**Lexical frequency and target language input in the primary languages classroom**

Rosamond Mitchell, University of Southampton, UK

<https://orcid.org/0000-0003-0325-528X>

R.F.Mitchell@soton.ac.uk

Florence Myles, University of Essex, UK

ORCID [0000-0003-4700-700X](http://orcid.org/0000-0003-4700-700X)

fmyles@essex.ac.uk

Abstract:

This paper presents an exploratory study of L2 input in the primary languages classroom. Teacher talk in 33 videorecorded lessons taught to Year 3 French beginners by a specialist teacher of French was analysed, in order to document the extent and nature of L2 vocabulary input. Further analyses compared the lexis of teacher talk in this primary setting with high frequency wordlists derived from two reference corpora designed to support the teaching of L2 French. Results showed that children were exposed to c45,000 words (tokens) of French, reflecting a teacher vocabulary of c650 lemmas. Intentionally taught lexis such as foods or animal names typically received massed practice, not always supplemented by distributed practice. The teachers’ partial but consistent use of French for classroom management contributed to a well balanced distribution of word classes. There was considerable overlap with both reference wordlists, and divergence could largely be explained by the teacher’s choice of topics, and incidental use of French to personalise and engage childrens’ individual interest, as well as for classroom management purposes. In discussion the need for careful planning and management of vocabulary instruction is highlighted, as well as the significant potential of L2 classroom management to enrich lexical input.

Keywords:

Instructed SLA, vocabulary, L2 input, learner corpora, early learners

Funding Details:

This work was supported by the Economic and Social Research Council under Grant number *RES-062-23-1545.*

Disclosure Statement:

The authors report there are no competing interests to declare.

Biographical Notes:

Rosamond Mitchell is Professor Emeritus of applied linguistics at the University of Southampton. She trained originally as a schoolteacher of Irish and French, and then undertook postgraduate studies in Scotland, where she worked as a researcher and lecturer before moving to Southampton to lecture in applied linguistics. She has research interests in SLA theory, in classroom language learning, and in study abroad and its impact on L2 learning and identity. Together with Florence Myles, she has for many years promoted corpus-based approaches to the study of L2 learning. She is co-author with Florence Myles and Emma Marsden of *Second Language Learning Theories* (4th edition, 2019).

Florence Myles is Professor Emerita of Second Language Acquisition at the University of Essex and Founding Chair of the Research in Primary Languages Network (RiPL, www.ripl.uk), which aims to bridge the gap between researchers, practitioners and policy makers with respect to primary foreign language learning and teaching. Her recent research has compared how children of different ages learn foreign languages in the classroom, in terms of grammar, vocabulary, attitudes and motivation, as well as classroom engagement, and how the age of the child impacts on learning processes. She is co-author with Rosamond Mitchell and Emma Marsden of *Second Language Learning Theories* (2019).

List of Figures

Figure 1 Caption: Distribution of selected ‘animals’ vocabulary in teacher talk (tokens per lesson)

Figure 1 Alt Text: Line graph plotting the frequency of occurrence of animal names vocabulary over the sequence of 33 lessons. A large peak is seen during lessons 14-16 (“pets” topic) and a minor one in Lesson 23.

Figure 2 Caption: Distribution of selected conjunctions in teacher talk (tokens per lesson)

Figure 2 Alt Text: Line graph plotting the frequency of occurrence of conjunctions over the sequence of 33 lessons. Conjunction *et* [= and] has several large peaks, otherwise conjunctions are evenly distributed.

Figure 3 Caption: Distribution of teacher lexis by *LLB* frequency bands (word types)

Figure 3 Alt Text: Bar chart showing distribution of teacher vocabulary over LLB frequency bands. There is a large peak for the 1k band, then a drop off to low values until another smaller peak for the “Beyond 5k” band.

Figure 4 Caption: Distribution of teacher lexis by *FLELex* CEFR levels (word types)

Figure 4 Alt Text: Bar chart showing distribution of teacher vocabulary over CEFR levels, s defined in FLELex. There is a large peak for the CEFR A1 band, and low values for all other bands.

**Introduction**

Target language (TL) input is accepted as a central element of classroom language learning, alongside varying proportions of explicit/deliberate learning, and fluency practice (depending on the learning stage and overall teaching approach). For early classroom learners in most settings, classroom input is likely to be the most significant encounter with the TL. While for learners of global English, extramural encounters are likely soon to provide rich additional sources of TL (de Wilde et al., 2020; Leona et al., 2021), learners of other languages are likely to remain dependent on classroom input for considerably longer. In a setting such as England, it is likely that a limited and variable amount of primary school provision, followed by a limited number of hours of secondary school foreign language instruction (Collen, 2022), will remain the prime source of input for most learners in mainstream compulsory education. It is therefore important to understand the characteristics of this input, so as eventually to promote its effectiveness in supporting learning.

In previous research in the primary French classroom, we have examined input factors most likely to promote the acquisition of target language vocabulary, including input frequency, and multimodality (Mitchell & Rule, 2022). However, our past research took account only of lexical frequency within a small classroom input corpus (the ‘Young Learners’ corpus: Myles et al, 2012; Myles, 2017). In that study, no account was taken of the likely longer term ‘usefulness’ of the words being heard and learned. Here, we examine the vocabulary of the same primary French classroom corpus from a different perspective, that of its relationship with lexical frequency in two published reference corpora of French. We hope this study can inform current debates around the role of frequency-based wordlists in classroom language teaching, with a particular focus on early learners.

