

What is a word?

Basic definition

A meaningful unit that can stand on its own, e.g. *boy, a, for, nice, work, walk, works, walked, I'll, John's*.

-s, -ed, -'ll and -'s are not meaningful units that can stand on their own, thus they are not considered words.

English

Words that use **hyphens** e.g. *I love chocolate ice cream* vs *I love ice-cream trucks*. (The first 'ice cream' is a noun, and the second is an adjective.)

Ice cream – 2 words

Ice-cream – 1 word

Co-owner (noun) – 1 word

Only use hyphens when they occur in the written convention of the word.

Affixes - We don't use hyphens unless they occur in the convention e.g. *composition* vs *non-alcoholic*

Contractions will not be **split up*** and are considered single words, e.g. *he's, she'd, won't, can't, couldn't, didn't* etc. (we will not split them up nor change the way they were uttered e.g. we hear *you can't*, we SHOULD NOT transcribe it as *you cannot*).

NOTE: we will use Standard British English for our spelling conventions.

If you are not sure, check <https://dictionary.cambridge.org/dictionary/english>

If in doubt, bring it up to Esh and Sera (who will check with Sham if need to)

***not the same as split-up**

Mandarin Chinese

Spacing at the end of each character to signify a word (in the chunk tier). E.g.

Utterance: “我喜欢吃汉堡包” (I like to eat hamburgers)

Chunk: 我 喜欢 吃 汉堡包

< https://www.seab.gov.sg/docs/default-source/documents/list_of_dictionaries_for_examination.pdf?sfvrsn=cc8b47b6_4 >

If in doubt, bring it up to Fei Ting & Sera.

Malay

Use **hyphens** for words that in formal writing have hyphens e.g. *jalan-jalan*, *anak-anak*, *hati-hati* etc. Words such as *matahari*, *apabila* etc. where we don't use hyphens will be written as one word. Similarly, words such as 'rumah merah' is made up of two words.

Dealing with **affixes** (imbuhan) such as *pe-*, *me-*, *ke-*, *se-*, *ber-*, *ter-*, *-an*, *-kan*, *-i*, *-wi* etc. We don't use hyphens for affixes (think 'writing the *karangan* for *cikgu*'), e.g. *menyapu*, *kebersihan*

Clitics don't stand on their own e.g. *anak**nya** cantik* (2 words), *otak**mu*** (1 word)

Bahasa Baku

< https://www.seab.gov.sg/docs/default-source/documents/list_of_dictionaries_for_examination.pdf?sfvrsn=cc8b47b6_4 >

<https://prpm.dbp.gov.my/>

If in doubt, bring it up to Sakinah (who will check with Shaza and Sham if need to).

Tamil

Gender, tense, case markers (e.g.him/her, me etc) and negation particles etc.

that are attached to root words are not considered a word, e.g. அவன்

ஒடினான் avan ōṭinān (2 words), அவள் ஒடினாள் aval ōṭināl (2 words)

Transcribe as Spoken Tamil.

< https://www.seab.gov.sg/docs/default-source/documents/list_of_dictionaries_for_examination.pdf?sfvrsn=cc8b47b6_4 >

Instances where it seems like more than 2 words are masquerading as one word or other doubts, bring them up to Esh and Sham.