

自然言語推論用データセット

SNLI

Stanford Natural Language Inference Corpus

sasakiy84

1. SNLI の文対と 3 つの推論関係
2. 文対の作成方法と収録規模
3. 自然言語推論の評価とモデル性能
4. SNLI の利用動向と後続データ
5. 文対データの入手方法と利用条件

1. SNLI の文対と 3 つの推論関係

SNLI の公式サイト



The screenshot shows the official website of the Stanford Natural Language Inference (SNLI) Corpus. The header features the Stanford NLP Group logo and navigation links: people, research blog, software, seminar, join, donate, local. The main content area is titled "The Stanford Natural Language Inference (SNLI) Corpus". It defines Natural Language Inference (NLI) as the task of determining the inference relation between two (short, ordered) texts: *entailment*, *contradiction*, or *neutral* (MacCartney and Manning 2008). The "The Corpus" section states that the SNLI corpus (version 1.0) is a collection of 570k human-written English sentence pairs manually labeled for balanced classification with the labels *entailment*, *contradiction*, and *neutral*. It aims to serve as a benchmark for evaluating representational systems for text, especially those induced by representation-learning methods, as well as a resource for developing NLP models of any kind. A citation for the paper "A large annotated corpus for learning natural language inference" by Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning is provided, along with links to the PDF and BibTeX files.

The Stanford Natural Language Processing Group

people research blog software seminar join donate local

The Stanford Natural Language Inference (SNLI) Corpus

Natural Language Inference (NLI), also known as Recognizing Textual Entailment (RTE), is the task of determining the inference relation between two (short, ordered) texts: *entailment*, *contradiction*, or *neutral* (MacCartney and Manning 2008).

The Corpus

The Stanford Natural Language Inference (SNLI) corpus (version 1.0) is a collection of 570k human-written English sentence pairs manually labeled for balanced classification with the labels *entailment*, *contradiction*, and *neutral*. We aim for it to serve both as a benchmark for evaluating representational systems for text, especially including those induced by representation-learning methods, as well as a resource for developing NLP models of any kind.

The following paper introduces the corpus in detail. If you use the corpus in published work, please cite it:

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. [pdf] [bib]

前提文と仮説文の関係を 3 種類に分類する

- 自然言語推論（Natural Language Inference、NLI）は、前提の記述をもとに別の記述が成り立つかを判断する課題である。
- SNLI では、状況を述べる前提文と、成り立つかを調べる仮説文を入力する。写真説明文から作られたデータだが、モデルの入力は 2 つの文だけである。
- 判断の選択肢は、仮説が導ける「含意」、真偽が決まらない「中立」、両立しない「矛盾」の 3 つである。

実例 1 : 前提から仮説が導ける含意

前提文

A man playing an electric guitar on stage.

仮説文

A man playing guitar on stage.

正解は含意

電気ギターを弾いているなら、ギターを弾いている。

実例 2：仮説の真偽が決まらない中立

前提文

A man reads the paper in a bar with green lighting.

仮説文

The man is reading the sportspage.

正解は中立

新聞のどの面を読んでいるかは、前提に書かれていない。

実例 3 : 前提と仮説が両立しない矛盾

前提文

3 young man in hoods standing in the middle of a quiet street facing the camera.

仮説文

Three people sit by a busy street bareheaded.

正解は矛盾

立っていることと座っていることなどが両立しない。

語の一致ではなく文が述べる関係を判定する

- 電気ギターを弾く人は、ギターを弾く人でもある。このように前提の情報から仮説が導ける場合が含意である。
- 新聞を読むという情報だけでは、スポーツ面を読むかは決まらない。情報が足りない場合が中立である。
- 同じ場面で立っていることと座っていることが両立しないように、前提と仮説が衝突する場合が矛盾である。

