

統語変形はコミュニケーションから創発するのか？

梶川康平¹ 窪田悠介² 大関洋平¹¹ 東京大学 ² 国立国語研究所

{kohei-kajikawa,oseki}@g.ecc.u-tokyo.ac.jp kubota@ninjal.ac.jp

概要

自然言語には合成性にもとづく統語構造が存在し、さらに統語構造を直接変形する操作（統語変形）が存在する。先行研究において、合成性はコミュニケーションと学習可能性に関する圧力のトレードオフとして創発することが示されている。しかしながら、自然言語になぜ統語変形が存在するのかは明らかではない。本研究では、統語変形を必要とする統語現象として等位接続に注目した上で、統語変形もまたコミュニケーションと学習可能性に関する圧力のトレードオフとして創発するのかを検証する。

1 はじめに

自然言語の普遍性を理解するためには、どのように普遍性を理論的に記述することができるのかに加え、なぜそのような普遍性が存在するのかという問いも決定的に重要である。そして近年、自然言語の普遍性は、コミュニケーションや学習可能性に関する圧力のトレードオフとして創発するという説明を試みる研究が注目を集めている [1, 2, 3, 4]。

Kirby らは、自然言語の普遍的な性質として合成性にもとづいた統語構造に注目し、合成性にもとづく統語構造が、コミュニケーションに関する圧力と学習可能性に関する圧力という相反する圧力が同時に働いたときに、両圧力のトレードオフとして創発するということを主張している [2, 5, 6]。たとえば、[2] では、文法の複雑さが異なる3種類の言語を想定し、それらの、コミュニケーションと学習可能性に関する圧力への適応力を検証した。ここでの文法の複雑さとは、文法規則全体を表現する最小のコード長、すなわち、文法自体の圧縮可能性を指す。3種類の言語とは具体的に、あらゆる意味を同一の言語表現で表現する縮退的言語 (degenerate)、合成性をもつ文法規則による合成的言語 (compositional)、あらゆる意味をそれぞれ単一の異なる言語表現で表現する総体的言語 (holistic) を想定しており、縮退的

言語から順に文法規則の数が増えている。つまり、順に文法の圧縮可能性が高くなっているという関係になっている。想定する言語の表現長はすべて同一であるという前提を置き、コミュニケーションに関する圧力により、より表現力の高い言語、すなわち、より曖昧性が少ない言語が求められるとする。一方、学習可能性に関する圧力により、より学習しやすい言語、すなわち、圧縮可能性の高い文法による言語が求められると考えられる。結果として、繰り返し学習モデル (iterated learning) によるシミュレーションにより、(i) コミュニケーションに関する圧力のみがかかる場合は、総体的言語が創発する、(ii) 学習可能性に関する圧力のみがかかる場合は、縮退的言語が創発する、(iii) 2つの圧力が同時にかかる場合は、合成的言語が創発する、ということを示した。

しかしながら、ここでの合成性にもとづく統語構造をもつ言語とは、文脈自由文法で表現することができる言語が想定されており、自然言語には文脈自由文法では表現しきれない複雑な統語現象が存在するという事実を十分に捉えられていない [7]。この事実は、理論言語学において、自然言語には統語構造を直接変形する操作（統語変形）が存在するという主張により直感的に捉えられてきた [8]。なお、全ての言語理論に「統語変形」という操作が存在するわけではないが、文脈自由文法では表現しきれない複雑な統語現象を捉えるための操作は、表現の仕方は違えど全ての言語理論に存在する [9, 10, 11, 12]。

本研究は、統語変形もまたコミュニケーションと学習可能性に関する圧力のトレードオフとして創発するのかを検証することを目的とする。具体的には、統語変形を必要とする統語現象として等位接続に注目した上で、文法の圧縮可能性が異なる3種類の言語による等位接続を、コミュニケーションの効率性と学習可能性の観点から比較する。結果として、統語変形のある言語では、2つの圧力のトレードオフが成り立っていることを示す。

2 統語変形

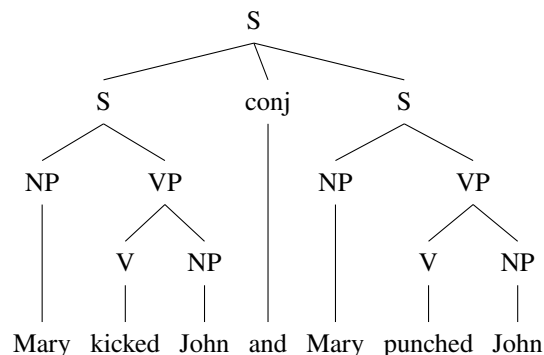
自然言語には合成性にもとづく統語構造が存在し、さらに統語構造を直接変形する操作である統語変形が存在すると主張されてきた [8]。統語変形を仮定する言語理論では、まず文脈自由文法にもとづき文の基底構造が生成され、必要に応じて統語変形で基底構造を変形することで最終的な統語構造を派生すると考えられている。例えば、受動文は統語変形を必要とする代表的な統語現象であり、目的語位置にある名詞句を文頭に移動させ、主語位置にある名詞句を文末に移動させる。ここで重要なのは、移動させる名詞句も、その名詞句を移動させる位置も、統語構造を参照して決定され、単語列の線形順序を参照することは決してない。

同様に、本研究で対象にする等位接続も、統語変形を必要とする統語現象であると伝統的に考えられてきた [8, 13, 14]。例えば、*John ran and swam.* という等位接続文は、*John ran and John swam.* という等位接続文から、*Mary kicked and punched John.* という等位接続文は、*Mary kicked John and Mary punched John.* という等位接続文から、それぞれ統語変形によって派生されると考えられる (図 1)。具体的には、ここでは以下のような統語変形が用いられていると考えられる：

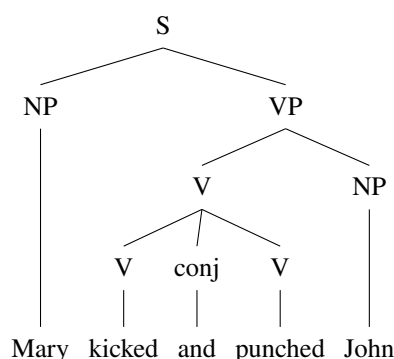
$$(Y + X_1 + Z) + \text{conj} + (Y + X_2 + Z) \\ \rightarrow Y + (X_1 + \text{conj} + X_2) + Z \\ \text{([8], p.113, 一部改変)}$$

X は任意の統語範疇、 Y, Z は任意の統語範疇もしくは文字列、 conj は任意の等位接続詞、 $+$ は連結を表す。統語変形を仮定することで、任意の統語範疇について、同一の統語範疇同士であれば等位接続が可能であるという事実を説明することができる。¹⁾

1) ここで示した「等位構造縮約変形 (Conjunction Reduction Transformation)」による等位接続の分析のナイーブな定式化は 70 年代に Partee によって批判され [15]、後の 80 年代の形式意味論研究において、等位接続詞の意味を統語変形を用いずに高階論理で記述する「一般化等位接続 (Generalized Conjunction)」が代案として提案された [16, 17]。「一般化等位接続」の手法は、証明論的な観点に立てば意味論レベルで統語変形を行っているものとみなすことができるため、等位接続の分析に (広い意味での) 統語変形が必要であることを否定するものではない。本研究で問題にしているのは、この後者の、単語列から意味表示を得る過程のいずれかの段階における「構造に対する操作」という意味での「統語変形」の概念であるため、「等位構造縮約変形」と「一般化等位接続」の差は以下の議論では実質的に問題とならない。



(a) 統語変形前



(b) 統語変形後

図 1: 等位接続の統語変形による分析

3 言語のデザインと圧力

本研究では、文法の圧縮可能性が異なる 3 種類の言語を検証する。これら 3 つの言語は、まず同一の文脈自由文法にもとづき生成した上で、その後 (1) 統語変形を適用しない、(2) 統語変形を適用する、(3) 非統語変形も適用することで構成している (表 1)：

1. no-transform 言語：統語変形が存在しない言語。文同士の等位接続のみ可能。
2. syntactic-transform 言語：統語変形が存在する言語。同一の統語範疇同士であれば等位接続が可能。
3. radical-transform 言語：統語変形による等位接続に加え、さらに「同一文中で既に使用した表現を削除する」という非統語変形が存在する言語。

言語	例文
no-transform	<i>Anna thought that John ran and Anna thought that John thought.</i>
syntactic-transform	<i>Anna thought that [John ran and thought].</i>
radical-transform	<i>Anna thought that [John ran and].</i>

