

This is a repository copy of *Creating and evaluating corpus-informed word lists for adolescent, beginner-to-low-intermediate learners of French, German, and Spanish*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/217354/>

Version: Published Version

Article:

Finlayson, Natalie Eloise, Marsden, Emma orcid.org/0000-0003-4086-5765 and Hawkes, Rachel (2024) *Creating and evaluating corpus-informed word lists for adolescent, beginner-to-low-intermediate learners of French, German, and Spanish*. *Language Teaching Research*. ISSN 1362-1688

<https://doi.org/10.1177/13621688241288877>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Creating and evaluating corpus-informed word lists for adolescent, beginner-to-low-intermediate learners of French, German, and Spanish

Language Teaching Research

1–57

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/13621688241288877

journals.sagepub.com/home/ltr

Natalie Finlayson 
University of York, UK

Emma Marsden 
University of York, UK

Rachel Hawkes
The Cam Academy Trust, UK

Abstract

This study explores some of the implications of policy changes relating to the composition and use of word lists for French, German, and Spanish as foreign languages in secondary schools in England. Against the backdrop of literature relating to word list creation and use, we review requirements for the vocabulary content of high-stakes examinations in these languages under current and new policy, and describe the methodological steps we took in collaboration with teachers and members of an awarding organization to create corpus-informed lists for adolescent, beginner-to-low-intermediate learners based on frequency, word-topic relatedness, and teacher judgments of usefulness, relevance and difficulty. Under current policy, awarding organizations provide educators with non-mandatory, topic-driven word lists structured around pre-determined themes. We analyse the content of lists compiled using each approach (corpus-informed or topic-driven) and examine their lexical coverage of four corpora designed to represent potential learning goals of adolescents: passing exams, further study, reading and discussing young adult literature, and engaging with web content. Despite being 36%–44% (foundation tier) and 11%–21% (higher tier) shorter, the new lists cover an average 11% (foundation tier) and 18% (higher tier) more of every corpus. Our further analyses suggest that these stark results can be attributed to (1) the nature of the content (rather than function)

Corresponding author:

Emma Marsden, D/M/107, Department of Education, University of York, York, YO10 5DD, UK

Email: emma.marsden@york.ac.uk

words, (2) negligible coverage benefits from multiword phrases on the current lists, and (3) a more balanced part-of-speech distribution in the new lists. Some of our methods were used by awarding organizations in England to develop accredited lists for the new examinations. Those lists share large numbers of lexical items with the lists reported here, suggesting that our findings have immediate implications for secondary school foreign language education in England. More generally, we demonstrate a replicable approach to developing short lists with high coverage, suggest some pedagogical applications, and discuss how our methods could be adapted for other contexts.

Keywords

adolescents, foreign language education, language testing, vocabulary, word lists

I Introduction

A solid grasp of lexis is fundamental for basic communication and progressing other areas of language ability. Vocabulary knowledge is a strong predictor of proficiency in all four basic language skills (In'nami et al., 2022; Jeon & Yamashita, 2022; Jeon et al., 2022; Kojima et al., 2022) and is necessary for advancing spoken fluency (Hilton, 2008), grammatical patterning (e.g. Francis, 1993; Hunston & Francis, 2000; Hunston et al., 1996, 1998), and pragmatics. The relationships between vocabulary knowledge and other aspects of learning highlight the importance of developing a robust, core vocabulary in the initial stages, and the need for a systematic approach to defining and sequencing lexis in curriculum content.

Nation (2007) proposed that a well-designed language course incorporates a balance of 'four strands' of learning activities: deliberate learning, meaning-focussed input, meaning-focussed output, and fluency development. While there is strong consensus that both deliberate and incidental encounters are necessary for acquisition and retention of vocabulary, deliberate learning seems to be more efficient and effective (Schmitt & Schmitt, 2020), with learning through flashcards and word lists, in particular, consistently leading to large gains in knowledge of form-meaning connections (Webb et al., 2020). In the secondary school context, Min (2008) found that intermediate learners of English in Taiwan retained more target words by combining reading with word-focussed activities than by supplementing reading with thematically related materials containing the target words. Additionally, Webb and Chang (2015) demonstrated that Taiwanese high school students with more prior vocabulary knowledge are likely to make greater vocabulary gains from extensive reading than students with smaller prior vocabularies. In the initial stages, then, explicit methods seem a useful means of developing the knowledge necessary to engage effectively with meaning-focussed input and output.

An essential resource for all four types of learning activity is a carefully designed word list that indexes target knowledge. Word lists focus attention on useful lexical items, set learning goals, support the development of achievement tests to monitor progress (Dang, 2019), and act as a benchmark for identifying and addressing the needs of learners with different experiences (e.g. little prior instruction; native or heritage speakers), which is especially useful at points of transition. Lists with headwords expanded

into lemmas or word families can be used with lexical profiling tools like *MultilingProfiler* (Finlayson et al., 2022) and *AntWordProfiler* (Anthony, 2024) to develop materials and tests that align with learners' assumed knowledge. While a comprehensive, 'one-size-fits-all' (Tschirner, 2019, p. 106) list for receptive knowledge and production is unfeasible, given the range of needs and interests represented in any cohort, a list representing a robust core of receptive knowledge can support learners in developing the autonomy necessary to create their own personal repertoires.

A strategically planned lexicon is especially important in contexts where very limited time is dedicated to languages. In England, the study of another language is compulsory from the ages of 7–14 years. Most students receive just 1–2 hours' instruction a week in the lower years (ages 11–14 years) of secondary school, and as little as 30 minutes a week in primary school (ages 7–11 years) (Collen, 2022). Primary school instruction is mainly delivered by non-linguists, and the quality of support for the primary to secondary transition varies greatly (Collen & Duff, 2024). With no obligatory word list available to date to inform teaching, secondary teachers have been unable to make assumptions about words that might already have been encountered. Further, many schools offer little to no international cultural or social engagement (Collen & Duff, 2024), and most students are not accessing the benefits (e.g. Peters et al., 2019; Sundqvist, 2009) of extramural exposure. These circumstances add yet more challenges to a non-compulsory subject that is already perceived to be difficult (Marsden, Dudley, & Hawkes, 2023), with the number of students taking a General Certificate of Secondary Education (GCSE) exam in a foreign language (FL) declining by almost 50% over the last 20 years (Churchward, 2019).

As part of a wider effort to make languages study more accessible to more students, the Department for Education (DfE, 2023) in England recently made it compulsory for commercial awarding organizations who create high-stakes national examinations to use corpus-informed word lists in the development of GCSEs in French, German, and Spanish, the three most widely taught languages (Collen & Duff, 2024). Students can choose to take a language GCSE at ages 15–16 years after five years of secondary school. By the current specifications, the skills required to pass the GCSE broadly align with A1–B1 in the Common European Framework of Reference for Languages (CEFR): A1/A2 for a standard pass at foundation tier,¹ A2 for an excellent pass at higher tier, and B1 for the highest possible mark (Curcin & Black, 2019). There is some variation within and across languages: foundation-tier listening skills seem to align with low-mid A2 for Spanish and low-mid A1 for French, while higher-tier students need low-mid B1 writing skills but only high A1-low A2 listening skills to achieve an excellent pass in French. Language-driven, cross-language standards for lexis could contribute towards redressing this balance.

The new policy, which will be implemented for teaching from 2024 and examination from 2026, is a first for foreign language education (FLE) in England and, as far as we know, for instruction and examination of French, German, and Spanish as FLs in secondary schools generally. Lexical syllabi for English have been integrated into the national curricula of schools in other countries such as Israel (Laufer, 2023a), Hong Kong (Education Bureau, 2023), and China (Ministry of Education, 2022). Word lists are used as standard to develop vocabulary tests for first-year students of

French in Swedish universities (Lindqvist & Ramnäs, 2023), and a word list for Welsh is being used to inform the revision of A1 and A2 level materials for adult courses offered by the National Centre for Learning in Wales (Knight et al., 2023). Wales is another country set to introduce word lists for French, German, and Spanish as FLs in schools, having announced an initiative to be introduced for teaching from 2025 (Qualifications Wales, 2024).

We are not aware of any studies that have tested the potential of word lists developed for national curricula to provide learners with the theoretical knowledge needed to comprehend texts of potential relevance. The current article, therefore, makes two important contributions: (1) a replicable and adaptable approach for researchers and educators to co-create corpus-informed word lists for adolescent, beginner-to-low-intermediate learners, and (2) an evaluation of the potential of these lists to cover material that represents the likely learning goals of adolescents. We begin with a summary of current and new requirements for GCSE vocabulary content, and consider key aspects of each policy against the backdrop of literature on word list creation and use (focussing, where possible, on issues relevant to our languages and learners). In the main part of the article, we describe methods for creating and evaluating exemplar lists within the constraints of a large-scale, fast-paced policy initiative. Some steps were taken in collaboration with teachers, and some with colleagues from Eduqas (an awarding organization sitting within the Welsh Joint Education Committee; WJEC). At the time of our list creation project,² Eduqas was in the process of obtaining accreditation by England's Office of Qualifications and Examinations Regulation (Ofqual) to permit them to examine the new GCSE subject content³ (DfE, 2023) in England. It therefore made sense to use Eduqas' (2019a, 2019b, 2019c) current, topic-driven word lists as the point of comparison in our evaluation study. These topic-driven lists were developed in line with current policy in England (DfE, 2022; original 2014 version no longer available online), which is relevant for GCSE examinations taken across 2018–25 and was introduced for teaching in 2016.

While the exemplar lists presented in this article were produced to align with the new curriculum content in England (DfE, 2023), we believe that both the process and outputs are adaptable and applicable in many contexts (including, potentially, the development of new word lists for the Welsh qualification mentioned above). Our findings have implications for educational stakeholders working in the context of French, German, and Spanish in secondary schools in England, and for educators and researchers considering approaches to creating lists for use in other FLE contexts. Our research questions are:

- Research question 1: How do the current and new GCSE lists compare in terms of (a) size, and proportions of (b) grammar and content words, (c) single-word items and multiword phrases, and (d) parts of speech?
- Research question 2: How much lexical coverage of materials that align with the likely learning goals of adolescents do the current and new lists provide?
- Research question 3: How is lexical coverage affected by proportions of (a) grammar and content words, (b) single-word items and multiword phrases, and (c) parts of speech?

II Background

Here, we briefly summarize the current and new policies on requirements for the vocabulary content of GCSEs and discuss how these relate to literature in four areas relevant to word list creation: implications of coverage for comprehension, types of word lists, appropriate list size, and units of word counting. These considerations draw on previous work, such as creating an essential word list for beginners (Dang & Webb, 2016) and designing a lexical syllabus for high school (Laufer, 2023a).

I Current and new policies about GCSE word lists

Under current policy, which will continue to determine the content of GCSE examinations until 2025, the three accredited awarding organizations – AQA, Pearson Edexcel, and Eduqas – are not required to use word lists. However, all provide topic-driven lists as guidance to teachers and materials developers. The current subject content does not regulate the size or content of these lists, and little direction is given on their intended purpose. Because awarding organizations have been obliged to create examinations that cover specific and wide-ranging topics (e.g. ‘family’; ‘home’; ‘technology’), word lists have historically been arranged in topic-related clusters. Prior to 2022,⁴ awarding organizations were also required to include in their exams unlisted words that they deemed to be ‘common or familiar’ at foundation tier and ‘less common or familiar’ at higher tier (e.g. Eduqas, 2019a, p. 16).

Under the new policy (DfE, 2023; on which all authors advised 2019–24), 100% of spoken material used in GCSE examinations (i.e. the target language text in listening exams) must be covered by the word lists. Texts for the reading exam may include a small number of ‘off-list’ words. Reading comprehension tasks may each include up to 2% cognates with English, an additional 2% glossed words, and proper nouns. Dedicated inferencing activities must include unlisted words as target items only. For oral and written production, any appropriately used words are rewarded (regardless of whether they are on or off list), but it must be possible to complete all production tasks with on list words. The following requirements lay out the minimum number and type of lexical items students are expected to know by the end of the course (DfE, 2023, pp. 5–8):

- 1,200 single-word items at foundation tier, and a further 500 at higher tier.
- Up to 30 additional short phrases (or compounds) of five words or fewer.
- Up to 20 additional single-word items or short phrases that refer to geographical places or cultural events.

In terms of selection criteria:

- At least 85% of single-word items must be among the 2,000 most frequent words in the most widely used standard forms of the language according to one or more large, multi-genre corpus/corpora of contemporary spoken and written texts.
- Words referenced in the compulsory grammar content must be included.
- Exemplar words to illustrate each regular grammar pattern specified in the compulsory grammar content must be included.

- A limited number of broad themes or topics relevant to the countries or communities where the language is spoken should be defined.

In terms of word counting (see Section II.4.b):

- Students are required to know words that can be inflected, and (for reading only) derived from listed lexical items using the specified grammar.
- Specified irregular word forms, and forms that follow regular patterns other than those included in the prescribed grammar, are counted as unique items.
- Additional meanings in different parts of speech are counted as unique items.
- All English equivalents that can be tested must be explicitly provided.

Having summarized key features of each policy that relate to word lists, we now consider the requirements in view of research relating to list creation and use.

2 Word lists, coverage, and comprehension

The requirement for assessments to use listed vocabulary is supported by findings about how much coverage (i.e. proportion of words known to learners) is needed for comprehension of texts. Coverage is a well-established predictor of text readability and comprehensibility, alongside factors such as background knowledge, syntactic complexity, and cohesion in written comprehension (see Crossley et al., 2023, for a summary), as well as speech rate, errors, corrections, and repetitions in spoken comprehension (O'Brien, 2014). Task-related variables (question format and text genre) have also been shown to affect the impact of lexical coverage on comprehension (Kremmel et al., 2023). While the 'optimal' proportion of known words depends on several factors, there is strong consensus that for adult, intermediate-to-advanced learners: (1) comprehension increases as unknown-word density decreases; (2) at least 90% coverage is needed to understand most types of texts in English (e.g. Kremmel et al., 2023; Noreillie et al., 2018; van Zeeland & Schmitt, 2013); (3) more coverage is needed to understand formal or academic genres than informal narratives and spoken informal narratives require less coverage than written ones (e.g. Schmitt et al., 2011; van Zeeland & Schmitt, 2013); and (4) substantial coverage is needed to infer the meanings of unknown words in written texts (e.g. Laufer, 2020), which is a particularly challenging skill for low-proficiency learners (Hamada, 2014). Inferencing in listening seems likely to be even more challenging because unknown words are difficult to locate in speech and cannot be revisited when processing aural input (as argued by Marsden, Dudley, & Hawkes, 2023), though few studies have investigated this.

Coverage thresholds for French, German, and Spanish are less well established, perhaps because word lists compatible with lexical profiling tools have been less readily available for these languages until recently (Finlayson et al., 2023). The two studies of which we are aware provide varying findings: intermediate learners of French needed less than 90% coverage to adequately comprehend spoken text (at least 86% for an average score of 74%; Noreillie et al., 2018), while beginner learners of Spanish needed more than 90% to comprehend graded narrative (at least 95% for average recall scores of 70% in literal comprehension and 66% in inferencing; Herman & Leiser, 2022). These results may depend to some extent on the nature of the tests used in the research, as Kremmel et al. (2023) note.

We are not aware of any studies that have investigated how these figures might differ for adolescents, whose L1 and L2 literacy may be different to that of adults. Given these unknowns, along with the consensus that comprehension increases with coverage and high coverage is necessary for inferencing, the finding that beginners may need higher coverage for adequate comprehension (Herman & Leiser, 2022), and the added challenges associated with aural inferencing, the approach adopted in the new policy (100% coverage of spoken text by the word list and a little less for reading comprehension and inferencing) seems to give these beginner-to-low-intermediate students the best chance of achieving adequate comprehension to pass the GCSE.

3 What type of word list serves the needs of adolescents?

On a cline of specialization from general service lists representing the ‘crucial starting point of L2 vocabulary learning’ (Dang, 2019, p. 289) to specific lists that meet bespoke technical needs, pedagogical lists for adolescents in mainstream education sit towards the general end. General word lists for pedagogical purposes are normally classified as one of three types: topic-driven, knowledge-based, or corpus-informed, though most incorporate more than one approach and can also draw on other principles. Here, we discuss key features of each list type, focusing on adolescent, beginner-to-low-intermediate learners.

Topic-driven lists tend to serve communicative approaches and have been popular in courses aligning with CEFR levels, particularly in the instruction of German (whereas English word list research and development have been more likely to be corpus-informed; Tschirner, 2019). This method of list creation aims to equip students with the words necessary to interact in practical situations deemed relevant to everyday life. The approach is exemplified by the methods used by awarding organizations in England to date. For example, Eduqas employed a team of specialist exam writers familiar with the target learner group to draw on their classroom and subject experience to select the current lexicon in line with pre-defined (Eduqas, 2019a, 2019b, 2019c) GCSE themes and topics (H. Potter, subject specialist, personal communication, 7 February 2023) deemed “of interest and relevance” (e.g. Eduqas, 2019a, p. 5) to adolescents. Topic-driven approaches have been used to create several influential vocabulary guides and workbooks for German as a foreign language (for a summary, see Bonazzi, 2017), as well as in CEFR-aligned word lists for French (e.g. Beacco & Porquier, 2007; Beacco et al., 2008), German (e.g. Glaboniat et al., 2005; Glaboniat et al., 2016), and Spanish (Instituto Cervantes, n.d.) used in language proficiency examinations offered by internationally recognized language organizations. Part of the core list used in French education in Swedish universities (Lindqvist & Ramnäs, 2023) was also developed based on teacher intuition about relevant topics. One potential advantage of topic-driven lists in materials development (Bonazzi, 2017) has been as a source of semantic clusters of words for textbooks arranged into topic-focussed chapters. There is some debate over the value of semantic clustering, however; while there is evidence to suggest that there are some benefits to learning words in semantically related sets, the effects of these are often short-lived or only observable with words referring to physically unlike entities, and other studies have reported negative effects (see Marsden, Dudley, & Hawkes, 2023, for further discussion). Another potential weakness of lists based on subjective choices is that important high-frequency words that are not clearly or directly connected with topics can be overlooked (Tschirner, 2009,

2019). The consequences are low coverage (e.g. Bonazzi, 2017; Marsden, Dudley, & Hawkes, 2023; Kusseling & Lonsdale, 2013) and a risk of skewing washback effects towards teaching low-frequency topic-bound nouns that are never or rarely reencountered (e.g. Häcker, 2008; Horst, 2013; Marsden & David, 2008). Indeed, an average 70% of words on current, topic-based GCSE lists have been used only once or never across four sets of French, German, and Spanish exams (Dudley & Marsden, 2024).

Knowledge-based word lists are based on learner input and output. The approach shares some similarities with the topic-driven method in that items are assigned to levels using pre-defined criteria, for example, ‘can-do’ statements describing structures and scenarios of a pre-determined difficulty – though the criteria are arguably more objective. Data about learner knowledge are derived from corpora of learner language (e.g. Capel, 2010), teacher judgments (Robles-Garcia et al., 2023), and/or vocabulary tests (e.g. Brysbaert et al., 2021; Schmitt et al., 2021). A potential limitation of this approach is the circularity involved: the more learners are exposed to learner-oriented vocabulary, the more likely they are to know it. Frequency in learner corpora reflects learners’ confidence and ability to produce lexical items and, potentially, how straightforward items are to learn and use in class or tests (Nation, 2016). However, it does not tell us anything about usefulness. Another possible limitation of lists based on data from a varied pool of informants is that they do not account for individual differences; see Schmitt et al. (2021), who addressed this issue by creating knowledge-based vocabulary lists with three different formats for learners with different L1s of varying degrees of linguistic proximity to English.

The corpus-informed lists specified in the new policy take a language-driven approach, aiming at high lexical coverage. By Zipf’s law, coverage is mainly provided by a very small number of words with high frequency in general language. In English, around 50% of any text is covered by the 100 most frequent words, most of which are function words. Around 80% is covered by the 2,000 most frequent, with just a 1%–2% increase for each additional 1,000 words thereafter (figures for French, German, and Spanish seem to be similar, if not higher; see Table 1). The reason for the skew is that high-frequency words often have many senses, and readily occur in different collocations and semantic patterns that create extended units of meaning (Sinclair, 1996, 1998). As a result of their ubiquity, high-frequency words are likely to (1) provide insights into cultural contexts (e.g. Kilgarriff, Charalabopoulou, et al., 2014), (2) carry lexical, syntactic, or semantic complexities which merit instructional attention, and (3) offer cognitive advantages for learners in terms of the speed and accuracy of recognition processes (e.g. Ellis, 2002; Tschirner, 2009). While emphasis on coverage is traditionally associated with receptive skills, studies in English (Laufer, 1998) and French (Caltabellotta et al., 2024) indicate that even advanced learners with vocabulary sizes of 5,000 words or more tend to prioritize words from the 2,000 band in their writing, suggesting that the policy’s focus on high-frequency words in the initial stages may also be beneficial for production. High-frequency words have been variously defined as the 2,000 (e.g. Nation, 2022) or 3,000 (Schmitt & Schmitt, 2014) most frequent in English. The new policy adopts the former definition, partly because the limited exposure to the language restricts the total number of words that can feasibly be learned (i.e. fewer than 2,000). Selecting from the most frequent 2,000 guarantees the very most useful will be included and aligns with the findings in Section II.4.a about the potentially greater coverage power of words in French and Spanish.

Table 1. Coverage by high-frequency words across languages and genres as reported in Nation (2001), Cobb & Horst (2004), Ramnäs (2019), Tschirner (2009), and Davies (2005).

K	Unit	Fiction				News				Academic				Spoken			
		En	Fr	Ge	Sp	En	Fr	Ge	Sp ^a	En	Fr ^b	Ge	Sp	En	Fr	Ge	Sp
1K	Lemma		85.3	74.5	79.6			67.4	76.0			67.6	–		–	85.2	87.8
	WF6	82.3				75.6	77.6			73.5	71.2			84.3			
2K	Lemma		90.8	80.0	86.1			73.9	84.0			74.7	–		–	89.2	92.7
	WF6	87.4				80.3	86.2			78.1	81.3			90.3			
3K	Lemma		93.9	82.7	89.6			77.3	88.2			78.5	–		–	90.9	94.0
	WF6						88.2				84.8						
	2KWF6+AWL	89.1			84.2					86.6				92.2			

Notes. En=English; Fr=French; Ge=German; Sp=Spanish; K=frequency band; WF6=word family level 6; AWL=academic word list (Coxhead, 2000). A dash (–) indicates no data available. ^aFigures are for the whole 'non-fiction' category.

^bIncludes medical texts only.

Given the context-bound nature of teaching, few pedagogical word lists use frequency as the only selection principle. A common approach is to choose one or more source frequency lists and adjust them in line with inclusion or exclusion criteria pertaining to usefulness, relevance, and difficulty. For example, West's (1953) General Service List and its update, the New General Service List (Browne et al., 2023) contain high-frequency words that are also considered 'learnable' and useful for personal and professional development. West (1953) specifically targeted words relevant to the needs and interests of 12–18-year-olds (Gilner, 2011, p. 69). Nation's (2017) BNC/COCA lists are organized by frequency and range (the number of corpus segments in which the word appears at least once), with adjustments made for words that are common in spoken language or part of lexical sets (e.g. numbers, months). The CEFR-aligned *English Profile Wordlists* (Capel, 2010) are based on frequency data from both general and learner corpora and draw on word lists from coursebooks and other materials. Other examples include Laufer (2023a) and Knight et al. (2023), who worked closely with curriculum planners and teachers to select words for their English and Welsh curriculum lists. The *Basic German vocabulary* (James & James, 1991) was developed by consolidating existing topic-driven and corpus-informed lists and using factors like familiarity and usefulness for everyday conversation as selection criteria. In French, Gougenheim et al. (1967) supplemented frequency data from an oral corpus with words frequently mentioned in surveys on specific topics, but the list is very out of date. Noting the 'sporadic' use of high-frequency words in textbooks for beginners, Antes (2023) has called for creation of a new general service list for French based on frequency and dispersion in contemporary corpora.

We are not aware of any previous examples of corpus-informed, pedagogical word lists designed for adolescent, beginner-to-low-intermediate learners of French, German, or Spanish. The new policy establishes parameters for creating such lists that combine the strengths of the three approaches just described: an objective, language-driven approach in the main, but with expert input to make judgments about relevance, topic-specificity, and/or learner knowledge. This approach seems appropriate for the context of mass education of anglophone adolescents.

4 Word list composition

Well-developed lexical syllabi represent two dimensions of target knowledge: the number of words learners know 'in some sense' (breadth) (Read & Dang, 2022, p. 2) and what they know about those words (depth). In the new policy, the former dimension is addressed by the requirements for list size (Section II.4.a) and the latter to an extent by requirements about form and meaning that inform a nuanced approach to defining word counting. Nation (2022) provides a comprehensive overview of aspects of form, meaning, and use involved in knowing a word. Some of these, relevant to beginner-to-low-intermediate learners, are reviewed in Section II.4.b.

a How many words do adolescent, beginner-to-low-intermediate learners of French, German, and Spanish need to know? Four factors to consider when determining appropriate list length are: (1) how many words are needed to achieve a target coverage threshold, (2)

Table 2. Mean (*SD*) unique lemmas in Eduqas GCSE papers in French, German, and Spanish.

	Foundation	Higher
2018	569.3 (71.8)	726.0 (80.6)
2019	570.7 (33.2)	720.0 (51.6)
2022	549.0 (75.0)	746.7 (71.0)
All years	1121.0 (136.7)	1460.3 (164.8)
Mean	563.0 (<i>SE</i> = 7.0)	730.9 (<i>SE</i> = 8.1)

Notes. *SD* = standard deviation; *SE* = standard error. The variability across languages can be attributed to the higher counts for German, which is likely due to the prevalence of compounding. For figures for individual languages, see Appendix A.

how many words learners of a given proficiency typically know, (3) words deemed not to require explicit instruction, and (4) contextual and curricular constraints.

The number of high-frequency words required to reach a specific coverage threshold varies across languages and genres. Table 1 shows coverage of four genres of adult text by the 1,000, 2,000, and 3,000 most frequent words in English (Nation, 2001), French (Cobb & Horst, 2004; Ramnäs, 2019), German (Tschirner, 2009), and Spanish (Davies, 2005). Although these are not replication studies (the corpus content is not balanced across languages, and different units of counting are used), some trends emerge regardless. For all languages, each band covers more speech and fiction than news and academic text, with the 2,000-band covering approximately 90% of spoken material and at least 80% of written material, with the exception of formal texts in German. Further, French and Spanish bands seem to provide more coverage than their German and English counterparts. The lowest coverage figures are observed in German, which may be due to (i) the prevalence of compounds that cannot be easily recognized by lexical profiling tools, (ii) the use of lemmas as the unit of counting compared with word families in English, and (iii) the use of erudite language in formal written genres (Jones, 2006). The last point likely explains the notably lower coverage of German news and academic texts (beyond the scope of beginner-to-low-intermediate learners).

Taken together, these findings about high-frequency words and coverage suggest that the top three frequency bands in French, German, and Spanish give ‘greater returns’ on texts that are potentially relevant to GCSE students than the same bands in English, supporting a focus on the 2,000-band in the new policy. However, the new word lists are not solely composed of high-frequency words, and as their primary function is to inform the creation of graded texts for the GCSE, high coverage is guaranteed. Indeed, another way of thinking about list size and coverage is to consider how many words are needed to create diverse reading and listening materials over a prolonged period. To investigate this, we counted the number of unique lemmas in all Eduqas GCSE papers released since 2018. Results (Table 2) suggest that students who learn 1,250/1,750 lexical items (equivalent to 1,121/1,581 lemmas; Table 6), would be amply prepared to comprehend the target language in exams that consist of 563 (foundation) and 731 (higher) lemmas on average. Further, lists of the proposed size contain as many lemmas as were needed to compile all exam papers released across three years, suggesting they offer a plentiful

source of words from which to create varied texts (for similar findings for AQA and Pearson Edexcel, see Dudley & Marsden, 2024).

Estimating the ‘typical’ vocabulary size of GCSE students is difficult, due to limited data and considerable variation in results. We are only aware of four studies with GCSE students, all conducted in French. Three of these (David, 2008; Milton, 2006, 2015) used the *X_Lex* (Meara & Milton, 2003), a lexical decision test estimating reported receptive knowledge of the form-meaning link in mid-high-frequency (< 5,001) words. Participants in these studies reported recognizing a mean 564 ($n=26$, $SD=352$, range: 0–1,650), 852 ($n=49$, $SD=440$, range: 0–1,800), and 775 ($n=16$, $SD=341$, range: 350–1,250) lemmas, respectively. Results from the fourth study, which used a meaning recognition test focusing on high-frequency (< 2,001) words from the current GCSE word list (Dudley et al., 2024), estimate the highest performing students could recognise a mean 1,480 lemmas ($n=220$, $SD=309$, 95% CIs [1,439–1,521], range: 295–1,993). Students of French at schools in Greece and Spain taking the *X_Lex* reported even higher estimations. For students in Greece at CEFR levels A1, A2, and B1, the mean estimations were 1,125 ($n=35$, $SD=620$, range: 0–2,550), 1,756 ($n=8$, $SD=398$, range: 1,500–2,500) and 2,422 ($n=11$, $SD=517$, range: 1,800–3,400), respectively. For students in Spain, the estimations were 894 ($n=18$, $SD=604$, range: 350–2,850), 1,700 ($n=9$, $SD=841$, range: 500–2,750), and 2,194 ($n=9$, $SD=717$, range: 1,100–3,100) across the same levels (Milton & Alexiou, 2009).

In the absence of equivalent research with school students learning German and Spanish school students, we reference a study of the relationship between reading proficiency and receptive knowledge of high-frequency words by students of German and Spanish at universities in the U.S. and Germany (Tschirner et al., 2018). For German, CEFR level A1⁵ predicted mastery of 83.7% of the 1,000-band, A2 the 1,000-band plus 64% of the 2,000-band, and B1 the 2,000-band plus 44.3% of the 3,000-band ($n=77$, $n=36$, $n=3$, respectively). For Spanish, the equivalent figures were 11.4% of the 1,000-band (A1), 97% of the 1,000-band (A2), and the 1,000-band plus 82.6% of the 2,000-band (B1), based on a small sample ($n=7$, $n=8$, $n=2$). Although these estimates of words known by learners at CEFR levels A1–B1 are based on studies conducted in somewhat different contexts, the general finding that A1 students know close to 1,000 lemmas, A2 students 1,000–2,000, and B1 students 2,000–3,000 supports a list size of 1,250 items (1,121 lemmas) at foundation (A1/A2) and 1,750 items (1,581 lemmas) at higher (A2 for all but the highest performers). These figures may seem ambitious considering the findings reported by David (2008) and Milton (2006, 2015), but it is important to note that their GCSE students followed a topic-driven approach that did not focus on the words tested in *X_Lex*. Results from Dudley et al.’s (2024) curriculum-aligned test show that the highest performing GCSE students can recognize as many words as other A1/A2 students on average.

Finally, the lists do not represent the totality of input students will receive. For example, under the new policy, students are required to develop the ability to recognize unlisted words in certain contexts. Choosing not to list words that are similar in the L1 or have meanings that can be deduced from word parts or context may also indirectly promote the development of such skills. In the context of the new GCSE, items that do not need listing are: (1) inflected forms of listed headwords that follow specified

Table 3. Ways of counting *focus*.

Unit of counting	Forms and senses covered
Type	focus (v): focus (v)
Lemma	focus (v): focus (v), focuses, focussing, focussed
Flemma	focus (v/n): focus (v), focuses, focussing; focussed, focus (n), foci
Word family	focus (v/n/adj): focus (v), focuses, focussing, focussed, focus (n), foci, refocus, refocuses, refocussing, unfocussed

Source. Adapted from Finlayson et al., 2023.

grammar patterns, and in reading only: (2) derived forms of listed headwords that follow specified grammar patterns, (3) target items in inferencing tests, (4) proper nouns, and (5) specified proportions of glossed words and cognates. Evidence relating to the requirements for inflectional and derivational morphology is discussed in Section II.4.b. Promoting cognate recognition skills in reading is justified in the initial stages, given English shares large numbers of cognates with French, German, and Spanish (e.g. Lubliner & Hiebert, 2011). Cognates can cause facilitation effects (e.g. Elgort, 2013; Schmitt et al., 2021; Vidal, 2011), and orthographical cognates are more transparent than phonological ones (Lubliner & Hiebert, 2011), justifying the restriction to testing unlisted cognates in reading. Additionally, students can (if desired) learn further unlisted words to personalize and extend the core. High-performing students could therefore learn to recognize numbers of words closer to those associated with CEFR B1.

In sum, the required core list size is likely to be appropriate for the context, given (1) the seemingly higher coverage provided by French, German, and Spanish words compared with English, (2) the findings about breadth of knowledge associated with CEFR levels A1–B1, and (3) the requirements promoting inferencing, recognition, and acquisition of unlisted words. Moreover, learning 1,250/1,750 items across five years seems to be a feasible target. For acquisition rates of around 3–4 words per contact hour (Milton, 2006) to occur, repeated exposure (e.g. Elgort & Warren, 2014; Pellicer-Sanchez & Schmitt, 2010; Vidal, 2011; Webb, 2007) across the four strands of practice types (Nation, 2017) is crucial. Marsden and Hawkes (2023) showed that words on a list of this size can be mapped to a practice schedule deliverable in the available curriculum time, ensuring that each receives attention for potential acquisition and retention.

b What is the most appropriate unit of counting for adolescent, beginner-to-low-intermediate learners of French, German, and Spanish? The ‘unit of counting’ by which a word list is organized represents learners’ assumed ability to recognize inflected forms, derived forms, and different meanings of listed headwords. Most lists categorize words as one of: types (headwords only), lemmas (headwords and inflected forms within the same part of speech), flemmas (headwords and inflected forms across parts of speech), or word families (headwords and inflected and derived forms across parts of speech). Table 3 illustrates ways in which forms and meanings of *focus* can be organized.

Compared to French, German, and Spanish, English is less morphologically rich and has fewer morphological irregularities in written form. As a result, few word lists for

English are organized by type, as it is assumed that learners at any level will be able to deal with variation in written inflectional forms (Bauer & Nation, 1993). However, evidence from psycholinguistics suggests that highly irregular forms, such as *drove* in English, are acquired and stored as distinct words by both adults and children (e.g. Kempley & Morton, 1982; Pinker, 1991; Ullman, 2001). Similarly, in German, both adults and children tend to access irregularly inflected verbs (e.g. *geschlafen* ‘slept’) and nouns (e.g. *Muskeln* ‘muscles’) holistically from the mental lexicon (Clahsen, 1999). Studies in French found slightly different evidence (Meunier & Marslen-Wilson, 2004), suggesting that speakers decompose inflected verb forms – including those with irregular stems but regular inflections (e. g. both *boi-* and *buv-* from *boire* ‘drink’) – prior to lexical access. These stems appear to have distinct representations in the lexicon, each stored as ‘fully regular verbs’ (Estivalet & Meunier, 2015, p. 1) with their own sub-lemmas of regularly inflected forms. Taken together, these findings suggest that some type-based counting may have a place in word listing and instruction, at least with beginner learners of these languages. The new policy requires that certain irregular forms are listed as individual lexical items, potentially reflecting evidence to date about storage and access for (highly) idiosyncratic forms.

Most word lists for general purposes (e.g. Brezina & Gablasova, 2015; Browne, 2014; Davies & Davies, 2018; Lonsdale & Le Bras, 2009; Tschirner & Möhring, 2020) or beginners (Dang & Webb, 2016) are organized by lemmas or flemmas. One reason to count lemmas rather than flemmas, especially in the initial stages, is that learning additional meanings of familiar words seems to be just as difficult as learning the primary meaning of unfamiliar words (González-Fernández & Webb, 2024). Homographs whose meanings also differ in part of speech have been rated as more ambiguous than homographs whose multiple meanings belong to the same grammatical class (Twilliey et al., 1994). Supporting this, Stoeckel et al. (2020) found that beginner and intermediate learners with some knowledge of a word in one part of speech understood it in another only 56% of the time.

Still, neither lemmas nor flemmas make assumptions about learners’ derivational morphological awareness, in this sense making either potentially suitable in beginner-to-low-intermediate contexts (Dang & Webb, 2016; McLean, 2018; Ward & Chuenjundaeng, 2009). There is evidence that learners’ receptive (Laufer et al., 2021) and productive (Iwaizumi & Webb, 2021, 2023) knowledge of derived words is closely associated with their vocabulary breadth. Similarly, learners’ ability to recognize affixes is linked to their vocabulary level, though this is not the case for the most infrequent (Mochizuki & Aizawa, 2000) and most difficult (Sonbul & El-Dakhs, 2024) affixes. These findings support the use of word families with advanced learners, for example in the creation of academic word lists (e.g. Coxhead, 2000; Dang et al., 2017; Dang, 2018), though Laufer (2023b) found that high school students of English in Israel were able to recognise derived forms with frequent affixes when contextual clues were provided. In sum, the appropriacy of different units of counting varies with purpose, and more work is needed to investigate the benefits of each for different contexts and uses (Webb, 2021).

The case of beginner-to-low-intermediate learners of French, German, and Spanish in England is interesting in this respect. The typological proximity of the three languages to English suggests that many learners may benefit from cross-linguistic orthographic transparency in the form and meaning of derivational affixes. Further, for languages with rich

inflectional systems, some derived forms may be more transparent than inflected forms. Verb lemmas in Spanish, for example, have over 50 unique inflected forms, some of which can be difficult to recognize. A bespoke unit of counting combining some inflectional and some derivational morphology as appropriate is needed for these languages. Bauer and Nation (1993) were the first to propose arranging affixes into seven levels of complexity. The word-family-based BNC/COCA lists for English (Nation, 2017) are available at level 3, which includes frequent, productive, predictable, and regularly derived forms, and level 6, which encompasses all derived forms except those with classical roots. More recently, Mochizuki and Aizawa (2000) proposed an affix acquisition order, and Sasao and Webb (2017) developed a levelled test of affix knowledge based on difficulty. Cobb and Laufer (2021) and Cobb et al. (2023) have developed lists in English and French based on ‘nuclear’ word families, which include only items meeting a certain within-family frequency threshold. These units, however, do not consider other aspects of complexity or usefulness. For example, words with affixes at level 6 in Bauer & Nation’s hierarchy are pushed out of some families but kept in others. We are not aware of any examples of counting units that combine specific selections of inflectional and derivational affixes, which is what the incoming policy proposes.

Finally, the policy’s focus on single-word items (rather than the large numbers of multiword phrases that have characterized GCSE lists to date) aligns with the idea of prioritizing the ‘commonest word forms in the language, their central patterns of usage, [and] the combinations which they typically form’ (Sinclair & Renouf, 1988, p. 148). Corpus linguistic research on lexical patterning has shown that all words have (variable) patterns in which they typically occur, and that those patterns carry aspects of meaning (Hunston et al., 1996, 1998; Sinclair, 1996, 1998). In other words, the policy supports awareness of the many multiword phrases (MWP) in which words frequently occur, rather than focussing on specific, pre-selected phrases. Although it has been suggested that learning formulaic language can support some learners in constructing rules by breaking chunks into their component parts (Myles et al., 1998), success in this can be very slow and highly variable depending on learners’ analytic ability. Further, under the current policy, it is not always clear whether MWPs are intended to be broken down. For example, in the current Spanish list (Eduqas, 2019c), *tomar prestado* (‘to borrow’) is listed, but *tomar* (‘to take’) in its general sense is not.

III Current study

The aims of this article are (1) to show how researchers, policymakers, awarding organizations, and teachers can collaborate to develop corpus-informed word lists, and (2) to evaluate the effectiveness of lists created using those methods. Addressing the first aim, this section describes the process. A quantitative approach to word list creation is relatively straightforward to replicate, whereas a more complex approach that integrates input from stakeholders requires careful reporting to be replicable in other contexts (Knight et al., 2023). Figure 1 summarizes our approach and the collaborators involved at each stage, serving as a road map for what follows.

All co-authors and the wider research team have teaching experience as former or current secondary school teachers and/or higher education language tutors. Our project partners included four colleagues from Eduqas/WJEC (a Qualifications Development

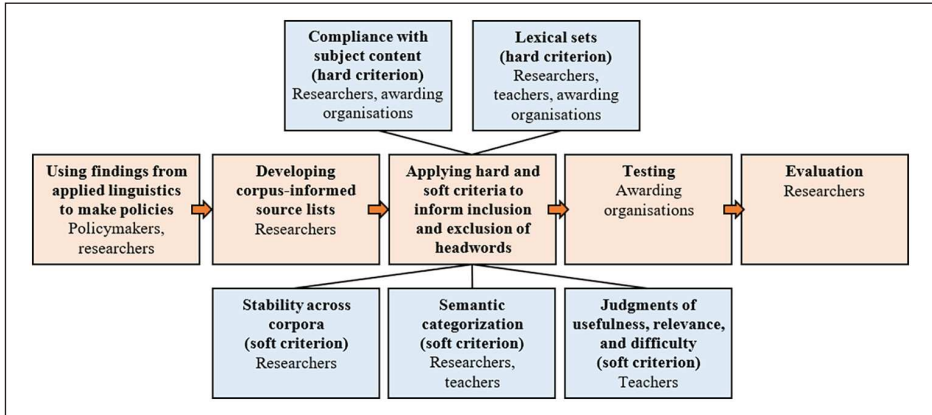


Figure 1. Stages of list creation and collaborators.

Officer and one Subject Officer for each language) and nine teachers (three per language) associated with the DfE’s National Centre for Excellence for Language Pedagogy (Marsden, Hawkes, et al., 2023). As teacher perceptions have been found to be better predictors of vocabulary knowledge (Dang et al., 2022; Robles-García et al., 2023) and lexical sophistication (Tidball & Treffers-Daller, 2008) than frequency-based criteria alone, we supplemented corpus data with ‘indigenous criteria’ (Knight et al., 2023), i.e. contextual experience that adds authenticity and practical relevance to research-informed resources. We chose to work with three teachers per language because it has been shown that the combined judgments of three (but not two) teachers can more accurately predict which words students know than frequency data alone (Robles-García et al., 2023).

As the literature review has already addressed the first stage of the process depicted in Figure 1 (policymakers’ decisions and research findings), this section focuses on the four other stages. Nation’s (2016) list of questions for critiquing English word lists informed some of our decisions. However, the processes were shaped by the constraints of a government-funded initiative operating within a two- to three-year timescale (from conceptualization, through public consultation, to approval) and a small research team.

1 Source lists

Any word list used for pedagogical purposes should ideally be derived from a corpus representing those purposes (e.g. Biber, 1993). However, because secondary school students’ future applications of languages are largely unknown (Hawkins, 1996; Nation & Sorell, 2016), defining the purpose of general service lists for secondary classrooms poses challenges. Therefore, we set ourselves the objective of developing lists that give optimum coverage of a range of texts adolescent learners may encounter in their immediate environments.

We refer to the frequency order and definitions in the Routledge dictionaries of French (Lonsdale & Le Bras, 2009), German (Tschirner & Möhring, 2020), and Spanish

(Davies & Davies, 2018), hereinafter referred to as ‘the Routledge lists’. These provide corpus-informed translations and other contextual information in English, and are developed from contemporary corpora meeting policy requirements: ‘internationally recognized ... at least 20 million words based on spoken and written language from a wide range of different contexts’ (DfE, 2023, p. 6). In this sense, they are more suitable than other available (i.e. accessible) frequency lists based on smaller, older, or mainly literary corpora or databases (e.g. Baudot, 1993; Verlinde & Selva, 2001; New et al., 2004). The Routledge corpora are similar in terms of size and sampling frame (Language-Driven Pedagogy, 2021a), though they use different dispersion measures⁶ and units of counting (flemmas in the French list, lemmas in Spanish, and lemmas plus names of countries and states in German). All lists consist of single-word items only. Multiword phrases were added manually in later stages.

The stability of general vocabulary across corpora has been questioned both in terms of the nature of shared items and their frequency order (Brezina & Gablasova, 2015). To investigate the stability of words on our source lists, we measured overlap between the 2,000 most frequent words in the Routledge lists and three comparison corpora of material representing potential learning goals. Our chosen genres for the comparison corpora were exam papers, web language, and, given that fewer words are needed to cover fiction than other written genres (Table 1), young adult literature. To create the exam papers corpora, we collected 160,000⁷ words of target language text (reading passages, listening transcripts, and questions) per language from GCSE and A/AS-level⁸ exam papers (see Tables B1 and B2 in Appendix B). These corpora are available to download from our OSF repository (<https://osf.io/5cxhq/>) and from IRIS (<https://doi.org/10.48316/sZm1K-WPKB8>). The young adult literature (YA) corpora each comprised 10 novels in a mix of original language and translations (see Table B3 in Appendix B). For web language, we used the TenTen Corpus Family in Sketch Engine (Kilgariff, Baisa, et al., 2014). We generated lemma-based frequency lists from each comparison corpus, extracted the 2,000 most frequent words sorted by average reduced frequency (ARF),⁹ and aligned them with the Routledge lists (e.g. by removing proper nouns and merging full forms and abbreviations). Following Brezina and Gablasova (2015), who used the same approach to identify a core general vocabulary for English, we identified the words common to each pair of lists and performed Spearman’s rank correlations on these shared words to establish whether they were distributed across corpora in a comparable way (Table 4).

Table 4. Pairwise comparison of adolescent-focussed lists with Routledge lists.

List	French			German			Spanish		
	Web	Exams	YA	Web	Exams	YA	Web	Exams	YA
Corpus size	15.2B	174K	553K	17.5B	164K	722K	17B	171K	697K
Overlap (raw)	1,636	1,423	1,229	1,634	1,323	1,164	1,615	1,407	1,389
Overlap (%)	81.8	71.2	61.5	81.7	66.2	58.2	80.8	70.4	69.5
r_s	.71	.56	.49	.77	.59	.58	.73	.42	.50
p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

Notes. K = thousand; B = billion; r_s = Spearman’s rho; YA = young adult novels.

Overall, the Routledge lists share large (at least 60%, accounting for compounding in German) proportions of high-frequency items with each corpus of adolescent-focussed texts. The very high (at least 80%) overlap with the web corpora is indicative of a strong stable core of general vocabulary, and the moderate to large (Cohen, 1988) correlations show that the rank orders in the Routledge lists and adolescent-focussed lists are comparable. Still, there are some differences, which speaks to the importance of using stability across word lists as an inclusion criterion for word selection (see Section III.2.b).

2 Inclusion/exclusion criteria

This section outlines the criteria by which words from the Routledge lists were selected for inclusion in the lists of 1,250 (foundation) and 1,750 (higher) items. A small proportion (less than 25%) of items were pre-determined based on hard criteria (see Section III.2.a). Remaining items were evaluated individually according to three soft criteria: (1) stability across corpora of adolescent-focussed material, (2) flexibility for communicating about different topics, and (3) teacher judgments of usefulness, relevance, and difficulty (Sections III.2.b–d). Items selected were assigned to either the foundation or the higher list, considering the requirements for each tier (Section III.2.e).

a Compliance with policy (the GCSE subject content) and lexical sets. A certain number of items (250–350 depending on language and tier) were pre-selected on account of their inclusion in the compulsory grammar content (DfE, 2023); that is, they are not at the discretion of awarding organizations. Exemplar words to illustrate each morphological pattern in the grammar content were also obligatory. A small number of (non-obligatory) items was then selected to complete five core lexical sets (Nation, 2016): days of the week, months, seasons, compass points, and numbers to 1,000. After this stage, the only hard criterion for the remaining words was a frequency ranking of 1–2,000 for at least 85% of all single-word items.

b Stability across corpora. This soft criterion indicates the likelihood of a word consistently featuring in the types of texts that adolescents may encounter. As we saw in Section III.1, we had access to four comparable lists of high-frequency words in corpora of potentially relevant material: general language, web language, exam papers, and young adult literature. To measure the stability of individual items, we counted the total number of lists (1–4) on which they appeared. This information, available to view in full in our OSF repository (<https://osf.io/5cxhq/>) and in IRIS (<https://doi.org/10.48316/W7DrC-vAFQf>), informed selection in two ways. For most (at least 85%) entries, priority was given to words with high frequency in the Routledge lists and at least two other corpora, as we considered high overlap a good indicator that words are encountered often in different contexts. Further, knowing which words are highly frequent in language written for adolescents (but not in general corpora) helped us select the words of ‘any frequency’ that can account for up to 15% of entries.

c Semantic categorization. This soft criterion considers the relatedness of lexical items with GCSE topics and themes, as judged by human raters. Using the Historical Thesaurus

Table 5. Mean (*SD*) inter-rater agreement across categories and languages.

	French	German	Spanish	Mean
Raw percentage	91.26 (5.63)	96.79 (1.97)	90.75 (5.34)	93.3 (<i>SE</i> = 1.96)
ACI	.89 (.09)	.96 (.02)	.89 (.07)	.92 (<i>SE</i> = .03)

Notes. *SD* = standard deviation; *SE* = standard error; ACI = Gwet's agreement coefficient. For breakdown by category, see Appendix C.

of English¹⁰ (Kay et al., 2021), we identified 52 semantic category headwords relevant to the GCSE, based on (but also expanding) themes commonly covered in textbooks. Following Chung and Nation (2004), who designed a scale for rating how closely related word meanings are to subject areas (that achieved inter-rater reliability of 95%), we trained two advanced speakers of each language (researchers and teachers) to assess how strongly high-frequency words on the Routledge list are related to each of the 52 topic headwords. We adapted Chung and Nation's example, which focused on anatomy, to create the following version of the scale with more general wording:

- 1 = no relationship with the topic (e.g. TRAVEL: BLACKBOARD)
- 2 = minimally related to the topic and could be used to talk about it in a general sense (e.g. TRAVEL: DURING)
- 3 = closely related to the topic, but also used in general language (e.g. TRAVEL: MOUNTAIN)
- 4 = specific to that topic (e.g. TRAVEL: SIGHTSEEING)

To calculate inter-rater agreement, we merged scores of 1–2 (no-to-weak relationship) and 3–4 (moderate-to-strong relationship) and analysed the resulting binary datasets in *Lancaster Stats Tools* (Brezina, 2018), using Gwet's AC1 to mitigate the effects of large numbers of zero values in each category (Zec et al., 2017). Agreement was very high overall (Table 5), with slightly higher consistency observed for concrete categories (e.g. *food, music*) than abstract ones (e.g. *friends, global issues*). The full dataset can be found in our OSF repository (<https://osf.io/5cxhq/>) and in IRIS (<https://doi.org/10.48316/SZyFK-1AkXT>).

Cases of disagreement were put to a third rater and all words with at least two positive values for a category were classified as 'topical'. In cases where fewer than five words had a moderate-to-strong relationship with a topic, that topic was merged with a closely related one (e.g. *food, drink, and eating out* in German) or removed. Because frequency can reflect concepts specific to cultures (Kilgarriff, Charalabopoulou, et al., 2014), category members and category names vary across languages (e.g. *food and eating out* in French, *food and drink* in Spanish; for topic-ordered groupings, see our OSF repository and IRIS). We then counted the total number of categories to which each word was assigned, creating a 'transferability index' ranging from 1 to the maximum number of categories for the language (up to 52). This index helped us select items likely to be topical for many categories.

d Teacher judgments. The third soft criterion was feedback from teachers on the suitability of items on the source lists. Teachers commented on the usefulness, relevance, and difficulty of target language words and translations, proposing alternatives where appropriate. Example suggestions for French included (1) removing overly formal words like *notamment* ('notably') and *élire* ('to elect'); (2) replacing formal words with more informal equivalents, for instance, substituting *usage* with *utilisation* ('use') and *exercer* with *pratiquer* ('to practise'); and (3) adjusting translations so that the meanings likely to be most relevant to adolescents were included, as these were not always the most frequent. For example, we added 'lessons' to the translation for *cours* ('course'). Teachers' suggestions were often mid-to-low-frequency words related to the education context that did not appear in the source lists. These recommendations helped identify the 15% of words that could be of 'any frequency'.

e Applying the criteria. Using the above criteria, two researchers per language allocated words to the foundation and higher tier lists until the respective target lengths of 1,250 and 1,750 items were reached. Three factors influenced our decisions about tier: (1) teachers' comments, (2) whether the grammar necessary to use the item creatively was included in the specifications for the tier, and (3) the relevance of the item to topics commonly covered at A-level (e.g. *in the news, social issues*), as students must take a higher tier paper to be eligible to continue study.

The other consideration at this stage was compliance with the bespoke counting units described in Section II.4.b. Because irregularly inflected forms and words with different meanings across different parts of speech are treated as unique entries, some of our chosen items occupied multiple slots on the lists, and careful decisions had to be made about how many irregular forms and meanings to include for each headword. For example, forms of *dire* ('to say, tell') in French follow a pattern specified in the prescribed grammar, with the exception of the second person plural *dites*. Therefore, the lemma occupies two of the 1,250/1,750 spaces. The two meanings of *historia* ('history, story') in Spanish are counted together, but the verbal and pronominal meanings of German *sein* ('to be' and 'his, its') are listed separately. Additional specifications for reading also had an impact. In some cases, cognates and derived forms of listed base words were excluded to prioritize more complex items, increasing lexical diversity in the lists.

3 Testing

At the testing stage, Eduqas colleagues used the custom list feature of *MultilingProfiler* (Finlayson et al., 2022), a vocabulary profiling tool optimized for French, German, and Spanish, in two ways: (1) to trial the feasibility of using the lists to write new material, and (2) to create lexical profiles of texts written for the current GCSE. To make the lists compatible with *MultilingProfiler*, we had to convert the headwords into bespoke units of counting following the inflectional and derivational grammar prescribed for each tier. This involved first expanding the headwords (including multiword phrases) into lemmas using *TreeTagger* (Schmid, 1994) and then removing inflected forms beyond the scope of the GCSE. In some cases, it was possible to remove whole (e.g. German present subjunctive) or partial (e.g. French plural imperfect) paradigms automatically

using the information in the tags, but irregular forms that were not listed as unique items had to be removed manually. Because derivational affixes can only be tested in reading, the units of counting are different for this skill. So, the final step was to create ‘reading-only’ versions of the lists by further expanding these bespoke lemmas into bespoke word families.

The testing phase brought some omissions (approximately 30 items per language) to our attention, most of which were mid-to-low-frequency ‘classroom’ items not included in the Routledge lists. Final adaptations were made accordingly.

4 Evaluation

To prepare the current word lists (Eduqas, 2019a, 2019b, 2019c) for analysis, we had to convert them into a format that could be compared with the new lists. Compiling definitive lists of headwords was not straightforward for three reasons. First, the current lists are structured by topic, so some headwords were listed multiple times and had to be merged. Second, the unit of counting was inconsistent; we found a mix of types, lemmas, flemmas, and multiword phrases, with some near-synonymous items listed together. We lemmatized the lists as far as possible and split multi-item headwords into their component parts. For example, *le goûter / le quatre-heures* (‘afternoon tea’), an entry on the French list comprising two near-synonyms, was separated into two lemmas. Similarly, *la boîte (de chocolats)* was divided into one lemma (‘box’) and one MWP (‘box of chocolates’). Finally, as function words required by the prescribed grammar are not itemized in the current subject content (DfE, 2022), we inferred them from Eduqas’ (2019a, 2019b, 2019c) grammar specifications. Our inferences were conservative in that we only included words that were unambiguously required. We did not make assumptions about items that might be covered by vague requirements like ‘common adverbial phrases’.

Lexical coverage plays an important role in the evaluation of corpus-informed word lists because of the close relationship between coverage and comprehension (see Section II.2). The more coverage a word list provides of a certain type of discourse, the more likely that list will help learners to comprehend that discourse (Schmitt et al., 2011). One of the primary purposes of our lists is, of course, to help students develop the vocabulary they need to pass the GCSE. In practice, the new lists are guaranteed to cover at least 96% of assessed reading material and 100% of listening material because the target language used in exams will be written to align with the lists. However, our lists should also help learners to understand other kinds of potentially relevant texts. Measuring lexical coverage of past exam texts (GCSE and A-level), young adult literature, and web language is an appropriate way of assessing the potential of the new and current, topic-driven lists for preparing learners to deal with authentic language. We note, again, that the current lists are not exhaustive; that is, they were not created with coverage in mind. However, teachers and textbook writers use them in much the same way as the new lists are intended (i.e. for structuring teaching and materials; see Marsden & Hawkes, 2023). Thus, the comparison is a timely and useful examination of the impact of the new policy.

To make the current lists compatible with lexical profiling tools, we had to expand the headwords into word families, as we did for the new lists (see Section III.4). The current content includes a broader and more extensive inflectional grammar than the new policy

and, thus, uses a different unit of counting. As this is a study of the impact of policy relating to vocabulary selection (rather than the cumulative impact of policy on both vocabulary and grammar), the unit of counting used in the coverage analysis must be constant. So, we expanded the headwords on the current lists following the parameters laid out in the new subject content made publicly available online on 14 January 2022.¹¹ However, the current subject content does not specify derivational morphology, so we did not carry out comparisons of the ‘reading-only’ lists. The analyses presented are for the word lists that are core to both listening and reading.

