人と言語モデルが捉える文の主題

藤原 吏生
東北大学 工学部 電気情報物理工学科

1 はじめに

(1) a.兄は部屋に向かった。
b.兄が部屋に向かった。

二つの文は論理的に同等であるが，文脈や何についてメッセージを伝えたいか（主題）といった観点から人はこの二文を使い分ける．例えば，「兄がさっき帰ってきた。」というような兄に関する話に続く文であれば例 1a の方が自然であるし，単に兄が何をしたか客観的に描写するのであれば例 1b の方が好まれるだろう．このような先行文脈に依存する選択において計算モデルと人の判断に乖離があるか，近年流暢な文章の生成が可能になったとされるニューラル言語モデルに焦点を当てて調査する．

本研究では，文章の形成や評価において重要な側面である**文の主題**を取り上げる [1, 2]．日本語の係助詞「は」は主題であることを表し [3]，例 1 の二つの文は「兄」が主題か否かという点で異なる．ある要素を主題化するかの判断には，情報の新旧 [4] やセンタリング理論 [5] などの観点から先行文脈を考慮する必要があるとされている [3, 6]．**文の主題の観点で言語モデルが人と同様の選択を行えるか**を分析することで，言語モデルの文章生成における先行文脈の影響について考察する．

文の主題は英語では語順で示されるが [2]，日本語では係助詞「は」などによって明示され，その範囲を策定しやすいため日本語で研究を行う．はじめに，クラウドソーシング[1)]を活用し，NAIST テキストコーパス（NTC）[7] 中の主格に対して人の「は」と「が」の判断に関するアノテーションを行った．「は」と「が」の対立には主題と非主題といった対立以外にも，対比と排他，節末まで係るか文全体に係るかといった様々な軸がある [3, 6, 8]．NTC の共参照アノテーションや先行文脈を見せる/見せないことによる人の判断の変化から，「は」と「が」の選択

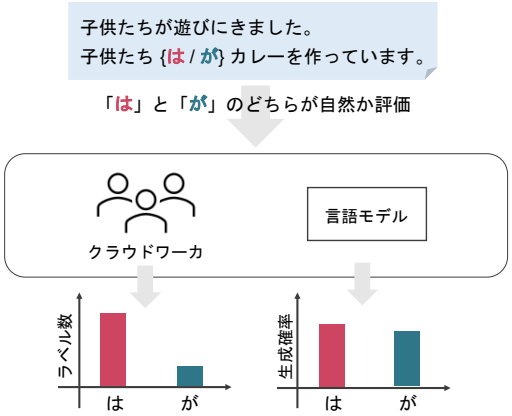


図 1 人と言語モデルによる「は」と「が」の選択

について特に文脈が依存するデータポイントを収集し，言語モデルの評価に用いた．

実験では，文脈依存の「は」と「が」の選択について，人とニューラル言語モデルの傾向を比較した (図 1)．人と言語モデルの選択の相関は高く，**言語モデルは人と近い選択を行うことができていた**．しかしながら，インスタンスごとの判断の不一致度 (agreement error) [9] や先行文脈の有無による判断の変化といった観点から分析を行うと，人と言語モデルの間に乖離が観察され，特に言語モデルは先行文脈を見なくても人と比べて不当に高精度な選択ができていた．これらの結果から，「は」と「が」の選択について言語モデルは人と近い選択が可能であるが，**判断の拠り所は両者で異なる**ということが示唆された．本研究で作成したデータセットは公開する[2)]．

2 関連研究

選択体系機能文法では，主題は「心理的主語」として「話者が伝えたいことの産出にとりかかる時にまず心の中にもつもの」と定義され [2]，文法的な主語とは異なるものである[3)]．文章の形成や評価にお

1) Yahoo!クラウドソーシング https://crowdsourcing.yahoo.co.jp/ を利用した．また，毎日新聞社に許可を得て NAIST テキストコーパス中の文をワーカに見せている．

2) https://github.com/rk-fujifuji/ntc_topicalization.git

3) 例えば「像は鼻が長い」のように，文法的主語と主題は必ずしも一致しない．

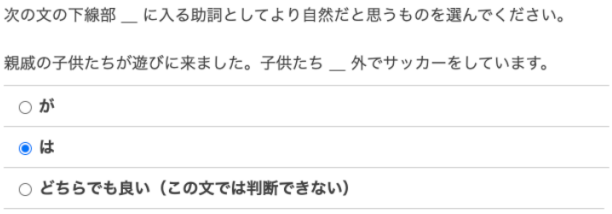


図 2　クラウドソーシングにおける設問の例.

