社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
IN2020-50 (2021-01)

発言者識別を用いた対話型キャラクタのセリフの違和感検出

森　康汰† 荒澤　孔明† 服部　峻††

†,††室蘭工業大学 ウェブ知能時空間研究室　〒 050–8585 北海道室蘭市水元町 27–1
E-mail: †{16024159,18096001}@mmm.muroran-it.ac.jp, ††hattori@csse.muroran-it.ac.jp

あらまし　近年，スマートフォンの普及に伴い若者を中心にソーシャルゲームが人気を集めている．そのソーシャルゲームの魅力として短いスパンでのイベントやストーリーのアップデートが一番の魅力であると考える．しかしながら，短いスパンで作成されたキャラクタのセリフにおいて，キャラクタに対し一貫性が無く違和感のある発言が作成されてしまうという問題がある．その問題を解決する方法として，キャラクタの個性の一部とも言える語尾・口調に着目した発言者識別を行う手法を提案し，語尾・口調に着目した違和感検出に対して正解率，特異度などの評価尺度を用いて評価実験を行う．また，発言者識別を用いたキャラクタのセリフの違和感検出を二次創作に使用した結果を定量的に評価し，語尾・口調における違和感検出のシステムの開発を目指す．

キーワード　自然言語処理，違和感検出，特徴抽出，機械学習

Jerkiness Detection in Spoken Lines of Interactive Characters Using Text-based Speaker Recognition

Kota MORI†, Komei ARASAWA†, and Shun HATTORI††

†,†† Web Intelligence Time-Space (WITS) Laboratory, Muroran Institute of Technology
27–1 Mizumoto-cho, Muroran, Hokkaido 050–8585, Japan
E-mail: †{16024159,18096001}@mmm.muroran-it.ac.jp, ††hattori@csse.muroran-it.ac.jp

Abstract　In recent years, social-network games have become more and more popular among young people due to the spread of smartphones. The most attractive features of social-network games are frequent events and story updates. However, there is a problem that the utterances of characters created in a short span of time are sometimes inconsistent and uncomfortable. To solve such a problem, this paper proposes a method for identification of speakers based on their utterances' ending and tone of voice, which are a part of characters' personality, and evaluates a system for jerkiness detection based on their utterances' ending and tone of voice using the method, with respect to several evaluation criteria such as accuracy rate, and specificity. In addition, this paper aims to develop a system for jerkiness detection based on utterances' ending and tone of voice, by quantitatively evaluating the results of applying jerkiness detection in characters' utterances using the method for identification of speakers based on their utterances' ending and tone of voice to characters' dialogues in derivative works.

Key words　Natural Language Processing, Jerkiness Detection, Feature Extraction, Machine Learning

1. まえがき

現在の日本において，スマートフォンの普及に伴い若者を中心にソーシャルゲームが人気を博している．ソーシャルゲームの魅力としては，空いた時間で気軽に遊べること，基本無料で遊べるということ，飽きさせないよう短いスパンでイベントやストーリーをアップデートすることが挙げられ，その中でも短いスパンでのイベントやストーリーのアップデートがソーシャルゲームにおける一番の魅力であると考える．

しかしながら，短いスパンで作成されたキャラクタのセリフには，作成時間の短さ故にしばしばキャラクタに対して違和感のある発言が作成されてしまうという問題がある．実際にあった例として，“バンドリ”というソーシャルゲームに“今井リサ”というキャラクタがいるが，このキャラクタはストーリー上で不自然な形で家族に弟がいるという設定が追加され，ユーザに混乱を招き炎上し結果としてゲーム運営会社が謝罪したという事例がある [1]．このような問題解決のため図 1 の様なシステムを構築することによって，キャラクタのセリフの違和感

