創発言語でも Harris の分節原理は成り立つのか？

上田亮 [1]　石井太河 [1]　宮尾祐介 [1]
[1] 東京大学
{ryoryoueda, taigarana, yusuke}@is.s.u-tokyo.ac.jp

概要

本論文の目的は創発言語において Harris の分節原理が成り立つかを検証するというものである．創発言語とは，シミュレーションにおいてエージェント間で生じる人工的な言語のことを指し，近年注目を集めつつある研究対象である．一方 Harris の分節原理とは，自然言語の系列データにおいて，データの意味を知らずとも統計的情報のみから分節境界が得られるという性質である．創発言語においても自然言語の性質が観察されるかどうかを検証した研究はいくつか存在するが，Harris の分節原理に着目した研究は今のところない．実験の結果，創発言語は Harris の分節原理が成立するためのいくつかの前提条件を満たすが，統計的な情報から得られる分節境界が必ずしも意味的に妥当なものであるとは限らない可能性が示唆された．

1　はじめに

ニューラルネットワークで表されたエージェントにコミュニケーションをさせることにより，言語の創発をシミュレーションする創発コミュニケーション（Emergent Communication）という分野が注目されつつある．創発コミュニケーションによって生じたプロトコルは創発言語（Emergent Language）と呼ばれる．この分野は人間と交流できる AI の開発を目的として生まれたが [1, 2, 3]，そこから派生して創発言語そのものの性質を調べる研究も現れた．特に，自然言語に見られる普遍性質が創発言語においても観察され得るかを調べたものが多く，例えば，構成性 [4, 5, 6]，Zipf 短縮 [7, 8, 9]，エントロピー最小化 [10] に関する研究などがある．

本論文の目標は，創発言語において Harris の分節原理 [11, 12, 13, 14] が成り立つかを検証することである．Harris の分節原理とは，自然言語の系列データにおいて，データのもつ意味を参照せずとも統計的な情報のみから分節境界が得られるという普遍的な性質のことである．ここでいう分節とは，音素の系列に対する単語 (形態素)，単語の系列に対する節等を指す．単語や意味が事前に与えられるわけではない創発言語において，統計的性質から分節境界を得る可能性を模索することは重要である．

本論文では，創発コミュニケーションで頻繁に用いられる Lewis シグナリングゲーム [15] の実験パラダイムを採用する．このゲームにはスピーカ S とリスナ L という 2 人のエージェントのみが登場し，スピーカ S からリスナ L への単方向通信のみが許される．各試行において，スピーカ S は入力集合 I から情報 $i \in I$ を得て，それをメッセージ $m = S(i)$ に変換する．メッセージ m を受け取ったリスナ L が i を復元することができればゲームは成功である．ここで，メッセージのデータ $\{S(i)\}_{i \in I}$ が創発言語とみなされる．実験の結果，創発言語は Harris の分節原理を成立させるための前提条件を満たすものの，統計的性質から得た境界が必ずしも意味のある境界とは限らないという示唆が得られた．他方，人間が予め与えた意味に関係なく境界のようなものが生じたという興味深い結果の意味するところは，今後の課題として残されている．

2　背景: 創発言語

言語を創発させるには，エージェントを取り巻く環境やエージェントのアーキテクチャ，最適化の手法などを決める必要がある．本論文では [5] による設定「属性値組集合を用いたシグナリングゲーム」を導入する．

2.1　属性値組集合

$a, v \in \mathbb{N}\ (a, v > 0)$ とする．a, v をそれぞれ**属性数**，**値数**と呼ぶことにする．**属性値組集合** (attribute-value set) D_v^a とは以下のように定義される順序組の集合である:

$$D_v^a = \{(d_0, \ldots, d_{a-1}) \mid d_i \in \{0, \ldots, v-1\}\}. \quad (1)$$

