

Research evidence in relation to a frequency-based approach to vocabulary teaching and learning

AQA Research

21 April 2021

Executive Summary

Current research into frequency-based approaches to vocabulary acquisition outline the benefits that such an approach can have for teachers and learners of additional languages. The present document upholds the strengths of a frequency-based approach to vocabulary while presenting evidence from academic literature concerning the ways in which a frequency-based approach to vocabulary could be modified to better fit the context of GCSE MFL teaching and learning.

The five main points covered are:

1. A frequency-based approach to vocabulary could be highly useful for teachers and learners of GCSE MFL with appropriate modifications to best suit the pedagogical context. When using a frequency-based approach to create a vocabulary list, a distinction should be made between words (individual lexical items), lemmas (a group of closely related lexical items that share meaning) and word families (a group of closely related lexical items that share meaning and common derivative formations) in order to create a list which would be most useful for teachers and learners (Nation, 2017). The recommended lexical coverage for GCSE level learners should include at least 3,000 lexemes or word families (Nation, 2017; Schmitt, 2014) and, therefore, a selection of fewer than 2,000 individually specified words (as proposed in the subject consultation documents) would be a very low amount of lexical items for use with learners who are working towards acquiring communicative abilities.
2. The potential benefits of a frequency-based approach to vocabulary for learning and for assessment would be significantly diminished due to the proposed 90% to 10% ratio of high-frequency to low-frequency lexical items as outlined in the consultation document. Findings from prior research into the ratios of high-frequency to low-frequency items suggests that the ratio can vary depending on the genre of spoken or written material, with some research suggesting that learner texts could feature a ratio of 80% to 20% high- to low-frequency vocabulary items at the least (Gardner, 2013; Milton, 2009; Nation, 2001; Nation, 2008; NCELP, 2021¹). Using a 90:10 ratio of high- to low-frequency lexical items in assessment design could harm the validity of examinations by introducing high levels of problematic predictability in written assessments, by diminishing the semantic and syntactic appropriateness of spoken and written resources used for assessment, and by limiting item writers' abilities to discriminate between students' differing ability levels.
3. Further clarification would be useful to guide the creation of proposed vocabulary lists to function alongside outlined lists of syntactic constructions (grammar lists), the number and content of which should be outlined and enumerated separately from the vocabulary list of lexical items. These two lists should be operationalised in order to allow for students to learn the appropriate number of lexical items specified while also serving to uphold the

¹ The document created by NCELP (2021) that outlines the rationales behind a frequency-based approach to vocabulary that has been proposed in the consultation documents itself acknowledges the 80% high-frequency to 20% lower frequency ratio of vocabulary items that is recommended across academic literature.

number of syntactic constructions students need to learn to gain communicative ability (Gardner, 2013).

4. A frequency-based approach can be combined with a thematic approach in order to make best use of extant resources that draw upon authentic materials to appeal to students' interests in alignment with pedagogical practice (Nation, 2017). Across academic literature, research recommends that frequency-based vocabulary lists be created in consultation with the relevant teachers and learners within the pedagogical context. (Dang, Webb and Coxhead, 2020).
5. Overall, a frequency-based approach could be signalled during KS2 and KS3 in advance of GCSE and from GCSE onward to A Level. Introducing a frequency-based approach across stages of learning would assist with sequencing, and would make best use not only of students' knowledge of high-frequency vocabulary items but also of students' metalinguistic awareness and morphosyntactic skills. Throughout their language learning, such skills are needed for them to apply their knowledge of high-frequency lexical items strategically towards making meaning of unknown lower frequency lexical items (Gardner, 2013).

Introduction

Research into vocabulary acquisition has highlighted that a frequency-based approach can be useful for language learners, provided that frequency-based vocabulary lists are generated from corpora that are relevant to the pedagogical context and that such lists are then edited and refined to be a best fit for the relevant L2 teachers and learners with input from teachers in particular (Dang, Webb and Coxhead, 2020; Nation, 2017).

1. Distinction between words and word families

For the creation of a vocabulary list, the distinction must be made as to whether the vocabulary list will be comprised of headwords, lemmas or word families. For the purposes of clarifying relevant terminology, the following are commonly agreed upon definitions of headwords, lemmas and word families:

Headword: A headword is the form of the word that appears in a dictionary entry. For the purposes of vocabulary list generation, the headword should also be the most frequently occurring form of a lemma (Dang, Webb and Coxhead, 2020).

Examples: “book”
 “walk”

Lemma: “The smallest unit in the meaning system of a language that can be distinguished from other similar units ... It can occur in many different forms in actual spoken or written sentences, and is regarded as the same lexeme even when inflected” (Richards and Schmidt, 2010: 210).