We carried out our coverage analysis using samples from the comparison corpora described in Section III.1 and *MultilingProfiler*, which can currently handle texts of up to 100,000 words. Our samples were (1) the full GCSE corpus (foundation and higher papers combined), (2) the full A/AS-level corpus, (3) 100,000 words of young adult literature (created from 10,000-word samples from each of the ten novels; three taken from the beginning of the text, four from the middle, and three from the end), and (4) 100,000 words of web text from the 2015 European Union web corpora (Goldhahn et al., 2012), as corpora in the TenTen Family are not available for download. Nation (2016) stresses that a fair lexical coverage study should use material taken from corpora other than the corpus from which the lists were made. As these adolescent-focussed corpora were used to verify (rather than make) the Routledge source lists, we feel our approach is justified.

IV Results of the evaluation study

Here, we present the findings from our comparative studies of list size and composition (Section IV.1) and lexical coverage (Section IV.2). In the tables and figures, the current and new lists are labelled by their years of initial implementation in teaching: 2016 and 2024, respectively. A breakdown of descriptive statistics by language can be found in Appendix D.

I List statistics

Research question 1 asked how the current and new lists compare in terms of (1) size, and (2) proportion of grammar and content words, single-word items and multiword phrases, and parts of speech. For this analysis, we lemmatized the lists by grouping irregularly inflected forms together with their headwords (i.e. converting the bespoke units of counting into partial lemmas) so that the figures could be compared with estimates of vocabulary size and coverage from previous studies that use lemmas (see Section II).

As is evident from Table 6, the new lists are much shorter, particularly at foundation tier.¹² The differences are especially striking when we consider that these figures are a conservative reflection of the number of words that could theoretically be encountered in an exam. Though the requirement to test ‘off-list’ words was removed as a legacy of Covid-19 adaptations (such that off-list vocabulary can now be glossed), it is inevitable that some such words continue to be included in exams because the current word lists are not comprehensive. Further, the size disparity would likely have been even greater had

Table 6. Number of lemmas in current and new GCSE word lists.

List		French	German	Spanish	Mean
2016	Foundation	1,976	1,789	1,752	1,839
	Higher	1,981	1,790	1,758	1,843
2024	Foundation	1,106 (-44.0%)	1,141 (-36.2%)	1,116 (-36.3%)	1,121 (-39.0%)
	Higher	1,563 (-21.1%)	1,592 (-11.1%)	1,589 (-9.6%)	1,581 (-14.2%)

Note. Percent length difference is shown in brackets.

Table 7. Mean (*SD*) grammar and content words in current and new GCSE word lists.

List		Grammar words		Content words*	
		Raw	As percentage of list	Raw	As percentage of list
2016	Foundation	153.7 (15.9)	.08	1685.3 (125.3)	.92
	Higher	158.0 (13.9)	.09	1685.0 (125.6)	.91
2024	Foundation	90.7 (2.9)	.08	1030.3 (18.9)	.92
	Higher	114.7 (20.1)	.07	1466.7 (23.7)	.93

Notes. *SD* = standard deviation. *The exemplar words required to illustrate each regular grammar pattern were included as content words, as they are selected by awarding organizations.

we used the bespoke unit of counting specified in the new policy, rather than partial lemmas. Unlike the new lists, the current lists do not itemize irregular forms, implying that *all* irregular forms can be tested. These forms would need to be listed as unique items under the new policy, further increasing the difference between the current lists and the new ones, which include only a limited number of irregular forms.

Table 7 compares the number of itemized grammar words in the new lists with the estimated (see Section III.2.e) numbers required by the compulsory grammar content in the current specifications. Note that we refer to these items as ‘grammar words’ to distinguish them from ‘function words’, as not all function words are required by grammar patterns, and some required words are arguably lexical (e.g. time adverbials). Overall, it seems that the effect of the policy changes on proportions of grammar and content words is minimal, though the difference in raw figures reflects a heavily reduced grammar content in the new policy, especially for German (see Table D1 in Appendix D). In contrast, we found that a substantially smaller proportion of items in the new lists are MWPs or compounds (Table 8). Note that the proportion of MWPs in the current lists varies between 15%–24% across languages (see Table D2 in Appendix D).

We also found a marked decrease in the proportion of nouns in the new lists and, relatedly, an increase in the proportion of every other part of speech (Figure 2). As redistributing part of speech proportions was not an explicit aim of the list creation process, these findings support the view that the topic-driven approach to vocabulary list development encourages a focus on nouns (Häcker, 2008), which may be one reason why Year 9 students of French and Spanish were found to produce a higher proportion of nouns in oral production than those in Year 13 (Marsden & David, 2008).

Table 8. Mean (*SD*) single-word items and multiword phrases/compounds in current and new GCSE word lists.

List		Single-word items		Multiword phrases/compounds	
		Raw	As percentage of list	Raw	As percentage of list
2016	Foundation	1472.7 (71.8)	.80	367.0 (99.6)	.20
	Higher	1475.0 (71.1)	.80	368.0 (99.5)	.20
2024	Foundation	1081.3 (23.8)	.96	39.7 (14.4)	.04
	Higher	1530.7 (22.7)	.97	50.7 (22.4)	.03

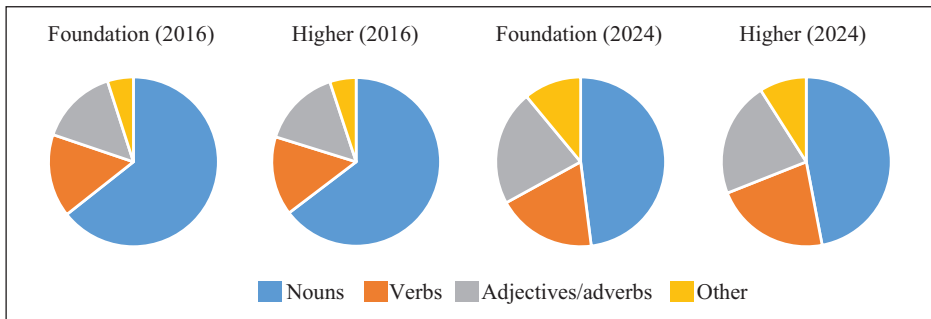


Figure 2. Part of speech breakdown in current and new GCSE word lists.

Note. For descriptives, see Table D3 in Appendix D.

In summary, the new lists (1) are substantially shorter than the current lists; (2) contain proportions of grammar words and content words that are similar to the current lists; (3) include far fewer MWPs; and (4) have more balanced part of speech proportions.

2 Lexical coverage

Research questions 2 and 3 ask what lexical coverage of material relevant to adolescents the current and new lists provide, and how much of this coverage can be attributed to grammar and content words, single-word items and multiword phrases, and different parts of speech.

Answers to research question 2 are summarized in Figure 3, which shows the means across languages. The headline finding is that despite being 36%–44% (foundation) and 11%–21% shorter (higher) than the current lists, the new, corpus-informed lists cover 9.1%–12.6% (foundation) and 14.7%–17.6% (higher) more of every genre, including the GCSE papers for which the current lists were developed. In line with other studies of lexical coverage (Section II.4.a), coverage by the German lists was a little lower on average than by the French and Spanish lists (see Table D4 in Appendix D). This was not the case for the GCSE exams, however, which suggests we may have been correct in our assumption that this cross-linguistic difference is (at least in part) attributable to erudite language typical of literature and, perhaps, some web articles. Impressively, coverage of (albeit

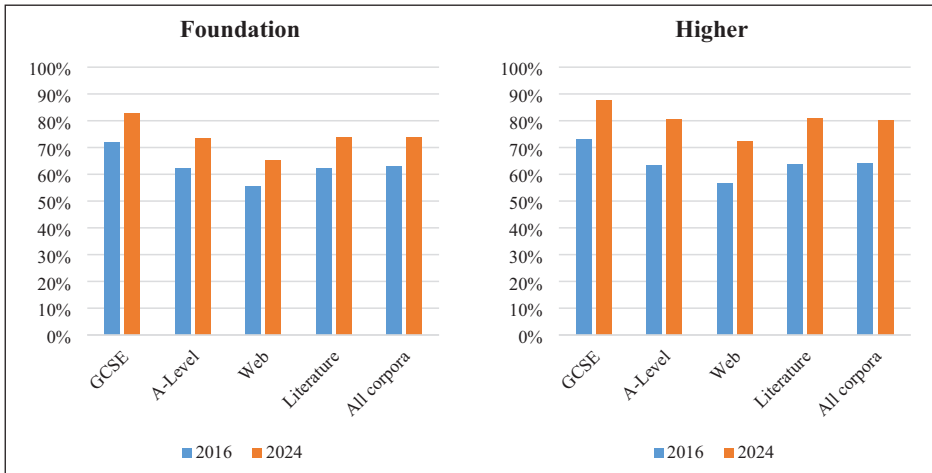


Figure 3. Lexical coverage by current (2016) and new (2024) lists.

Note. For descriptives, see Table D4 in Appendix D.

young adult) literature by the new higher lists for German (76%), Spanish (82%), and to a lesser extent French (82%) closely aligns with the values reported by Tschirner (2009), Davies (2005), and Rannäs (2019) for coverage of fiction by the 2,000 most frequent lemmas (Table 1). Given the new higher lists only comprise an average of just 1,581 lemmas, some of which have mid-to-low frequency, it seems our inclusion criteria did indeed result in the creation of a list highly relevant to adolescents. Coverage of literature and A-level exams by all lists was comparable to the mean across all corpora, with coverage of web language (the only genre not specifically written for adolescents) a little lower.

As it is well-established that function words account for the majority of lexical coverage, we wanted to explore whether the greater coverage by the new lists could be attributed to the items prescribed in the grammar specifications. To this end, we repeated the coverage analysis using separate lists of content and grammar words. Unsurprisingly, grammar words provide more coverage than content words across genres. However, this difference is notably more pronounced for the current lists compared to the new ones (Figure 4). The new lists perform considerably better in terms of content words, thus reducing this difference. Coverage by grammar words in the new lists is slightly lower across genres, but coverage of A-level texts, literature, and web language by content words more than doubles. This is remarkable when we consider that the new lists contain 41% (foundation) and 27% (higher) fewer grammar words and 39% (foundation) and 13% (higher) fewer content words than the current lists (Table 7). Clearly, the better performance by the new lists is unrelated to the grammar content, pointing to an effective approach to content word selection.

Our third analysis examined the effects of constraints on the number of items that can be MWPs or compound words. Interestingly, the substantial reduction (20% of the current lists compared with 3–4% of the new lists; Table 8) makes little to no difference in terms of coverage (Figure 5). MWPs on the current lists do cover slightly more of the GCSE corpus

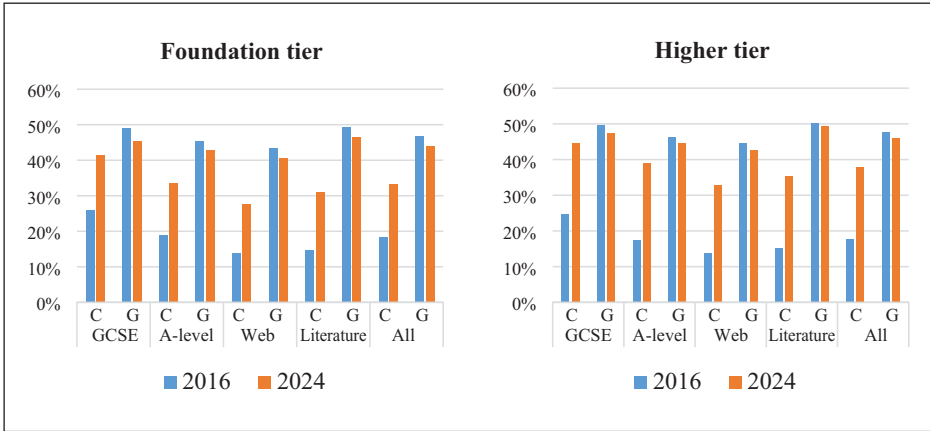


Figure 4. Lexical coverage by content (C) and grammar (G) words.
 Note. For descriptives, see Table D5 in Appendix D.

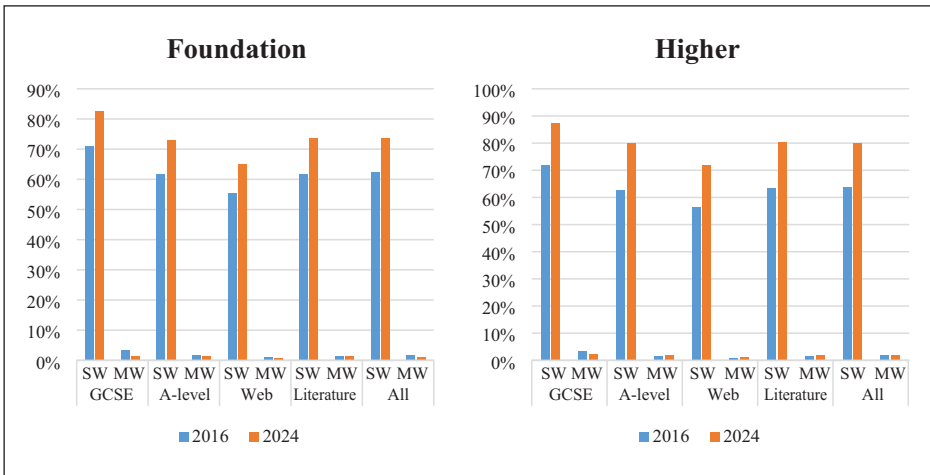


Figure 5. Lexical coverage by single-word items (SW) and multiword phrases/compounds (MW).
 Note. For descriptives, see Table D6 in Appendix D.

than other corpora, indicating that a small number have been included in exams that the lists were designed to support. We note that we may have very slightly underestimated the figures for MWPs because *MultilingProfiler* cannot recognize them when they are split by intervening words. For example, a listed MWP like *faire les magasins* (‘to go shopping’) would not be counted when split by an adverb, as in the case of *faire souvent les magasins* (‘to go shopping often’). However, we suspect that the difference is negligible.

Finally, we investigated implications of the change in part of speech distribution. Although the new lists include far fewer nouns (Table 8), these covered more than the

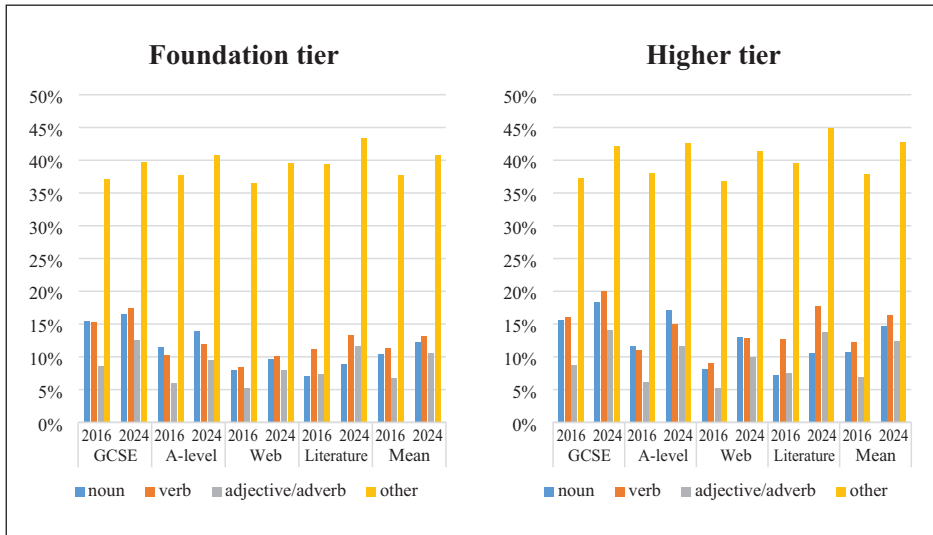


Figure 6. Lexical coverage by parts of speech.

Note. For descriptives, see Tables D7 and D8 in Appendix D.

larger number on the current lists, again pointing to a successful approach to content word selection. For each other part of speech, the more balanced distribution is reflected by increased coverage in the new lists (see Figure 6).

We wondered whether the observed differences might be influenced by the decision to use the bespoke counting unit described in the *new* policy, given that word selections for the current lists were (presumably) made with the *current* grammar content in mind. To check for this, we carried out two further analyses on the GCSE and young adult literature corpora using (1) bespoke counting units that comply with the grammar specified in the *current* subject content and (2) full lemmas. As shown in Tables D9 and D10 in Appendix D, the new lists cover far more of both genres, regardless of the unit of counting. Importantly, the considerably more extensive inflectional morphology in the current policy yields only negligible gains.

V Discussion

Returning to our broader aims – illustrating collaborative word list creation methods involving researchers, policymakers, teachers, and awarding organizations and testing the lists' coverage of adolescent-focussed material – we now reflect on some of our experiences of working with diverse stakeholders. We also consider implications of our evaluation study, present some pedagogical applications of the word lists, and discuss limitations of our approach. Finally, we draw implications of our work in the context of secondary schools in England for developing lexical components of French, German, and Spanish curricula in other beginner-to-low-intermediate settings.

1 Researchers and practitioners as partners in word list creation

Working with policymakers, teachers and awarding organizations has been mutually beneficial, improving our understanding of the context and educators' understanding of the research informing the lists (and policy). Achieving shared understanding was crucial for an initiative that could challenge established approaches. Knight et al. (2023, p. 1) note that the 'real-world challenge' of introducing a major new resource into a longstanding curriculum is doing so in a way that makes best use of existing resources for material and curriculum design while remaining transparent about principles driving the changes. To support practitioners who lack time and resources to engage with research but can be positively disposed to it (e.g. Marsden & Kasprovicz, 2017; Sato & Loewen, 2019), we ran professional development (e.g. Language-Driven Pedagogy, 2021b) for teachers and awarding organizations, including those not involved in the present study, about key concepts and practices in word list research. In these sessions, participants read non-technical summaries from the Open Accessible Summaries in Language Studies (OASIS) database (Marsden et al., 2018), completed activities to consolidate knowledge of technical terms like 'lemma', 'polysemy', and 'part-of-speech', and undertook training in profiling texts in *MultilingProfiler*. In turn, the sessions helped us understand what support might be needed for the introduction of compulsory word lists. Many of the resources discussed in Section V.3 were developed to help practitioners adapt existing materials and methods to situate them within the parameters of the new policy.

Since running the sessions, we have been providing ongoing support to the two awarding organizations who will offer the new GCSE – AQA and (indirectly via Ofqual and the DfE) Pearson Edexcel – as they prepared their accredited lists. AQA adopted similar methods to ours, selecting items from the Routledge lists and based on word-topic relatedness and incorporating feedback from teacher partners. Accordingly, overlap between the word lists presented here and AQA's accredited lists is very high: 82–93% across languages and tiers. Pearson Edexcel's accredited lists overlap with our examples by 73%–80%, though we know less about their approach. The high degrees of overlap suggest that our findings, and the implications for practice discussed in the following section, are relevant to both organizations' materials, potentially impacting hundreds of thousands of adolescent learners in England. We hope collaboration will continue, for example, on curriculum and resource development and professional development for teachers on effective use of word lists.

