

簡素なモデルでの創発言語の接触による クレオール単純化の再現

加藤大地¹ 上田亮¹ 宮尾祐介¹¹ 東京大学

{daichi5967, ryoryoueda, yusuke}@is.s.u-tokyo.ac.jp

概要

「異なる言語同士の接触で生まれる言語は、単純化される傾向にある」という性質を**クレオール単純化**という。先行研究には、ニューラルネットワークで構成されたエージェント同士のやり取りとして発生する人工的言語、**創発言語**を用いて、クレオール単純化の再現に成功した例がある。しかし利用されたモデルが煩雑で、クレオール単純化の本質的原因が、言語同士の接触自体にあると断定できるものではなかった。そこで我々は、最大限まで簡素化された創発言語の接触のモデルを設計し、クレオール単純化の再現を試みた。複数の評価指標を用いて言語の複雑さを評価し、簡素なモデルでもクレオール単純化が現れることを実験的に示した。

1 はじめに

異なる言語を話す人間同士が意思疎通を図る中で新たに発生する言語は、往々にして単純化される傾向にある。言語のこの性質を**クレオール単純化**と呼ぶ [13, 15, 1]。クレオール単純化がなぜ、どのように生じるかは、言語学の未解決な問題である。

この問題に対する切り口の1つとして、**創発言語**の枠組みを使う研究が存在する。創発言語とは、ニューラルネットワークで構成された人工エージェント同士の間に生じる、人工的なプロトコルのことである。人間の言語の発現・進化の歴史の解明につながる可能性があり、昨今の技術向上により現実に近い形式で言語創発を機械的に模倣できるようになってきたことも相まって、創発言語は注目を集めている分野である [9]。

Graesser ら [6] は、シミュレーションの中で人工的に創発言語同士を接触させ、実際にクレオール単純化を発現させることに成功した。しかし、クレオール単純化の再現のみに焦点を当てた場合、彼ら

のモデルはかなり煩雑である。ゆえに、クレオール単純化が起こる本質的原因が、本当に言語の接触自体にあるのか、それとも単に彼らの煩雑なエージェントのアーキテクチャの構造や実験設定が、偶然クレオール単純化を発生させるに至っただけなのか、判別が困難である。また、この研究の後続研究としてクレオール単純化を議論する際に、基盤となるモデルが簡素¹⁾なものでないと、さらに踏み込んだ設定での実験が難航する、という問題点もある。

そこで我々は、「煩雑なモデルではなく、言語同士の接触自体にクレオール単純化の本質が眠っているはずである」という仮説を立てた。この仮説の検証のため、最大限まで簡素化されるようにデザインされた創発言語の接触モデルを使って、それでもなおクレオール単純化が生じるか実験した。

我々は、Graesser ら [6] が用いた**エントロピー**のみならず、**N-gram**、**汎化性能**、**学習のしやすさ**、**TopSim** のような指標も用いて、言語の複雑さを計測した。実験の結果、我々の簡素なモデルでも、様々な評価指標でクレオール単純化が実際に発現することが確認された。これにより、クレオール単純化の再現には、先行研究のような煩雑なモデルを用いる必要性はないことが示唆された。今後、この簡素化されたモデルをベースに、創発言語の立場からクレオール単純化の議論が進むことが期待される。

2 背景

2.1 クレオール単純化

異なる言語同士が接触した際に、言語にどのような影響が出るかを研究する分野として **Contact Linguistics** [14, 17] がある。その中で、最も研究さ

1) 本論文では、「言語の簡単さ」「モデルの簡単さ」という2種類の「簡単さ」が登場する。混乱を避けるため、明示的に「言語が簡単(単純)」「モデルが簡素」と区別して語用する。また、同様に「難しさ」についても、「言語が難しい(複雑)」「モデルが煩雑」と明示的に区別する。

れている対象の一つとして、異なる言語を話す集団同士が接触して生じる新しい言語である、**クレオール言語**が挙げられる。クレオール言語の顕著な特徴として、他の自然言語に比べて、より簡単になる傾向があることが知られており [13, 15, 1]、本稿ではこれを**クレオール単純化**と呼ぶ。

