この言語リソースは、2012 年から 2019 年の間に Lang-8 ソーシャルネットワーキングサービスのユーザーによって書かれた英語のエッセイと添削文です。元のエッセイは他のユーザー（英語または日本語のネイティブスピーカーが多いですが必ずしもそうではありません）によってレビューおよび修正されました。Lang-8 ではユーザーが相互に文章を修正するため、全ての文／エラーがレビュー／修正されているわけではなく、いくつかのエラーが含まれている可能性があります。

ファイル名の形式： ファイル名は、3 つの情報でその内容を示します：

1. 年
2. 書かれた言語（1 は日本語、2 は英語）
3. 年の四半期（0-3）

    0. 1 月～3 月の終わり
    1. 4 月～6 月の終わり
    2. 7 月～9 月の終わり
    3. 10 月～12 月の終わり

例：2012.L1.Q0.dat は、2012 年 1 月～3 月の間の日本語の修正データを含んでいます。

データ形式： エッセイは改行で区切られた配列で表されます。各配列には以下の要素が含まれています：

1. ライターの母語を示す番号 - 下記の言語リストを参照してください。なお、L1 情報はユーザーによって報告されたものであり、場合によっては正確ではない可能性があります。
2. エッセイが書かれた言語（1 は日本語、2 は英語）
3. 元のエッセイを表す文字列
4. 元のエッセイを個々の文に分割した配列。つまり、この配列の i 番目の要素はエッセイの i 番目の文に対応します。
5. エッセイの各文に対するユーザーからの修正を表す配列の配列。つまり、i 番目の配列の j 番目の要素はエッセイの i 番目の文に対する j 番目の修正に対応します。修正にはマークアップが含まれる場合がありますが、これらは Lang-8 上のユーザー間コミュニケーションのために各投稿者が追加したものであり、そのまま提供されます。修正が行われていない行は空の配列となります。

言語: 1 日本語 2 英語 3 簡体中国語（中国） 4 韓国語 5 スペイン語 6 フランス語 7 イタリア語 8 ドイツ語 9 オランダ語 10 ロシア語 15 トルコ語 20 タイ語 21 ベトナム語 30 アラビア語 31 ヒンディー語 32 ベンガル語 33 ポルトガル語 34 フィンランド語 35 スウェーデン語 36 デンマーク語 37 ノルウェー語 38 エストニア語 39 ギリシャ語 40 ハンガリー語 41 チェコ語 42 ブルガリア語 43 リトアニア語 44 ルーマニア語 45 ウルドゥ語 46 ペルシャ語 47 フィリピノ語 48 パンジャブ語 49 ジャワ語 51 テルグ語 52 タミル語 54 アイスランド語 55 アゼリ語 56 ラトビア語 57 モンゴル語 58 インドネシア語 59 ポーランド語 60 チベット語 61 サンスクリット語 62 アイヌ語 63 ヘブライ語 64 ハワイ語 65 マレーシア語 66 アイルランド語 67 ウェールズ語 68 ジョージア語 69 ラオ語 70 ラテン語 71 ウクライナ語 72 ハイチ語 73 アルバニア語 74 アルメニア語 76 バスク語 77 ボスニア語 78 クロアチア語 79 トンガ語 80 セルビア語 82 スロバキア語 83 スワヒリ語 84 スラブ語 85 マケドニア語 86 マラーティー語 87 クメール語 88 ウズベク語 90 カタロニア語 91 広東語 92 イディッシュ語 93 スロベニア語 94 アフリカーンス語 95 ポルトガル語 96 エスペラント語 98 ミャンマー語 100 その他 101 フェロー語 102 ブルトン語 103 ウドムルト語 104 ロジバン語 105 パシュトー語 106 マルタ語 107 ウイグル語 108 ロマ語 109 ナバホ語 110 カンナダ語 111 オセット語 112 クルド語 113 トルクメン語 114 ズールー語 115 キルギス語 116 オランダ語（フラマン語） 300 繁体中国語

Short description:
This language resource is a collection of English essays written by users of Lang-8 social networking service between 2012 and 2019. The original essays were reviewed and corrected by other users (expectedly native speakers of English or Japanese, but not necessarily). Note that not all sentences/errors are reviewed/corrected and they might still contain some errors since users mutually revise their writings in Lang-8.

File naming scheme:
A filename indicates its contents with three pieces of information:
1. The year
2. The written language (1 for Japanese, 2 for English)
3. The quarter of the year (0-3).
   0. January ~ end of March
   1. April ~ end of June
   2. July ~ end of September
   3. October ~ end of December

Example: 2012.L1.Q0.dat contains Japanese correction data for January ~ end of March 2012.

Data format:
Essays are represented arrays separated by newlines. Each array contains the following elements:
1. A number indicating the writer's native language - please refer to the languages list below. Note the L1 information is reported by the user and it may be incorrect in some cases.
2. The language in which the essay is written, which is either 1 (Japanese) or 2 (English)
3. A string representing the original essay
4. An array representing the original essay segmented into individual sentences. In other words, the i-th element in this array corresponds to the i-th sentence in the essay.
5. An array of arrays representing any user-contributed corrections to each individual sentence of the essay. In other words, the j-th element of the i-th array corresponds to the j-th correction to the i-th sentence in the essay. Note that corrections may contain markup - these are added by each contributor for the sake of user-to-user communication on Lang-8 and are provided as-is. Note that lines that have no correction will have an empty array.

Languages:
1 Japanese

2 English
3 Simplified Chinese (China)
4 Korean
5 Spanish
6 French
7 Italian
8 German
9 Dutch
10 Russian
15 Turkish
20 Thai
21 Vietnamese
30 Arabic
31 Hindi
32 Bengali
33 Portuguese
34 Finnish
35 Swedish
36 Danish
37 Norwegian
38 Estonian
39 Greek
40 Hungarian
41 Czech
42 Bulgarian
43 Lithuanian
44 Romanian
45 Urdu
46 Persian
47 Filipino
48 Punjabi
49 Javanese
51 Telugu
52 Tamil
54 Icelandic
55 Azeri

56 Latvian
57 Mongolian
58 Indonesian
59 Polish
60 Tibet
61 Sanskrit
62 Ainu
63 Hebrew
64 Hawaiian
65 Malaysian
66 Irish
67 Welsh
68 Georgian
69 Laotian
70 Latin
71 Ukrainian
72 Haitian
73 Albanian
74 Armenian
76 Basque
77 Bosnian
78 Croatian
79 Tongan
80 Serbian
82 Slovak
83 Kiswahili
84 Slavic
85 Macedonian
86 Marathi
87 Khmer
88 Uzbek
90 Catalan
91 Cantonese
92 Yiddish
93 Slovenian
94 Afrikaans

95 Portuguese
96 Esperanto
98 Burmese
100 Other
101 Faroese
102 Breton
103 Udmurt
104 Lojban
105 Pashto
106 Maltese
107 Uyghur
108 Romani
109 Navajo
110 Kannada
111 Ossetic
112 Kurdish
113 Turkmen
114 Zulu
115 Kyrgyz
116 Flemish
300 Traditional Chinese