2 Implications of findings for practice

The mean size of the new lists – 1,121 (foundation) and 1,581 (higher) lemmas – aligns with the number of words typically known by (1) the highest performing students of GCSE French (Dudley et al., 2024), and (2) students working at CEFR levels A1 and A2 respectively (Section II.4.a). In contrast, the (conservatively estimated) size of the current lists is closer to the number of words known by B1 students, which seems an unrealistic target in the currently available curriculum time. One to two hours of contact time a week over five years is unlikely to be sufficient for teachers to provide the necessary conditions for reliable acquisition of over 1,800 items for both receptive and productive

use. On the other hand, repeated exposure to 1,121 and 1,581 items through language-focussed, meaning-focussed, and fluency-based activities seems feasible. Further, the new lists address an issue of seemingly higher expectations for current students of French, as there are around 200 more items in the current lists for French than for German and Spanish (Table 6). Considering the other inconsistencies observed across languages and tiers to date (Section I), it seems that the new requirements may help increase cross-language parity in this regard.

The new higher lists provide very high (87.9%) coverage of corpora of GCSE materials. Allowing for unlisted proper nouns, cognates and derived forms in reading materials, it seems likely that a student with a good grasp of words on the new higher list would be able to understand most of the language in a current exam, working to the 90% coverage proposed as necessary for adequate comprehension of written materials (though see the caveats to this figure discussed Section II.2). The slightly lower average coverage by the new foundation lists (82.5%) is to be expected, given the corpus includes papers from both tiers. As noted earlier, the new lists will be used to inform the writing of exams, so their coverage of papers for 2026 onwards will be much higher. Importantly, though, coverage figures for the other corpora demonstrate the value of the lists beyond the exam. Higher-tier students would likely be well prepared to commence A-level study and could, with some inferencing and the appropriate tools, feasibly tackle age-appropriate fiction. This has implications for materials developers and test writers in that young adult literature may need relatively little adaptation to be suitable for use with GCSE students, paving the way for the use of more (adapted) authentic texts as curriculum content.

The very low coverage provided by MWPs compared to single-word items raises questions about curriculum design and pedagogy choices on the balance between phrase-based and single-word teaching. Single-word items cover more because they appear in many kinds of structures and have a broad semantic range (Section II.4.b), a flexibility unlikely to be captured through the rote learning of many subjectively chosen ‘set’ phrases. Considering words as ‘nodes’ with a few prototypical meanings that sit in longer meaning units is likely to better prepare students to recognize the words in different contexts. There are, of course, cases where phrase-based teaching is important, especially in the initial stages where students do not have the knowledge and skills necessary to break down important MWPs into their component parts. Examples include the French *s’il vous / te plaît* (‘please’), in which a form of *plaire* (‘to please, delight’) is frequent on account of the specific meaning conveyed by the phrase, and the German *ich möchte* (‘I would like’), in subjunctive mood. Still, our findings suggest the types of phrases included in GCSE lists – often highly-specific, topic-bound concepts such as *Fernfahrer* (‘long-distance lorry driver’) and *lotería nacional* (‘national lottery’) – provide little boost to comprehension, and that the ‘spots’ most of these phrases occupy are more usefully allocated to single-word items.

Our analysis of coverage by grammar and content words questions the inclusion of an extensive grammar content at beginner levels, given that (1) the substantially fewer grammar words in the new content provide almost as much coverage as those in the current content, and (2) the results in Tables D9 and D10 (Appendix D) suggest that vastly constraining the required inflectional morphological patterns, especially at foundation

(for an overview of required grammar in current versus new policy, see Language-Driven Pedagogy, 2023a), is not to the detriment of comprehension. Rather, reducing the focus on infrequent, complex grammatical forms could increase the resources that can be invested in other aspects of language development.

The more balanced distribution of parts of speech in the new lists seems to be desirable in a core word list for beginner-to-low-intermediate learners. While some low-frequency topical nouns are a crucial part of a personalized lexical repertoire, it does not seem efficient to include them in large quantities on a core list for all. Notably, the considerable reduction in the proportion of nouns in the new lists did not decrease their coverage, whereas the increased proportions of verbs, adjectives, and adverbs resulted in more coverage by each. This is, perhaps, indicative of a shortcoming of the current lists in equipping students with vocabulary for comprehending a range of genres, especially those that are more descriptive (e.g. narrative fiction).

The potential implications discussed here assume that word lists have washback effects on teaching, learning, and curriculum design practices. Of course, the extent and nature of this washback depend on factors such as teachers' and learners' perceptions of tests, how teachers respond to test changes, and the influence of tests on instructional materials (McKinley & Thompson, 2018). We hope that any washback effects induced by the new GCSE lists will be positive, given that the lists reflect general language use, complement the curriculum, and inform exam content. Under current policy, the potential washback effects of lists are perhaps less clear, as word lists had not been used directly to develop exams. Still, the current lists have influenced textbook content, affecting vocabulary selection and sequencing (Marsden, Dudley, & Hawkes, 2023).

3 Applications of the word lists

Word lists can be used to create and adapt materials and tests that align with learner knowledge, and developers of such materials would be more likely to embrace this approach if it could be made easier to implement. As Marsden et al. (2023) note, one reason for the underuse of the current GCSE word lists to date has been the lack of lexical profiling tools – and, we add, the lack of versions of the lists expanded into bespoke families – available for awarding organizations and materials developers to use to check alignment.

The lists described in this study are available from our OSF repository (<https://osf.io/5cxhq/>) and IRIS (<https://doi.org/10.48316/ikldP-BFsdt>) in versions compatible with any lexical profiling tool, as a headword-only version at Language-Driven Pedagogy (2024), and as embedded files in *MultilingProfiler* alongside the AQA and Pearson Edexcel accredited lists ('GCSE: LDP', 'GCSE: AQA', and 'GCSE: Edexcel' in the 'List type' dropdown, respectively). Listening (bespoke lemmas) and reading (bespoke word families) versions of both the foundation and higher lists are available for all three list types. An advantage of using *MultilingProfiler* to analyze texts for the GCSE (or any course aligned with CEFR A1–B1) is that its features are designed to support accurate profiling of texts for beginner-to-low-intermediate learners (Finlayson et al., 2023). Users can (1) add proper nouns, glossed words, and cognates to an 'Extended' version of the list to include them in the profile, (2) automatically count and highlight words that

UNIVERSITY of York

Home **MultilingProfiler** Word Families FAQ About Contact

MultilingProfiler

Select the *list type* and the related options (if any) you want to use to profile your text.

List type: GCSE: LDP list Language: Spanish Tier: Foundation Modality: Reading

Extended List: Guti, Alicia

Profile window:

La religión no es importante para algunos jóvenes, pero yo soy **católico** y será **maravilloso** tener una **ceremonia** religiosa en la iglesia de mi pueblo. Me llevo genial con los padres y hermanos de Alicia y su abuelo es **encantador**. Me **fastidian** sus tíos y primos. Prefiero no invitarlos. (Guti)

Mis padres quieren celebrar nuestra boda con una fiesta **enorme**, pero como soy **tímida** eso será muy **incómodo**. Prefiero invitar a poca gente. En muchas bodas los **invitados** toman demasiado **alcohol**, se **emborrachan** y se **comportan** mal. En nuestra cena vamos a tomar bebidas sin **alcohol**. (Alicia)

Enable highlighting of cognates with English (in blue)
[Includes false cognates.]

Disable colour highlighting (recommended for very large texts >50,000 words)
[Orange indicates words that are not in your chosen list. Visit the [FAQ](#) page for more information.]

Profile Text

Hide Extended List Copy Results Download Stats (.csv)

Global Coverage	
Total number of words covered by all lists (including multi-word units)	86 88.7%

Word Statistics	
Total number of words in the text from your chosen list	81 83.5%
Total number of words in the text from your extended list	3 3.1%
Total number of cognates from outside your chosen list	2 2.1%
Total number of words in the text (chosen list + extended list + cognates)	86 88.7%
Total number of words in the text	97

Figure 7. Profile of a text in a GCSE Spanish foundation reading paper (AQA, 2020) using the relevant new list in *MultilingProfiler*.

are complete cognates (i.e. same spelling, ignoring diacritics) with English, (3) identify false cognates, and (4) adapt texts in the ‘Profile window’ if desired. The text in Figure 7 would need slight adjustment for use in a test aligned with the selected foundation reading list, which covers 88.7% of the content (when the 2% of cognates permitted in reading are included). To adapt it, 2% of the orange ‘off-list’ words could be glossed, and others replaced with on-list alternatives. Alternatively, the text could serve for inferencing or dictionary practice. AQA (2023) and Pearson Edexcel (2024) created accredited sample assessment materials using *MultilingProfiler* and this approach.

Another use of lists is in curriculum design. Language-Driven Pedagogy (2023b; Marsden & Hawkes, 2023) have shown how the items on the lists described in this study can

be introduced and revisited on a 0–3–9-week practice schedule, embedded in wider schemes of work over five years. New words are introduced at a rate of approximately 10 words per week (increasing to 15 words per week for higher tier in the last two years¹³) and revisited three and six weeks later. Very similar schemes of work for AQA and Pearson Edexcel syllabi are in development as part of ongoing work by the Oak National Academy (Harrison, 2024), and these are accompanied by cumulative word lists representing the lexical (and lexico-morphological) content introduced each week. Teachers and materials developers can use these lists to create syllabus-aligned activities and achievement tests and use authentic texts at appropriate points in the curriculum (Figure 8a and b). Adding a given week's word list to a flashcard app like *Quizlet* or *Anki* creates bespoke practice activities.

4 Limitations of our concept and approach

We have reported on a novel study carried out by a relatively small team across three languages, within the time and budget constraints of a two- to three-year, government-funded initiative. While creating the ideal resources from scratch was not always feasible, we made efforts to adapt existing data and tools to suit our purposes. Some limitations, however, were unavoidable. We discuss these here and suggest how they could be mitigated in future work.

Our source lists were derived from general corpora of texts considered representative of language likely to be encountered in various situations by a wide range of users, arguably mainly adults (of all ages). As with any general frequency lists derived from a single corpus (or a selection of subcorpora), the 'flavour' of corpus content started becoming apparent after the 1,000 most frequent words. In our case, some words in our initial selection pool may have been determined by the inclusion of academic textbooks and legal proceedings, perhaps at the expense of more items of potential relevance to our learners. As discussed in Section III, we addressed this possible limitation by removing less relevant words using inclusion criteria and human scrutiny, following the approaches of West (1953), Nation (2016), Knight et al. (2023) and others. An alternative approach could be to build a set of comparable corpora specifically for adolescent learners. This would require an extensive needs analysis and the collection of data from multiple sources, for example, narrative and audiovisual materials written for adolescents, spoken conversation, social media samples, and learning materials and tests. Comparable word lists could be generated to address potential effects of different dispersion measures and other issues relating to comparability of source corpora (see Section III.1). Access to a larger corpus of exam papers would also increase our confidence about the rank order of words in that genre, as raw frequencies dropped to single figures after around the 1,500 most frequent words.

Assigning words to semantic categories is an inherently subjective process. Our ratings are based on the judgments of just two or three individuals per language, all young professional linguists based in the UK. The similarity of their backgrounds may account for the very high inter-rater reliability scores observed. In our case, word-topic relatedness was just one of several soft selection criteria, rather than a hard cut-off. For studies using categorization as a hard criterion, a greater number of raters from diverse backgrounds is recommended for more reliable and generalizable results. If rating long lists, raters could work on subsections to reduce fatigue and drop-out.

(a) Home **MultilingProfiler** Word Families FAQ About Contact

MultilingProfiler

Select the *list type* and the related options (if any) you want to use to profile your text.

List type: LDP: KS3 list Language: German Year: Year 8 Term: Term 1.1 Week: Week 2

For accurate results, split compounds in your texts by adding a space between the words of which they consist (e.g., *Sommerferien* → *Sommer Ferien*).

Extended List

Rotkäppchen

Profile window


Wie **nun** Rotkäppchen in den **Wald kam**, **begegnete ihm der Wolf**. Rotkäppchen aber **wusste nicht**, was das **für ein böses Tier war**, und **fürchtete sich nicht vor ihm**. "Guten Tag, Rotkäppchen!" sprach er. "Schönen **Dank, Woll!**" - "Wo **hinaus so früh**, Rotkäppchen?" - "Zur **Großmutter**." - "Was **trägst du unter der Schürze?**" - "Kuchen und **Wein**. Gestern haben wir **gebacken**, da soll sich die **krankte und schwache Großmutter** etwas **zugut tun** und sich **damit stärken**." - "Rotkäppchen, wo **wohnt deine Großmutter?**" - "Noch eine **gute Viertelstunde weiter** im **Wald**, **unter den drei großen Eichbäumen**, da **steht ihr Haus**, **unten sind die Nusshecken**, **das wirst du ja wissen**," **sagte Rotkäppchen**.

Enable highlighting of cognates with English (in blue)
[Includes false cognates.]

Disable colour highlighting (recommended for very large texts >50,000 words)
[Orange indicates words that are not in your chosen list. Visit the [FAQ](#) page for more information.]

Profile Text

Hide Extended List Copy Results

(b)  UNIVERSITY of York

Home **MultilingProfiler** Word Families FAQ About Contact

MultilingProfiler

Select the *list type* and the related options (if any) you want to use to profile your text.

List type: LDP/OAK: KS4 AQA list Language: German Year: Year 10 Term: Term 3.1 Week: Week 2(H)

For accurate results, split compounds in your texts by adding a space between the words of which they consist (e.g., *Sommerferien* → *Sommer Ferien*).

Extended List

Rotkäppchen

Profile window

Wie **nun** Rotkäppchen in den **Wald kam**, **begegnete ihm der Wolf**. Rotkäppchen aber **wusste nicht**, was das **für ein böses Tier war**, und **fürchtete sich nicht vor ihm**. "Guten Tag, Rotkäppchen!" sprach er. "Schönen **Dank, Woll!**" - "Wo **hinaus so früh**, Rotkäppchen?" - "Zur **Großmutter**." - "Was **trägst du unter der Schürze?**" - "Kuchen und **Wein**. Gestern haben wir **gebacken**, da soll sich die **krankte und schwache Großmutter** etwas **zugut tun** und sich **damit stärken**." - "Rotkäppchen, wo **wohnt deine Großmutter?**" - "Noch eine **gute Viertelstunde weiter** im **Wald**, **unter den drei großen Eichbäumen**, da **steht ihr Haus**, **unten sind die Nusshecken**, **das wirst du ja wissen**," **sagte Rotkäppchen**.

Enable highlighting of cognates with English (in blue)
[Includes false cognates.]

Disable colour highlighting (recommended for very large texts >50,000 words)
[Orange indicates words that are not in your chosen list. Visit the [FAQ](#) page for more information.]

Profile Text

Hide Extended List Copy Results

Global Coverage

Total number of words covered by all lists (including multi-word units)	88	85.4%
---	----	-------

Figure 8. (a) Profile in *MultilingProfiler* of a sample text from *Rotkäppchen* ('Little Red Riding Hood') with the Oak National Academy AQA-aligned list for the start of Year 8, Term 1. (b) Profile in *MultilingProfiler* of a sample text from *Rotkäppchen* ('Little Red Riding Hood') with the Oak National Academy AQA-aligned list for the start of Year 10, Term 3 in *MultilingProfiler*.

There are three caveats regarding the *MultilingProfiler* output. First, we did not use a stop list, which means that proper nouns are included in the profiles and so coverage by the lists is very slightly underestimated (though *MultilingProfiler* does automatically exclude numerical figures and symbols from analysis). Second, *MultilingProfiler* does not currently perform semantic or part-of-speech tagging, so unlisted words sharing orthographic forms with listed words are included in the profile. This may very slightly overestimate coverage; for example, any instances of the French noun *pas* ('footstep') would have been added to the coverage for the high-frequency adjective *pas* ('not'). Finally, compounds like *Sommerferien* ('summer holidays') are not recognized unless they (or all their component parts) are listed words. Compound splitting software (e.g. *SMOR*, Schmid et al., 2004; *CharSplit*, Tuggener, 2016) may be embedded in future versions.

As with all coverage research, arguments about comprehension that draw on coverage rely on the assumption that students can recognize every target meaning of every listed word. In practice, this is unlikely to be the case, especially for beginner-to-low-intermediate learners. Further, as noted in Section II.2, the coverage construct does not consider other factors related to comprehension (e.g. test-related variables, phonological skills, syntactic knowledge, inferencing skills, educational factors, and individual differences), so evaluation studies based on coverage alone cannot provide a complete picture. Still, given the relationships that have been observed elsewhere between vocabulary knowledge and many of these factors, a high-coverage word list is likely to be a suitable source of words that can be used in curriculum and materials design to develop these (and other) areas of language proficiency.

VI Conclusions

In sum, the methods reported in this article have been successful in producing new, corpus-driven word lists for testing French, German, and Spanish that (1) operationalize new policy requirements, (2) are shorter and more powerful in terms of coverage than the current, topic-driven lists developed for similar purposes, and (3) are broadly replicable and adaptable for use in other contexts. Importantly, our methods have been adapted by awarding organizations who will offer high-stakes national examinations in these languages from 2026, with the result that their accredited word lists are similarly concise and should provide equally high coverage of various text types with potential relevance to adolescents. It is too early to speculate over the potential implications of a more structured lexis and better-aligned teaching and testing practices for motivation and exam performance. Nevertheless, it is likely that principles underpinning the new policy and lists will washback into classroom practice, and early impact has been observed in surveys of teacher perceptions of the new GCSE. Secondary school teachers rank vocabulary as the top priority in planning for the new curriculum (Collen, 2023); 25% of teachers expect the new GCSE to have a positive impact on GCSE uptake, with others neutral on this (Collen & Duff, 2024); and 80% of primary school teachers would welcome similar language-specific lists of minimum vocabulary and grammar (Collen & Duff, 2024).

Because they are language driven, our list creation methods are broadly replicable and can be adapted for use in different contexts (e.g. other languages, proficiency levels, and L1s). Corpora of the sizes used in this study are relatively straightforward to build if

electronic materials can be obtained. Words can be assigned to any relevant semantic category using the methods we have described, and teacher instructions can be adapted to elicit more specific feedback. *MultilingProfiler* and *AntWordProfiler* support lexical profiling with custom word lists and texts in any language (though are optimized for French, German, and Spanish, and English, respectively). To develop bespoke training for list creators, summaries of language research from different contexts can be selected from OASIS (<https://oasis-database.org>).

We hope that our study inspires further work on developing corpus-informed word lists for beginner or low-intermediate learners of languages other than English, particularly in low-exposure environments. As we have seen, this is a neglected area of research, even though word lists have informed teaching and testing in this context for decades.

Acknowledgements

We would like to thank the following colleagues at the former National Centre for Excellence for Language Pedagogy (now Language-Driven Pedagogy) for their invaluable support and contributions: Lauren Smith and Elin Graves (Research Assistants) for their meticulous work in preparing and organizing the exam papers corpus, compiling word lists for comparative study, and leading on lexical profiling; Morag Kewell and Francis Flynn (Research Assistants) for their assistance in preparing the exam papers and support with lexical profiling; Amber Dudley (Research Associate) for calculating overlap between the lists in this study and the accredited lists from AQA and Pearson Edexcel, as well as for serving as a rater; and Inge Alferink, Nick Avery, Louise Bibbey, Louise Caruso, Amanda Izquierdo, Charlotte Moss, Catherine Salkeld, and Peter Watson (Resource Developers) for their roles as raters and for their work in identifying stable lexical items across corpora and contributing to the final selection of lexical items.