Examples: “book, books” (single lexeme with headword *book*)
 “walks, walking, walked” (single lexeme with headword *walk*)

Word family: “A word family consists of a headword, its inflected forms, and its closely related derived forms” (Gardner and Davies, 2014: 307). A word family consists of the headword of the lexical form, the lexeme of the lexical forms, and derivations of the lexical form that can be different parts of speech.

Examples: “proceed, proceeds, proceeding, proceeded, procedure, procedures,
 procedural, proceedings” (single word family with headword *proceed*)

As per the subject content consultation document, the vocabulary list is to be created as set out in Paragraph 20:

“20. Words will be listed in the basic form commonly found in dictionaries. Where different parts of speech are derived from a single root, each part of speech will be listed separately. Further derived and inflected forms of each part of speech should not be separately listed, provided that the defined grammar covers their formation. Inflected and derived forms which follow no regular pattern should be listed as separate items under a section called ‘Highly irregular inflected forms’.”

Given this approach, the vocabulary list as proposed would appear to be comprised of lemmas. However, the inclusion of irregular inflected forms as distinct lexical items within the recommended vocabulary list would consume a high percentage of the vocabulary list by presenting multiple lexical forms of the same lemma, with the result that students would be learning a small number of words, lemmas and word families overall.

2. Ratio of high-frequency to low-frequency items

Nation (2008, 2017) has conducted extensive research into the vocabulary frequency of resource materials for language learning, with the estimated ratio of high- to low- frequency items appropriate for learner texts cited as approximately 80% to 20%, while also acknowledging variation and flux in this ratio depending on texts' topics and contexts. Gardner (2013) establishes that the genre and context of resource materials has a significant impact on the ratio of high- to low-frequency vocabulary items present in a spoken or written text, and that language learners would be accustomed to texts with varying ratios of high- to low-frequency vocabulary items depending on the topic at hand.

In order to design resources for teaching and for assessment, materials would have to be constructed using profilers and concordances to align percentages of high- to low-frequency vocabulary items with the proposed ratios. While this reconfiguration of resources could be plausible for assessment design, such reconstruction of resources would be less immediately feasible for centres and teachers with access to a limited amount of predetermined resource materials.

Qualitative evidence gathered from a two hour feedback session with Chief Examiners, Chairs and experienced item writers from French, Spanish and German at AQA highlighted potential impacts that a restricted ratio of high- to low-frequency vocabulary could have upon the design of items that would draw upon students' skills of inferencing, i.e., students' application of metalinguistic awareness and strategies to use their knowledge of known high-frequency vocabulary items to gather the meaning of unknown low-frequency vocabulary items.

The participants from AQA detailed that low-frequency items were used by design in tasks to assess students inferencing strategies and metalinguistic awareness, skills that were flagged by the participants as being crucial to the ability to be able to discriminate between differing students' abilities. An additional significant concern raised by AQA participants was that of problematic predictability, in that students would potentially be able to predict the vocabulary items that would appear in written examinations given the limited number of proposed vocabulary items. Tied to this issue, participants also stressed that the proposed ratio could harm the authenticity of resource materials used for assessment to a more significant extent than at present. They expressed strong concerns that the use of materials that have been edited to feature unnatural phrasing or syntax for assessment purposes could resultantly damage overall validity by hindering students' comprehension of texts.

3. Combined frequency-based and thematic approach

A frequency-based approach can generate a vocabulary list that omits certain vocabulary items which are needed for L2 learning (Nation, 2016; Dang, Webb and Coxhead, 2020). For the development of a frequency-driven vocabulary list (BNC/COCA2000) derived from a combination of the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), Nation (2016) included low-frequency items which fell into thematic clusters of, for example, days of the week, months, numbers and countries in order to improve the usage of frequency-based vocabulary lists for the purposes of L2 teaching and learning.

Dang, Webb and Coxhead (2020) further recommend that implementation of a frequency-based approach to vocabulary acquisition for L2 teaching and learning requires input from L2 teachers and learners in order to refine and to operationalise frequency-based vocabulary lists for use in the classroom.

4. Progression to and from GCSE

Prior to GCSE

As mentioned in the consultation document, assumptions of prior vocabulary learning at KS2 and KS3 should be viewed with caution. The lack of consistent and cohesive transitions between primary and secondary MFL provision (Collen, 2020; Chambers, 2014; Chambers, 2019) may result in a change in the taught language for many pupils. Consequently, many schemes of work start all pupils with the same preliminary content in Year 7.

