

	file_name	#gU	subj_rate	root_rate	int		file_name	#gU	subj_rate	root_rate	int
1	lab500su044.cha	1228	0.599	0.67	5	24	sgr385su057.cha	2409	0.625	0.692	5
2	stp125jg050.cha	2079	0.671	0.704	5	25	sgr195su127.cha	1746	0.592	0.676	5
3	adv700ju023.cha	1139	0.526	0.539	5	26	ofc320su153.cha	1102	0.51	0.543	5
4	sem475ju084.cha	1080	0.71	0.731	5	27	sgr175su123.cha	2969	0.546	0.621	5
5	sve999mx104.cha	3126	0.542	0.634	5	28	lab175su026.cha	2127	0.569	0.659	5
6	sgr565su144.cha	1461	0.63	0.667	5	29	ofc115su060.cha	3519	0.501	0.526	5
7	sgr999su146.cha	1654	0.645	0.675	5	30	stp450sg128.cha	1457	0.749	0.759	4
8	lab200ju018.cha	1246	0.624	0.676	5	31	sem545mg083.cha	2215	0.575	0.611	4
9	ofc195su116.cha	2760	0.505	0.532	5	32	lab500su089.cha	1122	0.721	0.719	4
10	ofc575mu046.cha	1159	0.568	0.592	5	33	stp545ju091.cha	1108	0.665	0.698	4
11	lab175su032.cha	1673	0.549	0.627	5	34	lel185su066.cha	1187	0.614	0.676	4
12	mtg999st015.cha	1617	0.582	0.596	5	35	sem140jg070.cha	2053	0.631	0.662	4
13	ofc270mg048.cha	2389	0.542	0.574	5	36	dis315ju101.cha	1291	0.659	0.692	4
14	sgr175mu126.cha	3158	0.563	0.618	5	37	tou999ju030.cha	1256	0.678	0.748	4
15	sgr999mx115.cha	2023	0.7	0.732	5	38	sem495su111.cha	1526	0.683	0.721	4
16	sve999mx148.cha	1553	0.537	0.643	5	39	sem475mx041.cha	1511	0.711	0.725	4
17	sgr200ju125.cha	2478	0.505	0.559	5	40	mtg270sg049.cha	1177	0.713	0.698	4
18	adv700ju047.cha	5992	0.645	0.645	5	41	ofc300ju149.cha	1045	0.571	0.619	4
19	ofc280su109.cha	1619	0.563	0.598	5	42	def270sf061.cha	1252	0.696	0.654	4
20	ofc150mu042.cha	2287	0.645	0.677	5	43	sem340jg072.cha	1417	0.766	0.771	3
21	sem300mu100.cha	1940	0.644	0.687	5	44	dis280su058.cha	1188	0.561	0.601	3
22	mtg425jg004.cha	1472	0.428	0.503	5	45	lel300su076.cha	1005	0.755	0.761	2
23	dis175ju081.cha	1023	0.545	0.595	5						

語率・動詞率が同程度であるが片や大規模講義、片やキャンパスツアーとかなり性質の異なる会話のデータとなっており、比較対照するのに適している。なお主語率・動詞率・疑問文率については、値そのものではなく、ファイル全体の主語率・動詞率・疑問文率との差をプロットしている。

全体の傾向としては主語率・動詞率と疑問文率・話者交代密度が逆相関していることが見てとれるだろう。また、大規模講義とキャンパスツアーで各指標の変動の仕方がかなり異なっているということも注目し得る。特に講義データにおいては基本的には教員が講義をしている場面が大半であり主語率・動詞率共に高い傾向があるが、中心付近で両指標の大きな「谷」が出現しており、その部分では疑問文率・話者交代密度共に高くなっているという点は興味深い。実際この区間は講義ではなく質疑応答・議論が行われており、それがデータに顕著に表れている。

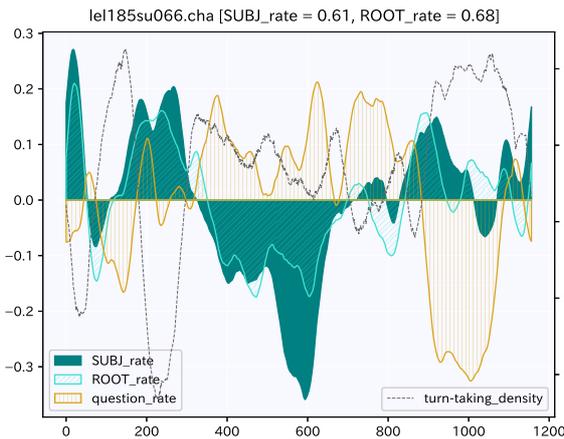


図 1: 大規模講義のデータ例 (lel185su066.cha)

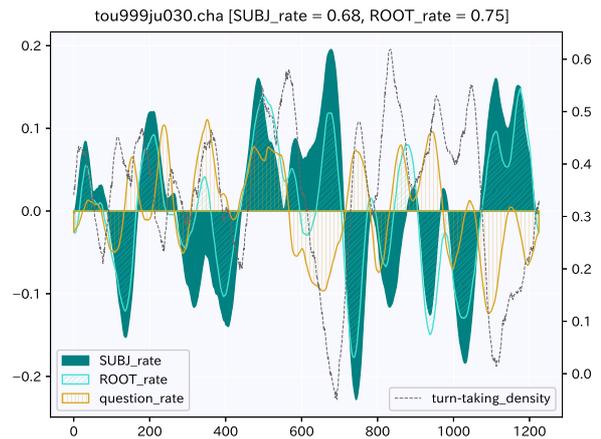


図 2: キャンパスツアーのデータ例 (tou999ju030.cha)