We are also grateful for the valuable insights provided by our teacher collaborators, and the contributions of our partners at Eduqas/WJEC, who played an integral role in the development and testing of the lists.

Lastly, we are grateful to the three anonymous reviewers for their thoughtful feedback, which greatly enhanced this article, and to Professor María del Pilar García Mayo and the *Language Teaching Research* editorial team for ensuring a smooth and supportive publication process.

Author contribution statement

Natalie Finlayson: conceptualization (lead); methodology (lead); validation (lead); formal analysis (lead); investigation (lead); resources (lead); data curation (lead); writing – original draft (lead); writing – review & editing (lead); visualization (lead); supervision (supporting); project administration (lead). Emma Marsden: conceptualization (lead); methodology (supporting); formal analysis (supporting); investigation (supporting); resources (lead); writing – original draft (supporting); writing – review & editing (lead); visualization (supporting); supervision (lead); project administration (lead); funding acquisition (lead). Rachel Hawkes: conceptualization (lead); methodology (supporting); investigation (supporting); resources (lead); writing – review & editing (supporting); supervision (supporting); project administration (supporting); funding acquisition (supporting).

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for the research, authorship, and publication of this article was provided by (1) the Department for Education for England awarded to the former

National Centre for Excellence for Language Pedagogy (2018–2023; MFL Hubs, 30365) and to Emma Marsden at the University of York (2023–2024); (2) Research England (York2019); (3) the Higher Education Innovation Funding and the Economic and Social Research Council Impact Acceleration Account (York2018); and (4) the University of York.

ORCID iDs

Natalie Finlayson  <https://orcid.org/0000-0002-9918-2371>

Emma Marsden  <https://orcid.org/0000-0003-4086-5765>

Notes

1. At higher tier, passing students are awarded a mark in the grade bracket 4–9, where 9 is the highest. At foundation tier, the bracket is 1–5, with a grade of 4 required to pass.
2. For reasons unrelated to vocabulary lists, Eduqas withdrew from the Ofqual accreditation process just prior to the current version of this article. They are currently working towards accreditation to offer the new ‘Made-for-Wales’ GCSE offered by Qualifications Wales. Under the Welsh policy, awarding organizations must provide core lists based on corpus frequency and everyday language and include words that are transferable across themes (Qualifications Wales, 2024).
3. ‘Subject content’ documents are issued by the DfE, laying out policy on expected knowledge, understanding, and skills for developing testing specifications for the named subject.
4. This requirement was removed during and after COVID19, though unlisted words can be (and are) still used in exams.
5. See report from the American Council on the Teaching of Foreign Languages (2024) on the correspondence between their assessments and CEFR ratings.
6. Raw frequency divided by Gries’ (2008) DP for French; Juilland’s D for Spanish; lexical range for German.
7. Enough to reflect the distribution of linguistic features in a register (Biber, 1993).
8. Students in England study for the Advanced (A-) level school-leavers exam (CEFR B1–B2) from ages 16–18 years. They can opt to take the AS-level after the first year.
9. ARF discounts multiple occurrences of words that appear closely together to prevent over-representation from a small number of texts (Savický & Hlaváčová, 2002).
10. To our knowledge, this is the only resource that has arranged every recorded word in a language into detailed hierarchies of meaning.
11. These lists are not exactly the same as the embedded versions in the *MultilingProfiler*. The latter incorporate some very minor potential changes (e.g. to the definition of irregular verbs in French), which, at the time of writing, were still being reviewed by the DfE and Ofqual. These differences would not meaningfully change our findings.
12. There is negligible difference between tiers in the current lists because the same topic-driven vocabulary lists are used with both. A few extra function words required for the higher grammar specifications account for the small differences.
13. Curriculum time can sometimes increase to (almost) three hours per week in the latter stages of GCSE courses.

References

- American Council on the Teaching of Foreign Languages. (2024). *Assigning CEFR ratings to ACTFL assessments*. American Council on the Teaching of Foreign Languages. Available at: https://www.actfl.org/uploads/files/general/Assigning_CEFR_Ratings_To_ACTFL_Assessments.pdf (accessed October 2024).

- Antes, T.A. (2023). A general service list for French? Teaching the vocabulary that matters. *Language Teaching*, 56, 570–573.
- Anthony, L. (2024). *AntWordProfiler: Version 2.2.1* [computer software]. Waseda University. Available at: <https://www.laurenceanthony.net/software> (accessed October 2024).
- AQA. (2020). GCSE Spanish Foundation Tier Paper 3 Reading. Available at: <https://cdn.sanity.io/files/p28bar15/green/52fc7a82bbfd7c22d82aafeb4a8c96f54a5b220b.pdf> (accessed November 2024).
- AQA. (2023). GCSE French 8652 Assessment resources. Available at: <https://www.aqa.org.uk/subjects/languages/gcse/french-8652/assessment-resources> (accessed October 2024).
- Baudot, J. (1993). *Fréquence d'utilisation des mots en français écrit contemporain [Frequency of use of words in contemporary written French]*. Presses de l'université de Montréal.
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6, 253–279.
- Beacco, J.-C., Lepage, S., Porquier, R., & Riba, P. (2008). *Niveau A2 pour le français—Un référentiel*. Editions Didier.
- Beacco, J.-C., & Porquier, R. (2007). *Niveau A1 pour le français – Un référentiel*. Editions Didier.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8, 243–257.
- Bonazzi, A. (2017). *Grundwortschatz Deutsch: Ein Korpusansatz [Core German Vocabulary: A Corpus Approach]*. Unpublished masters dissertation, Technische Universität Dresden. Available at: https://www.academia.edu/32091050/Grundwortschatz_Deutsch_Ein_Korpusansatz_Core_German_Vocabulary_A_Corpus_Approach (accessed November 2024).
- Brezina, V. (2018) *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36, 1–22.
- Browne, C., Culligan, B., & Phillips, J. (2023). The New General Service List 1.2. Available at: <https://www.newgeneralservicelist.org> (accessed November 2024).
- Brysbuert, M., Keuleers, E., & Mandera, P. (2021). Which words do English non-native speakers know? New supranational levels based on yes/no decision. *Second Language Research*, 37, 207–231.
- Caltabellotta, A., Van Steendam, E., Noreillie, A.-S., & Peters, E. (2024, March 19). *Vocabulary knowledge and vocabulary use: A cross-sectional comparison of L2 English and French* [paper presentation]. AAAL 2024.
- Capel, A. (2010). A1–B2 vocabulary: Insights and issues arising from the English Profile Wordlists project. *English Profile Journal*, 1, e3.
- Chung, T.M., & Nation, I.S.P. (2004). Identifying technical vocabulary. *System*, 32, 251–263.
- Churchward, D. (2019). *Recent trends in modern foreign language exam entries in anglophone countries*. Ofqual. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/844128/Recent_trends_in_modern_foreign_language_exam_entries_in_anglophone_countries_-_FINAL65573.pdf (accessed October 2024).
- Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991–1013.
- Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15–38). John Benjamins.
- Cobb, T., & Laufer, B. (2021). The nuclear word family list: A list of the most frequent family members, including base and affixed words. *Language Learning*, 71, 834–871.
- Cobb, T., Lindqvist, C., & Ramnäs, M. (2023). Enfin! Une liste de vocabulaire française fondée sur les familles de mots [Finally! A French vocabulary list based on word families] [paper

- presentation]. *Association Canadienne-Française pour l'avancement des sciences (ACFAS) 2023*. Available at: https://lexutor.ca/acfas_lex_23/liste_fams.pdf (accessed October 2024).
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Routledge.
- Collen, I. (2022). *Language trends 2022*. British Council. Available at: <https://www.britishcouncil.org/research-insight/language-trends-2022> (accessed October 2024).
- Collen, I. (2023). *Language trends 2023*. British Council. Available at: <https://www.britishcouncil.org/research-insight/language-trends-england-2023> (accessed October 2024).
- Collen, I., & Duff, J. (2024). *Language trends England 2024*. British Council. Available at: <https://www.britishcouncil.org/research-insight/language-trends-england-2024> (accessed October 2024).
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Crossley, S., Heintz, A., Choi, J.S., Batchelor, J., Karimi, M., & Malatinszky, A. (2023). A large-scaled corpus for assessing text readability. *Behavior Research Methods*, 55, 491–507.
- Curcin, M., & Black, B. (2019). *Investigating standards in GCSE French, German and Spanish through the lens of the CEFR*. Ofqual. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/844034/Investigating_standards_in_GCSE_French_German_and_Spanish_through_the_lens_of_the_CEFR.pdf (accessed October 2024).
- Dang, T.N.Y. (2018). A hard science spoken word list. *International Journal of Applied Linguistics*, 169, 44–71.
- Dang, T.N.Y. (2019). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 288–303). Routledge.
- Dang, T.N.Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, 67, 959–997.
- Dang, T.N.Y., & Webb, S. (2016). Making an essential word list for beginners. In I.S.P. Nation (Ed.), *Making and using word lists for language learning and testing* (pp. 153–167). John Benjamins.
- Dang, T.N.Y., Webb, S., & Coxhead, A. (2022). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*, 26, 617–641.
- Davies, M. (2005). Vocabulary range and text coverage: Insights from the forthcoming Routledge frequency dictionary of Spanish. In D. Eddington (Ed.), *Selected proceedings of the 7th Hispanic Linguistics Symposium* (pp. 106–115). Cascadilla Proceedings Project.
- Davies, M., & Davies, K.H. (2018). *A frequency dictionary of Spanish: Core vocabulary for learners* (2nd ed.). Routledge.
- David, A. (2008). Vocabulary breadth in French L2 learners. *The Language Learning Journal*, 36, 167–180.
- Department for Education (DfE). (2022). *GCSE modern foreign languages*. Department for Education. Available at: <https://www.gov.uk/government/publications/gcse-modern-foreign-languages> (accessed October 2024).
- Department for Education (DfE). (2023). *GCSE French, German and Spanish subject content*. Department for Education. Available at: <https://www.gov.uk/government/publications/gcse-french-german-and-spanish-subject-content> (accessed October 2024).
- Dudley, A., & Marsden, E. (2024). The lexical content of high-stakes national exams in French, German, and Spanish in England. *Foreign Language Annals*, 57, 311–338.
- Dudley, A., Marsden, E., & Bovolenta, G. (2024). A context-aligned two thousand test: Toward estimating high-frequency French vocabulary knowledge for beginner-to-low intermediate proficiency adolescent learners in England. *Language Testing*, 41, 759–791.
- Educalibre. (2019). Descargar PDF libros juveniles [Download PDF youth books]. Educalibre. Available at: <https://educalibre.info/descargar-pdf-libros-juveniles/> (accessed March 2023).

- Education Bureau. (2023). *Preamble to the development of the wordlists for the English language curriculum*. Education Bureau. Available at: https://www.edb.gov.hk/en/curriculum-development/kla/eng-edu/references-resources/Wordlists_preamble.html (accessed October 2024).
- Eduqas. (2019a). *WJEC Eduqas GCSE in French: Version 2*. Eduqas. Available at: https://www.eduqas.co.uk/qualifications/french-gcse/#tab_keydocuments (accessed October 2024).
- Eduqas. (2019b). *WJEC Eduqas GCSE in German: Version 2*. Eduqas. Available at: https://www.eduqas.co.uk/qualifications/german-gcse/#tab_keydocuments (accessed October 2024).
- Eduqas. (2019c). *WJEC Eduqas GCSE in Spanish: Version 2*. Eduqas. Available at: https://www.eduqas.co.uk/qualifications/spanish-gcse/#tab_keydocuments (accessed October 2024).
- Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing*, 30, 253–272.
- Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, 64, 365–414.
- Ellis, N. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–188.
- Estivalet, G.L., & Meunier, F.E. (2015). Decomposability and mental representation of French verbs. *Frontiers in Human Neuroscience*, 9, 1–10.
- Finlayson, N., Marsden, E., & Anthony, L. (2022). *MultilingProfiler: Version 3* [software]. Available at: <https://www.multilingprofiler.net> (accessed October 2024).
- Finlayson, N., Marsden, E., & Anthony, L. (2023). Introducing MultilingProfiler: An adaptable tool for analysing the vocabulary in French, German, and Spanish texts. *System*, 118, 103122.
- Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 137–156). John Benjamins.
- Gilner, L. (2011). A primer on the General Service List. *Reading in a Foreign Language*, 23, 65–83.
- Glaboniat, M., Müller, M., Rusch, P., Schmitz, H., & Wertenschlager, L. (2005). *Profile Deutsch: A1–C2: Version 2.0*. Langenscheidt.
- Glaboniat, M., Perlmann-Balme, M., & Studer, T. (2016). *Goethe-Zertifikat B1. Deutschprüfung für Jugendliche und Erwachsene. Wortliste [Goethe Certificate B1: German Exam for Young People and Adults. Word List]*. Goethe-Institut.
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In N. Calzolari, K. Choukri, T. Declerck, et al. (Eds.), *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)* (pp. 759–765). European Language Resources Association (ELRA).
- González-Fernández, B., & Webb, S. (2024). How well are primary and secondary meanings of L2 words acquired? *Studies in Second Language Acquisition*, 46, 818–840.
- Gougenheim, G., Michéa, R., Rivenc, P., & Sauvageot, A. (1967). *L'élaboration du français fondamental (1er degré) [The development of basic French (1st level)]*. Didier.
- Gries, S. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13, 403–437.
- Häcker, M. (2008). Eleven pets and 20 ways to express one's opinion: The vocabulary learners of German acquire at English secondary schools. *The Language Learning Journal*, 36, 215–226.
- Hamada, M. (2014). The role of morphological and contextual information in L2 lexical inference. *Modern Language Journal*, 98, 992–1005.

- Harrison, A. (2024, March 27). *Introducing our languages and MFL curriculum partner*. Oak National Academy. Available at: <https://www.thenational.academy/blog/introducing-our-languages-and-mfl-curriculum-partner> (accessed October 2024).
- Hawkins, E. (1996). *Thirty years of language teaching*. CILT.
- Herman, E., & Leeser, M.J. (2022). The relationship between lexical coverage and type of reading comprehension in beginning L2 Spanish learners. *The Modern Language Journal*, 106, 284–305.
- Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *The Language Learning Journal*, 36, 153–166.
- Horst, M. (2013). Mainstreaming second language vocabulary acquisition. *Canadian Journal of Applied Linguistics*, 16, 171–188.
- Hunston, S., & Francis, G. (2000). *Pattern grammar*. John Benjamins.
- Hunston, S., Francis, G., & Manning, E. (1996). *Collins COBUILD Grammar Patterns 1: Verbs*. HarperCollins.
- Hunston, S., Francis, G., & Manning, E. (1998). *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. HarperCollins.
- In'nami, Y., Koizumi, R., Jeon, E., & Arai, Y. (2022). L2 listening and its correlates. In E. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 235–283). John Benjamins.
- Instituto Cervantes. (n.d.). Niveles de referencia. Índice [Reference levels. Index.]. Available at: https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/indice.htm (accessed November 2024).
- Iwaizumi, E., & Webb, S. (2021). Measuring L1 and L2 productive derivational knowledge: How many derivatives can L1 and L2 learners with differing vocabulary levels produce? *TESOL Quarterly*, 56, 100–129.
- Iwaizumi, E., & Webb, S. (2023). To what extent do learner- and word-related variables affect production of derivatives? *Language Learning*, 73, 301–336.
- James, C., & James, C. (1991). *Basic German vocabulary*. Langenscheidt.
- Jeon, E., In'nami, Y., & Koizumi, R. (2022). L2 speaking and its external correlates. In E. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 339–367). John Benjamins.
- Jeon, E., & Yamashita, J. (2022). L2 reading comprehension and its correlates. In E. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 29–86). John Benjamins.
- Jones, R.L. (2006). An analysis of lexical text coverage in contemporary German. In A. Wilson, P. Rayson, & D. Archer (Eds.), *Corpus linguistics around the world* (pp. 115–120). Brill.
- Kay, C., Alexander, M., Dallachy, F., et al. (Eds.). (2021). *The historical thesaurus of English: Version 5.0* (2nd ed.). University of Glasgow. Available at: <https://ht.ac.uk> (accessed October 2024).
- Kempey, S.T., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, 73, 441–454.
- Kilgarriff, A., Baisa, V., Bušta, J., et al. (2014). The sketch engine: Ten years on. *Lexicography*, 1, 7–36.
- Kilgarriff, A., Charalabopoulou, F., Gavrilidou, M., et al. (2014). Corpus-based vocabulary lists for language learners for nine languages. *Language Resources and Evaluation*, 48, 121–163.
- Kremmel, B., Indrarathne, B., Kormos, J., & Suzuki, S. (2023). Unknown vocabulary density and reading comprehension: Replicating Hu and Nation (2000). *Language Learning*, 73, 1127–1163.

- Knight, D., Fitzpatrick, T., Morris, S., et al. (2023). Corpus to curriculum: Developing word lists for adult learners of Welsh. *Applied Corpus Linguistics*, 3, 100052.
- Kojima, M., In'nami, Y., & Kaneta, T. (2022). L2 writing and its external correlates. In E. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 159–211). John Benjamins.
- Kusseling, F., & Lonsdale, D. (2013). A corpus-based assessment of French CEFR lexical content. *Canadian Modern Language Review*, 69, 436–461.
- Language-Driven Pedagogy. (2021a). *Vocabulary lists: Rationales and uses*. Language-Driven Pedagogy. Available at: <https://resources.ldpedagogy.org/concern/resources/t722h880z> (accessed October 2024).
- Language-Driven Pedagogy. (2021b). *NCELP professional development audio screencasts*. Language-Driven Pedagogy. Available at: <https://resources.ldpedagogy.org/collections/4x51hj27z> (accessed October 2024).
- Language-Driven Pedagogy. (2023a). *Comparison of inflectional morphology specified in 2016 and 2024 GCSE Subject Content*. Language-Driven Pedagogy. Available at: <https://resources.ldpedagogy.org/concern/resources/zp38wg94d> (accessed October 2024).
- Language-Driven Pedagogy. (2023b). *Teaching materials and schemes of work*. Language-Driven Pedagogy. Available at: <https://ldpedagogy.org/ncelp-schemes-of-work> (accessed October 2024).
- Language-Driven Pedagogy. (2024). *GCSE word lists for French, German, and Spanish*. Language-Driven Pedagogy. Available at: <https://resources.ldpedagogy.org/concern/resources/sq87bw293>
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19, 255–271.
- Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL Quarterly*, 54, 1076–1085.
- Laufer, B. (2023a). From research to a national curriculum: The case of a lexical syllabus. In G. Erickson, C. Bardel, & D. Little (Eds.), *Collaborative research in language education* (pp. 151–164). De Gruyter.
- Laufer, B. (2023b). Understanding L2-derived words in context: Is complete receptive morphological knowledge necessary? *Studies in Second Language Acquisition*, 46, 200–213.
- Laufer, B., Webb, S., Kim, S.K., & Yohanan, B. (2021). How well do learners know derived words in a second language? The effect of proficiency, word frequency and type of affix. *International Journal of Applied Linguistics*, 172(2), 229–258.
- Lindqvist, C., & Ramnäs, M. (2022). Élaboration d'une liste de vocabulaire dans un contexte FLE [Developing a vocabulary list in a French as a foreign language context]. *Studia Romanica Posnaniensia*, 49, 65–74.
- Lonsdale, D., & Le Bras, Y. (2009). *A frequency dictionary of French: Core vocabulary for learners*. Routledge.
- Lubliner, S., & Hiebert, E.H. (2011). An analysis of English–Spanish cognates as a source of general academic language. *Bilingual Research Journal*, 34, 76–93.
- Marsden, E., Alferink, I., Andringa, S., et al. (2018). *Open accessible summaries in language studies (OASIS)* [database]. Available at: <https://www.oasis-database.org> (accessed October 2024).
- Marsden, E., & David, A. (2008). Vocabulary use during conversation: A cross-sectional study of development from year 9 to year 13 among learners of Spanish and French. *The Language Learning Journal*, 36, 181–198.
- Marsden, E., Dudley, A., & Hawkes, R. (2023). Use of word lists in a high-stakes, low-exposure context: Topic-driven or frequency-informed. *The Modern Language Journal*, 107, 669–692.