2.2 創発言語の発生手法

創発言語分野では、言語を機械的に再現するために、**ゲーム**と呼ばれるタスクを設定し、その成功率が上がるようモデルを学習する、という形で実験を行うことが多い。ゲームは、Lewis[11]による**シグナリングゲーム**をもとに設計されることがほとんどである。本節で、ゲームの流れを大まかに説明する。

ゲームには、**sender**、**receiver** と呼ばれる 2 種類のエージェントが登場する。まず sender は、入力に属する空間 I からランダムに抽出された入力 i を受け取る。 i のデータ形式はゲームの設定によって異なり、画像、テキスト、数字列など多岐にわたる。sender はこの入力を見て、記号列 (**メッセージ**) m を生成し、receiver に送信する。receiver はこのメッセージのみを頼りに、sender の入力 i が I 中のどれにあたるかを予測し、その推定入力 \hat{i} を出力として生成する。その後、 i と \hat{i} が一致しているかに応じてゲームの成功・失敗が判定され、その結果をもとにエージェントの内部状態が更新されて、学習が進んでいく。メッセージだけが 2 つのエージェントを接続する架け橋になることを考えると、学習が進みゲームの成功率が上がったとき、そのメッセージは sender、receiver にとって共通に理解できる、意味を持つ記号列になるはずである。それを言語と見做し、その性質を解析するのが、創発言語という分野である。

2.3 創発言語の評価指標

構成性 言語には**構成性**と呼ばれる性質があることが知られている。自然言語で使われる単語はそれほど多くないが、ルールに従って単語を組み合わせることで、膨大な事象を表現できるようになる。例えば、「黄色のバナナ」と「緑色のバナナ」を表現する際、これらに特別な単語を割り当てるのではなく、色を表す単語と「バナナ」という対象を表す単語を組み合わせることで対応する。**Topographic Similarity (TopSim)** [2, 10] は、言語の構成性を測る指標としてよく知られている。入力が属する空間を I 、メッ

セージが属する空間を M とする。TopSim は I における 2 つの入力の距離と、 M におけるこれらの入力に対応する 2 つのメッセージの距離との相関係数として定義される。TopSim が高いほど、構成性が高いことを表す。この指標は「入力同士の意味が近いときには、対応するメッセージ同士も似たようなものになるはずである」という直感に基づく。

不確実性 Graesser ら [6] は言語の複雑さを計測する指標として、sender のメッセージ生成の際の**エントロピー**を用いた。彼らは「ある事象に対し、その表現方法の不確実性が大きいほど、難しい言語と言えるはずである」という直感の下、エントロピーが高いほど、言語の複雑さが大きいと評価した。

2.4 創発言語の接触に関する先行研究

Graesser ら [6] の研究は、創発言語同士の接触により、クレオール単純化の再現に初めて成功した研究である。しかし、彼らはクレオール単純化のみならず、接触による言語への他の影響についても同時に再現を試みており、実験設定は 2.2 節でのゲームと以下のような点で異なっていて、煩雑である：

- 各エージェントが sender、receiver の両方の役割を果たし、1 ゲームで複数回メッセージを交換するマルチステップな設定にしている。
- 分割された画像の半分が入力として与えられ、もう一方のエージェントの入力画像と合わせたものが全入力のうちどれに当たるかを推定する必要があり、一方通行の意思疎通が起こらないようにしている。
- エージェントが 2 人より多く登場する。

また、メッセージとして単なる 1 記号を採用しており、創発言語分野で重要な構成性 (2.3 節) の評価を行うことができない、という問題点もあった。

3 実験手法

ゲームの基本的な流れは 2.2 節で説明したものに沿う。各入力 i は n_{att} 個の値を連ねたタプルであり、各値は 0 から $n_{val} - 1$ の n_{att} 種類の値を取り得る。また、 c_{voc} 種類の相異なるシンボルからなる集合 V から、シンボルを c_{len} 回選び (重複も許される)、並べたものをメッセージ m として定義する。

創発言語の接触を、次のように発生させる (図 1):

1. 2 つの sender-receiver のペア $P_1 = (S_1, R_1)$, $P_2 = (S_2, R_2)$ を用意し、それぞれ成功率が 99% 以上