いて，主題の構造は結束性などと並び重要な側面であるとされている [1, 2]．また，例 1 のような「は」と「が」の使い分けは日本語学習者にとって難しいとされており [10]，計算モデルが自然に使い分けられるのであれば，ライティング支援，学習者支援にも繋がる．更に日本語では主題が省略される傾向もあり，どの要素を主題とするのが自然かといった自動評価は，どの要素を省略すると自然かといった研究・評価にも繋がると考えられる．

言語モデルによる文章の容認性判断に関する既存研究では，文レベルの文法性判断や [11, 12, 13, 14]，論理的な能力 [15, 16] に焦点を当てた分析が多い．一方で，言語モデルの分析の文脈で文の主題に直接焦点を当てた研究は我々の知る限り存在しない．

3　データ作成

3.1　クラウドソーシング

NTC [7] 中の以下の条件を全て満たす主格をアノテーション対象とした．

- 文内で最も後方に出現する動詞の主格である
- 格助詞「が」または係助詞「は」を伴う
- 各文章の先頭から 2-4 文目に出現する

上の条件を満たした主格について付随する助詞を隠し，クラウドワーカーに「は」と「が」のどちらを使用するのが適切か聞いた（図 2）．「どちらでもよい (この文では判断できない)」という選択肢も用意した．それぞれの主格について，(i) 同一文章内の先行文脈を全て見せる場合と (ii) 文脈を見せない場合（特定の文脈を想像しない）の 2 つの設定でアノテーションをした．両設定では異なるワーカがアノテーションをしている．また，ある問題について文脈ありの設問を解いた後，同一の文脈なしの設問を解くということが生じないようにした．

事前に同様のタスクを複数回実施して優秀なワーカを選別し，さらにチェック設問を設定して不適切なラベルを付与するワーカを適宜除外した．この時点で 234 人のワーカから，各主格，各設定（文脈の有無）に対して 8 人のラベルを得た．さらに統計的後処理[4] により信頼度下位 30%のワーカを除外し，結果的に 3 人以下のラベルしか付与されていないデータポイントも除外した．最終的に 1,562 の主格に対して，文脈の有無の両設定それぞれで平均 6 人から，計 18643 ラベルが付与された．得られたデータセットにおけるアノテーション一致度はクリッペンドルフの α で 0.694 であり，信頼可能なデータの一致度の下限とされる 0.667 を超えている [18]．

表 1　サブセットの統計

サブセット	ラベル付けした主格の数	ラベル数
文脈依存の「は」	437	5,205
文脈非依存の「は」	437	5,255
「が」全体	688	8,183
計	1,562	18,643

3.2　サブセットの作成

1 節で言及したとおり，「は」と「が」の対立には主題か非主題かという軸だけでなく，対比を表すか排他を表すか，判断文であるか現象文であるか，どこまで係るかなどの様々な対立が内包されている [3, 8]．先行文脈で既に言及された要素は主題になりやすくなる（関連の主題）という，文脈依存の傾向が反映されたデータを収集するため，以下の条件を全て満たす主格を**文脈依存の「は」**として収集した．

- NTC 上で「は」を伴っていた．
- 共参照アノテーション上で，主格の要素が先行文脈において既出である．
- 文脈を見せることで，「は」と答える割合が 15%よりも上昇した．

上記条件を満たさずコーパス上で「は」を伴っていたデータポイントの集合については文脈非依存の「は」，コーパス上で「が」を伴っていたデータポイントの集合については「が」全体と呼ぶ．サブセットの統計を表 1 に示す．

作成したサブセットごとにデータの例を示す．また，必要に応じて先行文脈を灰色で，アノテーション対象の主格を太字で載せる．

文脈依存の「は」の例

(2)十万円が当たった。暮れに息子が買ったジャンボ宝くじの二十枚連番の一枚である。昨日、息

4)　MACE [17] を用いて各アノテータの信頼度を計算した．

子**は**テレビでの当選番号と照らし合わせながら、「三千六百円になるから、お母さんにあげる」と言っていた。

毎日新聞記事データ集 1995 年版

例 2 では，主格の要素である「息子」が先行文脈に出現している．このように先行文脈に出現した要素やそれに関連するものは「は」を伴って主題になりやすいことが知られている [3].