**Literature review**

***Vocabulary in instructed SLA***

Vocabulary development is central to successful L2 development overall (David et al, 2009), and optimal progression among classroom learners requires a mix of focused vocabulary instruction and informal exposure (Butler, 2019; Webb, 2020). Classroom learning of vocabulary is generally recognised to be a slow process (Miralpeix, 2020); thus for example, Orosz (2009) estimated that Hungarian young learners on average learned 1,500 words of English over four years (Grade 3 to Grade 6), or a mean of 375 words per year, with 277 hours of classroom input. In the United Kingdom, poor progress in vocabulary development among school-aged learners is a longstanding concern (Milton, 2008). Various reasons for this have been suggested:

* Classroom time is limited (Collen, 2022);
* The lexical content of textbooks has declined (Tschichold, 2012; Milton & Hopwood, 2021);
* Vocabulary instruction is unsystematic (Marsden & Hawkes, 2023);
* Young instructed Anglophone learners learning other languages in school do not typically engage in ‘extramural’ L2 learning, unlike young ESL learners elsewhere in Europe (e.g. Cadierno et al, 2022; de Wilde et al, 2020).

It seems likely that all of these factors play some role, and it is important to maximise the effectiveness of what is acknowledged to be a very limited curriculum space for L2 learning. Instruction in one foreign language is included as part of the curriculum in England from age 8-11 (Key Stage 2), but at this level, a large majority of schools offer the language for only 30-60 minutes per week (Collen, 2023). At secondary school level, it is compulsory to study one foreign language for the first three years (Key Stage 3), and just over half of all students continue to do so until age 16 (Key Stage 4), receiving c 400-450 hours’ instruction in total, an overall FL experience which has in recent decades been the most limited in Europe (for comparison, see figures in European Education and Culture Executive Agency, 2023).

Given these time limitations, a systematic approach to classroom vocabulary instruction is important and is promoted in varying ways by current policymakers and methodologists. The use of frequency-based wordlists has a long history going back e.g. to West’s 1953 *General Service List of English Words* in the case of English (Howatt, 1984), or to another 1950s project *Le français fondamental* in the case of French (Coste, 2006; Gougenheim et al., 1964; Michéa, 1953). In England, wordlists provided by the various examining boards are a longstanding aspect of the 16+ General Certificate of Secondary Education (GCSE) examination syllabus. The latest redesign of this syllabus now requires the use of centrally prescribed L2 wordlists primarily based on frequency in contemporary reference corpora (DfE 2022a, 2022b); students at Foundation level are intended to master the commonest 1,200 items in these corpora, students at Higher level to master 1,700. The work of the National Centre for Languages Pedagogy (NCELP) has promoted a systematic approach to the introduction, practice and recycling of this frequent vocabulary (Marsden & Hawkes, 2023), grounded in skill acquisition theory (DeKeyser, 2015), and international research provides evidence of the effectiveness of spaced recycling for vocabulary learning (Bloom & Shuell, 1981; Kim & Webb, 2022; Nation, 2020; Zhang, 2022). Other current methodologists place more emphasis on modelling/practice/recycling of whole sentence patterns but also acknowledge the need for regular revisiting of “must-learn” vocabulary (e.g. Conti, 2022; Smith & Conti, 2016). Research exploring the classroom treatment of vocabulary at early levels, other than textbook analysis (e.g. Tschichold, 2012; Milton & Hopwood, 2021), is scant in the UK setting. However the limited international research identified by Butler (2019) stresses the significance of classroom L2 input for vocabulary learning in the primary school.

***Classroom input and L2 vocabulary development***

When analysing classroom input more closely from the perspective of vocabulary learning, it is necessary to take account of the distinction between incidental and intentional learning (Hulstijn, 2001). It has repeatedly been shown that rich exposure to English outside the formal classroom (through e.g. TV, music or video games) can lead to impressive incidental learning among young learners, even before any formal instruction begins (de Wilde et al., 2021). Within the classroom, teacher management talk in the L2, ‘communicative’ tasks and CLIL all offer opportunities for incidental learning over time, and recent studies confirm the positive overall influence on children’s classroom-based L2 learning of teacher target language use (e.g. Erk & Pavičić Takač, 2021; Shabaka-Fernández, 2023). However, incidental vocabulary learning requires numerous repetitions in the input (Kormos, 2020), and cannot be the main route to vocabulary development with early learners in time-limited classrooms. As we saw above, research on classroom learning has repeatedly shown the greater effectiveness of intentional learning involving systematic practice including spaced practice. A focus on intentional vocabulary instruction and practice is accordingly also found in the recommendations of early learning methodologists such as Pinter (2006) or Sharpe (2001).

Given this likely mix of intentional instruction and incidental exposure (plus of course the added complication of classroom codeswitching between L2 and L1), classroom L2 input must be viewed as complex, and its relationship with vocabulary acquisition even more so.