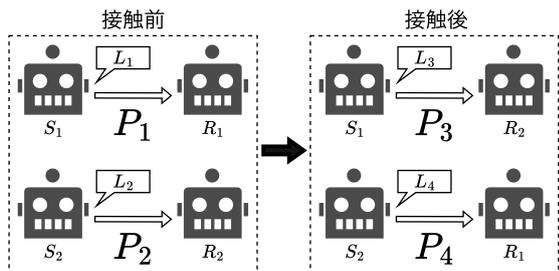


図1 創発言語の接触を生じさせる流れを表した簡略図

表1 sender の内部状態で決まるカテゴリカル分布のエントロピーの値を、プロトコルごとに計算したもの。値は四捨五入、括弧の中の数字は標準誤差を表す。

L_1	L_2	L_3	L_4
0.443	0.449	0.327	0.328
(±0.0043)	(±0.0041)	(±0.0097)	(±0.0091)

になるまで、独立に学習を行う。ここで創発する2つのプロトコルをそれぞれ L_1, L_2 と呼ぶ。

- receiver の「交換」を行う。つまり、新たなペア $P_3 = (S_1, R_2), P_4 = (S_2, R_1)$ を作成する。
- P_3, P_4 で正解率が99%以上になるまで、独立に学習を行う。ここで創発する2つのプロトコルをそれぞれ L_3, L_4 と呼ぶ。

L_1 を話す sender と L_2 を聞く receiver によるクレオール L_3, L_2 を話す sender と L_1 を聞く receiver によるクレオール L_4 が作成できる。 L_1, L_2 と L_3, L_4 を比較して、クレオール単純化が発現するか様々な評価指標で検証するのが、本実験の主旨である。

4 実験結果とその解釈

我々は複数のパラメータで実験を行った。 $n_{att} = 2, n_{val} = 100, c_{voc} = 10, c_{len} = 6$ の設定の結果を掲載するが、他のパラメータ設定でも同様の考察ができる結果を得た。掲載する結果は、同じパラメータに対し random seed のみを変更しながら100回実行を行い、平均を取ったものである。

4.1 エントロピー

学習時、sender は内部のカテゴリカル分布からメッセージをサンプルする。学習の最後の epoch におけるカテゴリカル分布のエントロピーを表1に示す。接触後言語の方がエントロピーが低く、Graesserら[6]の研究と同様にクレオール単純化が発現した。

4.2 N-gram

テスト時、sender は内部状態に応じて決定的にメッセージを排出する。あり得る入力全てに対して

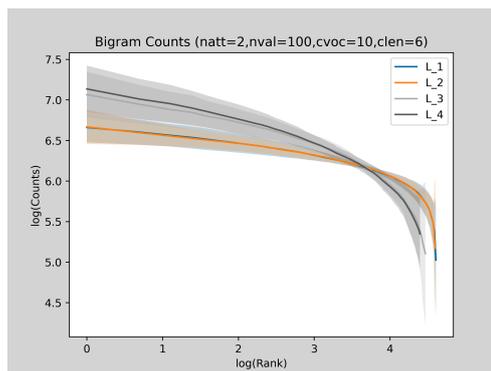


図2 各言語プロトコルにおける、単語ペアの頻度の順位と頻度の関係を表したグラフ。縦横軸ともに対数表示。薄い色で塗られた部分は標準誤差によるばらつきを表す。

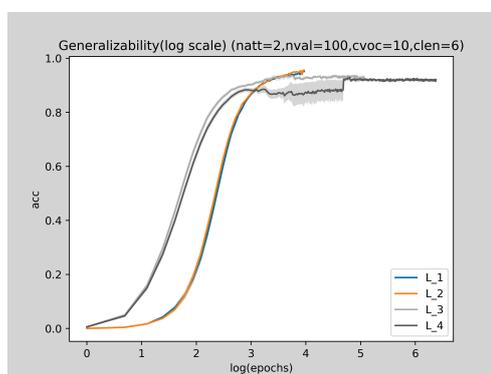


図3 各言語プロトコルにおける、epoch ごとのテストデータに対するゲームの成功率の推移。横軸のみ対数表示。薄い色で塗られた部分は標準誤差によるばらつきを表す。

sender にメッセージを出力させることで、sender が話す言語のコーパスのようなものを作成できる。このコーパスを用いて、各単語 (unigram)、各単語ペア (bigram) の出現頻度の分布を作成する。代表して bigram の結果を載せる (図2)。横軸が各単語ペアが現れる頻度が高い順に並び替えた順位、縦軸がその頻度自体を表す。標準誤差区間がやや重なっていて、明らかな違いがあるとは言い難いが、接触後言語の方が頻度の分布に偏りがある結果になっていると言える。分布に偏りがあることは、エントロピーが小さいことを表し²⁾、クレオール単純化が発現している結果であると言える。