「片手」と「両手」の違いは矛盾か

前提文の一部

with one hand holding a skinny but long stick

仮説文の一部

with both hands holding a skinny but long stick

正解は矛盾

「片手だけ」と読むなら、両手で持つ仮説と矛盾する。片手という表現が、他方の手を排除するかは曖昧である。

2. 文対の作成方法と収録規模

先行する RTE と SNLI の規模の違い

- 含意を判定する先行ベンチマークとして RTE があった [1]。
 - RTE 第 3 回（2007 年）は、開発 800 組＋テスト 800 組だった。
- それに対して SNLI（2015 年）は 570,152 組を収録した。
 - 合計 1,600 組の RTE 第 3 回に対し、収録数は約 356 倍となる。
- その結果、ニューラルモデルを学習しやすくなった。

[1] Giampiccolo et al. (2007). ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. The Third PASCAL Recognizing Textual Entailment Challenge. §2.
<https://aclanthology.org/W07-1401/>

[2] Bowman et al. (2015). EMNLP, pp. 632–642. A large annotated corpus for learning natural language inference. <https://aclanthology.org/D15-1075/>

大規模な人手作成の文対で学習と評価を行う

- 先行する含意判定課題 RTE に対し、SNLI はニューラルモデルを学習できる規模の文対を用意した。
- 写真の説明文を前提とし、人が 3 種類の関係に対応する仮説を書いた。写真の同じ状況を想定して関係を判断する。
- Stanford Natural Language Inference Corpus として、570,152 組を収録し、2015 年の EMNLP で発表された。

写真説明文をもとに 3 種類の仮説を書く

- 写真に英語の説明文を付けたデータ Flickr30k と、写真の物体や関係を注釈する Visual Genome の初期データを素材にした。
- 作業者には写真を見せず、その説明文だけを提示した。同じ写真の状況について、確実に真・真かもしれない・確実に偽の文を各 1 つ書いた。
- この仮説文の作成には約 2,500 人が参加した。各文対には、作成した仮説に対応するラベルが付いた。

5 人の票から正解ラベルを決める

判定者	含意	中立	矛盾
作成者＋追加の 4 人	3 票	1 票	1 票
正解ラベル	含意	—	—

3 票以上が一致すれば、そのクラスを正解とする。3 票に達するクラスがなければ「-」を付ける。票の配分は規則を示す例。

追加検証の合意がない文対を 3 分類から除く

- 56,951 組には追加の 4 人が判定を行い、作成者のラベルと合わせて 1 組 5 票を得た。
- 検証対象の 98% では 3 票以上が一致し、 58.3% では全員が一致した。検証は約 30 人の信頼された作業者が担当した。

[1] Bowman et al. (2015). EMNLP, pp. 632–642. A large annotated corpus for learning natural language inference. <https://aclanthology.org/D15-1075/>

[2] <https://nlp.stanford.edu/projects/snli/>

収録規模と 3 分類に使う文対数

分割	元の収録数	正解ラベル付き
学習用	550,152	549,367
開発用	10,000	9,842
テスト用	10,000	9,824
合計	570,152	569,033