But most research on classroom discourse is interaction-focused (e.g. Gardner, 2019; Kunitz et al., 2021; Loewen & Sato, 2018), and empirical research on classroom talk from an input perspective is quite limited. To study classroom talk as input, searchable corpora of L2 lessons are required, and these are relatively rare. Notable examples with adult learners include the Portland corpus of adult ESL lessons (Reder et al., 2003), where learner speech has been the main focus of analysis (e.g. Hellerman, 2008; Eskildsen & Wagner, 2015), and the Corpus of English for Academic and Professional Purposes (CEAPP), where Hall (2022) has analysed the “linguistic designs” of teacher questions. For younger learners, an important example is the corpus of ESL lessons in French-medium schools in Canada developed by Collins et al. (2009). Like Hall, these researchers analysed teacher speech as input for aspects of English morphosyntax, specifically the *his/her* distinction, simple past *-ed* and progressive *-ing*. They found that high token frequency did not by itself lead to acquisition; the most ‘easy to learn’ morpheme was progressive *-ing*, which was characterised by saliency and occurrence in varied contexts as well as high frequency.

As for corpus-based research on classroom input from the perspective of vocabulary learning, in our own earlier work on the ‘Young Learners’ lesson corpus we examined how far a subset of lexical items received multimodal support (including gestures, images or written reinforcement), which reflected teacher’s intentionality and is likely to have promoted saliency (Mitchell & Rule, 2022). Similarly, Plonsky and Loewen (2013) analysed a small corpus of Spanish lessons at university level. For their seven target lexical items, they tracked frequency of occurrence, and examined qualitatively any focus-on-form discussion connected with them. They concluded that no single variable was predictive of lexical development, and argue for mixed method approaches to both researching and teaching L2 vocabulary. To develop a more complete picture of the overall vocabulary input made available through classroom talk, further research is needed, beyond such small and focused studies.

***Wordlists and reference corpora***

It is not controversial that words taught to early L2 learners should be ‘useful’ (Nation, 2020). But given the slow rate of vocabulary acquisition in any classroom (Miralpeix, 2020), choices and prioritisation are needed. In any real world curriculum there is likely to be a mix of high frequency vocabulary found in any register of the target language, and more topic-specific words relevant to students’ presumed interests. Thus for example, vocabulary instruction in primary classrooms commonly reflects topic choices such as ‘food’ and ‘animals’ (Szpotowicz, 2009). But what is an appropriate balance between these two dimensions? And how easy is it to achieve this? It has been shown that the lexical choices made by L2 textbook writers, for example, may be quite idiosyncratic, even when they claim to be informed by reference corpora of the relevant language (see e.g. Cheng & Lam, 2022; Suquet, 2018). Nation (2016) discusses the varied issues involved in creating frequency-based target word lists for the language classroom, starting with the challenge of defining the ‘words’ which will be counted, and the selection – or development – of appropriate reference corpora. Jones and Durrant (2022) concur that corpora which guide wordlist development should reflect appropriate registers of language, and also suggest the use of judgement to complement corpus-based frequency, e.g. in teaching lexical ‘sets’ such as days of the week or seasons, where corpus frequency can be quite variable. (*Le français fondamental* supplemented frequency-based lexis on similar principles, promoting the idea of *disponibilité* or ‘availability’ to include names of common objects etc. which might nonetheless be relatively rare in corpora: Michéa, 1953). A complication when developing wordlists for young learners is the lack of specialised reference corpora (unlike for example in the world of English for Academic Purposes: Coxhead, 2020). However, the Council of Europe has promoted the development of ‘Reference Level Descriptions’ (RLDs) for the Common European Framework of Reference for Languages (CEFR), including wordlists which draw on learner corpora and curricula at the different levels of the Framework, and where Level A1 is clearly relevant for beginners and early learners. Thus for example, the *English Profile* produced by Cambridge English Language Assessment (CELA: 2015) provides wordlists for learners at all CEFR levels, including one of just under 800 items for Level A1. These English wordlists are derived from learner corpora created through the assessment programmes of CELA itself. Beacco and associates have created RLDs for French from A1 to B2 (see e.g. Beacco & Porquier, 2007 for Level A1, which proposes a list of c 750 lexical ‘elements’), though these French RLDs are based on expert judgement rather than on corpus data (Pinter & François, 2020).

For the purposes of our study, we selected two corpus-based wordlists which have been developed to support the teaching and learning of L2 French. The first is the *Frequency dictionary of French* of Lonsdale and Le Bras *(LLB*: 2009). This dictionary is based on a specially constructed corpus of 23 million words (half spoken, half written), using a range of everyday adult material such as recorded conversations, parliamentary debates, film subtitles, literature, and news media. The corpus was tagged for parts of speech and lemmatised; the eventual list of ‘the 5,000 top lemmas in French’ (p. 6) was derived primarily from the summed frequency of lemmas’ variant forms, moderated by their dispersion across different domains within the corpus (p. 5). The wordlists recently proposed for GCSE French are largely based on the most frequent lemmas in the *LLB* list.

The second wordlist is *FLELex*, a resource developed by researchers at the University of Louvain (François et al., 2014; Pintard & François, 2020). This list is based on a much smaller 770,000-word corpus extracted from textbooks and graded readers designed for learners of L2 French. In terms of difficulty, the materials included in the corpus were intended for learners at all six levels of the CEFR. As in *LLB*, the texts contributing to *FLELex* were tagged for parts of speech and lemmatized, producing a resulting total list of just over 14,000 lemmas (TreeTagger version). Frequency was calculated similarly to *LLB,* taking account of dispersion as well as raw frequency. Of course, these lemmas appeared with varying frequency at different CEFR levels (see Table 1 for examples).