4.3 汎化性能

Chaabouni ら [3] と同様に、学習データには含まれない入力を使って成功率を計算し、その推移を見ることで、汎化性能を計測する (図3)。接触後言語の

2) 実際、最尤推定による確率分布でエントロピーを計算したところ、接触後言語の方が小さくなった。

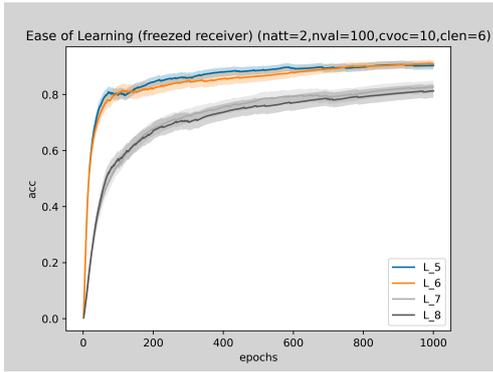


図4 receiverの状態をfreezeさせてsenderのみを学習させたときの、学習中のゲーム成功率の推移。L₅, L₆はL₁, L₂のfreezeされたreceiverをもとに学習した言語、L₇, L₈はL₃, L₄のfreezeされたreceiverをもとに学習した言語である。薄い色で塗られた部分は標準誤差によるばらつきを表す。

方が素早く成功率が向上している³⁾。「汎化性能が高い方が、少ない単語でより多くの事象を表現できるため、簡単な言語である」という直感を頼りにすると、接触後言語の方が汎化性能が高いことは、クレオール単純化が発現している結果であると言える。

4.4 学習のしやすさ

本実験では、**学習のしやすさ**として下記の3種類を用いる。

1. 学習の中での成功率の推移
2. senderのパラメタをfreezeさせ、新たに初期化したreceiverを学習する際の、成功率の推移
3. receiverのパラメタをfreezeさせ、新たに初期化したsenderを学習する際の、成功率の推移

2はChaabouniら[3]が利用した指標を参考に行っている。彼らは成功率の推移のグラフの下側の面積を定量的に比較したが、我々はグラフ自体をそのまま比較する。この指標は、「親が子に言語を教えること」に類推できる実験であると言える。1,2については特筆すべき点はないので、図の掲載は省略する。学習しやすい言語の方が簡単、つまり素早く成功率が上がる言語の方が簡単と直感できるが、実際に、接触後言語の方が素早く成功率が上がるという結果が得られた。

3も前段落の「親が子に言語を教える」類推が同

- 3) 実行ごとに学習を打ち切るepochが異なっており、epochが大きくなるにつれ学習が継続されている実行が減るため、epochが大きい部分では値のブレが大きくなっている。このブレは現象の平均をとって見る際には非本質的な結果なので、epochが小さい部分について議論するのが望ましい。

表2 TopSimの値をプロトコルごとに計算したもの。値は四捨五入、括弧の中の数字は標準誤差を表す。

L ₁	L ₂	L ₃	L ₄
0.0957	0.0949	0.1417	0.1383
(±0.0034)	(±0.0038)	(±0.0047)	(±0.0044)

様にできるが、本実験で唯一、クレオール単純化に基づく直感に反する結果を得た(図4)。異なるパラメタ設定でも、接触前と後で差異がない、もしくは接触前の言語の方が素早く成功率が上がるという結果になった。結果から、sender、つまり言語を発する側にとっては、クレオール単純化が本質的に起こらず、むしろクレオールが、クレオールではない言語と同等もしくはそれ以上の難しさを孕んでいる可能性が示唆された。ただ、強化学習により学習が進むsenderの学習の促進のために、損失関数に足された「エントロピー増大」のための項が、結果に不自然な働きかけをしているだけで、実際のクレオールに本質的に備わる性質ではない可能性を否定できないため、将来の研究で更なる検証が求められる。