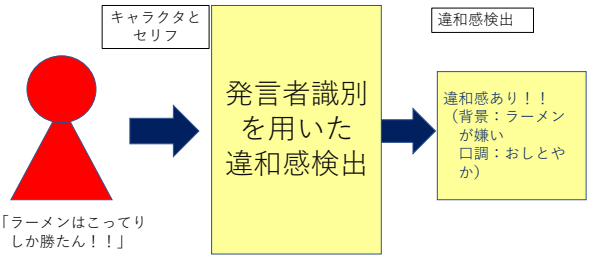

図 1　提案システムのイメージ

検出を機械が行うことになり，キャラクタのセリフ作成の時間削減や，セリフにおけるキャラクタの一貫性などの面で活躍が見込まれると考える．

本稿では，あるキャラクタのセリフ中の語尾・口調などに対してテキスト解析を行い，セリフ中に違和感が存在するかを検出する手法について提案する．

2. 違和感について

ストーリーライターを補助するような先行研究として“セリフに基づく個性を考慮した話者の分散表現に関する考察”[2] などがすでに存在するが，違和感を検出する研究については不十分である．違和感には様々な種類のものがあるが，本稿における違和感を以下のように定義する．

2.1 状況との不一致

キャラクタが発言する際の状況とセリフの内容に対する考慮しなくてはならない違和感について以下の通り定義する．

- 対話相手との関係性
- 発言者の感情
- 対話相手の感情
- 発言する場面の空気感

2.2 個性との不一致

セリフの内容とキャラクタの個性との間の考慮しなくてはならない違和感について以下の通り定義する．

- キャラクタの語尾・口調
- キャラクタの背景・過去

上記で挙げた違和感についてどれもキャラクタのセリフにおける一貫性の保持のために必要なファクタではあるが，本稿ではキャラクタの個性が顕著に表れ，定量的に評価を出しやすいため，上記中のキャラクタの語尾・口調について焦点を当てるものとする．

3. 提 案 手 法

2 種類の提案手法の概観を図 2 に示し，詳述していく．

3.1 提案手法 1

提案手法 1（図 2 の上部）として，キャラクタの語彙に着目した識別器を用いた違和感検出について提案する．その詳細は図 3 に示す．本稿では，python のオープンソース機械学習ライブラリの scikit-learn [3] 中の SVM を用いて違和感の有無について分類する．そのセリフの特徴ベクトルとして BoW (Bag of Words) を使用する．ただし，その特徴ベクトルには，形態素解析ソフト MeCab を使用し，それらの単語の基本形を用いる．

