

# Thematic Categorisation of Highly Frequent Vocabulary

## What is this resource and what is it for?

This resource contains highly frequent (<2001) content words in French, German, and Spanish sorted into topics and themes that are likely to be relevant to school students. ([See here for a rationale for the choice of frequency list](#)). The resource lists words under every topic headword with which they are considered to have a strong relationship.

## Why is this resource important?

Many of the 2,000 most frequent words in a language are highly frequent because they are culturally, socially, linguistically, and personally useful words. For this reason, most of them fit into a range of topics and themes. It is important for students to connect words with **more than one topic** so that they can recognise and produce them in a **range of contexts**.

## How was this resource made?

First, NCELP made a long list of topics. We used Kay et al. (2021) as a source of topic headwords, because, to our knowledge, this is the only existing resource which arranges every recorded word in a language into detailed hierarchies of meaning. We identified 52 relevant topic headwords in total. This same long list was used for all three languages.

Next, groups of at least three people ('raters') were asked to evaluate the strength of the relationship between each of the 2,000 most frequent words and each topic headword, separately for each language. We used an adapted version of [Chung and Nation's \(2004\)](#) 4-point rating scale. (Note, Chung & Nation used this method to identify technical vocabulary for advanced learners). The raters were asked to decide whether each word: i) has no particular relationship with the topic (e.g., TRAVEL: BLACKBOARD); ii) is minimally related to the topic and could be used to talk about it in a general sense (e.g., TRAVEL: DURING); iii) is closely related to the topic, but also used in general language (e.g., TRAVEL: MOUNTAIN); and iv) has a meaning specific to only that topic (e.g., TRAVEL: SIGHTSEEING). Words deemed by the raters to fall into the latter two categories (iii and iv) were considered to have strong relationships with the topic headwords. To help reduce the impact of different personal experiences on associations with words, we required at least two raters to agree that the connection between a word and topic headword was strong (in category iii or iv) for that word to be counted as a topic word.

Finally, we organised the topic lists into five broad themes – Identity Forming; Everyday Living; Communication; People and Places; Global and Social Issues – to provide an *illustration* of how the 52 topics might be organised.

**Notes.** The resources are different in each language, as highly frequent words reflect the cultures in which the language is spoken, as represented in the corpora from which the frequency lists were generated. One consequence of this is that some of the 52 topics can

appear under different broad themes across languages (or, if very few examples of topic words were identified, those topics may not appear at all).

To obtain more reliable thematic categorisations, more raters could be used, with a representative balance of, for example, different experiences and interests.

### References

[Chung, T. M., & Nation, J. S. P. \(2004\)](#). Identifying technical vocabulary. *System*, 32(2), 251–263.

Kay, C., Alexander, M., Dallachy, F., Roberts, J., Samuels, M., & Wotherspoon, E. (Eds.) (2021). *The Historical Thesaurus of English* (2nd ed., version 5.0). University of Glasgow. <https://ht.ac.uk/>.