右列は、正解ラベルが付いた SNLI 1.0 の文対数。

文対数と前提文字列の種類数

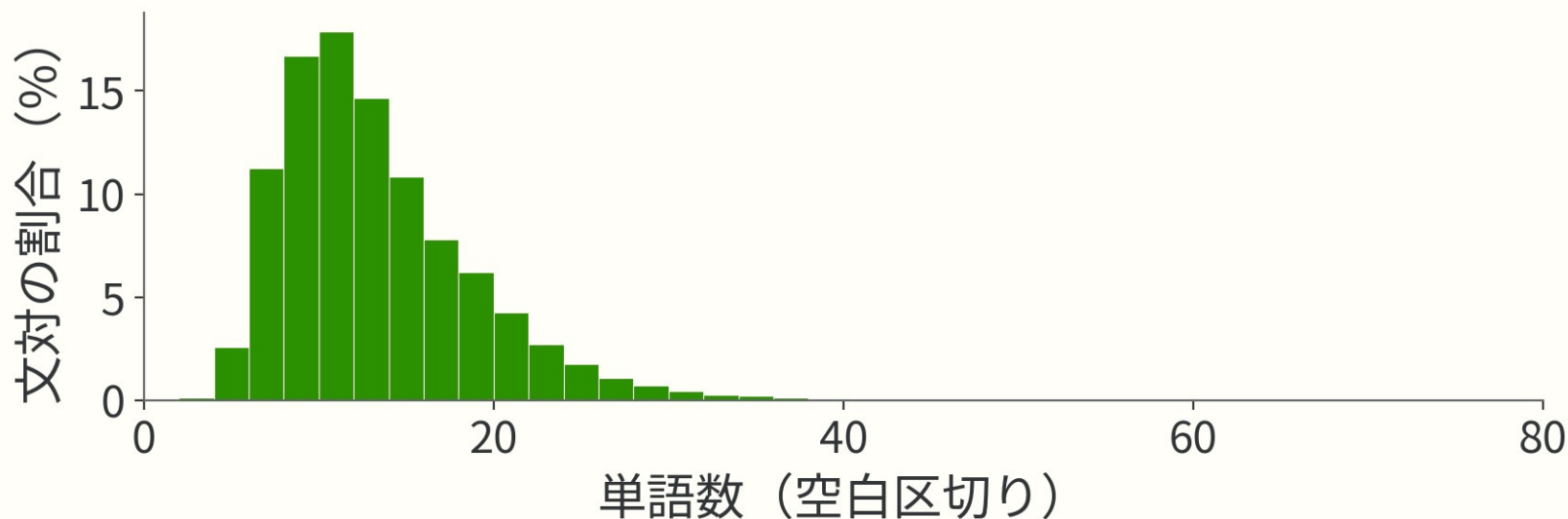
分割	正解付き文対	異なる前提文字列
学習用	549,367	150,736
開発用	9,842	3,319
テスト用	9,824	3,323

同じ前提から複数の仮説が作られるため、前提は繰り返し現れる。種類数は完全一致の文字列数。

前提文・仮説文の単語数と集計単位

- 含意の例では、前提の “A man playing an electric guitar on stage.” は 8 語、仮説の “A man playing guitar on stage.” は 6 語である。
- 単語数は空白で区切って数える。前提と仮説は役割が異なるため、それぞれの分布を調べる。
- 全体の集計では正解付き 569,033 文対を対象とし、同じ前提が複数の文対に現れる場合も文対ごとに数える。

前提文の単語数の分布



平均 12.89 語、中央値 12 語、最大 78 語。

前提が最長の実例

正解付き文対中で 402 文字・78 語。

A man wearing a helmet, red pants with white stripes going down the sides and a white and red shirt is on a small bicycle using only his hands while his legs are up in the air, while another man wearing a light blue shirt with dark blue trim and black pants with red stripes going up the sides is standing nearby, gesturing toward the first man and holding a small figurine of one of the seven dwarves.

仮説は “The people who are shown are not all doing the exact same thing.” 正解は含意。

仮説文の単語数の分布



平均 7.42 語、中央値 7 語、最大 56 語。

3. 自然言語推論の評価とモデル性能

文対 3 組の予測から正解率を計算する

正解	予測	得点
含意	含意	1
中立	矛盾	0
矛盾	矛盾	1

正解した 2 組を評価した 3 組で割り、 $2 \div 3 \times 100 = 66.7\%$ となる。

正解率は正解付き文対について集計する

- 各文対では、予測クラスと正解ラベルが一致すれば 1、不一致なら 0 とする。
- その得点を評価対象の文対全体で平均し、百分率で報告する。正解ラベルがない「-」の文対は除く。
- SNLI のテスト用では、正解付き 9,824 組を評価対象とする。

2015 年の LSTM の正解率は 77.6% だった

原著のモデル	テスト正解率
単語埋め込みの和	75.3%
単純な再帰型ネットワーク	72.2%
長期依存を扱う LSTM	77.6%

いずれも SNLI の学習用で学習し、テスト用で評価した [1]。

[1] Bowman et al. (2015). EMNLP, pp. 632–642. A large annotated corpus for learning natural language inference. Table 6 <https://aclanthology.org/D15-1075/>