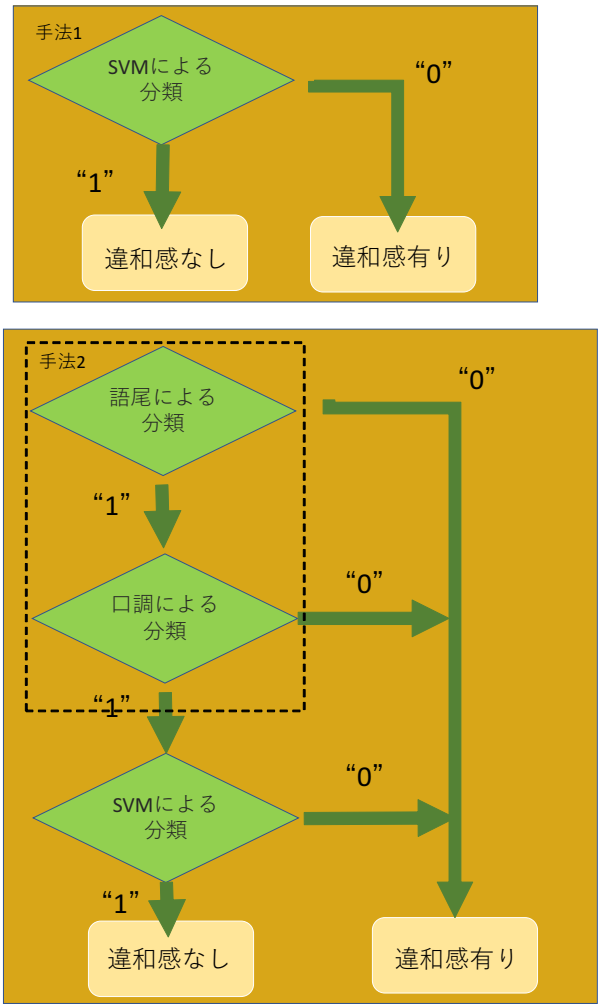

図 2　2 種類の提案手法の概観

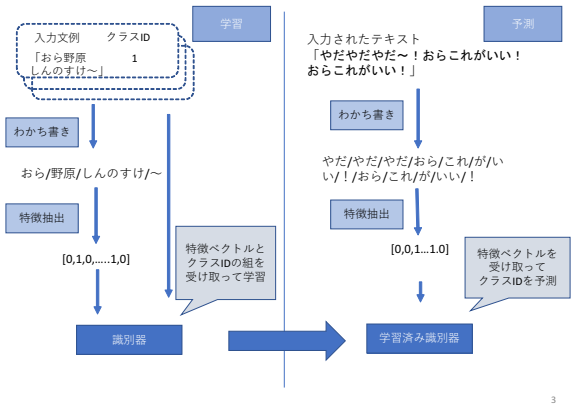

図 3　語彙に着目した発言者識別

3.2 提案手法 2

提案手法 2 はキャラクタの語尾に着目した識別器とキャラク

タの口調に着目した識別器を用いた違和感検出について提案する．図 2 に記述されているフローチャートに従い，入力されたテキストを対象に語尾についての識別を行い，“違和感なし”と識別されたものに，口調についての識別を行う．その過程で“違和感あり”と一度でも識別されたものは“違和感あり”と予測され，“違和感なし”と予測されたもののみに提案手法 1 を使用し分類させるという手法である．

3.2.1 語尾についての識別

キャラクタの語尾を識別するために，語尾リストを用いる．理想として，この語尾リストにはあるキャラクタとそのキャラクタがよく使用する語尾のペアが複数，キャラクタ分格納されている．本稿では簡易的に以下のような語尾リストを作成した．まず，様々なキャラクタがよく使用する語尾のみをウェブ上のニコニコ大百科 [4] からおよそ 200 件取得する．次に，本稿では特定のキャラクタ c に限定してセリフの違和感検出を行うことを予め定めておき，200 件のデータに対して，そのキャラクタ c の語尾であるか否かのラベルを著者らが付与した．

具体的な違和感検出の処理は図 4 の通りである．まず，あるセリフ s に対して形態素解析を行い，以下の条件を満たす形態素を語尾とする．

- その形態素がそのセリフ s の最後の形態素で，且つ助動詞であるとき
- その形態素が最後から K 番目の形態素で，且つ助動詞であり，$K+1$ 番目から最後までの形態素がすべて記号であるとき

次に，そのセリフ s の語尾がキャラクタ c の語尾リストの中に含まれていた場合，そのセリフ s がキャラクタ c の発言として違和感があると分類する．ただし，4 章の評価実験では語尾リストに含まれる 200 件は予め形態素解析のユーザ辞書として登録した．

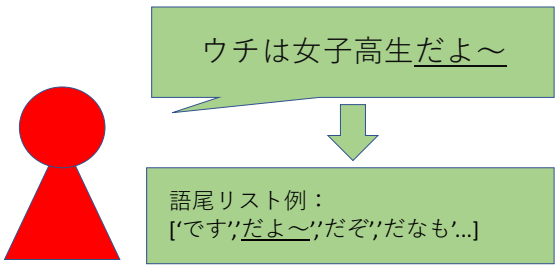

図 4　語尾に着目した違和感検出

3.2.2 口調についての識別

例えば，普段「お姉さん」とよく発言するキャラクタが，急に「お姉ちゃん」と発言したり，自分の事を「オラ」と発言するキャラクタが，急に「ぼく」と発言したりする時，我々はそれらに違和感を感じるであろう．このように，あるキャラクタの口調に違和感を覚えてしまう状況の 1 つとして，そのキャラクタが何度も使用する言葉があるにも関わらず，その言葉が使われていない場合が挙げられる．

そこで本稿では，あるキャラクタが発言したセリフ中で「置き換えた方がそのキャラクタの発言として自然である」という別の単語が認識された時，そのセリフはそのキャラクタの発言として違和感があると識別する手法を提案する．

具体的に，セリフ s がキャラクタ c の発言として違和感があるか否かを識別する処理について説明する．まず，そのセリフ s に出現する単語群を w_i^s $(i=1,2,...)$ とする．この時「キャラクタ c の発言として，単語 w_i^s を別の単語 w に置き換えるべき」であるという状況は，その置き換え先の単語 w が以下の条件を満たしている時であるという仮説を立てた．

- 単語 w はキャラクタ c が頻繁に使う単語である．
- 単語 w を単語 w_i^s と置き換えてもセリフの意味は変わらない．（単語 w と単語 w_i^s の意味的に類似する）