[insert Table 1 about here]

For pedagogical purposes, the Louvain researchers wanted to assign the lemmas of *FLELex* to a particular CEFR level, i.e. to create level-specific wordlists. Initially approaches were tried in which a lemma was assigned the level of the text(s) in which it first occurred, or the level at which it had highest frequency. These approaches are somewhat problematic, e.g. the first approach would place the French verb *abandonner* [to abandon] at A1, while the second would place it at B1. In later work, however, a more sophisticated approach has been adopted. Pintard and François (2020) took the Beacco and Porquier (2007) Reference Level Descriptors as a starting point, and used machine learning to infer CEFR levels for all lemmas in *FLELex* from their frequency distributions. This current version of *FLELex* is downloadable from [https://cental.uclouvain.be/cefrlex/*FLELex*/download/](https://cental.uclouvain.be/cefrlex/flelex/download/) . The machine learning procedure described by Pintard and François (2020) allocates lemmas to levels in a more sophisticated way (so that e.g. *abandonner* is allocated to the intuitively more satisfactory level A2). However it leads to uneven allocation of lemmas to particular CEFR levels, as seen in Table 2. From the perspective of this research project, the main question of interest was how the lexis of teacher speech would map onto the various CEFR levels identified in *FLELex*, as well as to the frequency bands identified in *LLB*.

[Insert Table 2 about here]

**Research questions**

The research questions addressed in this paper are:

1. What are the characteristics of the L2 lexis made available to early learners of French in their teacher’s classroom talk?
2. How far does the lexis of classroom talk reflect the lexis of frequency-based wordlists grounded in corpora designed to inform the teaching/ learning of L2 French?
3. What factors other than L2 frequency drive teachers’ lexical selection?

**Materials and procedure**

The ‘Young Learners’ corpus comprises 33 L2 French lessons taught to an intact class of 7-8 year old monolingual Anglophone learners in an English primary school. The learners were complete beginners and had no contact with French outside the classroom. The corpus was created as part of the longitudinal study ‘Learning French from ages 5, 7 and 11: An investigation into starting ages, rates and routes of learning amongst early foreign language learners’ (Myles et al., 2012). The study received overall ethical approval from the Research Ethics Committee of the Faculty of Humanities and Social Sciences, University of Newcastle. The lessons were delivered by a specialist visiting teacher, using an oracy-led bilingual approach. (That is, English was used substantially alongside French, as an additional means of classroom management and metalinguistic commentary.) All lessons were videorecorded and fully transcribed in CHAT format (MacWhinney, 2000). The teacher’s utterances in French were part-of-speech (POS) tagged using CLAN software with the current French version of the MOR program, available at <https://talkbank.org/morgrams/>, plus manual correction.

To address Research Question 1, the total number of tokens for each lemma identified and POS-tagged by French MOR in the transcribed teacher speech was counted. Distributions of items within selected word classes were examined over the complete lesson sequence. Any changes in the distribution of word classes over time were explored by comparing selected lessons (Lessons 1, 10, 20, 30).

To address Research Question 2, the teacher lexis was compared first with the reference corpus of Lonsdale and Le Bras (*LLB*), and secondly with *FLELex* (TreeTagger version). To prepare the data for comparison with *LLB*, the teacher wordlist was slightly revised so as to match parts of speech with the lexical units used by *LLB.* (For example, French MOR distinguishes between the use of *avoir* and *être* as main verbs and as auxiliary verbs, however *LLB* does not do this.) The analysis tool MultiLingProfiler developed by NCELP (Finlayson et al., 2021) was used to assign items from this revised teacher wordlist to the five frequency bands (1k to 5k) of *LLB.* For the comparison with *FLELex*, the teacher wordlist was annotated with the CEFR levels assigned to individual items by *FLELex*.

To address Research Question 3, the influence of curriculum topics and of classroom management language on lexical selection was explored through additional qualitative analysis.

**Results**

***Characteristics of teacher lexis***

The teacher taught bilingually, using both French and English for classroom management and for linguistic explanations. In French, over the course of the 33 lessons, she produced 652 lemmas (word types)/44,316 word tokens*.* Table 3 shows the distribution of these tokens in the total teacher corpus, for selected word classes.

[insert Table 3 about here]

 In her speech, frequency for individual tokens was highest for certain function words:

* Definite article *le/la/les/l’*
* Indefinite article *un/une/des*
* Pronoun *je* [I]
* Verb *avoir* [to have]
* Verb *être* [to be]
* Discourse marker/exclamation *bien* [good, well].

As far as content words were concerned, high frequency items were typically connected to curriculum topics such as animals, the body, colours, food and drink etc. Clusters of words connected to these topics were the focus of intentional instruction (see e.g. Figure 1 for examples). Other items owed their frequency to repeated use in songs and stories such as *La chenille qui fait des trous* [The very hungry caterpillar].

A considerable group of items had only one token (n=98), or two tokens (n=56). Such items sometimes occurred to provide variation in work on particular topics (e.g. *araignée* [spider], during the ‘animals’ topic). Others were not a focus of instruction but might be used incidentally, for example to personalise exchanges with individual students:

*Tatouage* [ tattoo] (Lesson 16, 3 tokens)

TEAF: *bonjour Albertine\** [hello Albertine].

CHIF: *bonjour xxx* [hello xxx].

TEAF: *tu as un tatouage* [have you got a tattoo]?

CHIE: xxx .

TEAF: *fais voir le tatouage* [let’s see the tattoo]!

TEAF: *Albertine je peux voir ton tatouage s’il te plaît* [Albertine can I see your tattoo please]?