文脈非依存の「は」の例

(3)民法九〇〇条**は**、非嫡出子の相続分について「嫡出子の二分の一とする」と定めている。

毎日新聞記事データ集 1995 年版

例 3 は主格の要素である「民放九〇〇条」について解説をする文となっている．このように主格名詞について解説 (措定) するときには「は」を伴う [19].

「が」全体の例

(4)ペソの暴落でメキシコ経済の混乱が続いている。北米自由貿易協定、中南米諸国だけでなく、日本や欧州企業にも影響**が**及び始めた。

毎日新聞記事データ集 1995 年版

例 4 は「メキシコ経済の混乱によって影響が及ぶ」という現象を表した文である．現象をありのまま，判断の加工を施さないでそのまま表現した文を現象文と呼び，「が」を伴う [20]．他にも排他の「が」など，様々な用法の「が」がこのサブセットに属する．

4 実験設定

言語モデル パラメータ数の異なる 2 つ Transformer ベースの言語モデル [21]（400M パラメータの Trans-L と 55M パラメータの textscTrans-s）と LSTM ベースの言語モデル [22]（LSTM）について，「は」と「が」の選択の傾向を分析する．言語モデルは文章レベルであり，学習データは約 300 万の新聞記事から成る[5]．言語モデルへの入力は，JUMAN[23] で形態素に分割したのち sentencepiece[24] でサブワードに分割した[6].

指標 3 節で作成した各インスタンスは，$(c, s_{は}, s_{が})$ の三つ組で表される．ただし，$s_{は}$ と $s_{が}$ は分析の対象となる主格が「は」または「が」を伴う文であり，c は同一文章内の先行文脈である．各インスタンスにおける「は率」$r_{は}$ を人と言語モデルに対して以下のように計算する．

$$r_{は} = \frac{p(s_{は})}{p(s_{は}) + p(s_{が})} \tag{1}$$

ただし $p(s_{は})$ および $p(s_{が})$ は，人の場合何人中何人が「は」または「が」が自然であると判断したかの割合[7]であり，言語モデルの場合 $s_{は}$ および $s_{が}$ に対する生成確率である．$r_{は}$ が大きいほど「が」ではなく「は」が自然であると判断した割合が高いという点で，$r_{は}$ はある種の「は」と「が」の選択における確信度の指標と考えられる．

また，文脈 c を見せて（入力して）得られた $p(s_{は})$ および $p(s_{が})$ を元に計算した「は率」を $r^{\text{ctx}}_{は}$，文脈を見せない場合で得られた「は率」を $r^{\text{wo\_ctx}}_{は}$ と区別する．人および各言語モデルについて，各インスタンスにおける文脈の影響を $\Delta_{は} = r^{\text{ctx}}_{は} - r^{\text{wo\_ctx}}_{は}$ とする．$\Delta_{は}$ が大きいインスタンスほど，文脈を見ることにより「は」を伴うと判断するようになった度合いが高いインスタンスであると言える．

5 実験 1: 文脈依存の「は」の傾向

はじめに，3.2 節で作成した文脈依存の「は」のサブセットについて人と言語モデルの傾向を分析する (表 2)．各言語モデルが各インスタンスに付与した $r^{\text{ctx}}_{は}$ と $\Delta_{は}$ について，ワーカから得られた値との順位相関係数を報告する．$r^{\text{ctx}}_{は}$ の相関は，「は」か「が」を選択する際の確信度の強さにおける人らしさを示す．$\Delta_{は}$ の相関は，「は」と「が」の選択に及ぼす文脈の影響が人と言語モデルで似ているかを示す．

また，NAIST テキストコーパス上で主格に対して用いられていた助詞を正しい助詞[8]とみなした場合の正解率も報告する．人の正解率は各インスタンスに付与されたラベルのうち正しい助詞（本実験では「は」）を選択できたものの割合とし，言語モデルの正解率は正しい助詞に対して高い確率を付与できたインスタンスの割合とした．文脈を見せた場合と見せない場合でそれぞれ正解率を計算した．

文脈ありの正解率を見ると人と言語モデルが同程度の値を示しており，およそ 90%という高い精度で「は」と「が」の選択が可能であった．一方で表 2 中の文脈なしの正解率と文脈ありの正解率の比較より，人は文脈を見ることで正解率が 12 ポイント