- Marsden, E., & Hawkes, R. (2023). Situating practice in a limited-exposure, foreign languages school curriculum. In Y. Suzuki (Ed.), *Practice and automatization in second language research: Perspectives from skill acquisition theory and cognitive psychology* (pp. 89–118). Routledge.
- Marsden, E., Hawkes, R., Earnshaw, L., & Hobson, V., & NCELP team. (2023). *NCELP final report*. National Centre for Excellence for Language Pedagogy. Available at: https://eprints.whiterose.ac.uk/200637/1/NCELP_Final_Report_9_April_2023.pdf (accessed October 2024).
- Marsden, E., & Kasprovicz, R. (2017). Foreign language educators' exposure to research: Reported experiences, exposure via citations, and a proposal for action. *The Modern Language Journal*, 101, 613–642.
- McKinley, J., & Thompson, G. (2018). Washback effect in teaching English as an international language. In J.I. Liontas, M. DelliCarpini, & S. Abrar-ul-Hassan (Eds.), *The TESOL encyclopedia of English language teaching*. Wiley.
- McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, 39, 823–845.
- Meara, P.M., & Milton, J. (2003). *X_Lex: The Swansea vocabulary levels test*. Express Publishing.
- Meunier, F., & Marslen-Wilson, W. (2004). Regularity and irregularity in French verbal inflection. *Language and Cognitive Processes*, 19, 561–580.
- Milton, J. (2006). Language Lite? Learning French vocabulary in school. *Journal of French Language Studies*, 16, 187–205.
- Milton, J. (2015). French lexis and formal exams in the British foreign language classroom. *Revue Française de Linguistique Appliquée*, 20, 107–119.
- Milton, J., & Alexiou, T. (2009). Vocabulary size and the Common European Framework of Reference for languages. In B. Richards, M.H. Daller, & D.D. Malvern, et al. (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 194–211). Palgrave.
- Min, H.-T. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning*, 58, 73–115.
- Ministry of Education of the People's Republic of China. (2022). *English curriculum standards for compulsory education: 2022 Version*. Beijing Normal University Press.
- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28, 291–304.
- Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48, 323–364.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I.S.P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1, 2–13.
- Nation, I.S.P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.
- Nation, I.S.P. (2017). *The BNC/COCA Level 6 word family lists: Version 1.0.0 [datasets]*. Available at: <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists> (accessed October 2024).
- Nation, I.S.P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.
- Nation, I.S.P., & Sorell, J. (2016). Corpus selection and design. In I.S.P. Nation (Ed.), *Making and using word lists for language learning and testing* (pp. 95–107). John Benjamins.
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36, 516–524.
- Noreillie, A.S., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign

- languages: An approximate replication study of Stæhr (2009). *International Journal of Applied Linguistics*, 169, 212–231.
- O'Brien, M.G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, 64(4), 715–748.
- Pearson Edexcel. (2024). *Specification and sample assessments*. Pearson Edexcel. Available at: <https://qualifications.pearson.com/en/qualifications/edexcel-gcses/french-2024.coursematerials.html> (accessed October 2024).
- Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, 22, 31–55.
- Peters, E., Noreillie, A.-S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning*, 69, 747–782.
- Pinker, S. (1991). Rules of language. *Science*, 253, 530–535.
- Qualifications Wales. (2024). *GCSE approval criteria*. Qualifications Wales. Available at: <https://qualifications.wales/regulation-reform/reforming/qualified-for-the-future/made-for-wales-gcses/gcses-approval-criteria> (accessed October 2024).
- Ramnäs, M. (2019). Étendue du vocabulaire et compréhension écrite: Le français à l'université en Suède [Vocabulary breadth and reading comprehension: French at university in Sweden]. *Bergen Language and Linguistics Studies*, 10, 12.
- Read, J., & Dang, T.N.Y. (2022). Measuring depth of academic vocabulary knowledge. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688221105913>
- Robles-García, P., Stewart, J., Nicklin, C., et al. (2023). 'The wisdom of crowds': When teacher judgments outperform word-frequency as a predictor of students' vocabulary knowledge. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688231176067>
- Sasao, Y., & Webb, S. (2017). The Word Part Levels Test. *Language Teaching Research*, 21(1), 12–30. <https://doi.org/10.1177/1362168815586083>
- Sato, M., & Loewen, S. (2019). Do teachers care about research? The research-pedagogy dialogue. *ELT Journal*, 73, 1–10.
- Savický, P., & Hlaváčová, J. (2002). Measures of word commonness. *Journal of Quantitative Linguistics*, 9, 215–231.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing: Volume 12*, Manchester.
- Schmid, H., Fitschen, A., & Heid, U. (2004). SMOR: A German computational morphology covering derivation, composition, and inflection. In M.T. Lino, M.F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the fourth international conference on language resources and evaluation* (pp. 1263–1266). European Language Resources Association.
- Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, L., & Kremmel, B. (2021). Introducing knowledge-based vocabulary lists (KVL). *TESOL Journal*, 12, e622.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95(1), 26–43.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47, 484–503.
- Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching* (2nd ed.). Cambridge University Press.
- Sinclair, J. (1996). The search for units of meaning. *Textus*, 9, 75–106.
- Sinclair, J. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive lexical semantics* (pp. 1–24). John Benjamins.

- Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 140–158). Longman.
- Sonbul, S., & El-Dakhs, D.A.S. (2024). L2 knowledge of English prefixes: The effect of vocabulary breadth and exposure. *Language Teaching Research*, 28, 2111–2135.
- Stoeckel, T., Ishii, T., & Bennett, P. (2020). Is the lemma more appropriate than the flemma as a word counting unit? *Applied Linguistics*, 41(4), 601–606.
- Sundqvist, P. (2009). *Extramural English matters: Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary*. Doctoral dissertation, Karlstad University. DiVA. Available at: <https://www.diva-portal.org/smash/get/diva2:275141/FULLTEXT03.pdf> (accessed October 2024).
- Tidball, F., & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French Language Studies*, 18, 299–313.
- Tschirner, E. (2009). Korpuslinguistik und Fremdsprachenunterricht [Corpus linguistics and foreign language teaching]. *Germanica Wratislaviensia*, 129, 117–132.
- Tschirner, E. (2019). Der rezeptive Wortschatzbedarf im Deutschen als Fremdsprache [The receptive vocabulary need in German as a foreign language]. In E. Peyer, T. Studer, & I. Thonhauser (Eds.), *IDT 2017: Band 1: Hauptvorträge* (pp. 98–111). Erich Schmidt.
- Tschirner, E., Hacking, J., & Rubio, F. (2018). The relationship between reading proficiency and vocabulary size: An empirical investigation. In P. Ecke & S. Rott (Eds.), *Understanding vocabulary learning and teaching: Implications for language program development* (pp. 58–77). Heinle & Heinle.
- Tschirner, E., & Möhring, J. (2020). *A frequency dictionary of German: Core vocabulary for learners* (2nd ed.). Routledge.
- Tuggener, D. (2016). Incremental coreference resolution for German. Unpublished doctoral dissertation, University of Zurich. Available at: https://www.cl.uzh.ch/dam/jcr:b2212d28-6248-47dc-a4e3-04206ff4c6db/tuggener_diss.pdf (accessed October 2024).
- Twilley, L.C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition*, 22(1), 111–126.
- Ullman, M.T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, 30, 37–69.
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34, 457–479.
- Verlinde, S., & Selva, T. (2001). Corpus-based vs. intuition-based lexicography. Defining a word list for a French learners' dictionary. In R. Rayson, A. Wilson, T. McEnery, A. Hardie, & S. Khoja (Eds.), *UCREL Technical Papers Volume 13 Special Issue. Proceedings of the Corpus Linguistics 2001 conference* (pp. 594–598). Lancaster University. Available at: <https://ucrel.lancs.ac.uk/publications/cl2003/CL2001%20conference/papers/verlinde.pdf> (accessed October 2024).
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61, 219–258.
- Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, 37, 461–469.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28, 46–65.
- Webb, S. (2021). Word families and lemmas, not a real dilemma: Investigating lexical units. *Studies in Second Language Acquisition*, 43, 973–984.

- Webb, S., & Chang, A.C.-S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37, 651–675.
- Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary-learning activities? A Meta-analysis. *The Modern Language Journal*, 104(4), 715–738.
- West, M. (1953). *A general service list of English words*. Longman.
- Zec, S., Soriani, N., Comoretto, R., & Baldi, I. (2017). High agreement and high prevalence: The paradox of Cohen's Kappa. *Open Nursing Journal*, 11, 211–218.

Appendix A

Number of unique lemmas needed in Eduqas GCSE papers

Table A1. Number of lemmas needed in Eduqas GCSE exams (2018–22).

	French		German		Spanish		Mean (SD)	
	F	H	F	H	F	H	F	H
2018	534	676	652	819	522	683	569.3 (71.8)	726 (80.6)
2019	566	717	606	773	540	670	570.7 (33.2)	720 (51.6)
2022	532	781	631	794	484	665	549.0 (75.0)	746.7 (71.0)
All years	1,100	1,448	1,267	1,631	996	1,302	1121.0 (136.7)	1460.3 (164.8)
Mean (SD)	544.0 (19.1)	724.7 (52.9)	629.7 (23.0)	795.3 (23.0)	515.3 (28.6)	672.7 (9.3)	563.0	730.9

Notes. SD = standard deviation; SE = standard error. F = foundation; H = higher; Proper nouns are not included. The slightly higher numbers for German are likely due to compounding.

Appendix B

Composition of corpora created for the study

Table B1. Size (words) and composition of GCSE corpora.

Awarding organization	Tier	Papers included	French	German	Spanish
AQA	F	2018–21, specimen and sample	11,793	10,730	10,359
AQA	H	2018–21, specimen and sample	16,206	14,884	15,396
Edexcel	F	2018–21 and sample	12,341	10,539	10,851
Edexcel	H	2018–21 and sample	15,174	14,776	14,954
Eduqas	F	2018, 2019, 2022 and sample	6,732	7,934	6,282
Eduqas	H	2018, 2019, 2022 and sample	9,979	10,745	8,893
Total			72,225	69,608	66,735

Notes. F = foundation tier; H = higher tier.

Table B2. Size (words) and composition of AS/A-Level corpora.

Awarding organization	Tier	Papers included	French	German	Spanish
AQA	AS	2018–20	16,993	13,209	13,434
AQA	A	2018–21 and samples 1–2	26,493	24,492	26,220
Edexcel	AS	2017–20*	14,758	13,932	13,250
Edexcel	A	2018–21 and sample 1	19,842	18,740	20,976
Eduqas	AS	2017–19, 2022 and sample	9,907	10,509	10,636
Eduqas	A	2018–19, 2022 and sample	13,984	13,443	19,462
Total			101,977	94,325	103,978

Notes. *Edexcel AS papers for Spanish are not available for 2020.

Table B3. Size (words) and composition of young adult literature corpora.

<i>French:</i>	553,423
Beauvais, C. (2010). <i>Les petites filles top-modèles</i> . Talents Hauts. https://www.numilog.com/1050546/1050546.ebook	13,391
Chartres, M. (2020). <i>L'âge des possibles</i> . L'École des Loisirs. https://www.numilog.com/1214938/1214938.ebook	51,327
Guilbert, N. (2021). <i>Old soul</i> . Courtes et Longues. https://www.numilog.com/1293291/1293291.ebook	77,358
Lenne-Fouquet, M. (2022). <i>Corps de fille</i> . Talents Hauts. https://www.numilog.com/1450394/Corps-de-fille.ebook	32,765
Loyer, A. (2021). <i>Filles uniques</i> . Slalom. https://www.numilog.com/1275547/1275547.ebook	54,888
Mondiot, V. (2020). <i>Les derniers des branleurs</i> . Éditions Actes Sud. https://www.numilog.com/1205536/Les-derniers-des-branleurs.ebook	92,376
Monnier, C. (2020). <i>Je ne voulais pas vous faire pleurer</i> . Slalom. https://www.numilog.com/1130525/1130525.ebook	23,596
Murail, L., & Murail, M. (2021). <i>Angie</i> . L'École des Loisirs. https://www.numilog.com/1269216/1269216.ebook	95,474
Rowling, J.K. (1999). <i>Harry Potter à l'École des Sorciers</i> . (J. Ménard, Trans.) [Kindle]. Pottermore Publishing. (Original work published 1997.)	87,060
Witek, J. (2021). <i>J'ai 14 ans et ce n'est pas une bonne nouvelle</i> . Éditions Actes Sud. https://www.numilog.com/1281428/1281428.ebook	25,188
<i>German:</i>	722,313
Aydemir, F. (2017). <i>Ellbogen</i> . Hanser. https://www.blume-buch.de/shop/item/9783446255951/ellbogen-von-fatma-aydemir-e-book-epub	64,414
Funke, C. (2011). <i>Drachenreiter</i> . Dressler. https://www.blume-buch.de/shop/item/9783862722853/drachenreiter-l-von-cornelia-funke-e-book-epub	99,759
Gier, K. (2009). <i>Rubinrot: Liebe geht durch alle Zeiten</i> . Arena. https://www.buecher.de/shop/fantasy/rubinrot-liebe-geht-durch-alle-zeiten-bd-l-ebook-epub/gier-kerstin/products_products/detail/prod_id/37349771/	71,194

(Continued)

Table B3. (Continued)

Haddon, M. (2011). <i>Supergute Tage oder die sonderbare Welt des Christopher Boone</i> (S. Hübner, Trans). Blessing. (Original work published in 2003.) https://www.blume-buch.de/shop/item/9783641063566/supergute-tage-oder-die-sonderbare-welt-des-christopher-boone-von-mark-haddon-e-book-epub	57,641
Herrndorf, W. (2011). <i>Tschick</i> . Rowohlt. https://www.blume-buch.de/shop/item/9783644107816/tschick-von-wolfgang-herrndorf-e-book-epub	58,311
Herwig, J. (2023). <i>Halber Löwe</i> . Gerstenberg. https://www.blume-buch.de/shop/item/9783836992008/halber-lowee-von-johannes-herwig-e-book-epub	51,938
Poznanski, U. (2013). <i>Erebos</i> . (2nd ed.). Lion. https://www.blume-buch.de/shop/item/9783732000968/erebos-von-ursula-poznanski-e-book-epub	123,878
Preußler, O. (2012). <i>Krabat</i> . Thienemann-Esslinger. https://www.blume-buch.de/shop/item/9783522620536/krabat-roman-von-otfried-preuler-e-book-epub	59,934
Rowling, J.K. (2005). <i>Harry Potter und der Stein der Weisen</i> . (K. Fritz, Trans). [Kindle]. Pottermore Publishing. (Original work published 1997.)	81,692
Schäuble, M. (2017). <i>Endland</i> . Hanser. https://www.blume-buch.de/shop/item/9783446257412/endland-von-martin-schauble-e-book-epub	53,552
Aydemir, F. (2017). <i>Ellbogen</i> . Hanser. https://www.blume-buch.de/shop/item/9783446255951/ellbogen-von-fatma-aydemir-e-book-epub	64,414
Funke, C. (2011). <i>Drachenreiter</i> . Dressler. https://www.blume-buch.de/shop/item/9783862722853/drachenreiter-l-von-cornelia-funke-e-book-epub	99,759
<i>Spanish:</i>	697,045
Dahl, R. (2016). <i>Las brujas</i> . (M. De Juan Gruyat, Trans). Santillana Educación. (Original work published 1983.)	33,525
Green, J. (2019). <i>Mil veces hasta siempre</i> . (N. Sobregués Arias, Trans). Nube de Tinta. (Original work published 2017.)	62,875
Levy, N. (2019). <i>Lo que no nos contaron</i> . (I. González-Gallarza Granizo, Trans). HarperCollins. (Original work published 2008.)	95,841
Ness, P. (2019). <i>De hombres a monstruos</i> . (R. Gil Giner, Trans.). Nube de Tinta. (Original work published 2010.)	130,315
Ortiz, E. (2019). <i>Todo saldrá bien</i> . Duomo Ediciones.	114,628
Preston, N. (2019). <i>El anónimo</i> . (N. Navarro Díaz, Trans). Crossbooks. (Original work published 2018.)	70,364
Punset, A. (2016). <i>Juntas, of course</i> . Montena.	34,074
Rowling, J.K. (1999). <i>Harry Potter y la piedra filosofal</i> . (A. Dellepiane Rawson, Trans). [Kindle]. Pottermore Publishing. (Original work published 1997.)	77,775
Saint Exupéry, A. (n.d.). <i>El Principito</i> . (Biblioteca Digital del ILCE, Instituto Latinoamericano de la Comunicación Educativa, Trans.). (Original work published 1943.) http://bibliotecadigital.ilce.edu.mx/Colecciones/index.php?clave=ObrasClasicas	13,008
Salar, O. (2019). <i>Enamorarse: clases prácticas</i> . Harlequin Ibérica.	64,640

We purchased a set of contemporary (published within the last 25 years), popular novels in French and German by authors of different genders and nationalities. These were eBooks with Digital Rights Management (DRM)-encryption removed. For Spanish, our choice was more restricted as we could not find books without DRM-encryption. We sourced our Spanish texts (mostly translated from English or French) from Educable (2019), where they are available as PDFs.