そこで本稿では，セリフ s の，キャラクタ c の発言としての違和感の有無を示す $\mathrm{hasJerk}(s,c)$ の定義式を以下の通り提案する．

$$\mathrm{hasJerk}(s,c)=\begin{cases}1 & (\ ^{\exists}i,\ ^{\exists}j,\ \mathrm{sim}(w_i^s,w_j^c)\geqq t_s \\ & \qquad \text{and } \mathrm{N}(w_j^c)\geqq t_n) \\ 0 & (\text{otherwise})\end{cases}$$

ただし，w_j^c $(j=1,2,...)$ は，キャラクタ c の過去のセリフに出現した単語群，$\mathrm{N}(w_j^c)$ $(j=1,2,...)$ はその出現回数を表している．そして，キャラクタ c の過去のセリフに頻出し，かつセリフ s の中のある単語 w_i^s の意味を変えないような単語 w_j^c が 1 つでも存在する時，単語 w_i^s は単語 w_j^c に置き換えた方が良く，セリフ s はキャラクタ c の発言として違和感があると認識することになる．

本稿では，4 章の評価実験の際に単語間の類似性について，WordNet [7] を使用する．また，類義語として抽出される単語には類似性があるものとする．使用例を図 5 に示す．

4. 評価実験

本章では，実際にそれぞれの提案手法が優れている度合いを評価するために，実際に違和感検出を行い結果を考察する．

4.1 データセット

本実験では，学習データとして NETFLIX [5] にて公開されているアニメーション作品の字幕データを使用し，特定のキャラクタか，それ以外かでラベル付けを行い学習モデルを作成する．表 1 に示す 2 つのアニメーション作品の字幕データを使用する．

表 1　評価実験用データセット

作品	発言者	ジャンル	真偽の割合
クレヨンしんちゃん	野原しんのすけ	日常	3:2
鬼滅の刃	竈門炭治郎	アクション	9:1

テストデータとして二次創作の小説まとめサイト [6] から学習データに使用した作品の小説を使用する．データ単位につい

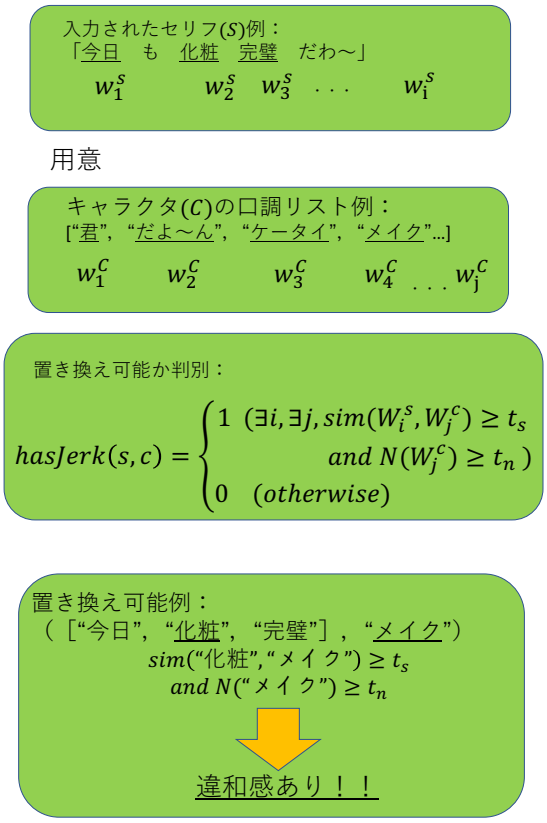

図 5　口調に着目した違和感検出

て，小説中で改行が発生するまでを 1 データとする．二次創作の小説にはアニメーションで発言されないセリフ（違和感があるセリフ）があるものとし，違和感の有無をラベル付けしたデータを実験に使用する．