\*Children’s names are pseudonyms

Others might be used in classroom management talk:

*Colle* [glue] (Lesson 24, 2 tokens)

TEAE: do you have a glue stick I could borrow for a second? we're going to need glue sticks but I'm going to borrow one just while I model what we're doing, sorry, I wasn't very well organised.

TEAF: *merci, vous prenez le bâton de colle d'accord* [thanks, you take the glue stick right]?

CHIE: glue stick.

TEAF: [demonstrating] *oui, vous mettez la colle dans le centre du fruit, comme ça regardez, comme ça, et puis vous tirez un tout petit bout de papier comme ça et* ... [yes, you put the glue in the middle of the fruit, like that look, like that, and then you take a very small bit of paper like that and …]

CHIE: scrunch it up?

TEAF: *oui Gabriel, il faut coller comme ça* [yes Gabriel, you must stick it like that].

A check was carried out to explore whether the distribution of word classes changed over time, by examining four separate lessons (1, 10, 20, 30). Findings are presented in Table 4, for selected word classes. The table shows that nouns and verbs are the most common syntactic categories throughout the lessons. Variations in the frequency of different word classes between lessons are partly due to topic/ text selection; thus for example in lesson 1, a particular song was rich in adjectives, and in lesson 10 numerals were being practised systematically.

[insert Table 4 about here]

Figure 1 further illustrates the relationship between topic and lexical frequency, showing the distribution of a particular noun group. In lessons 13-16, ‘pets’ were a topic focus; these lessons also introduced a song about pets, which was sung again in lessons 23 and 33. The figure shows how these items were modelled intensively when in topic focus, but with no systematic further practice once the topic focus changed. In contrast, Figure 2 shows the distribution of selected conjunctions over the complete lesson sequence, showing much more even patterns of occurrence. These items were rarely the focus of intentional instruction, but inevitably occurred frequently in the teacher’s L2 management language and in texts of all types (songs, stories etc).

[Insert Figure 1 about here]

[Insert Figure 2 about here]

***Comparisons with reference corpora***

As explained earlier, the wordlist generated by the CLAN program FREQ from the classroom corpus of teacher speech and tagged for parts of speech using the program MOR was revised to match the lemmas of the corpus-based *LLB* reference wordlist. This process produced a teacher wordlist of 652 lemmas. The analysis tool MultiLingProfiler was used to tag each lemma on this revised list with the frequency band to which it is assigned by *LLB.* Figure 3 shows the resulting distribution.

[Insert Figure 3 about here]

Figure 3 shows that 389/652 lemmas found in teacher speech (or 59.6%) fell in the combined 1k and 2k bands of *LLB.* Given the high frequency with which many of these lemmas were used, 87.4% of the individual tokens in teacher speech also fell within these bands. Further smaller groups of lemmas fell in the 3k, 4k and 5k bands, so that 82.5% of the lemmas used by the teacher overall (and 95.1% of tokens) can be found in *LLB.* However 115 lemmas (17.6%) do not appear at all in *LLB.*

What is the nature of these ‘rarer’ words? Table 5 shows that many of them are nouns which relate to specific curriculum topics, notably food and drink, where no less than 60 different lemmas are found in teacher speech, but only 22 of these occur in *LLB* at all (and only 7 are found in *LLB* 1k-2k). Others relate to the classroom setting, and groups of ‘rarer’ adjectives deal with colours on the one hand, and emotions on the other.

[Insert Table 5 about here]

A number of these non-*LLB* ‘rarer’ words used by the teacher can be related to Michéa’s (1953) concept of *disponibilité* [availability], i.e. they name common everyday objects but are likely to be underrepresented in reference corpora; this may apply to colour terms such as *brun* or *marron*, or to items in the ‘food’ grouping such as *frites* [chips] for example. The context of the primary school classroom and age-appropriate activities have their own register and associated everyday vocabulary, e.g. the verbs *colorier* [to colour] or *tricher* [to cheat]. In several categories, inspection of the contexts where ‘rarer’ words occur shows that they are extensions of topic vocabulary which were introduced in response to children’s personal interests and characteristics, e.g. to give authentic descriptions of hair and eye colour (e.g. *châtain, roux, noisette*).

A second comparison of the 652-item teacher wordlist was then carried out with the *FLELex* wordlists for different CEFR levels. Figure 4 shows the outcome.

[Insert Figure 4 about here]

The analysis showed that 417 lemmas or 63.9% of the teacher wordlist were tagged as occurring in the CEFR A1 wordlist (of 1,247 words). A further 39 lemmas (6.0%) were found in the CEFR A2 list (of 678 words); overall therefore, just under 70% of the word types used by the teacher could be related to CEFR A1 or A2. A recalculation for tokens rather than types showed that 82.7% of the tokens in teacher speech belonged to these levels. A further 22.2% of the word types (comprising 10.3% of tokens) was tagged as belonging to other CEFR levels, leaving 52 word types (8.0%) as absent from *FLELex*.