分析の結果は図 3, 4 に示した通りである。図から明らかなように、主語率・動詞率共に、値が高くなると疑問文率との相関が負の値に偏り、値が低くなると正の値に偏ることが見て取れる。話者交代密度については図の右上あたりにも多少分布があり、主語率・動詞率共にほぼ相関がないと思えるようなファイルがいくつか存在しているが、全体の傾向としてはやはり右肩下がりで、話者交代密度が低ければ主語率・動詞率が高くなり、またその逆も然り、という傾向が見て取れる。相関の相関 (ピアソンの相関係数) については、疑問文率が主語率 $r = -0.52$ ($p = 0.000$), 動

詞率 $r = -0.56$ ($p = 7.374e - 05$) で、話者交代密度は主語率 $r = -0.49$ ($p = 0.001$), 動詞率 $r = -0.28$ ($p = 0.061$) となっており、話者交代密度と動詞率の逆相関については有意傾向で、やや低いものという結果になった。

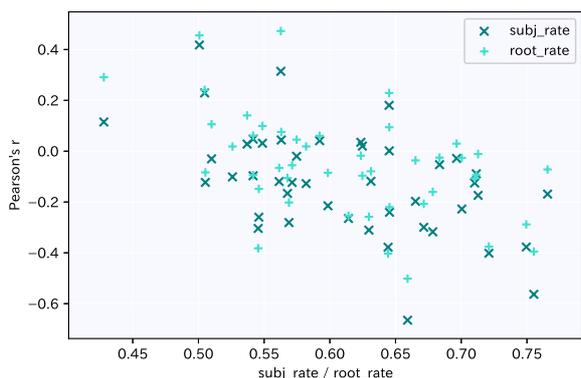


図 3: 主語率・動詞率と疑問文率との相関係数 (Pearson's r)

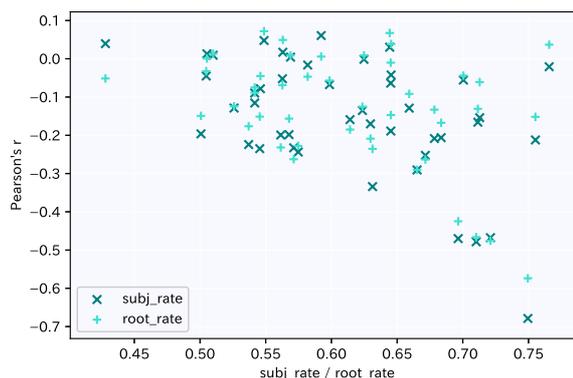


図 4: 主語率・動詞率と話者交代密度との相関係数 (Pearson's r)

3. まとめと考察・課題

分析の結果は以下のようにまとめられる:

- (5) a. 主語率・動詞率共に、会話の「インタラクティブ性」と逆相関する、つまり、インタラクティブな会話ほど主語・動詞ともに用いられない傾向がある
- b. 主語率・動詞率共に、高ければ高いほど、「疑問文率」「話者交代密度」と逆相関する、つまり、主語・動詞が用いられる発話が多い会話ほど、質問-応答のやり取りが行われるなど、話者交代が頻繁におこる場面では主語・動詞が用いられない傾向がある

主語/動詞を含むかどうかというのは文法関係の重要性を知る上では粗い近似であり、(1b) のような定型性の高い発話をうまく扱うことができていないという点は今後の課題であるが、この傾向から、会話であっても一人の参加者が比較的長いターンを取り詳細な情報を説明したり語ったりする場面では文法関係が重要な役割を果たす、という可能性が示唆される。今後はこのような数量的な分析だけでなく、主語率・動詞率の低い場面や高い場面においてどのような性質の発話が用いられているかといった意味・機能に着目した質的な分析を進めていくことで、会話と文法の関係についてさらなる洞察を深めていきたい。

参考文献

- Du Bois, J. W. (1985). Competing motivations. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 343–365). Amsterdam: John Benjamin.
- Hopper, P. J. (2001). Grammatical constructions and their discourse origins: Prototype or family resemblance. In M. Pütz, S. Niemeier, & R. Dirven (Eds.), *Applied cognitive linguistics i: theory and language acquisition* (pp. 109–129). Berlin; New York: Mouton de Gruyter.
- Iwasaki, S. (2015). A multiple-grammar model of speakers' linguistic knowledge. *Cognitive Linguistics*, 26(2), 161–210.
- MacWhinney, B., & Wagner, J. (2010). Transcribing, searching and data sharing: The clan software and the talkbank data repository. *Gesprachsforschung*.
- Newmeyer, F. J. (2010). What conversational english tells us about the nature of grammar: A critique of thompson's analysis of object complements. In K. Boyle & E. Engberg-Pedersen (Eds.), *Language usage and language structure* (pp. 3–44). Berlin; New York: Walter de Gruyter.
- 大野剛・サドラー美澄・第十早織. (2020). 日本語の日常会話における「基本的」他動詞節現実的な文法理論の構築を目指して. 中山俊秀・大谷直輝 (編), *認知言語学と談話機能言語学の有機的接点: 用法基盤モデルに基づく新展開* (pp. 165–182). 東京: ひつじ書房.
- Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (1999). *The michigan corpus of academic spoken english*. Ann Arbor, MI: The Regents of the University of Michigan.
- Thompson, S. A. (2002). “object complements” and conversation: Towards a realistic account. *Studies in Language*, 26(1), 125–163.