SNLI で報告されたモデルの正解率

モデル	方式	テスト正解率
ESIM 2017 [1]	双方向 LSTM + 注意	88.0%
SemBERT 2020 [2]	BERT + 意味役割	91.9%

ESIM は文間の語の対応に注意を向け、SemBERT は誰が何をしたかを表す意味役割を BERT へ加える。

[1] ESIM 公開実装 . <https://github.com/coetaur0/ESIM>

[2] SemBERT 著者実装 . <https://github.com/cooelf/SemBERT>

文を数値表現へ変換して関係を予測する

- ニューラルモデルは、まず単語を数値のベクトルへ変換し、単語列から文の特徴を計算する。
- LSTM は、語を順に処理して離れた位置の情報を保持する再帰型ネットワークである。ESIM（2017）は双方向 LSTM と、文間の語の対応を重み付けする注意機構を使う。
- BERT は大量の文で事前学習した言語モデルである。SemBERT（2020）は、誰が何をしたかなどの意味役割を BERT へ加える。

[1] Bowman et al. (2015). EMNLP, pp. 632–642. A large annotated corpus for learning natural language inference. <https://aclanthology.org/D15-1075/>

[2] <https://github.com/coetaur0/ESIM>

[3] <https://github.com/cooelf/SemBERT>

仮説文だけで正解率約 67% に達する

- 仮説文だけでもラベルを推測できる注釈の偏りがある [1]。
 - 仮説だけを使う分類器で約 67% という報告がある。

[1] Gururangan et al. (2018). NAACL-HLT, pp. 107–112. Annotation Artifacts in Natural Language Inference Data. <https://aclanthology.org/N18-2017/>

[2] SNLI dataset card. <https://github.com/huggingface/datasets/blob/1.18.4/datasets/snli/README.md>

4. SNLI の利用動向と後続データ

MultiNLI による 10 ジャンルへの対象拡張

- その後、多ジャンルの推論データ MultiNLI が登場した [1]。
 - 2018 年の論文で発表された。
- SNLI は写真説明を扱った。これに対し、会話・小説など 10 ジャンルへ広げた [2]。
- さらに、未知のジャンルへの汎化も評価する。

[1] Williams et al. (2018). NAACL-HLT, pp. 1112–1122. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference.
<https://aclanthology.org/N18-1101/>

[2] MultiNLI 著者リポジトリ . <https://github.com/nyu-mll/multiNLI>

e-SNLI は関係ラベルに判断理由を加えた

- e-SNLI (2018) は、SNLI の文対へ自然言語で書いた判断理由を追加した拡張データである。
- 含意・中立・矛盾というラベルに加えて、なぜその関係になるかを文章として収録する。

ANLI はモデルが誤る文対を反復して集める

- Adversarial NLI （ ANLI 、 2020 ） は、人がモデルを誤らせる文対を作る推論データである。
- モデルの予測を作業者へ返し、難しい例を集める。この例でモデルを学び直す工程を繰り返す。
- 多ジャンルへ広げた MultiNLI 、説明を加えた e-SNLI とは、推論例を増やす方法が異なる。

推論ラベルを文の類似度の学習にも使う

- 文埋め込みは、文全体を 1 つの数値ベクトルで表す方法である。近い意味の文が近い位置になるように学習する。
- SimCSE （ 2021 ） の教師あり学習では、前提と含意の仮説を近付ける正例、前提と矛盾の仮説を区別する難しい負例として使う。
- この利用では、3 分類の正解率を上げるだけでなく、文同士の類似度に使える表現を学習する。

SNLI と MultiNLI をまとめて学習に使う

- AllNLI は、SNLI と MultiNLI をまとめた学習データである。写真説明文と多ジャンルの文対を含む。
- 文埋め込み用ライブラリ Sentence Transformers は、2026-10-15 の確認時点でも AllNLI による学習例を公開している。
- その学習例には、比較する負例を選び直す GISTEmbed（2024）の損失関数を用いる構成も含まれる。