4.5 TopSim

2.3節で説明したTopSimを計算した結果が表2である。Chaabouniら[3]は、TopSimの高さとsenderを固定したときのreceiverの学習のしやすさが強く相関することを指摘しており、簡単な言語が学習しやすいという直感から推論すると、簡単な言語はTopSimが高くなる、という仮説が立つ。実際、接触後言語の方が明らかにTopSimが高くなっており、クレオール単純化が発現している結果とみることができる。

5 おわりに

本論文では、簡素な創発言語接触のシミュレーションモデルを利用し、クレオール単純化が再現することを、多様な評価指標を用いて検証した。実際に多くの指標においてクレオール単純化が発現していると解釈できる結果が得られたが、receiverを固定してsenderを学習するという設定においてのみ、異なる結果が得られた。この点に関しては、学習の方法を変えて再度実験する等の検討が必要になるだろう。しかし、煩雑なモデルを使わなくとも、数多くの手法でクレオール単純化を確かめられたため、今後この簡素なモデルをベースに、創発言語の立場からクレオール単純化の議論が進むことを期待できる。

参考文献

- [1] Peter Bakker, Aymeric Daval-Markussen, Mikael Parkvall, and Ingo Plag. Creoles are typologically distinct from non-creoles. **Journal of Pidgin and Creole languages**, Vol. 26, No. 1, pp. 5–42, 2011.
- [2] Henry Brighton and Simon Kirby. Understanding linguistic evolution by visualizing the emergence of topographic mappings. **Artif. Life**, Vol. 12, No. 2, pp. 229–242, 2006.
- [3] Rahma Chaabouni, Eugene Kharitonov, Diane Bouchacourt, Emmanuel Dupoux, and Marco Baroni. Compositionality and generalizability 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.
- [4] 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.
- [5] Kyunghyun Cho, Bart van Merriënboer, Çağlar 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.
- [6] Laura Graesser, Kyunghyun Cho, and Douwe Kiela. Emergent linguistic phenomena in multi-agent communication games. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019**, pp. 3698–3708. Association for Computational Linguistics, 2019.
- [7] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In **5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings**. OpenReview.net, 2017.
- [8] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, **3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings**, 2015.
- [9] Angeliki Lazaridou and Marco Baroni. Emergent multi-agent communication in the deep learning era. **CoRR**, Vol. abs/2006.02419, , 2020.
- [10] Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark. Emergence of linguistic communication from referential games with symbolic and pixel input. In **6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings**. OpenReview.net, 2018.
- [11] David Lewis. **Convention: A philosophical study**. John Wiley & Sons, 2008.
- [12] Chris J. Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. In **5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings**. OpenReview.net, 2017.
- [13] John H McWhorter. The worlds simplest grammars are creole grammars. 2001.
- [14] Carol Myers-Scotton, et al. **Contact linguistics: Bilingual encounters and grammatical outcomes**. Oxford University Press on Demand, 2002.
- [15] Mikael Parkvall, et al. The simplicity of creoles in a cross-linguistic perspective. **Language complexity: Typology, contact, change**, pp. 265–285, 2008.
- [16] John Schulman, Nicolas Heess, Theophane Weber, and Pieter Abbeel. Gradient estimation using stochastic computation graphs. In Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, editors, **Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada**, pp. 3528–3536, 2015.
- [17] Sarah Grey Thomason and Terrence Kaufman. **Language contact**. Edinburgh University Press Edinburgh, 2001.
- [18] Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. **Mach. Learn.**, Vol. 8, pp. 229–256, 1992.