2.2 シグナリングゲーム

言語を創発させる環境は，Lewis シグナリングゲーム [15] に基づいて定式化される．**シグナリングゲーム** G は四つ組 (I, M, S, L) から成る．ここで，I は入力集合，M はメッセージ集合であり，関数 $S : I \to M$ はスピーカエージェント，関数 $L : M \to I$ はリスナエージェントと呼ばれる．ゲーム G の目標は各入力 $i \in I$ について $i = L(S(i))$ を成り立たせることである．スピーカ S とリスナ L は学習対象であり，目標に近づくよう最適化される．[5] と同様，本論文では属性値組集合を入力集合とする: $I = D_v^a$. さらに，メッセージ集合 M は固定長離散記号列の集合とする．即ち，有限のアルファベット A 及びメッセージ長 $\mathrm{len} \in \mathbb{N}$ $(\mathrm{len} > 0)$ を用いて

$$M = A^{\mathrm{len}} = \{a_1 \dots a_{\mathrm{len}} \mid a_i \in A \text{ for } i = 1, \dots, \mathrm{len}\} \quad (2)$$

と定義する．また，以降の議論では，**創発言語**とはシグナリングゲーム $G = (I, M, S, L)$ から得られるデータ $\{S(i)\}_{i \in I}$ のことを指すものとする．創発言語データ $\{S(i)\}_{i \in I}$ は，自然言語でいえば $|I|$ 個の独立した発話のデータと捉えることができる．

3 背景: Harris の分節原理

Harris の仮説によれば，自然言語の音素系列における単語境界は，後続し得る音素の種類数が増大する点に現れるという．本節では [12, 13] による Harris の仮説の情報理論的な定式化，及びそれに基づいた境界検出アルゴリズムを導入する．

3.1 定式化

$\mathfrak{X}$ をアルファベット，$\mathfrak{X}^*$ を $\mathfrak{X}$ 上の系列の集合，$\mathfrak{X}^n$ を $\mathfrak{X}$ 上の n-gram の集合とする．このとき，系列 $s = x_0 \cdots x_{|s|-1} \in \mathfrak{X}^*$ に対する**分岐エントロピー**を以下のように定義する:

$$\begin{aligned} h(s) &\equiv \mathcal{H}(X_{|s|} \mid X_0 \cdots X_{|s|-1} = s) \\ &= -\sum_{x \in \mathfrak{X}} P(X_{|s|} = x \mid X_0 \cdots X_{|s|-1} = s) \\ &\qquad \times \log_2 P(X_{|s|} = x \mid X_0 \cdots X_{|s|-1} = s). \end{aligned} \quad (3)$$

ここで，$|s|$ は系列 s の長さ，各 X_i は $\mathfrak{X}$ 上の確率変数，$P(X_{|s|} = x \mid X_0 \cdots X_{|s|-1} = s)$ は系列 $s \in \mathfrak{X}^*$ が出現した直後に記号 $x \in \mathfrak{X}$ が出現する確率を表す．分岐エントロピー $h(s)$ は，特定の記号列 s が出現したときに，次に来る記号の不確かさの度合いを表している．また，長さ n の系列に対する**条件付きエントロピー**を以下のように定義する:

$$\begin{aligned} H(n) &\equiv \mathcal{H}(X_n \mid X_0 \cdots X_{n-1}) \\ &= \sum_{s \in \mathfrak{X}^n} P(X_0 \cdots X_{n-1} = s) h(s). \end{aligned} \quad (4)$$

ここで，$P(X_0 \cdots X_{n-1} = s)$ は n-gram $s \in \mathfrak{X}^n$ が出現する確率である．H は分岐エントロピー h の n-gram に関する平均と捉えることもできる．

自然言語のデータにおいては，$H(n)$ は n に関して単調に減少することが知られている．つまり，自然言語データの部分系列 $x_0 \cdots x_n \in \mathfrak{X}^{n+1}$ が与えられたときに，平均的には $h(x_0 \cdots x_{n-1}) > h(x_0 \cdots x_n)$ となる．一方で，その大域的な傾向に反して h が増大する点もある．この自然言語データにおける分岐エントロピー h の増減に関する，以下の普遍性質のことを **Harris の分節原理**と呼ぶ[1]:

> $\mathfrak{X}$ の要素を単位とする自然言語データにおいて，ある部分系列 $x_0 \cdots x_n \in \mathfrak{X}^{n+1}$ が存在して
>
> $$h(x_0 \cdots x_{n-1}) < h(x_0 \cdots x_n) \quad (5)$$
>
> となるとき，x_n はより大きな単位の系列における分節境界である．

3.2 境界検出アルゴリズム

系列データを $s = x_0 \cdots x_{|s|-1}$ とし，その部分系列を $s_{i,j} = x_i \cdots x_{j-1}$ と表すことにする．境界判定アルゴリズムはパラメータ $\mathrm{max\_len} \in \mathbb{N}, \mathrm{threshold} \in \mathbb{R}$ を伴って以下のような手順で実行される:

1. $i := 0$; $w := 1$; とする．
2. $i \geq |s|$ ならばプログラムを終了する．
3. $h(s_{i,i+w})$ を計算する．
4. $w > 1$ かつ $h(s_{i,i+w}) - h(s_{i,i+w-1}) > \mathrm{threshold}$ ならば，$i + w$ を境界点と判定する．
5. $w < \mathrm{max\_len}$ かつ $i + w < |s| - 1$ ならば $w := w + 1$; として 2 に戻る．さもなくば $i := i + 1$; $w := 1$ として 2 に戻る．

また，創発言語においては自然言語に見られるような分節境界が存在するとは限らないため，以降では境界検出アルゴリズムによって得られた仮の分節境界のことを**仮説境界**と呼ぶことにする．

4 実験設定

1) [12, 13] では，このことを *Harris's hypothesis* と呼んでいたが，同著者の近年の出版物 [14] に倣い Harris の分節原理と呼び改めた．

4.1 問題設定

創発言語において Harris の分節原理が成立するかどうかを調べるにあたり，我々は以下の 3 つの問いに答えなければならない．

問 1. H（式 4）は単調減少するか？
問 2. h（式 3）は増大点（式 5）をもつか？
問 3. 仮説境界は分節境界を意味するか？

問 3 は，創発言語において Harris の分節原理が成立するかという問いに他ならない．しかし，H が単調に減少し h が所々増加するという性質が満たされなければ問 3 はそもそも意味をなさない．創発言語においてはこのような性質が成り立つかどうかさえも自明ではないため，まず問 1, 問 2 に答えなければならない．また，問 1, 問 2 に答えるには単に H と h を計算すればよいが，問 3 に答える方法は自明でない．創発言語に分節境界があるとしても，それがどんなものであるのか事前に知る術がないからである．そこで以下の仮定を置く:

> ゲーム $G = (D_v^a, A^{\mathrm{len}}, S, L)$ の創発言語 $\{S(i)\}_i$ において，仮説境界が分節境界を意味するならば，属性数 a が大きくなるほど仮説境界の数も多くなるはずである． (A)

属性値組集合 D_v^a は，元々創発言語の構成性を測るために導入された設定 [4] を，さらに一般化したものである [5]．例えば *color*，*shape* という 2 属性をもつ入力を用いたときに，創発言語が十分に構成的であるならば，*color* の値を指す記号と *shape* の値を指す記号が別々に出現し，それらが複合して 1 つのメッセージを成すだろうと [4] は考えた．これは創発言語の構成性を測るうえで基本的な考え方となっている．今回の設定でいえば，分節が複合的なメッセージを構成するための単位であると想定したときに，指し示すべき属性が増えれば分節もそれに合わせて多くなるだろうと考えられる．仮定 (A) はそのような考えに基づいている．問 3 を仮定 (A) に従って以下のように言い換える:

問 3'. 属性数 a が大きくなるほど仮説境界も多くなるか？

4.2 パラメータの設定

属性値組集合 問 3' に答えるため属性数 a にはいくつかの値を取らせたい一方，ゲームの複雑さを揃えるために属性値組集合のサイズ $|D_v^a| = v^a$ はできるだけ等しくしたい．そこで $v^a = 4096$ として a, v を以下のように設定する:

$$(a, v) \in \left\{ \begin{array}{l} (1, 4096), (2, 64) \\ (3, 6), (4, 8), (6, 4), (12, 2) \end{array} \right\}. \quad (6)$$

メッセージ集合 メッセージ集合 $M = A^{\mathrm{len}}$ を定義するには，メッセージ長 len とアルファベットのサイズ $|A|$ を決めてやればよい．今回の実験では $\mathrm{len} = 32, |A| = 8$ とした．

アーキテクチャと最適化 エージェントのアーキテクチャ及び最適化手法についても [5] に従う．ここではアーキテクチャについて簡単に触れる．各エージェントはエンコーダ・デコーダモデルで表される．スピーカのデコーダとリスナのエンコーダは GRU[16] とし，スピーカのエンコーダとリスナのデコーダは順伝播型ニューラルネットワークとする．GRU の隠れ状態のサイズは [5] に倣い 500 とした．

境界判定アルゴリズム 境界判定アルゴリズム（節 3.2）にはパラメータ $\mathrm{max\_len}, \mathrm{threshold}$ が伴う．各メッセージが固定長 $\mathrm{len} = 32$ であることに合わせて $\mathrm{max\_len} = \mathrm{len} - 1 = 31$ とする．threshold の設定は自明ではないため，いくつかのパターンを試す:

$$\mathrm{threshold} \in \{0, 1/4, 1/2, 3/4, 1, 5/4, 3/2, 7/4, 2\}. \quad (7)$$

4.3 試行回数とデータの妥当性

式 6で設定した各 (a, v) に関して，異なるランダムシードで 4 回エージェントを学習させることとする．学習の結果，入力集合 I のうち 99% 以上の入力 $i \in I$ に対して $i = L(S(i))$ となるに至ったエージェントから得られる創発言語を**妥当な創発言語**と呼ぶことにする．

5 実験結果

$(a, v) = (12, 2)$ について 3 つの妥当な創発言語が得られ，その他の (a, v) についてはそれぞれ 4 つの妥当な創発言語が得られた．

5.1 条件付きエントロピーは単調減少する

問 1 に答えるため，H（式 4）が単調減少するかどうかを調べた．図 1にその結果を示す．妥当な創発言語における $H(n)$（赤実線）は，n に関して単調減少していることが見て取れる．従って妥当な創発言語において問 1 は成立する．なお，$H(n)$ の単調減少性はエージェントの学習前から成り立つわけではない．図 1に青破線で示したグラフを見ると，学習