5） 言語モデルの訓練に用いたデータは，3.1 節で作成した評価データと重複しない．

6） Unigram モデル [25] を用いた（character coverage=0.9995, vocab size=100000）

7） 「どちらでもよい」という判断をした人数も合わせて，「は」または「が」が自然であると判断した人数の割合とした．

8） NAIST テキストコーパスの文章は新聞記事であり，十分に推敲された文章であると考えられる．

表 2　文脈依存の「は」の選択傾向

モデル	正解率 (文脈なし)	正解率 (文脈あり)	$r^{\text{ctx}}_{\text{は}}$ の相関	$\Delta_{\text{は}}$ の相関
人	77.9	**89.9**	-	-
TRANS-L	**88.3**	**89.9**	0.312	-0.003
TRANS-S	86.5	86.7	0.364	-0.115
LSTM	85.1	85.6	**0.415**	**0.058**

表 3　「は」と「が」の選択傾向 (データセット全体)

モデル	正解率 (文脈なし)	正解率 (文脈あり)	$r^{\text{ctx}}_{\text{は}}$ の相関	$\Delta_{\text{は}}$ の相関
人	84.0	85.9	-	-
TRANS-L	**88.1**	**88.3**	**0.796**	-0.024
TRANS-S	88.0	88.0	0.794	-0.007
LSTM	85.8	85.7	0.791	**0.023**

上がっているにも関わらず，言語モデルは先行文脈をみても正解率がほとんど変わらないことが確認された．特に文脈を見ない場合の言語モデルの正解率が人よりも 10 ポイント程度高く，**人が認識していないような文内の何らかの特徴量を手がかりに言語モデルが「は」と「が」の選択を行っている**ことが示唆された[9]．また，$r^{\text{ctx}}_{\text{は}}$ の相関が弱いことから人と言語モデルの間のある種の選択の確信度が異なり，$\Delta_{\text{は}}$ の相関がないことからも文脈の考慮の仕方に乖離があることが分かる．つまり，言語モデルは正しく「は」と「が」の選択が行えているものの，その選択をする要因は人と異なることが示唆される．このような乖離は，言語理解タスクにおいて回答に十分な情報が提供されなくても解けてしまう問題 [26] や，人に認識不可能な特徴量（敵対的事例）の存在 [27] と類似する問題と考えられる．

本実験で確認された「は」と「が」の選択における乖離は，クラウドワーカによる複数人のアノテーションや文脈が存在しない場合の判断といったコーパス上に表出されない情報によって明らかになった結果であり，本データセットの作成に意義があったことを強調したい．なお言語モデル間の差については，$r^{\text{ctx}}_{\text{は}}$ の相関において LSTM が比較的人に近いという観察が得られた．

6　実験 2: コーパス全体の傾向

5 節では文脈依存の「は」のサブセットについて分析したが，最後に 3 節で作成したデータセット全体についても，人と言語モデルの傾向を分析する (表 3)．文脈なしの正解率と文脈ありの正解率を比較すると，人も言語モデルも文脈の有無による正解率の変化は小さく，データセット全体としては「は」と「が」の選択は文内の情報のみで可能であることが示唆された．この結果は「は」と「が」の使い分けにしばしば情報の新旧といった談話レベルの概念が持ち込まれることと対照的であり，談話レベルの判断が必要な状況は実際のコーパス上では少数である可能性，文脈を見せない設定においても文内の情報から人が適当な文脈を復元できている（例えば代名詞の使用から情報の新旧の予測がつくなど）可能性を含めて今後さらに調査したい．

$r^{\text{ctx}}_{\text{は}}$ の相関はどの言語モデルでも 0.79 以上と，データセット全体においては人と言語モデルの選択の確信度の強さに高い相関があることが確認された[10]．表 2 の $r^{\text{ctx}}_{\text{は}}$ の相関は弱かったことを踏まえると，言語モデルは「は」と「が」のどちらが自然かという粒度の選択においては人と近い傾向がみられるが，「は」を選択する際にどれくらいの確信度で選択しているかという側面では人と傾向に乖離があると考えられる (付録図 3)．$\Delta_{\text{は}}$ の相関はなく，5 節の結果と一貫して，人と文脈の考慮の仕方に乖離があることが確認された．

7　おわりに

本研究では，談話レベルの文脈に依存する系列的な選択における言語モデルと人の判断について分析を行い，乖離があることを示した．「は」か「が」という選択においては言語モデルは人よりも正確な選択が行えるが，先行文脈の情報に依存した選択を行わないという点で人と乖離がみられた．これは言語モデルと人の「は」と「が」の選択の拠り所が両者で異なるということである．今後は，言語モデルが文内および先行文脈のどの部分の情報を頼りに系列的な選択を行うのかを調査していきたい．

謝辞．

本研究は JSPS 科研費 P19H04162 の助成を受けたものです．また，データセットの作成にあたり毎日新聞記事データ集 1995 年版の使用を許可していただいた毎日新聞社に感謝致します．

本研究を進めるにあたり，ご指導，ご助言を頂いた乾健太郎教授，鈴木潤教授に心より感謝致します．また，日頃より研究活動や論文執筆を直接指導

9）　言語モデルだけが文脈を見ずに正解できたインスタンスを付録に示す．

10）　人と言語モデルの $r^{\text{ctx}}_{\text{は}}$ の相関が高かったインスタンスを付録に示す．

してくださいました栗林樹生さんに心より感謝致します．さらに，日々の議論の中で多くのご助言を頂きました研究室の皆様に感謝致します．

参考文献

[1]Michael Alexander Kirkwood Halliday and Ruqaiya Hasan. Cohesion in english. *Longman*, 1976.

[2]Michael Halliday, Christian MIM Matthiessen, and Christian Matthiessen. *An introduction to functional grammar*. Routledge, 2014.

[3]野田尚史. 「は」と「が」. くろしお出版, 1996.

[4]松下大三郎. 標準日本口語法. 中文館書店, 1930.

[5]Barbara J Grosz, Aravind K Joshi, and Scott Weinstein. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 1995.

[6]砂川有里子. 文法と談話の接点. くろしお出版, 2005.

[7]Ryu Iida, Mamoru Komachi, Kentaro Inui, and Yuji Matsumoto. Annotating a japanese text corpus with predicate-argument and coreference relations. In *Proceedings of the linguistic annotation workshop*, pp. 132–139, 2007.

[8]日本語記述文法研究会. 現代日本語文法 5 とりたて・主題. くろしお出版, 2009.

[9]Suhas Arehalli and Tal Linzen. Neural language models capture some, but not all, agreement attraction effects, Feb 2020.

[10]孟玲秀. 『日本語教育における「は」と「が」の教授法』——中国人学習者に対する日本語教育の場合——. 2004.

[11]Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. Assessing the ability of lstms to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, Vol. 4, pp. 521–535, 2016.

[12]Jey Han Lau, Alexander Clark, and Shalom Lappin. Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science*, Vol. 41, No. 5, pp. 1202–1241, 2017.

[13]Jey Han Lau, Carlos S Armendariz, Shalom Lappin, Matthew Purver, and Chang Shu. How furiously can colourless green ideas sleep? sentence acceptability in context. *arXiv preprint arXiv:2004.00881*, 2020.

[14]Alex Warstadt, Alicia Parrish, Haokun Liu, Anhad Mohananey, Wei Peng, Sheng-Fu Wang, and Samuel R Bowman. Blimp: The benchmark of linguistic minimal pairs for english. *Transactions of the Association for Computational Linguistics*, Vol. 8, pp. 377–392, 2020.

[15]Xuhui Zhou, Yue Zhang, Leyang Cui, and Dandan Huang. Evaluating commonsense in pre-trained language models. In *AAAI*, pp. 9733–9740, 2020.

[16]Gregor Betz, Christian Voigt, and Kyle Richardson. Critical thinking for language models, 2020.

[17]Dirk Hovy, Taylor Berg-Kirkpatrick, Ashish Vaswani, and Eduard Hovy. Learning whom to trust with mace. In *Proceedings of NAACL*, pp. 1120–1130, 2013.

[18]Klaus Krippendorff. *Content Analysis: an Introduction to its Methodology*. Sage publications, 2004.

[19]三尾砂. 國語法文章論. 三省堂, 1948.

[20]三上章. 現代語法序説ーシンタクスの試みー. 刀江書院, 1953.

[21]Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, pp. 5998–6008, 2017.

[22]Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.