As described earlier, the two reference wordlists *LLB* and *FLELex* are derived from different corpora (general for *LLB*, textbook-specific for *FLELex*). The wordlist of *LLB* has been limited to the 5,000 most frequent lemmas found in the corpus, divided into 1k bands, while *FLELex* ranks all 14,000+ lemmas found in its corpus into 6 different bands; these internal bands have been derived in somewhat different ways (frequency plus dispersion for LLB, frequency plus dispersion plus input from the French RLD of Beacco et al. for *FLELex*). The most relevant overall comparison relates to the two starting bands for each wordlist (the 2,000 words of the 1k and 2k bands for LLB, and the 1,925 words of the A1 and A2 levels for *FLELex*). For these bands, the teacher wordlist shows considerable correspondence with both lists. It does appear that the lemmas of *FLELex* are a somewhat better match for the teacher wordlist than *LLB*, given that 69.8% of word types in teacher speech are found in *FLELex* A1/A2, while only 59.6% is found in *LLB* 1k/2k. This is however not very surprising, given that *FLELex* uses a reference corpus more closely tailored to the learner group. This point is illustrated in Table 6, which lists all animal names occurring in teacher talk (n=30), together with their frequencies. (The most frequent items on the list were mostly linked to the curriculum topic of ‘pets’ and the focus of intentional instruction; exceptions such as *chenille, papillon* or *tortue* were found in frequently recycled stories. The least frequent items were mostly used as distractors in a multiple choice guessing game.) The table shows that of this list, 11 items are found in *FLELex* A1-A2, while only three are found in *LLB* 1-2k. Once token frequency is taken into account, however, the apparent difference between the two sets of reference lists is very much reduced, given that LLB 1k/2k and FLELex A1/A2 each contribute well over 80% of tokens in teacher speech.

[Insert Table 6 about here]

**Discussion**

It is sobering to appreciate the limited nature of the target language input made available to learners through a sequence of hour-long lessons equivalent to a full year of instruction with a specialist French teacher, in the primary classroom. The lesson sequence analysed here exposed the learners to just c650 lemmas/44,500 tokens of French in total. The need to maximise the effectiveness/usefulness of this exposure is evident, and the curriculum and pedagogy adopted in several ways promoted this. The focus on topics thought to be of interest for children, and intentional multimodal instruction in vocabulary relevant to these topics, were very clear. Songs and stories were also revisited, offering repeated exposure to relevant lexis (including functional lexis) in a wider context. The teacher’s consistent use of French alongside English for some classroom management purposes and to personalise interactions with individual children also enriched the L2 input with consistent exposure to common function words (see Figure 2) and also to context-relevant content words. There were some tensions however between the richness of input, and opportunity for recycling/ spaced encounters which should be expected to promote learning (Marsden & Hawkes, 2023); this can be seen for example in the treatment of animal vocabulary, where some items received only massed practice (see Figure 1), and others occurred with frequencies too low for much learning to be expected. Was it worthwhile, for example, to introduce *araignée* [spider] in a planned activity, but only twice? (Of course, the cognate nature of some L2 vocabulary must also be acknowledged, and this may boost receptive knowledge of items such as *gorille, lion* or *dinosaure*, even if not recycled: Cobb, 2000; Szpotowicz, 2009.)

Comparisons with two corpus-based wordlists designed in somewhat different ways to support the L2 learning of French showed that a substantial amount of teacher lexis comprised a ‘common core’ of high frequency content and function words, which were also found in the lowest two bands of each reference list (i.e. the 1k and 2k bands in the case of *LLB*, and the CEFR A1 and A2 levels of *FLELex*). Thus c 60% of lemmas in teacher speech were also found in *LLB* 1k-2k, and c70% were also found in *FLELex* A1-A2. (And as previously noted, once token frequency was taken into account, these figures rose to 87.4% in the case of LLB 1k-2k, and to 82.7% for FLELex A1-A2.) The closer match of the teacher wordlist to *FLELex* A1-A2 is clearly due to its underlying, more specialised reference corpus of L2 textbook content, in contrast to the texts of the corpus underlying *LLB* which were aimed at an adult L1 French audience. *FLELex* is also influenced by its links with the French RLD, one consequence being somewhat greater weight given to *disponibilité*; thus for example, 11 of the 15 colour words used by the teacher are found in the *FLELex* AI-A2 lists. A considerable gap remains however between the teacher wordlist and *FLELex* A1-A2; again this can largely be explained by curriculum focus, e.g. on foods, where only 27/60 teacher wordlist items were found in *FLELex* A1-A2. There is clearly a discussion to be had, in the future design of wordlists and curriculum planning, around the balance to be struck in early vocabulary instruction between children’s immediate interests and the positive motivation associated with these, and longer term ‘usefulness’ beyond the classroom of that vocabulary which is intentionally taught.

This study has a number of implications for curriculum planning, the design of wordlists, and for vocabulary pedagogy in the primary classroom, which can be summarised as follows:

* Frequency is an important principle, which should influence the design of wordlists from the earliest stages of instruction;
* Corpora underpinning any wordlists should be chosen to reflect as nearly as possible, learners’ age and interests (Jones & Durrant, 2022), and general corpus data could helpfully be triangulated with datasets of teacher talk;
* A topic-based curriculum including ‘rarer’ topic-specific vocabulary and lexical sets is not incompatible with strong exposure to high-frequency words;
* The vocabulary associated with a particular topic should include a not-too-large core of a size which can be both explicitly taught and then regularly re-visited;
* Supplementary, personalised vocabulary should be available, but children need to develop individual responsibility for learning/ retaining it;
* Cognates are a bonus in vocabulary learning (Szpotowicz, 2009), which should be explicitly noticed and capitalised on;
* Classroom management through the target language is of great importance for vocabulary development, a) because it provides regular incidental exposure to general high frequency vocabulary across all major word classes; and b) because it introduces and routinely rehearses important, age-appropriate vocabulary (as we saw in this study, for classroom vocabulary and also for emotions vocabulary).