5. 文対データの入手方法と利用条件

配布データでは文・クラス・識別子を対応させる

- Hugging Face 版では premise が前提文、hypothesis が仮説文、label が正解クラスを表す。
- 公式 ZIP は 1 行に 1 件を格納する JSONL 形式で、文対 ID と作業者の判定も含む [1]。
- カード掲載の容量は取得時約 20.4 MB、展開後約 68.4 MB である [2]。
- データの利用条件は CC BY-SA 4.0 である [2]。

[1] <https://nlp.stanford.edu/projects/snli/>

[2] Stanford NLP, SNLI — Dataset Card. <https://huggingface.co/datasets/stanfordnlp/snli>

公式 ZIP には文対 ID と作業者の票も入る

- sentence1 は前提文、 sentence2 は仮説文、 gold_label は正解ラベルである。
- pairID は文対の識別子、 annotator_labels は作業者による判定の一覧を表す。
- 正解ラベルが「-」の文対は、 3 分類の学習・評価から除く。数値でラベルを返す配布サービスでは、未確定のラベルは負の値で表される。

Hugging Face から文対を取得する

```
from datasets import load_dataset

ds = load_dataset("stanfordnlp/snli")
valid = ds.filter(lambda row: row["label"] >= 0)

print(valid["test"][0])
print(ds["train"].features["label"].names)
```

データ配布サービスの Hugging Face を利用する。前提は premise、仮説は hypothesis、クラスは label に入る。

SNLI の目的・文対の特徴と利用動向

- 英語の文対を、含意・中立・矛盾に分類する。
- 2015 年に、大規模な学習用データとして作られた。
- 写真説明文から人が仮説を作り、一部を再検証した。
- MultiNLI はジャンル、 e-SNLI は説明を拡張した。
- 現在も、文埋め込みの教師データとして使われる。

Appendix

1. 開発者の研究と発表先
2. 前提・仮説と推論ラベルの追加例
3. 文対数・文長の統計と極値の実例
4. 推論モデルの評価と文対の配布・作成仕様
5. 自然言語推論の関連文献

SNLI の開発者と研究分野

- SNLI は 2015 年 9 月の EMNLP で発表された（リスボン） [1]。
- 第一著者の Samuel R. Bowman は自然言語推論を研究する。
 - SNLI に加え、多ジャンルの MultiNLI の共著者である [2]。
- 最終著者の Christopher D. Manning は単語表現などを研究する。
 - 単語ベクトルを学習する GloVe の共著者である [3]。

[1] Bowman et al. (2015). EMNLP, pp. 632–642. A large annotated corpus for learning natural language inference. <https://aclanthology.org/D15-1075/>

[2] Williams et al. (2018). NAACL-HLT, pp. 1112–1122. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. <https://aclanthology.org/N18-1101/>

[3] Pennington et al. (2014). EMNLP. GloVe: Global Vectors for Word Representation. <https://aclanthology.org/D14-1162/>

追加例 1 含意

前提文

This church choir sings to the masses as they sing joyous songs from the book at a church.

仮説文

The church is filled with song.

正解は含意

合唱していることを、教会が歌で満たされていると言い換える。

追加例 2 中立

前提文

A woman with a green headscarf, blue shirt and a very big grin.

仮説文

The woman is young.

正解は中立

服装や表情だけでは、若いかどうかは判断できない。

追加例 3 含意

前提文

An old man with a package poses in front of an advertisement.

仮説文

A man poses in front of an ad.

正解は含意

advertisement と ad は同じ内容を表す。

追加例 4 中立

前提文

A statue at a museum that no seems to be looking at.

仮説文

The statue is offensive and people are mad that it is on display.

正解は中立

見られていない理由や人々の感情は、前提から分らない。

追加例 5 矛盾

前提文

A land rover is being driven across a river.

仮説文

A sedan is stuck in the middle of a river.