表 1: 文法の圧縮可能性が異なる 3 種類の言語

学習可能性 人間の一般認知能力として、同一のデータから学習されるパターンの中では最も単純なパターンが選好される [18]。以上の3つの言語の中では、no-transform 言語から順に1つずつ変形規則が増えているため、no-transform 言語が最も単純である。従って、学習可能性に関する圧力の下では、より文法長（同一のデータを記述するのに使う規則数）が短い圧縮可能性の高い文法が生成する言語が好まれると考えられる：**no-transform 言語 > syntactic-transform 言語 > radical-transform 言語**。

コミュニケーションの効率性 圧縮可能性の高い文法による言語が、コミュニケーションに関する圧力の観点からも選好されるとは限らない。具体的には、no-transform 言語は、文法の圧縮可能性こそ高いが、統語変形が無いため生成される言語表現は長く冗長なものとなる。一方、radical-transform 言語は、文法の圧縮可能性こそ低いが、生成される言語表現は短く簡潔なものとなる。

先行研究では、コミュニケーションに関する圧力として、効率的なコミュニケーションを実現するため話し手もしくは聞き手にかかる圧力を想定することが多い [19, 20, 21, 22]。効率的なコミュニケーションとは、話し手が最小の労力を払いつつ、聞き手が最大の理解を得られる状況のことを指す [4]。Kirby らは、表現長がすべて同一である言語を比較し、聞き手が高い理解を得られる言語、つまり曖昧性が低い言語がコミュニケーションに関する圧力によって創発するとした [2]。しかしながら、曖昧性は多くの場合に文脈によって解消されると仮定すると、より短い言語表現が要求される [1]。そこで、本研究では、この短く簡潔な言語表現への圧力をコミュニケーションに関する圧力として想定する。従って、コミュニケーションに関する圧力の下では、より表現長（同一の意味を表現するのに使う単語数）が短い言語が好まれると考えられる：**radical-transform 言語 > syntactic-transform 言語 > no-transform 言語**。

4 実験

4.1 データ生成

まず、文の基底構造を生成する文脈自由文法を仮定した上で、その文法から文をランダムにサンプリングし、さらに統語変形や非統語変形を適用し、3

S	$\rightarrow NP VP \mid S conj S$
VP	$\rightarrow TV_{that} \bar{S} \mid TV NP \mid IV$
\bar{S}	$\rightarrow that S$
TV_{that}	$\rightarrow thought \mid said \mid believed \mid explained \mid knew$
TV	$\rightarrow liked \mid kicked \mid knew \mid tought \mid called \mid admired \mid met$
IV	$\rightarrow ran \mid swam \mid laughed \mid slept \mid thought$
NP	$\rightarrow John \mid Mary \mid Ken \mid Sue \mid James \mid Robert \mid Susan \mid Ana$
$conj$	$\rightarrow and$

図 2: 基底構造を生成する文脈自由文法

基底構造	対応する意味
<i>John ran</i>	ran(j)
<i>John liked Mary</i>	liked(j, m)
<i>John thought Mary ran</i>	thought(j, ran(m))
<i>John ran and Mary ran</i>	ran(j) & ran(m)

表 2: 基底構造と対応する意味

種類の言語を生成した。実際に用いた文脈自由文法を図 2 に示す。また、基底構造を生成する際、等位接続や埋め込みの影響で生成される文が指数関数的に増大することを防ぐため、同一文中に *and* は1つまで、*that* は2つまでという制約を加えた。²⁾さらに、radical-transform 言語は、原理的に曖昧性が生じてしまうため、曖昧性が生じている文については、部分的に一部の単語にインデックスを付与し、曖昧性が解消されるように設計した。加えて、基底構造に対応する意味は、表 2 に示した通り、述語論理によって表現した。

4.2 モデル

モデルによる統語構造の理解を評価するため、文から対応する意味への翻訳タスクを実施し、その精度を測るという方法がある [24]。そこで、各言語の学習可能性を評価するため、アテンション付き RNN によるエンコーダ・デコーダモデル [25] を用いて、各言語における文から対応する意味への翻訳タスクを実施した。単語埋め込みは 200 次元とし、エンコーダ・デコーダは入力層、隠れ層ともに 256 次元の RNN を用いた。最適化は学習率を 5.0×10^{-5} に設定した Adam [26] で行った。

2) 加えて、*John liked John.* や *John said that John ran.* といった束縛条件 [23] に違反した文は生成しないように制約を加えた。

