EwD Corpora Information

Summary

> Pre-Reading Note

Pre-Reading Note

Access corpora_info_in_ewd.md to read this article in EwD L0.

1. The Corpus Row

1.1. Row Format

> SE | L0 ⦙ L1 < L2 > L3 ¦ Full

1.2. Explaining the Columns: SE and the EwD Variants

SE stands for *Standard English*. This is our term for words or texts written in the regular English spelling, without diacritics or any other sort of pronunciation guidance. This article is written in SE.

L0, **L1**, **L2**, **L3** and **Full** are the five main *variants* of **English with Diacritics (EwD)**, the four "L"s also meaning the didactic levels. Below is the explanation for each of them:

- [TODO: complete this explanation] Level 0 (L0) — EwD's **absolute-beginner** level. {Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.} *When omitted on a corpus row, the L0 word is equal to the L1 word.*

- Level 1 (L1) — EwD's **beginner** level. It includes only the indications that are essential for newcomers. It avoids the specifics of unstressed syllables and some nuances of stressed syllables. *When omitted on a corpus row, the L1 word is equal to the SE word.*

- Level 2 (L2) — EwD's **intermediate** level. It provides a more detailed guidance, utilizing a decent amount of applicable indications to guide learners towards a holistic understanding of English pronunciation. In this level, learners will acknowledge the patterns of unstressed syllables and explore nuances of stressed syllables not covered in the previous levels. *When omitted on a corpus row, the L2 word is equal to the L1 word.*

- Level 3 (L3) — EwD's **advanced** level. It is also referred to as the *Minimal* variant. After exposure to extensive indications of the previous level, this level retains only the most essential ones. Thus, learners will understand which of EwD's indications are actually unnecessary (or not very necessary) as they gradually grasp the patterns of spoken English. *When omitted on a corpus row, the L3 word is equal to the L1 word.*

- Full variant — The **most phonetic** presentation of EwD. It is not a level, but a variant meant to be used on *dictionary entries* or the like. As such, it goes very specific and ships all applicable marks, with the exception of a few beginner level (L0 or L1) ones, as well as some marks deemed very unnecessary in any EwD variant. *When omitted on a corpus row, the Full variant word is equal to the **L2** word.*

It's important to mention that learners are *not* meant to move from variant L3 to Full. On the contrary, the next step after L3 is the transition to reading SE texts without phonetic assistance, that is, gradually exiting EwD.

2. The Accept-or-Except Strategy (AES)

The Accept-or-Except Strategy (AES) in EwD is a key mechanism that balances the integrity of SE spelling with the need for phonetic clarity. This strategy primarily focuses on *retaining standard spellings*, and on applying *diacritical marks* only when necessary to assist in pronunciation — this is the **Accept side**[1].

[TODO: increment this paragraph, briefly explain the word distinction method used] Conversely, when standard spellings obscure the correct pronunciation, and diacritics would complicate them more than clarify, EwD shifts to the **Except side**, adopting so-called *fallback graphemes* that more accurately reflect spoken English.

[TODO: improve or change this paragraph] This approach eases the learning process and facilitates the transition from reading with diacritical assistance to reading independently. By preserving standard spellings as much as possible, EwD helps learners recognize and understand the spelling-to-speech patterns of English, thereby gradually building their confidence and proficiency in pronouncing English words without phonetic assistance.

2.1. AES Sides

Below is a brief explanation for the two sides of the Accept-or-Except Strategy:

- **Accept side** — The collective of words left intact (without diacritic) and words that are treated only with diacritic(s). The words from this side are referred to as *accepted words*. About **97%** of the 1000 most frequent English words are on this side. We may also call this the "A" side.

- **Except side** — The collective of words for which diacritics are deemed not enough or not suitable to indicate pronunciation. The words from this side are referred to as *exceptions*. About **3%** of the 1000 most frequent English words are on this side. We may also call this the "E" side.

2.2. The "Not Disclosed / Disclosed" Approach

[TODO: continue this explanation] For exceptions, the EwD material employs a "not disclosed / disclosed" approach in order to hide or clarify their pronunciation according to the necessity. The "not disclosed" form is a given exception with all letters preserved, while the "disclosed" form is the same word with the difficult grapheme replaced by its corresponding *fallback grapheme*. {Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.}

3. Acronyms & Abbreviations

[TODO: add explanation]

3.1. Stress and Accent Variations

[TODO: add explanation]

- str: stressed
- un: unstressed
- GA: General American accent

- GB: General British accent (*aka* Received Pronunciation – RP)

- CP: Common Pronunciation

- LCP: Less Common Pronunciation

- SP: Spelling Pronunciation

- AP: Academic Pronunciation

3.2. Row Tags

[TODO: add explanation]

3.2.1. AES Groups

[TODO: add explanation and examples]

- [A]: Accepted — The group of words on the **Accept side**[1]. It is made up of subgroups [Au] and [Awd]. The words on this group are referred to as *accepted words*.

    - [Ain]: Accepted, intact — [A] words that *don't* receive diacritics in any level.

    - [Awd]: Accepted, with diacritics — [A] words that receive *diacritic(s) in at least one level.*

        * [Awd^n]: Accepted, with diacritics, n — [Awd] words that receive *diacritic(s) only at given level(s).* The "n" corresponds to the level number(s).

- [E]: Exception — The group of words on the **Except side**. It is made up of subgroups [Eap] and [Esp]. The words on this group are referred to as *exceptions.*

    - [Eap]: Exception, all pronunciations — [E] words whose *all* used pronunciations require fallback graphemes to be clarified. In the corpora these words are always tagged accordingly.

        * [Eap^n]: Exception, all pronunciations, n — [Eap] that are treated with fallback graphemes only at given levels. The "n" corresponds to the level number(s). In the corpora these words are always tagged accordingly.

    - [Esp]: Exception, some pronunciations — [E] words whose *some* of its used pronunciations (but now all of them) require fallback graphemes to be clarified. In the corpora these words are always tagged accordingly.

3.2.2. **Didactics Tags**

[TODO: add explanation and, maybe, examples]

-
-

3.2.3. **SE Spelling Tags**

[TODO: add explanation and examples]

- AS: American spelling

- BS: British spelling

- FE: Word with *false etymology*

Appendix

Notes:

1. EwD strives to keep as many words as possible on the Accept side, with about 97% of the 1000 (one thousand) most frequent English words falling under this side (up to EwD v2.7).

Abbreviations:

- aka: also known as

Glossary: (TODO: add definitions)

- corpus:

- corpora: