

——ローカル LLM による生成表現の改善——

指導教員：“Hodošček Bor 准教授”，“今尾 康裕教授”

1. Introduction	3. New section
2. 先行研究	4. Results

本研究は、調理に関する中国語動画を対象とし、字幕および音声からコーパスを構築し、その中で用いられる動詞の分析を通じて、ローカル LLM による生成表現の改善を目指すものである。調理動画は日常的かつ実用的な語彙を含んでおり、自然言語処理における意味役割ラベルや生成制御に活用できる可能性を持つ。

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = a + 2b + 3c \quad (2)$$

[illegible]

$$\begin{aligned} \sum_{k=0}^n k &= 1 + \dots + n \\ &= \frac{n(n+1)}{2} \end{aligned} \quad (4)$$

1

voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequaleam animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore.

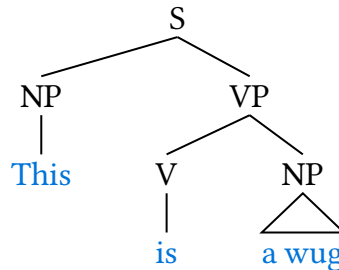


図 1: What is a wug?

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequaleam animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

4

日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。
日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。
日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。
日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。
日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。
日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。
日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。
日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。
日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。
日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。日本語で論文を書く。

我想用日语写一篇论文。我想用日语写一篇论文。道可道非常道名可名非常名无名天地之始有名万物之母故常无欲以观其妙常有欲以观其徼此两者同出而异名同谓之玄玄之又玄众妙之门天下皆知美之为美斯恶已皆知善之为善斯不善已故有无相生难易相成高下相倾音声相和前后相随不尚贤使民不争不贵难得之货使民不为盗不见可欲使民心不乱是以圣人之治虚其心实其腹弱其志强其骨常使民无知无欲使夫智者不敢为也为无为则无不治道冲而用之或不盈渊兮似万物之宗挫其锐解其纷和其光同其尘湛兮似或存

nǐ hao3, shì jiè ! Typst de zhōng wén pái bǎn fēi cháng kù !

2. 先行研究

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

(Benjamin, 2012), (Hodošček ほか, 2012), (DuBay, 2007), (Pitler & Nenkova, 2008), (Sato ほか, 2008), (近藤 ほか, 2008)

Date	°No	Description
24/01/03	813	Filtered participant pool
24/01/03	477	Transitioned to sec. regimen
24/01/11	051	Cycled treatment substrate

3. データ

節 2 に書いたように... Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari

voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.



3. 1. データの詳細

表 1: 表の説明

n°	CID Ligando	Nombre Ligando	Afinidad	RMSD (l.b.)	RMSD (u.b.)
1	234523	LoremIpsum	234	0	0
2	2345	LoremIpsum	2365	0	0
3	3453	LoremIpsum	45634	0	0
4	83452	LoremIpsum	2456	0	0
5	210	LoremIpsum	245	0	0
6	3417	LoremIpsum	45634	0	0
7	4345	LoremIpsum	3456	0	0
8	4334	LoremIpsum	3456	0	0

4. 方法

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius.

```
import spacy
```

```
nlp = spacy.load("ja_ginza_electra")

with open("corpus.txt") as f:
    text = f.read()

lemmas = [token.lemma_ for token in nlp.pipe(text)]

nlp("これは例文です。")
```

5. 結果

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequae doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguere possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torquem illum hosti detraxisse, ut aliquam ex eo est consecutus? – Laudem et caritatem, quae sunt vitae sine metu degendae praesidia firmissima. – Filium morte multavit. – Si sine causa, nollem me ab eo delectari, quod ista Platonis, Aristoteli, Theophrasti orationis ornamenta neglexerit. Nam illud quidem physici, credere aliquid esse minimum, quod profecto numquam putavisset, si a Polyaeno, familiari suo, geometrica discere maluisset quam illum etiam ipsum dedocere. Sol Democrito magnus videtur, quippe homini erudito in geometriaque perfecto, huic pedalis fortasse; tantum enim esse omnino in nostris poetis aut inertissimae segnitiae est aut fastidii delicatissimi. Mihi quidem videtur,

inermis ac nudus est. Tollit definitiones, nihil de dividendo ac partiendo docet, non quo ignorare vos arbitrer, sed ut.

6. 考察

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequale doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguere possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torquem illum hosti detraxisse, ut aliquam ex eo est consecutus? – Laudem et caritatem, quae sunt vitae sine metu degendae praesidia firmissima. – Filium morte multavit. – Si sine causa, nollem me ab eo delectari, quod ista Platonis, Aristoteli, Theophrasti orationis ornamenta neglexerit. Nam illud quidem physici, credere aliquid esse minimum, quod profecto numquam putavisset, si a Polyaeno, familiari suo, geometrica discere maluisset quam illum etiam ipsum.

7. 終わりに

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aequale doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguere possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et.

参考文献

- Benjamin, R. G. (2012 年). Reconstructing Readability: Recent Developments and Recommendations in the Analysis of Text Difficulty. *Educ Psychol Rev*, 24, 63–88. <https://doi.org/10.1007/s10648-011-9181-8>
- DuBay, W. H. (2007 年). *Smart Language: Readers, Readability, and the Grading of Text*. Impact Information.
- Hodošek, B., Abekawa, T., Murota, M., & Nishina, K. (2012 年). *Readability of example sentences in writing assistance tool Natsume*. 1–4.
- Pitler, E., & Nenkova, A. (2008 年). Revisiting Readability: A Unified Framework for Predicting Text Quality. *EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 186–195.
- Sato, S., Matsuyoshi, S., & Kondoh, Y. (2008 年). Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus. *LREC-08*, 654–660.
- 近藤陽介, 松吉俊, & 佐藤理史. (2008 年). 教科書コーパスを用いた日本語テキストの難易度推定. 言語処理学会第 14 回年次大会発表論文集, 1113–1116.