本文で説明しきれなかったモデルのアーキテクチャや、学習に使うアルゴリズムについて紹介する。本実験は Chaabouni ら [3] の研究を基にしており、以下で紹介するアーキテクチャやアルゴリズムは Chaabouni ら [3] のものと同じものであるため、合わせて参考にされたい。

A モデルのアーキテクチャ

sender、receiver ともに、内部に GRU [5] による層を持っている。これらのエージェントのアーキテクチャを以下に簡潔に説明する。

A.1 Sender

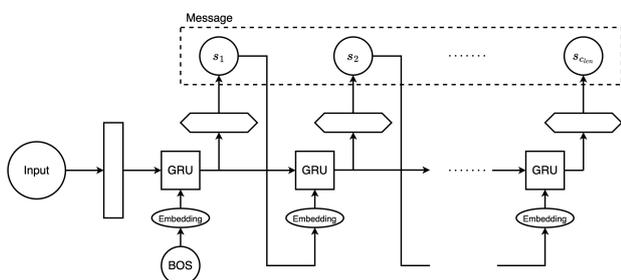


図 5 sender エージェントのアーキテクチャの概略図。ただし、 c_{len} はメッセージの長さを表す。

sender はまず、受け取った入力を GRU セルの隠れ状態ベクトルとして使う。BOS の埋め込みベクトルと同時に入力し、そこから出てくる新たな隠れ状態ベクトルから、メッセージの最初のシンボル s_1 を生成する。学習時は、学習の探索を促すため、隠れ状態ベクトルから作られるパラメタを基にした分布に従って、確率的にシンボルを生成するが、テスト時は、パラメタの中でもっとも値が大きいシンボルを貪欲に選択してシンボルを生成する。さらに、この隠れ状態ベクトルと、生成されたメッセージは、次の GRU セルに渡され、後続のシンボルの生成に再帰的に利用される。

A.2 Receiver

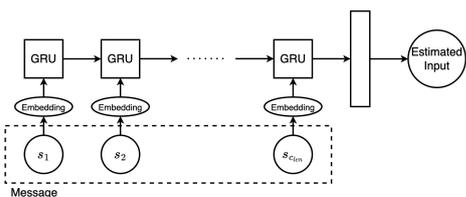


図 6 receiver エージェントのアーキテクチャの概略図。ただし、 c_{len} はメッセージの長さを表す。

receiver は、sender から受けとったメッセージを、

1 シンボルずつ GRU セルで消費する。全てのシンボルを消費し終わった後に出てくる隠れ状態ベクトルを線形層で整形し、それを推定入力として出力する。

B 学習

創発言語のほとんどのゲームにおいて、メッセージは離散的であるため、学習時に通常利用される誤差逆伝播法を利用することができない。これに対するアプローチとして、

- Gumbel-Softmax [12, 7]：離散分布であるカテゴリカル分布を Gumbel-Softmax 分布と呼ばれる別の連続分布で近似する手法。
- REINFORCE [18]：強化学習における方策勾配法の一種。

を利用することが多い。しかし、Chaabouni ら [3] の研究では、新たな“hybrid method”として、確率的誤差逆伝播法と REINFORCE を組み合わせた学習アルゴリズム [16, 4] を利用して学習を行った。

θ_S, θ_R を、それぞれ sender、receiver の内部状態を表すパラメタであるとする。また、 $\mathcal{L}(i, \hat{i}(m; \theta_R))$ を、入力 i と receiver の推定入力 $\hat{i}(m; \theta_R)$ との cross-entropy loss であるとする。学習の目標を、この cross-entropy loss の期待値 $E(\mathcal{L}(i, \hat{i}(m; \theta_R)))$ を最小化することと定める。Chaabouni ら [4] は、代替関数を、

$$\mathcal{L}(i, \hat{i}(m; \theta_R)) + \left(\{\mathcal{L}(i, \hat{i}(m; \theta_R))\} - b \right) \log p_S(m | i; \theta_S) \quad (1)$$

のように定め、この代替関数の勾配の期待値が、cross-entropy loss の期待値の勾配に等しいことを利用して、学習を進めた。ただし、 $\{\cdot\}$ は stop-gradient 演算子、 $p_S(m | i; \theta_S)$ は入力 i を受け取った sender がメッセージ m を返す確率、 b はベースラインである。(1) 式の第一項は receiver を学習するのに使われる。この部分は単純に誤差逆伝播法を利用すれば最適化できる。また、(1) 式の第二項は sender を学習するのに使われる。この部分は REINFORCE によく似たアルゴリズムを利用することで最適化できる。このように、全体のモデルを 2 つのアルゴリズムを組み合わせたような形で利用して学習を行うことから、Chaabouni ら [3] はこの手法を“hybrid method”と呼んでいる。

なお、optimizer として、Adam [8] を利用する。