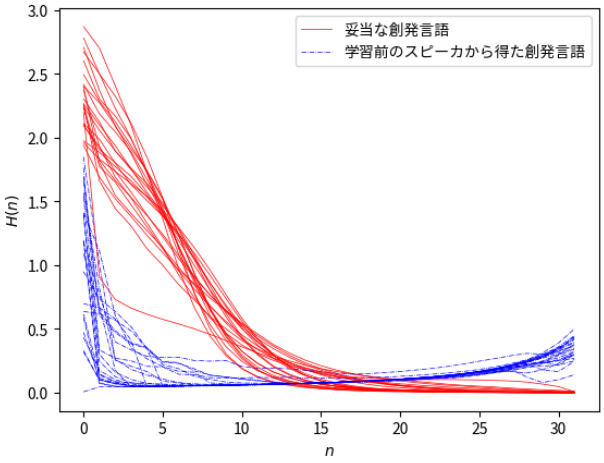

図 1　$H(n)$（式 4）のプロット．妥当な創発言語に関する結果（赤実線）と学習前のスピーカから得た創発言語に関する結果（青破線）を示してある．

前の創発言語においては $H(n)$ が単調に減少するとは限らないことが見て取れる．つまり学習の過程で $H(n)$ の単調減少性が生じる．

5.2　分岐エントロピーは所々増大する

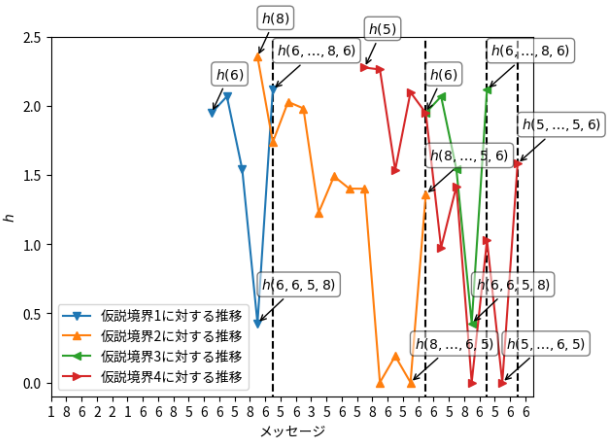

図 2　設定 $(a, v) = (2, 64)$ の下での創発言語における，あるメッセージ上での h の推移の様子．threshold = 5/4 として仮説境界 $1, \ldots, 4$ が得られた（黒破線）．

次に，問 2 に答えるため，妥当な創発言語における h の推移を調べた．例として実際の h の推移の様子を図 2に示す．縦軸には h の値，横軸には $(a, v) = (2, 64)$ として得られた妥当な創発言語からサンプルしたメッセージ “$1, 8, 6, 2, \ldots$” を取った．さらに，threshold = 5/4 として得られた 4 つの仮説境界の位置を黒破線で示してある．各折れ線グラフが，対応する仮説境界点で threshold を超える増大を示した h の推移を表している．なお，h の増大点（式 5）の根拠となる部分系列は，メッセージの先頭や一つ前の仮説境界から始まるとは限らないことに注意されたい．図 1で見たような大域的な傾向がある一方で，図 2では h が増減を繰り返していることが見て取れる．また，メッセージあたりの仮説境界の数を調べた結果を図 3に示す．図 3は，各 threshold（式 7）を用いたときに，1 メッセージあたりにいくつ仮説境界が含まれているかを示している．threshold < 2 においては，どの (a, v) にも仮説境界が存在していることが見て取れる．以上より，妥当な創発言語において問 2 は成立する．

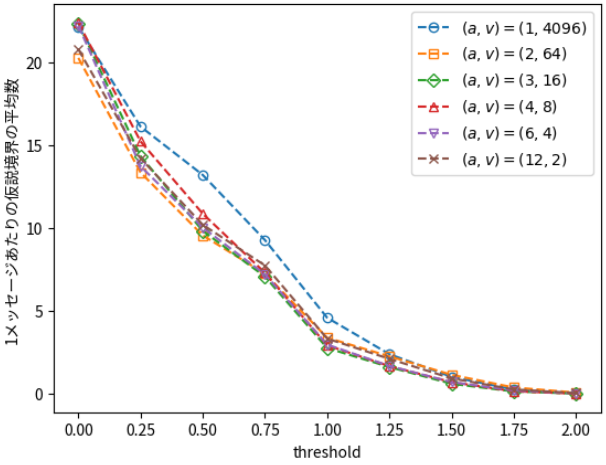

図 3　メッセージあたりの仮説境界の平均数．いくつかの threshold を試してある．

5.3　仮説境界は分節境界ではないかもしれない

再び図 3に着目すると，やや見づらいがどの threshold においても属性数 a に対して単調に仮説境界が増えるわけではないことが分かる[2)]．故に問 3’は成立しない．$a = 1$ のときに最も仮説境界が多くなる傾向があるようにさえ見える．

6　考察と今後の展望

実験の結果から，創発言語は Harris の分節原理が成立するためのいくつかの前提（問 1, 問 2）を満たすものの，仮説境界が分節境界を意味しない可能性が示唆された．創発言語に意味のある仮説境界をもたせるにはどうしたらよいか．創発言語の構成性を向上させる手法 [6] の適用が 1 つの方法として考えられる．一方，人間が与えた入力の意味に関わらず創発言語に仮説境界が生じた事実はそれはそれで興味深いことでもある．この仮説境界は何を意味するのか．これも今後解決していくべき課題である．

2)　単調に減るわけでもない．

参考文献

[1] Serhii Havrylov and Ivan Titov. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, **Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA**, pp. 2149–2159, 2017.