At both KS2 and KS3, the amount of time allocated to MFL may also be limited (Collen, 2020). This could be a barrier to set vocabulary being covered during these phases. Furthermore, it is not uncommon for younger pupils to be removed from MFL to focus on English literacy and other core skills (Collen, 2020), which reduces instruction time for them even further. This may be especially true for learners of English as an additional language (EAL) (Collen, 2020), many of whom contribute to the candidature taking minority MFLs at GCSE. Therefore, while the status and timetabled provision of MFL at KS2 and KS3 remains limited, prescribing a list that is assumed learnt at the start of KS4 may be problematic.

Although a reliance on guaranteed vocabulary learnt at KS2 and KS3 may not be feasible for GCSE studies, the language learning skills from this period can be drawn upon. Exposure to foreign language learning can aid the development of metalinguistic skills in young learners across languages (Reder, Marec-Breton, Gombert, & Demont, 2013) and these can be utilised in later study.

By including word families in GCSE MFL vocabulary lists and encouraging learners to implement morphological manipulation to create inflection or derivational forms from the root, pupils can draw on these resources. The recognition of cognates is a further skill that may be introduced to younger learners (Jones & Coffey, 2009). Again, this skill can be used in GCSE studies without relying on taught vocabulary at the lower Key Stages.

In moving from GCSE to A Level, A Level specifications would also have to be modified to reflect the nature of any modified frequency-based vocabulary lists.

Conclusion

In summary, research evidence confirms the utility of a frequency-based approach to vocabulary teaching and learning, but with acknowledgement that learners require exposure to a flexible number of high- and low-frequency items in resource materials, both spoken and written, in order to improve upon both their vocabulary knowledge and their metalinguistic, morphosyntactic and inference generation skills. A frequency-based approach could be most beneficial if vocabulary and grammar lists are made in relation to the specific pedagogical contexts of teaching and learning in GCSE MFLs, and ideally with input from teachers and learners.

References

Chambers, G. (2014). Transition in modern languages from primary to secondary school: the challenge of change. *The Language Learning Journal*, 42, 242-260.

Chambers, G. (2019). Pupils' perceptions of Key Stage 2 to Key Stage 3 transition in modern foreign languages. *The Language Learning Journal*, 47, 19-33.

Collen, I. (2020). *Language Trends 2020: Language teaching in primary and secondary schools in England*. British Council.

Dang, T.N.Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (ed.), *The Routledge handbook of vocabulary studies*. New York: Routledge.

Dang, Webb and Coxhead (2020) Evaluating lists of high-frequency words: teachers' and learners' perspectives. *Language Teaching Research*, 24, 1-25.

Dang, T.N.Y., & Webb, S. (2016). Evaluating lists of high-frequency words. *ITL – International Journal of Applied Linguistics*, 167, 132–158.

Dang, T.N.Y., and Webb, S. (2014) The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66–76.

Driscoll, P., Macaro, E. and Swarbrick, A. (2013) *Debates in modern languages education*. London: Taylor and Francis Group.

Gardner, D., and Davies, M. (2014) A new academic vocabulary list. *Applied Linguistics*, 35, 305–327.

Gardner, D. (2013) *Exploring vocabulary: language in action*. London: Taylor and Francis Group.

-
- Jones, J., & Coffey, S. (2009). *Modern Foreign Languages 5-11: A Guide for Teachers (Primary 5-11)*. London: David Fulton.
- Milton, J (2009) *Measuring Second Language Vocabulary Acquisition*. Bristol, Multilingual Matters.
- Milton, J. (2011) The role of classroom and informal vocabulary input in growing a foreign language lexicon. *Journal of Applied Linguistics*, 26, 59-80.
- Nation, I.S.P. (2017) *Learning vocabulary in another language*. 2nd edition. Cambridge: Cambridge University Press.
- Nation, I.S.P. (2016). *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins.
- Nation, I.S.P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.
- NCELP (2021) *Vocabulary lists: rationales and uses*. Available online via: <https://resources.ncelp.org/concern/resources/t722h880z?locale=en>.
- Reder, F., Marec-Breton, N., Gombert, J., & Demont, E. (2013). Second-language learners' advantage in metalinguistic awareness: A question of languages' characteristics. *British Journal of Educational Psychology*, 83, 686-702.
- Richards, J. C. and Schmidt, R. (2010) *Longman dictionary of language teaching and applied linguistics*, 4th ed. London: Longman.
- Schmitt, N. (2014) Size and depth of vocabulary knowledge: what the research shows. *Language Learning*, 64 (4), 913-951.