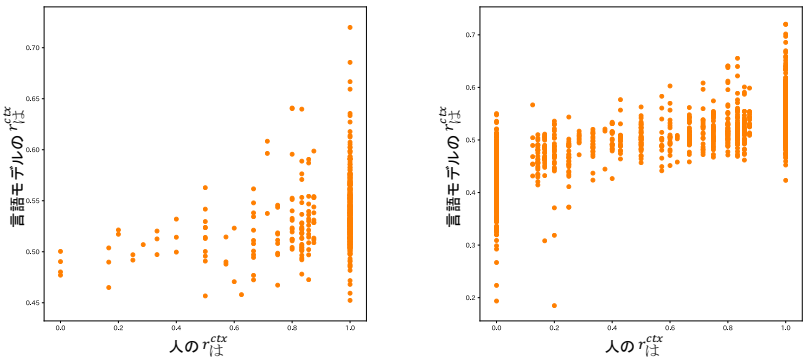
[23]Daisuke Kawahara and Sadao Kurohashi. Case frame compilation from the web using high-performance computing. In *Proceedings of LREC*, Genoa, Italy, May 2006. European Language Resources Association (ELRA).

[24]Taku Kudo and John Richardson. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *EMNLP*, pp. 66–71, 2018.

[25]Taku Kudo. Subword regularization: Improving neural network translation models with multiple subword candidates. *Proceedings of ACL*, 2018.

[26]Suchin Gururangan, Swabha Swayamdipta, Omer Levy, Roy Schwartz, Samuel R Bowman, and Noah A Smith. Annotation artifacts in natural language inference data. *arXiv preprint arXiv:1803.02324*, 2018.

[27]Andrew Ilyas, Shibani Santurkar, Dimitris Tsipras, Logan Engstrom, Brandon Tran, and Aleksander Madry. Adversarial examples are not bugs, they are features. In *Advances in Neural Information Processing Systems*, pp. 125–136, 2019.



(a) 文脈依存の「は」　　(b) データ全体

図 3　人と言語モデル (TRANS-L) の $r_{\text{は}}^{ctx}$ の相関．1 プロットが 1 インスタンスを表す．

表 4　人と言語モデルの「は」と「が」の選択傾向

サブセット	モデル	正解率 (文脈なし)	正解率 (文脈あり)	$r_{\text{は}}^{\text{ctx}}$ の相関	$\Delta_{\text{は}}$ の相関
データセット全体	人	84.0	85.9	-	-
	TRANS-L	88.1	88.3	0.796	-0.024
	TRANS-S	88.0	88.0	0.794	-0.007
	LSTM	85.8	85.7	0.791	0.023
文脈依存の「は」	人	77.9	89.9	-	-
	TRANS-L	88.3	89.9	0.312	-0.003
	TRANS-S	86.5	86.7	0.364	-0.115
	LSTM	85.1	85.6	0.415	0.058
文脈非依存の「は」	人	90.2	84.3	-	-
	TRANS-L	88.6	87.9	0.392	-0.102
	TRANS-S	86.3	87.2	0.359	-0.016
	LSTM	81.7	81.7	0.340	-0.086
「が」全体	人	83.9	84.3	-	-
	TRANS-L	87.6	87.5	0.531	-0.025
	TRANS-S	90.0	89.2	0.523	0.054
	LSTM	88.8	88.2	0.520	0.023

A　インスタンス例

人と言語モデルの $r_{\text{は}}^{\text{ctx}}$ の相関が高かったインスタンス

(5)受験生に試練の季節がやってきた。十四日、日本海側が大雪に見舞われたなか始まった大学入試センター試験。今年の受験者**は**女子が大幅に増えるなど過去最多となった。

毎日新聞記事データ集 1995 年版

(6)三日未明から夕方にかけて北海道、東北地方で三陸はるか沖地震の余震とみられる地震が四回あった。午前一時四十二分ごろ、東北地方の一部で地震**が**あった。

毎日新聞記事データ集 1995 年版

言語モデルだけが文脈を見ずに正しい助詞を選択できたインスタンス

(7)４歳の孫を連れておそば屋さんへ入りました。鍋焼きうどんを取って２人で食べました。孫**は**必死になって食べましたが、とても食べ切れません。

毎日新聞記事データ集 1995 年版

(8)七日は、無病息災を願う「春の七草」の日。大阪・キタの阪神百貨店では、午前八時から七草がゆの無料サービスを行い、サラリーマンや若い女性たち、剣道の早朝練習帰りの中学生らが季節の味覚を楽しんだ。用意した五百食**は**約一時間でなくなった。

毎日新聞記事データ集 1995 年版