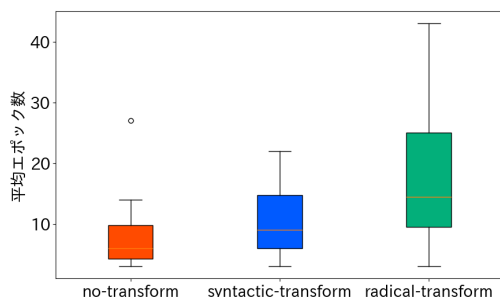


図 3: 各言語の学習可能性

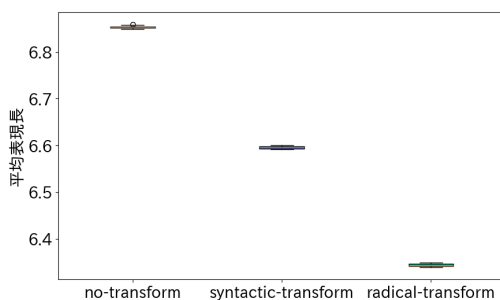


図 4: 各言語のコミュニケーションの効率性

4.3 評価

学習可能性 4.1 節で導入した文法にもとづき、3 節で導入した 3 種類の言語である no-transform 言語、syntactic-transform 言語、radical-transform 言語をそれぞれ 50,000 文ずつ生成する。その後、train:test=8:2 に分割し、[27] を踏襲した以下の条件で、各言語の学習可能性を評価した：

- モデルの出力した意味表現と、正解の意味表現が完全一致したときのみ正解とする。
- test データについて、正解率が 0.99 を上回るまで、あるいは 50 エポック回すまで学習を続ける。
- 正解率が初めて 0.99 を超えたエポック数を、その言語の学習可能性とする。
- 上記のプロセスを、30 回行う。

コミュニケーションの効率性 各言語の平均表現長を計算することで、各言語のコミュニケーションの効率性を評価した。

5 結果・考察

各言語における学習可能性（平均エポック数）およびコミュニケーションの効率性（平均表現長）を、それぞれ図 3 および図 4 に箱ひげ図で示す。

各言語における学習可能性およびコミュニケーションに関する圧力に、統計的な有意差がある

比較群	平均エポック数 統計量	p 値	平均表現長 統計量	p 値
♣/◇	2.562	0.016	101.0	0.012
◇/♡	-3.059	0.005	103.0	0.013
♡/♣	4.564	< 0.001	39.5	0.002

表 3: 平均エポック数および平均文長について、それぞれ対応のある t 検定とウィルコクソンの符号付順位検定の結果。♣ は no-transform 言語、◇ は syntactic-transform 言語、♡ は radical-transform 言語を表す。有意水準は、多重比較を補正するため、Bonferroni 補正により $\alpha = 0.05/3 = 0.017$ とした。

かを確認するため、それぞれ対応のある t 検定とウィルコクソンの符号付順位検定を行なった結果、全ての言語間で統計的な有意差が認められた(表 3)。³⁾ まず、各言語の学習可能性は、高い順に、予測通り、no-transform 言語 > syntactic-transform 言語 > radical-transform 言語という順になっている。一方、各言語のコミュニケーションの効率性も、予測通り、高い順に、radical-transform 言語 > syntactic-transform 言語 > no-transform 言語の順になっている。結論として、学習可能性とコミュニケーションに関する圧力のトレードオフとして、統語変形による等位接続を持つ syntactic-transform 言語が創発する可能性が示唆された。

6 おわりに

本研究は、統語変形が、コミュニケーションと学習可能性に関する圧力のトレードオフとして創発するのかを検証することを目的とし、結果として、統語変形のある言語では、2つの圧力のトレードオフが成り立っていることを示した。しかしながら、本研究は、そのようなトレードオフが最適であるか、という観点や、そのようなトレードオフが実際に創発するか、という観点での検証はまだ行っていない。今後、現状の自然言語のシステムが2つの圧力下での最適解であるか、また、Kirby らの繰り返し学習モデルを用いて、統語変形のある言語が創発するのかを検証することを課題とする。

3) 3 言語における学習可能性およびコミュニケーションの効率性について、正規性の検定 ($\alpha = 0.05$) を行い、前者の 3 群は正規分布に従っているのに対して、後者の 3 群は正規分布に従っていないと判断し、それぞれパラメトリック、ノンパラメトリックな検定で分析することにした。