4.2　評 価 尺 度

本実験では，表 2 に示す混同行列を算出した後，正解率 (Accuracy)，特異度 (Specificity) を用いて，予測性能を評価する．

表 2　混 同 行 列

	違和感なしデータ	違和感ありデータ
違和感なしと予測	TP	FP
違和感ありと予測	FN	TN

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN)$$

$$\text{Specificity} = TN/(TN + FP)$$

4.3　提案手法 1 による実験結果

提案手法 1 を用いて実験した結果は表 3 の通りである．

4.4　提案手法 1 に対する考察

本稿では 2 つの作品について実験を行った．表 3 から，“野原しんのすけ”の発言データに対して正解率・特異度のどちらの点でも高い値を出力している．これは実験対象の発言データ中にキャラクタの個性が強く出ているからではないかと考えられる．“竈門炭治郎”の発言データに対しては作品のジャンルがアクションということもあり，主人公は戦闘の描写が多く出てくる語彙に偏りがあるため識別されにくいのではないかと推測される．また，ストーリーが進む際に出てくる技の名前や人の名前が未知語となってしまったため識別できなかったというのも要因の 1 つであると考える．

表 3　提案手法 1 の違和感検出の性能

評価尺度	クレヨンしんちゃん	鬼滅の刃	平均値
正解率	0.496	0.220	0.358
特異度	0.611	1.000	0.806

4.5　提案手法 2 による実験結果

図 2 中の破線部に当たる語尾・口調による違和感識別の結果が表 4 の通りである．

表 4　図 2 の破線部の違和感検出の性能

作品	実際に違和感あり / 分類器が違和感あり
クレヨンしんちゃん	0.512 (20/39)
鬼滅の刃	0.076 (1/13)

次に図 2 中の下部に当たる語尾・口調と語彙について違和感検出の結果が表 5 の通りである．

表 5　提案手法 2 の違和感検出の性能

評価尺度	クレヨンしんちゃん	鬼滅の刃	平均値
正解率	0.517	0.220	0.369
特異度	0.888	1.000	0.944

4.6　提案手法 2 に対する考察

語尾の識別について，表 4 の 2 つの作品の実験結果の平均値が提案手法 1 の値よりも高いため，提案手法 2 の語尾・口調について違和感検出を行うことで精度が向上したと考えられる．“野原しんのすけ”の発言データに対してはキャラクタの語尾がはっきりしているため違和感が検出しやすく判別結果もかなり良いものとなった．“竈門炭治郎”の発言データに対しては語尾に個性が出てこないため有効ではなく，口調による判断しかできなかった．

口調の識別について，“野原しんのすけ”の発言データに対しては正解率を一定の数値に保ったまま，違和感がある文章に対して正しく予測することができたため，口調の識別として良い精度が出ていると考えられる．“竈門炭治郎”の発言データに対しては，SVM での結果と同じ結果となってしまった．これは表 3 からもわかる通り“竈門炭治郎”というキャラクタが語尾・口調に個性が強く出てこないことが原因ではないかと考え，提案手法 1 の改善案としては使用するデータの種類により精度が異なるのではないかと考える．

提案手法 2 で抽出できなかった例を表 6 に示す．提案手法 2 の口調についての t_s 値と t_n 値について，本稿では t_s については“WordNet に類義語が存在するかどうか”，t_n 値については t_n=1 にて実験を行い，その結果として表 6 が失敗データとして取得したものである，この結果から，t_s 値と t_n 値につい

表 6　類義語として抽出された例

作品	抽出された単語	類義語
クレヨンしんちゃん	暗い	重い
クレヨンしんちゃん	助け	世話
鬼滅の刃	仲間	相手

てのパラメータについて閾値を算出することが重要であると考える．

4.6.1　出現回数の諸検討

3.2.2 節で定義した口調リスト中の単語群，$\mathrm{N}(w_j^c)$ について数値の検討を行う．本実験中の学習データにおける置き換え候補になりうる単語の出現回数は表 7 の通りである．

表 7　出現回数の諸検討

出現回数	1	2	3	4 以上
単語数	16	3	2	5

単語群，$\mathrm{N}(w_j^c)$ に対してここでは t_n の値で検討する．$t_n \in \{1,2,3\}$ の時の評価尺度について表 8 に示す．表 8 から，本実験では正解率と t_n の値の変化にはあまり大きな相関が見られなかった．特異度に着目すると t_n=1 の時が最大となっているが，特異度の精度をより向上させるためには本実験では SVM の精度ではなく語尾・口調での違和感検出が求められることがわかった．