[2] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. Multi-agent cooperation and the emergence of (natural) language. In **5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings**. OpenReview.net, 2017.

[3] Angeliki Lazaridou, Anna Potapenko, and Olivier Tieleman. Multi-agent communication meets natural language: Synergies between functional and structural language learning. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020**, pp. 7663–7674. Association for Computational Linguistics, 2020.

[4] Satwik Kottur, José M. F. Moura, Stefan Lee, and Dhruv Batra. Natural language does not emerge 'naturally' in multi-agent dialog. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017**, pp. 2962–2967. Association for Computational Linguistics, 2017.

[5] Rahma Chaabouni, Eugene Kharitonov, Diane Bouchacourt, Emmanuel Dupoux, and Marco Baroni. Compositionality and generalization in emergent languages. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020**, pp. 4427–4442. Association for Computational Linguistics, 2020.

[6] Yi Ren, Shangmin Guo, Matthieu Labeau, Shay B. Cohen, and Simon Kirby. Compositional languages emerge in a neural iterated learning model. In **8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020**. OpenReview.net, 2020.

[7] Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. Anti-efficient encoding in emergent communication. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, **Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada**, pp. 6290–6300, 2019.

[8] Mathieu Rita, Rahma Chaabouni, and Emmanuel Dupoux. "lazimpa": Lazy and impatient neural agents learn to communicate efficiently. In Raquel Fernández and Tal Linzen, editors, **Proceedings of the 24th Conference on Computational Natural Language Learning, CoNLL 2020, Online, November 19-20, 2020**, pp. 335–343. Association for Computational Linguistics, 2020.

[9] Ryo Ueda and Koki Washio. On the relationship between zipf's law of abbreviation and interfering noise in emergent languages. In Jad Kabbara, Haitao Lin, Amandalynne Paullada, and Jannis Vamvas, editors, **Proceedings of the ACL-IJCNLP 2021 Student Research Workshop, ACL 2021, Online, JUli 5-10, 2021**, pp. 60–70. Association for Computational Linguistics, 2021.

[10] Eugene Kharitonov, Rahma Chaabouni, Diane Bouchacourt, and Marco Baroni. Entropy minimization in emergent languages. In **Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event**, Vol. 119 of **Proceedings of Machine Learning Research**, pp. 5220–5230. PMLR, 2020.

[11] Zellig S. Harris. From phoneme to morpheme. **Language**, Vol. 31, No. 2, pp. 190–222, 1955.

[12] Kumiko Tanaka-Ishii. Entropy as an indicator of context boundaries: An experiment using a web search engine. In Robert Dale, Kam-Fai Wong, Jian Su, and Oi Yee Kwong, editors, **Natural Language Processing - IJCNLP 2005, Second International Joint Conference, Jeju Island, Korea, October 11-13, 2005, Proceedings**, Vol. 3651 of **Lecture Notes in Computer Science**, pp. 93–105. Springer, 2005.

[13] Kumiko Tanaka-Ishii and Zhihui Jin. From phoneme to morpheme: Another verification using a corpus. In Yuji Matsumoto, Richard Sproat, Kam-Fai Wong, and Min Zhang, editors, **Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead, 21st International Conference, ICCPOL 2006, Singapore, December 17-19, 2006, Proceedings**, Vol. 4285 of **Lecture Notes in Computer Science**, pp. 234–244. Springer, 2006.

[14] 田中久美子. 言語とフラクタル. 東京大学出版会, 2021.

[15] David K. Lewis. **Convention: A Philosophical Study**. Wiley-Blackwell, 1969.

[16] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL**, pp. 1724–1734. ACL, 2014.