謝辞

本研究は JST さきがけ JPMJPR21C2、JSPS 科研費 21K00541、国立国語研究所共同研究プロジェクト「計算言語学的手法による理論言語学の実証的な方法論の開拓」の助成を受けたものです。

参考文献

- [1] Steven T. Piantadosi, Harry Tily, and Edward Gibson. The communicative function of ambiguity in language. **Cognition**, Vol. 122, No. 3, pp. 280–291, 2012.
- [2] Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. Compression and communication in the cultural evolution of linguistic structure. **Cognition**, Vol. 141, pp. 87–102, 2015.
- [3] Charles Kemp, Yang Xu, and Terry Regier. Semantic typology and efficient communication. **Annual Review of Linguistics**, Vol. 4, No. 1, pp. 109–128, 2018.
- [4] Edward Gibson, Richard Futrell, Steven T. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. How efficiency shapes human language. **Trends in Cognitive Sciences**, Vol. 23, No. 5, pp. 389–407, 2019.
- [5] Simon Kirby and James R. Hurford. **The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model**, pp. 121–147. Springer London, London, 2002.
- [6] Simon Kirby, Tom Griffiths, and Kenny Smith. Iterated learning and the evolution of language. **Current Opinion in Neurobiology**, Vol. 28, pp. 108–114, 2014. SI: Communication and language.
- [7] Stuart M Shieber. Evidence against the context-freeness of natural language. In **Philosophy, language, and artificial intelligence**, pp. 79–89. Springer, 1985.
- [8] Noam Chomsky. **Syntactic structures**. Mouton, 1957.
- [9] Aravind K. Joshi, Leon S. Levy, and Masako Takahashi. Tree adjunct grammars. **Journal of Computer and System Sciences**, Vol. 10, No. 1, pp. 136–163, 1975.
- [10] Carl Pollard and Ivan A Sag. **Head-driven phrase structure grammar**. University of Chicago Press, 1994.
- [11] Mark Steedman. **Surface structure and interpretation**. MIT press, 1996.
- [12] Edward Stabler. Derivational minimalism. In Christian Retoré, editor, **Logical Aspects of Computational Linguistics**, pp. 68–95. Springer Berlin Heidelberg, 1997.
- [13] Noam Chomsky. **The Logical Structure of Linguistic Theory**. Springer New York, NY, 1975 (=1955).
- [14] John R. Ross. **Constraints on variables in syntax**. PhD thesis, MIT, 1967.
- [15] Barbara Hall Partee. Negation, conjunction, and quantifiers: Syntax vs. semantics. **Foundations of Language**, Vol. 6, pp. 153–165, 1970.
- [16] Gerald Gazdar. A cross-categorial semantics for coordination. **Linguistics and Philosophy**, Vol. 3, No. 3, pp. 407–409, 1980.
- [17] Barbara Partee and Mats Rooth. Generalized conjunction and type ambiguity. **Formal semantics: The essential readings**, pp. 334–356, 1983.
- [18] Nick Chater and Paul Vitányi. Simplicity: a unifying principle in cognitive science? **Trends in Cognitive Sciences**, Vol. 7, No. 1, pp. 19–22, 2003.
- [19] George K. Zipf. **Human behavior and the principle of least effort**. Addison-Wesley, 1949.
- [20] Steven T. Piantadosi, Harry Tily, and Edward Gibson. Word lengths are optimized for efficient communication. **Proceedings of the National Academy of Sciences**, Vol. 108, No. 9, pp. 3526–3529, 2011.
- [21] Charles Kemp and Terry Regier. Kinship categories across languages reflect general communicative principles. **Science**, Vol. 336, No. 6084, pp. 1049–1054, 2012.
- [22] Shane Steinert-Threlkeld. Quantifiers in natural language: Efficient communication and degrees of semantic universals. **Entropy**, Vol. 23, No. 10, 2021.
- [23] Noam Chomsky. **Lectures on Government and Binding**. De Gruyter Mouton, Berlin, New York, 1981.
- [24] Najoung Kim and Tal Linzen. COGS: A compositional generalization challenge based on semantic interpretation. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 9087–9105, Online, November 2020. Association for Computational Linguistics.
- [25] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. **arXiv**, 2014.
- [26] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In **ICLR 2015**, San Diego, CA, USA, 2015. Conference Track Proceedings.
- [27] Shane Steinert-Threlkeld and Jakub Szymanik. Learnability and semantic universals. **Semantics and Pragmatics**, Vol. 12, pp. 1–39, 2019.