Regarding the limitations of this study and future research: This paper reports a re-analysis of a small naturalistic lesson corpus of beginner French. We have reported elsewhere on aspects of the children’s L2 development, i.e. the development of receptive vocabulary and the earliest emergence of L2 grammar (Mitchell & Myles, 2019). Among other findings, our earlier work showed the importance of both input frequency and multimodality for lexical development. These are clearly facilitated both by intentional instruction around motivating topics, and by a consistent commitment to L2 use (albeit alongside L1) for classroom management and interpersonal relations. However more focused, experimental designs would be required to address questions of how other characteristics of teachers’ L2 input such as lexical selection, or mass vs. distributed practice, impact on L2 development over time with classroom learners of this age. Given the increased focus on vocabulary development in policy discussions, we look forward to such studies which will provide essential underpinning for the development of both curriculum and pedagogy in primary languages.

**Acknowledgements**

We are grateful to the children, their French teacher and their school, who admitted us into their classroom world to collect the data used in this study. We also acknowledge the other researchers who contributed significantly to the project: Annabelle David, Christophe dos Santos, Kevin McManus, Sarah Rule. Our research depends on the public availability of two corpora, *LLB* and *FLELex* (University of Louvain, Belgium); our corpus analyses have been facilitated by the tools provided by the TalkBank project at Carnegie Mellon University, USA, and by the MultiLingProfiler project at York University, UK.

**References**

Beacco, J.-C. & Porquier, R. (2007). *Niveau A1 pour le français*. Paris: Didier.

Bloom, K.C. & Shuell, T.J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, 74(4), 245-248. DOI: [10.1080/00220671.1981.10885317](https://doi.org/10.1080/00220671.1981.10885317)

Butler, Y. G. (2019). Teaching vocabulary to young second- or foreign-language learners: What can we learn from the research? *Language Teaching for Young Learners, 1*(1), 4-33.

Cadierno, T., Hansen, M. B., Lauridsen, J. T., Eskildsen, S. W., Fenyvesi, K., & Hannibal Jensen, S. (2022). A multi-factor study of the development of English receptive skills by young Danish children. *Bilingualism: Language and Cognition, 25*(4), 566-578.

Cobb, T. (2000). One size fits all? Francophone learners and English vocabulary. *The Canadian Modern Language Review*, 57(2), 295-324.

Collen, I. (2022). *Language trends 2022: Language teaching in primary and secondary schools in England*. London: The British Council.

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *The Modern Language Journal, 93*(3), 336-353.

Conti, G. (2022). The real MARSEARS : How an EPI sequence truly unfolds (the "Conti method"). Retrieved from <https://gianfrancoconti.com/2022/05/18/the-real-marsears-how-an-epi-sequence-truly-unfolds/>

Coste, D. (2006). Français élémentaire, débats publics et représentations de la langue. *Documents pour l'histoire du français langue étrangère ou seconde, 36*.

Coxhead, A. (2020). Academic vocabulary. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 97-110). Abingdon: Routledge.

Dahl, A. (2015). Input and language competence in early start foreign language programmes. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 125-154). Berlin: Walter de Gruyter.

David, A., Myles, F., Rogers, V., & Rule, S. (2009 ). Lexical development in instructed L2 learners of French: Is there a relationship with morphosyntactic development?. In H. Daller, Malvern, D., Meara, P., Milton,J.,  Richards B., Treffers-Daller, J. (Ed.), *Vocabulary studies in first and second language acquisition: the interface between theory and application* (pp. 147-163). Basingstoke: Palgrave.

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94-112). New York: Routledge.

de la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary. *Studies in Second Language Acquisition,* *24*, 81-112.

Department for Education. (2022a). *French, German and Spanish: GCSE subject content*. <https://www.gov.uk/government/publications/gcse-french-german-and-spanish-subject-content>

Department for Education. (2022b). *GCSE subject content for modern foreign languages*. <https://www.gov.uk/government/publications/gcse-modern-foreign-languages>.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition, 23*(1), 171-185.

Erk, M., & Pavičić Takač, V. (2021). Teacher target language input and young learners' aural comprehension of English. *Explorations in English Language and Linguistics, 9*(1), 1-25.

European Education and Culture Executive Agency. (2023). *Key Data on Teaching Languages at School in Europe – 2023 Edition.* Brussels: Publications Office of the European Union. doi:doi/10.2797/529032.

Finlayson, N., Marsden, E., Anthony, L., Bovolenta, G., & Hawkes, R. (2021). MultiLingProfiler (Version 2) [Computer software]. University of York. <https://www.multilingprofiler.net/>

François, T., Gala, N., Watrin, P. & Fairon, C. [*FLELex*: A graded lexical resource for French foreign learners.](http://www.lrec-conf.org/proceedings/lrec2014/pdf/1108_Paper.pdf) In *the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavik, Iceland, 26-31 May.

Gardner, R. (2019). Classroom interaction research: the state of the art. *Research on Language and Social Interaction, 52*(3), 212-226.

González-Fernández, B., & Schmitt, N. (2017). Vocabulary acquisition. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 280-298). Routledge.

Gougenheim, G., Michéa, R., Rivenc, P., & Sauvageot, A. (1964). *L’élaboration du français fondamental (1er degré).* Paris : Didier.