Appendix C

Word-topic relatedness: Breakdown of inter-rater agreement measures by category

	Mean		French		German		Spanish	
	Raw percentage	ACI	Raw percentage	ACI	Raw percentage	ACI	Raw percentage	ACI
Animals	98.28	0.98	99.10	0.99	99.40	0.99	96.35	0.96
Appearance and body ^a	96.95	0.97	96.50	0.96	97.40	0.97	N/A	N/A
The arts	94.43	0.93	90.05	0.89	97.20	0.97	93.05	0.93
Charity and voluntary work ^b	89.82	0.88	90.10	0.89	94.60	0.94	84.75	0.81
Cinema and TV	94.97	0.94	94.90	0.94	98.35	0.98	91.65	0.91
Colour	99.00	0.99	N/A	N/A	N/A	N/A	99.00	0.99
Communication ^c	91.86	0.90	86.04	0.82	96.55	0.96	93.00	0.92
Countries and nationalities ^d	97.65	0.98	98.95	0.99	98.65	0.99	96.35	0.96
Cultural background	94.95	0.94	94.70	0.94	96.50	0.96	93.65	0.93
Customs and festivals ^e	91.45	0.90	91.85	0.91	95.25	0.95	87.25	0.85
Daily routine	97.50	0.98	97.00	0.97	98.00	0.98	N/A	N/A
Economy ^f	88.35	0.84	79.99	0.70	94.75	0.94	90.30	0.89
Education post-16	86.80	0.84	83.74	0.79	N/A	N/A	89.85	0.88
Emotion ^g	90.96	0.89	92.15	0.91	94.60	0.94	86.14	0.82
Environment ^h	87.08	0.84	84.44	0.80	92.70	0.92	84.10	0.79
Faith	95.08	0.96	94.10	0.94	97.55	0.97	93.60	0.93
Family	94.05	0.93	94.05	0.93	98.45	0.98	89.65	0.87
Food ⁱ	96.48	0.96	95.70	0.95	97.90	0.98	95.85	0.96
Free-time activities	90.45	0.89	93.40	0.93	96.95	0.97	81.00	0.76
Friends	94.15	0.93	95.75	0.95	98.30	0.98	88.40	0.86
Geographical features	96.40	0.96	96.35	0.96	96.45	0.96	N/A	N/A
Global issues	87.85	0.85	86.94	0.84	92.05	0.91	84.55	0.81
Healthy and unhealthy living ^j	89.33	0.87	86.79	0.84	96.55	0.96	84.65	0.80
Home ^k	94.07	0.93	96.30	0.96	98.50	0.98	87.40	0.85
In the news/journalism ^l	93.95	0.93	87.74	0.85	93.95	0.93	N/A	N/A
Jobs, career choices and ambitions	83.85	0.78	74.04	0.61	96.25	0.96	81.25	0.76
Language learning	93.92	0.93	89.34	0.88	98.50	0.98	N/A	N/A
Law ^m	91.13	0.89	83.04	0.77	94.40	0.94	95.95	0.96

(Continued)

Appendix C. (Continued)

	Mean		French		German		Spanish	
	Raw percentage	ACI	Raw percentage	ACI	Raw percentage	ACI	Raw percentage	ACI
Location and distance	94.60	0.94	N/A	N/A	N/A	N/A	94.60	0.94
Marriage and partnership	95.30	0.95	93.10	0.92	98.75	0.99	94.05	0.93
Medical	92.87	0.92	88.94	0.86	96.80	0.97	N/A	N/A
Military	93.81	0.93	87.59	0.85	97.60	0.97	96.25	0.96
Music	95.57	0.95	94.45	0.94	98.75	0.99	93.50	0.93
Numbers	96.30	0.96	N/A	N/A	N/A	N/A	96.30	0.96
Personal characteristics	94.15	0.94	92.85	0.92	95.45	0.95	N/A	N/A
Politics	86.18	0.81	80.04	0.70	92.65	0.91	85.85	0.83
Popular culture ⁿ	96.95	0.97	95.80	0.96	98.10	0.98	N/A	N/A
Poverty and homelessness	91.85	0.91	88.49	0.86	95.20	0.95	N/A	N/A
Region	94.20	0.94	94.20	0.94	N/A	N/A	N/A	N/A
Relationships	97.10	0.97	N/A	N/A	97.10	0.97	N/A	N/A
School and college ^o	87.21	0.84	82.74	0.77	96.85	0.97	82.05	0.77
Sciences and maths	87.90	0.84	81.44	0.74	94.35	0.94	N/A	N/A
Sexuality	97.72	0.98	96.90	0.97	99.30	0.99	96.95	0.97
Shopping	97.00	0.97	95.15	0.95	98.85	0.99	N/A	N/A
Social issues ^p	87.75	0.85	86.89	0.84	91.60	0.91	(84.75) ^{b,p}	0.81
Social media	94.28	0.94	94.95	0.95	97.90	0.98	90.00	0.89
Sport ^q	90.22	0.88	93.30	0.93	97.50	0.97	79.85	0.74
Technology	93.02	0.92	94.00	0.93	96.40	0.96	88.65	0.87
Textile and clothing ^r	98.02	0.98	96.70	0.97	99.05	0.99	98.30	0.98
Time ^s	96.48	0.96	96.00	0.96	98.15	0.98	95.30	0.95
Town ^t	95.00	0.95	95.00	0.95	98.00	0.98	87.50	0.84
Transport ^u	95.75	0.95	93.40	0.93	98.35	0.98	95.50	0.95
Travel and tourism ^v	91.69	0.92	89.44	0.87	97.00	0.97	89.05	0.86
Weather	98.87	0.99	98.65	0.99	99.20	0.99	98.75	0.99
Mean	93.22	0.92	91.26	0.89	96.79	0.96	90.75	0.89

Notes. The figure in brackets refers to a merged category that is included only once in the calculation of the mean. ACI = Gwet's agreement coefficient; ^aAppearance, physical description and body (German); ^bSocial issues and charity (Spanish); ^cCommunication and media (Spanish); ^dCommunication, informal and spoken (German); ^eNationalities (French); ^fCustoms, festivals and history (German); ^gEconomy, business and finance (German); ^hEmotion and opinion (German); ⁱEnvironment, natural world and climate change (German); ^jFood and eating out (French), Food, drink and eating out (German); Food and drink (Spanish); ^kHealthy/unhealthy living and lifestyles (German); Health and lifestyle (Spanish); ^lHome, home life and furniture (German); ^mJournalism (French); ⁿLaw and crime (German); ^o(Popular) culture (German); ^pSchool and college (French), School, college, uni and education (German); ^qSocial issues and rights (German), Social issues and charity (Spanish); ^rSport and leisure (Spanish); ^sTextile, clothing and fashion (German); ^tDays, months and time phrases (German), Time and time markers (Spanish); ^uTown and local area (German), Town and local area (Spanish); ^vJourneys and transport (Spanish); ^wTravel (French), Tourism (French), Tourism (Spanish), Travel and holidays (Spanish).

Appendix D

Word list statistics and coverage by language

Table D1. Proportion of grammar and content words.

	French				German				Spanish				Mean (SD)			
	G		C		G		C		G		C		G		C	
	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage
2016 F	146	.07	1830	.93	172	.10	1617	.90	143	.08	1609	.92	153.7 (15.9)	.08	1685.3 (125.3)	.92
2024 F	89	.08	1017	.92	89	.08	1052	.92	94	.08	1022	.92	90.7 (2.9)	.08	1030.0 (18.9)	.92
2016 H	151	.08	1830	.92	174	.10	1616	.90	149	.08	1609	.92	158.0 (13.9)	.09	1685.0 (125.6)	.91
2024 H	109	.07	1454	.93	98	.06	1494	.94	137	.09	1452	.91	114.7 (20.1)	.07	1466.7 (23.7)	.93

Notes. F=foundation tier; H=higher tier; Notes. G=grammar words; C=content words; SD=standard deviation.

Table D2. Proportion of single-word items and multiword phrases/compounds.

	French				German				Spanish				Mean (SD)			
	SW		MWP		SW		MWP		SW		MWP		SW		MWP	
	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage	Raw	Percentage
2016 F	1509	.76	469	.24	1519	.85	270	.15	1390	.79	362	.21	1472.7 (71.8)	.80	367.0 (99.6)	.20
2024 F	1077	.97	29	.03	1107	.97	34	.03	1060	.95	56	.05	1081.3 (23.8)	.96	39.7 (14.4)	.04
2016 H	1512	.76	469	.24	1520	.85	270	.15	1393	.79	365	.21	1475.0 (71.1)	.80	368.0 (99.5)	.20
2024 H	1524	.98	39	.02	1556	.98	36	.02	1512	.95	77	.05	1530.7 (22.7)	.97	50.7 (22.9)	.03

Notes. F= foundation tier; H= higher tier; SW= single-word items; MWP= multiword phrases; SD= standard deviation.

Table D3. Part of speech proportions (percentages).

	French				German				Spanish				Mean			
	n	v	a	o	n	v	a	o	n	v	a	o	n	v	a	o
2016 F	.65	.16	.14	.05	.65	.14	.14	.06	.64	.16	.15	.05	.65	.16	.15	.05
2024 F	.50	.18	.22	.11	.47	.18	.23	.12	.46	.21	.21	.11	.48	.19	.22	.11
2016 H	.65	.16	.14	.05	.65	.14	.14	.06	.64	.16	.16	.05	.64	.15	.15	.05
2024 H	.49	.21	.21	.09	.45	.21	.24	.09	.47	.23	.20	.10	.47	.22	.22	.09

Notes. F=foundation tier; H=higher tier; n=nouns; v=verbs; a=adjectives or adverbs; o=other.

Table D4. Coverage by full lists (percentages).

	Foundation					Higher				
	GCSE	A-Level	Web	Literature	Mean	GCSE	A-Level	Web	Literature	Mean
French 2016	73.3	63.5	57.0	64.6	64.6	74.1	64.1	57.7	65.6	64.1
French 2024	83.9	75.0	66.5	76.8	75.5	87.9	81.0	73.2	81.7	80.3
German 2016	71.6	59.6	51.1	59.2	60.4	73.3	61.8	53.1	62.2	65.4
German 2024	81.0	69.3	59.1	68.6	69.5	86.6	78.2	67.4	78.6	81.0
Spanish 2016	70.8	63.2	58.5	62.7	63.8	71.2	63.7	59.2	63.5	62.6
Spanish 2024	83.8	75.7	69.7	76.1	76.3	88.3	82.2	76.0	82.0	77.7
Mean 2016	71.9	62.1	55.5	62.1	62.9	72.9	63.2	56.7	63.8	64.4
Mean 2024	82.9	73.3	65.1	73.8	73.8	87.6	80.5	72.2	80.8	82.1

Table D5. Coverage by content words and grammar words (percentages).

	Foundation										Higher									
	GCSE		A-level		Web		Literature		Mean		GCSE		A-level		Web		Literature		Mean	
	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G
French 2016	27.4	50.0	20.7	46.0	16.1	43.3	16.4	51.0	20.1	47.6	27.7	50.4	20.5	46.4	15.9	43.8	16.6	51.7	20.2	48.1
French 2024	38.3	51.3	32.3	47.5	27.0	44.2	29.3	53.7	31.7	49.1	41.7	52.4	37.1	48.6	32.3	45.5	32.8	55.2	36.0	50.4
German 2022	24.1	50.2	16.4	44.7	10.9	41.0	11.4	48.8	15.7	46.2	25.0	51.5	17.0	46.4	10.7	43.2	13.9	50.2	16.7	47.8
German 2024	41.1	43.0	32.2	38.9	24.9	35.8	29.9	40.7	32.0	39.6	44.5	45.6	38.6	41.6	31.0	38.2	36.6	44.4	37.7	42.4
Spanish 2016	26.5	47.0	19.4	45.9	14.7	46.1	15.9	48.4	19.1	46.8	21.8	47.2	15.0	46.1	15.0	46.4	12.9	48.8	16.2	47.1
Spanish 2024	45.1	41.7	36.3	42.0	30.7	41.9	33.6	45.4	36.4	42.8	47.9	44.1	40.9	43.9	34.9	44.4	36.5	48.7	40.1	45.3
Mean 2016	26.0	49.1	18.8	45.5	13.9	43.5	14.6	49.4	18.3	46.9	24.8	49.7	17.5	46.3	13.9	44.5	15.2	50.2	17.7	47.7
Mean 2024	41.5	45.3	33.6	42.8	27.5	40.6	30.9	46.6	33.4	43.8	44.7	47.4	38.9	44.7	32.7	42.7	35.3	49.4	37.9	46.0

Notes. C = content words; G = grammar words.

Table D6. Coverage by single-word items and multiword phrases/compounds (foundation) (percentages).

	Foundation										Higher									
	GCSE		A-Level		Web		Literature		Mean		GCSE		A-Level		Web		Literature		Mean	
	SW	MP	SW	MP	SW	MP	SW	MP	SW	MP	SW	MP	SW	MP	SW	MP	SW	MP	SW	MP
French 2016	72.2	4.1	63.1	2.3	56.7	1.1	64.1	2.0	64.0	2.4	72.8	4.1	63.7	1.9	57.4	0.8	65.2	1.8	64.8	2.1
French 2024	83.2	2.6	74.6	1.9	66.3	0.6	76.5	1.7	75.1	1.7	87.5	2.7	80.7	2.0	73.0	0.8	81.4	2.2	80.6	1.9
German 2016	70.6	3.4	59.4	1.3	51.0	0.8	58.9	1.2	60.0	1.7	72.3	3.6	61.5	1.5	52.8	0.9	62.0	1.8	62.1	1.9
German 2024	80.9	1.2	69.2	0.7	58.9	0.4	68.4	0.6	69.4	0.7	86.3	1.5	77.7	0.8	66.6	0.7	78.2	1.0	77.2	1.0
Spanish 2016	69.9	2.7	62.8	1.4	58.3	1.0	62.5	1.2	63.4	1.6	70.4	2.8	63.3	1.6	59.0	1.2	63.3	1.4	64.0	1.7
Spanish 2024	83.8	0.0	75.5	1.6	69.4	1.2	76.0	1.4	76.2	1.1	88.1	3.1	82.0	2.6	75.8	2.4	81.8	2.9	81.9	2.8
Mean 2016	70.9	3.4	61.7	1.7	55.3	1.0	61.9	1.5	62.5	1.9	71.8	3.5	62.8	1.6	56.4	1.0	63.5	1.7	63.6	1.9
Mean 2024	82.6	1.3	73.1	1.4	64.9	0.7	73.6	1.3	73.6	1.2	87.3	2.4	80.1	1.8	71.8	1.3	80.4	2.0	79.9	1.9

Notes. SW = single-word items; MP = multiword phrases.

Table D7. Coverage by parts of speech in foundation lists (percentages).

		French				German				Spanish				Mean			
		n	v	a	o	n	v	a	o	n	v	a	o	n	v	a	o
GCSE	2016	17.5	15.7	7.6	38.6	13.3	14.9	9.1	38.5	15.4	15.2	9.3	34.5	15.4	15.3	8.7	37.2
	2024	17.7	17.0	11.9	41.2	14.3	16.2	15.2	39.8	17.8	19.1	10.4	38.2	16.6	17.4	12.5	39.7
A-Level	2016	13.1	10.6	5.6	38.7	10.1	9.4	6.2	37.0	11.1	10.8	6.4	37.6	11.4	10.3	6.0	37.8
	2024	15.1	11.5	9.1	44.4	12.4	10.7	11.6	37.8	14.3	13.7	7.9	40.2	13.9	12.0	9.5	40.8
Web	2016	10.5	7.6	4.2	37.4	5.7	7.0	5.0	35.1	7.5	10.6	6.3	36.9	7.9	8.4	5.2	36.5
	2024	11.5	8.5	7.8	42.5	7.3	8.2	9.6	35.9	10.2	13.7	6.5	40.2	9.7	10.1	8.0	39.5
Literature	2016	9.2	12.4	7.3	40.2	5.4	8.5	6.8	40.8	6.7	12.7	8.2	37.1	7.1	11.2	7.4	39.4
	2024	10.7	13.7	11.0	47.1	6.6	8.8	14.0	42.1	9.5	17.6	9.9	40.9	8.9	13.4	11.6	43.3
Mean	2016	12.6	11.6	6.2	38.7	8.6	10.0	6.8	37.8	10.2	12.3	7.5	36.5	10.5	11.3	6.8	37.7
	2024	13.7	12.7	10.0	43.8	10.1	11.0	12.6	38.9	13.0	16.0	8.9	39.6	12.3	13.2	10.5	40.8

Table D8. Coverage by parts of speech in higher lists (percentages).

		French				German				Spanish				Mean			
		n	v	a	o	n	v	a	o	n	v	a	o	n	v	a	o
GCSE	2016	17.5	16.3	7.6	38.6	14.0	16.1	9.3	38.6	15.4	15.8	9.5	34.5	15.6	16.1	8.8	37.2
	2024	19.5	18.8	14.4	44.2	16.1	19.8	16.2	40.4	19.3	21.5	11.9	41.7	18.3	20.0	14.1	42.1
A-Level	2016	13.1	11.1	5.6	38.9	10.9	10.5	6.4	37.4	11.1	11.4	6.5	37.6	11.7	11.0	6.2	38.0
	2024	18.6	13.8	12.1	44.9	15.7	15.0	13.2	38.9	17.2	16.3	9.5	43.9	17.1	15.0	11.6	42.5
Web	2016	10.5	8.2	4.2	37.6	6.3	7.7	5.1	35.8	7.5	11.3	6.5	37.0	8.1	9.1	5.3	36.8
	2024	15.2	11.0	10.6	42.9	10.3	11.3	11.6	37.2	13.4	16.5	7.6	44.0	13.0	12.9	9.9	41.4
Literature	2016	9.2	13.2	7.3	40.4	5.8	11.4	6.8	41.0	6.7	13.6	8.5	37.1	7.3	12.7	7.6	39.5
	2024	12.7	16.5	15.1	47.6	8.3	16.1	15.5	42.9	10.7	20.7	10.7	44.4	10.5	17.8	13.8	45.0
Mean	2016	12.6	12.2	6.2	38.9	9.2	11.4	6.9	38.2	10.2	13.0	7.7	36.5	10.7	12.2	6.9	37.9
	2024	16.5	15.0	13.0	44.9	12.6	15.5	14.1	39.8	15.2	18.7	9.9	43.5	14.7	16.4	12.4	42.7

Table D9. Coverage of GCSE and literature corpora using different units of counting (foundation).

	GCSE corpus			Young adult literature corpus		
	Lemma	2016 bespoke	2024 bespoke	lemma	2016 bespoke	2024 bespoke
French 2016	0.74	0.74	0.73	0.66	0.66	0.65
French 2024	0.85	0.85	0.84	0.78	0.78	0.77
German 2016	0.74	0.73	0.72	0.63	0.63	0.59
German 2024	0.85	0.84	0.81	0.76	0.75	0.69
Spanish 2016	0.71	0.71	0.71	0.64	0.64	0.63
Spanish 2024	0.85	0.85	0.84	0.78	0.77	0.76

Table D10. Coverage of GCSE and literature corpora using different units of counting (higher).

	GCSE corpus			Young adult literature corpus		
	lemma	2016 bespoke	2024 bespoke	lemma	2016 bespoke	2024 bespoke
French 2016	0.75	0.75	0.74	0.66	0.66	0.66
French 2024	0.88	0.88	0.88	0.83	0.82	0.82
German 2016	0.74	0.74	0.73	0.64	0.63	0.62
German 2024	0.87	0.87	0.87	0.80	0.79	0.79
Spanish 2016	0.71	0.71	0.71	0.65	0.65	0.64
Spanish 2024	0.89	0.89	0.88	0.83	0.83	0.82