正解は矛盾

車種と動作が、同じ場面の説明として両立しない。

追加例 6 中立

前提文

A blond-haired doctor and her African american assistant looking threw new medical manuals.

仮説文

A doctor is studying

正解は中立

正解は中立だが「勉強している」と読めるかは解釈に依存する。

追加例 7 矛盾

前提文

A young family enjoys feeling ocean waves lap at their feet.

仮説文

A family is out at a restaurant.

正解は矛盾

海の波を感じる場面と、レストランの場면을区別している。

追加例 8 含意

前提文

A couple walk hand in hand down a street.

仮説文

A couple is walking together.

正解は含意

手をつないで歩くことから、一緒に歩くことが導ける。

追加例 9 中立

前提文

Three firefighter come out of subway station.

仮説文

Three firefighters putting out a fire inside of a subway station.

正解は中立

消防士が駅を出ることから、消火しているとは限らない。

追加例 10 中立

前提文

This church choir sings to the masses as they sing joyous songs from the book at a church.

仮説文

The church has cracks in the ceiling.

正解は中立

合唱の説明からは、天井のひびの有無は分からない。

文長を集計した版・分割と文対数

- SNLI 1.0 の公式データに含まれる文対を集計した [1]。
 - 正解付き 569,033 組を使い、ラベルが「-」の組は除く。
- sentence1 と sentence2 を加工せずに数えた。
 - Python の len() で空白・句読点を含む Unicode 文字を数える。
 - str.split() で空白区切りの単語を数える。
- 前提は文対ごとに繰り返し数える。

[1] SNLI 1.0 公式データ . <https://nlp.stanford.edu/projects/snli/>

正解ラベル付きのクラス別件数

分割	含意	中立	矛盾	合計
学習用	183,416	182,764	183,187	549,367
開発用	3,329	3,235	3,278	9,842
テスト用	3,368	3,219	3,237	9,824

ラベルが付いた 569,033 組を集計。各クラスは約 3 分の 1 となる。

前提文・仮説文の単語数の記述統計

文	平均	標準偏差	最小	25%	中央	75%	95%	最大
前提	12.89	5.68	2	9	12	16	23	78
仮説	7.42	3.07	1	5	7	9	13	56

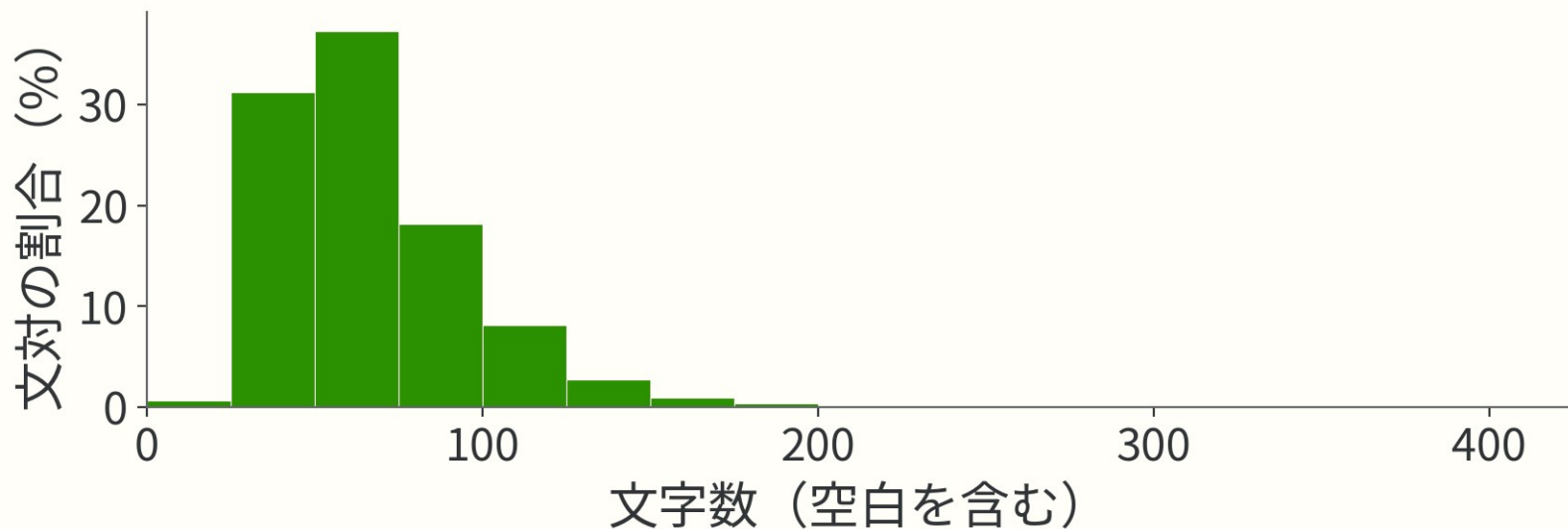
569,033 組を集計。標準偏差は母標準偏差（分母 N）。
前提を文対ごとに数える。