また，本稿の実験では置き換える単語を特定の品詞としているため，単語の出現回数と識別の精度の間に相関が見えないのではないかと考察することができる．

5.　まとめと今後の研究課題

5.1　ま　と　め

本稿では発言者識別を用いたキャラクタのセリフの違和感検出に関する研究を行ってきた．本研究は，先行研究の話者識別とは異なり，違和感検出を目的としているため，評価尺度の“正解率”，“特異度”の値の変化について着目した．表 3，表 5 から，提案手法 1，提案手法 2 共に特異度の値が高いことがわかり，また提案手法 1 と提案手法 2 を比較した時，正解率をほぼ一定にしたまま特異度が上がっていることがわかる．これは違和感検出の精度が向上したことを示していると考えられる．

本稿のアプローチとしては発言者識別と違和感検出に特化した識別器を用意し識別させるものであったが，別の観点から違和感を定義し学習させることによって，また違う結果が望めるのではないかと考えられる．

5.2　今後の課題

今後の課題として，提案手法 2 の改良案について考察する．

5.2.1　語尾について

語尾についての識別器の判断基準が“入力されたテキストに語尾が入っているか”と，“入力されたテキスト中の語尾が識別したキャラクタのものか”で判断していた．精度向上のためには“入力されたテキストに語尾が無かった場合，語尾を追加できるか”と考え，入力されたテキストに語尾になりえる単語を付属させ，違和感の識別を行うことができれば，より語尾の面で違和感を検出できるのではないかと考えた．

表 8　t_n による違和感検出性能の違い

評価尺度	t_n=1	t_n=2	t_n=3
正解率	0.517	0.517	0.524
特異度	0.888	0.814	0.814

5.2.2　口調について

口調についての識別器について，提案手法 1 の精度に多少なりとも依存しているため，口調についての識別器内の閾値を設定しないと提案手法 1 の結果から向上できなくなるのではないかと考える．

また，本稿では違和感について特にキャラクタの語尾・口調に着目したモデルについて提案したが，“野原しんのすけ”のような発言データ中の語尾・口調に個性が出てくるキャラクタに対しては良好的に違和感を検出できるが，“竈門炭治郎”のような個性が出てこないキャラクタに対してはあまり良い結果にならなかったため，他の違和感についての識別器の作成も検討の余地があると考える．本実験では口調の置き換え語に対して品詞すべてではなく，一部の品詞のみで行っていたが，今後は品詞に対する諸検討を行い，抽出される単語から個性や違和感が検出されるか研究を進めていこうと考えている．

本稿では単語間の類似性について WordNet から取得される類義語には類似性があるものとして使用していたが，今後としては単語間の分散表現を利用し，類似度に閾値を決め，単語間の類似度の閾値と違和感の検出結果の相関について研究していこうと考えている．

最後に，本稿では実験した語尾・口調などの識別器について，現段階では文章中に個性が顕著に表れているものに対しては良好な精度を出せているが，この研究が進み，個性が顕著に表れていない文章に対しても良好な精度が安定して出力することができれば，キャラクタのセリフについてだけでなく台本や，なりすまし，SNS のアカウント特定などの面で，研究の分野では著者推定や発話者予測などで活躍が期待されると考える．

文　　献

[1] AnimeAnime.jp, https://animeanime.jp/article/2019/10/23/49148.html (2020/12/18 参照).

[2] 佐藤 幸一, 福田 清人, 森 直樹, 松本 啓之亮, “台詞に基づく個性を考慮した話者の分散表現に関する考察,” 言語処理学会 第 24 回年次大会 発表論文集, pp.1280–1283 (2018). https://anlp.jp/proceedings/annual_meeting/2018/pdf_dir/C7-4.pdf (2020/12/18 参照).

[3] scikit-learn, https://scikit-learn.org/stable/ (2020/12/18 参照).

[4] ニコニコ大百科, https://dic.nicovideo.jp/a/%E8%AA%9E%E5%B0%BE%E3%81%AE%E4%B8%80%E8%A6%A7 (2020/12/18 参照).

[5] NETFLIX, https://www.netflix.com/ (2020/12/18 参照).

[6] 二次創作の小説まとめ, エレファント速報, http://elephant.2chblog.jp/ (2020/12/18 参照).

[7] 日本語 WordNet, http://compling.hss.ntu.edu.sg/wnja/ (2020/12/18 参照).