前提文・仮説文の文字数の記述統計

文	平均	標準偏差	最小	25%	中央	75%	95%	最大
前提	66.48	29.52	7	46	60	81	121	402
仮説	37.49	15.88	1	26	34	45	68	295

569,033 組を集計。標準偏差は母標準偏差（分母 N）。
前提を文対ごとに数える。

前提文の文字数の分布



平均 66.48 文字、中央値 60 文字、最大 402 文字。

[1] SNLI 1.0. <https://nlp.stanford.edu/projects/snli/>

正解付き 569,033 文対。同じ前提文が複数の文対に現れる場合も、それぞれ数えた。棒 1 本は 0 ~ 24 文字、25 ~ 49 文字…の区間を表す。

分割別の前提の単語数

分割	件数	平均	標準偏差	中央	95% 点
学習	549,367	12.85	5.66	12	23
開発	9,842	13.94	6.32	12	25
テスト	9,824	13.91	6.18	13	25

単語は Python の `str.split()` による空白区切り。

仮説文の文字数の分布



平均 37.49 文字、中央値 34 文字、最大 295 文字。

[1] SNLI 1.0. <https://nlp.stanford.edu/projects/snli/>

正解付き 569,033 文対の仮説文を集計した。棒 1 本は 0 ~ 24 文字、25 ~ 49 文字…の区間を表す。

分割別の仮説の単語数

分割	件数	平均	標準偏差	中央	95% 点
学習	549,367	7.42	3.07	7	13
開発	9,842	7.51	3.13	7	13
テスト	9,824	7.48	3.08	7	13

単語は Python の `str.split()` による空白区切り。

仮説の文字数が最大の実例

正解付き文対中で 295 文字。

The cat toys with the catastrophically injured yet alive mouse before putting it out of it's misery. Just a few more broken ribs and guts coming out the anus, and Kitty Kat will let you die, mouse. Then the cat will go leave a steaming pile of parasitic feces in the litterbox. Cats are cute!

前提は “A doberman bites a large blue ball.” 正解は矛盾。仮説には前提と無関係な描写が多い。

仮説が 56 語の実例の前提全文

仮説の単語数が最大の実例の前提文。

It is a photo of a young girl wearing pink pants and a jacket that's pink, red, blue and white the little girl is on one knee sitting on some big rocks smiling posing for a picture with one hand holding a skinny but long stick behind her there is what looks like a wooded area.

前提には“one hand”とある。正解ラベルは矛盾。

仮説が 56 語の実例の仮説全文

正解付き文対中で単語数が最大の仮説文。

It is a photo of a young girl wearing pink pants and a jacket that's pink, red, blue and white the little girl is on one knee sitting on some big rocks smiling posing for a picture with both hands holding a skinny but long stick behind her there is what looks like a wooded area.

仮説では“both hands”に変わっている。正解ラベルは矛盾。

単語の対応を用いる RE2 の評価結果

モデル	SNLI の正解率	報告条件
RE2	$88.9 \pm 0.1\%$	10 回の平均と標準偏差

ACL 2019 の手法。著者リポジトリ掲載の報告値。

公式 ZIP では文対 ID と投票を確認できる

```
import json
from urllib.request import urlretrieve
from zipfile import ZipFile

url = "https://nlp.stanford.edu/projects/snli/snli\_1.0.zip"
urlretrieve(url, "snli_1.0.zip")
with ZipFile("snli_1.0.zip") as z:
    path = "snli_1.0/snli_1.0_test.jsonl"
    row = json.loads(next(z.open(path)))
    print(row["pairID"], row["annotator_labels"])
```

公式の JSONL には sentence1 、 sentence2 、 gold_label などを含む。

作成時の指示と品質確認の補足

- 作業者は前提の説明に合う同じ写真を想定して仮説を書く。
 - 画像を見せず、3つの関係に対応する文を作る [1]。
- 規約を守らない作業者を外し、自動の大量投稿を拒否した。
 - 開発用とテスト用には複数の判定が付いた文対を使う。

[1] Bowman et al. (2015). EMNLP, pp. 632–642. A large annotated corpus for learning natural language inference. §2.1 <https://aclanthology.org/D15-1075/>

A large annotated corpus for learning natural language inference

Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). EMNLP, pp. 632–642.

SNLI の原著論文。文対の収集、3 種類の関係の注釈と検証、初期モデルの評価を報告する。

<https://aclanthology.org/D15-1075/>

A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference

Williams, A., Nangia, N., & Bowman, S. R. (2018). NAACL-HLT, pp. 1112–1122.

MultiNLI の原著論文。10 ジャンルの文対を収集し、分野内・分野間の汎化を評価する。

<https://aclanthology.org/N18-1101/>

Annotation Artifacts in Natural Language Inference Data

Gururangan, S., Swayamdipta, S., Levy, O., Schwartz, R., Bowman, S. R., & Smith, N. A. (2018). NAACL-HLT, pp. 107–112.

NLI の注釈に伴う偏りを分析した研究。仮説文だけでもラベルを予測できることを示す。

<https://aclanthology.org/N18-2017/>

Natural Language Inference

Sentence Transformers contributors (n.d.). Sentence Transformers documentation. 参照 2026-10-15.

Sentence Transformers の公式学習例。NLI 文対を使う文埋め込みの損失関数と実装を示す。

https://github.com/UKPLab/sentence-transformers/tree/master/examples/sentence_transformer/training/nli

Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

Reimers, N., & Gurevych, I. (2019). EMNLP-IJCNLP, pp. 3982–3992.

Sentence-BERT の原著論文。BERT から文埋め込みを生成し、文間の類似度を計算する。

<https://aclanthology.org/D19-1410/>

SimCSE: Simple Contrastive Learning of Sentence Embeddings

Gao, T., Yao, X., & Chen, D. (2021). EMNLP.

文埋め込みの対照学習手法を提案した論文。ドロップアウトと NLI 文対を正例の構成に用いる。

<https://aclanthology.org/2021.emnlp-main.552/>

Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference

McCoy, R. T., Pavlick, E., & Linzen, T. (2019). ACL, pp. 3428–3448.

HANS の提案論文。語の重なりや部分列などの統語的な近道に NLI モデルが依存するかを検査する。

<https://aclanthology.org/P19-1334/>

Adversarial NLI: A New Benchmark for Natural Language Understanding

Nie, Y., Williams, A., Dinan, E., Bansal, M., Weston, J., & Kiela, D. (2020). ACL.

ANLI の原著論文。人がモデルを誤らせる文対を作る反復的なデータ収集を提案する。

<https://aclanthology.org/2020.acl-main.441/>

e-SNLI: Natural Language Inference with Natural Language Explanations

Camburu, O.-M., Rocktäschel, T., Lukasiewicz, T., & Blunsom, P. (2018). Advances in Neural Information Processing Systems, 31, pp. 9539–9549.

SNLI に自然言語の説明を追加した論文。関係ラベルに加えて、その判断理由を収録する。

<https://github.com/OanaMariaCamburu/e-SNLI>