Hall, J. K. (2022a). L2 classroom interaction and its links to L2 learners’ developing L2 linguistic repertoires: A research agenda. *Language Teaching, 55*(1), 100-115. doi:10.1017/S0261444820000397

Hall, J. K. (2022b). L2 classroom input and L2 positionally sensitive grammars: The role of information-seeking question sequences. *The Modern Language Journal, 106*(S1), 113-131.

Hulstijn J.H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In Robinson P. (Ed.), Cognition and second language instruction (pp. 258–286). Cambridge: Cambridge University Press.

Jones, M., & Durrant, P. (2022). What can a corpus tell us about vocabulary teaching materials? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics. 2nd Edition* (pp. 341-357). Abingdon: Routledge.

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning, 72*(1), 269-319. doi:<https://doi.org/10.1111/lang.12479>

Kimura, D., Malabarba, T., & Kelly Hall, J. (2018). Data collection considerations for classroom interaction research: a conversation analytic perspective. *Classroom Discourse, 9*(3), 185-204.

Kormos, J. (2020). How does vocabulary fit into theories of second language learning? In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 207-222). Abingdon: Routledge.

Kunitz, S., Markee, N., & Sert, O. (2021). *Classroom-based conversation analytic research : Theoretical and applied perspectives on pedagogy*. Cham, Switzerland: Springer Nature.

Leona, N. L., van Koert, M. J. H., van der Molen, M. W., Rispens, J. E., Tijms, J., & Snellings, P. (2021). Explaining individual differences in young English language learners’ vocabulary knowledge: The role of extramural English exposure and motivation. *System, 96*, 102402.

Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching, 51*(3), 285-329.

Lonsdale, D. & Le Bras. Y. (2009). *A frequency dictionary of French: Core vocabulary for learners.* Routledge.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. 3rd Edition. Lawrence Erlbaum Associates.

Marsden, E., & Hawkes, R. (Forthcoming 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.

Michéa, R. (1953). Mots fréquents et mots disponibles: Un aspect nouveau de la statistique du langage. *Les langues modernes*, 47(4), 338–344.

Milton, J. (2008). French vocabulary breadth among learners in the British school and university system: comparing knowledge over time. *Journal of French Language Studies, 18*(3), 333-348.

 Milton, J. (2022). Vocabulary denial and the false god of structuralism in Ofsted’s 2021 Curriculum Research Review for languages. *The Language Learning Journal, 50*(2), 156-171.

Milton, J., & Alexiou, T. (2012). Vocabulary input, vocabulary uptake and approaches to language teaching. *The Language Learning Journal, 40*(1), 1-5.

Milton, J., & Hopwood, O. (2021). Vocabulary loading in the *Studio* textbook series: a 40% decline in the vocabulary input for French GCSE. *The Language Learning Journal*, 1-17.

Miralpeix, I. (2020). L1 and L2 vocabulary: Size and growth. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 189-206). Abingdon: Routledge.

Mitchell, R., & Myles, F. (2019). Learning French in the UK setting: Classroom engagement and attainable learning outcomes. *Apples: Applied Language Studies, 13*(1), 69-93.

Mitchell, R., & Rule, S. J. (2022). Learning vocabulary in the primary languages classroom: What corpus analysis can tell us. In K. McManus & M. Schmid (Eds.), *How special are early birds? Foreign language teaching and learning (EuroSLA Studies 6)* (pp. 37-61). Berlin: Language Science Press.

Myles, Florence. 2017. Learning foreign languages in primary schools: is younger better? *Languages, Society and Policy* 1(1). 1– 8.

Myles, F., & Mitchell, R. (2012). *Learning French from ages 5, 7 and 11: An investigation into starting ages, rates and routes of learning amongst early foreign language learners. ESRC End of Award Report, RES-062-23-1545*. Swindon: Economic and Social Research Council.

Nation, P. (2016). *Making and using word lists for language learning and testing*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Orosz, A. (2009). The growth of young learners' English vocabulary size. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 181-194). Bristol: Multilingual Matters.

Nation, P. (2021). Is it worth teaching vocabulary? *TESOL Journal, 12*(4), e564.

Pintard, A. and François, T. (2020). [Combining expert knowledge with frequency information to infer CEFR levels for words](https://aclanthology.org/2020.readi-1.13/). In *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI)* (pp. 85-92).

Pinter, A. (2006). Teaching young language learners. Oxford: Oxford University Press.

Plonsky, L., & Loewen, S. (2013). Focus on form and vocabulary acquisition in the Spanish L2 classroom. *Language, Interaction and Acquisition, 4*(1), 1-24.

Reder, S., Harris, K., & Setzler, K. (2003). The multimedia adult ESL learner corpus. *TESOL Quarterly, 37*(3), 546-557.

Shabaka-Fernández, S. (2023). The effect of teacher language use in Spanish EFL classrooms. *The Language Learning Journal, 51*(2), 161-177.

Sharpe, K. (2001). *Modern foreign languages in the primary school: The what, why and how of early MFL teaching*. London: Kogan Page.

Smith, S., & Conti, G. (2016). *The language teacher toolkit*: Createspace Book Publishing.

Suquet, P. (2018). Le contenu linguistique des manuels de français pour les apprenants du collège. [The language content of textbooks of French for lower secondary school students] *Synergies Europe,*(13), 139-146.

Szpotowicz, M. (2009). Factors influencing young learners’ vocabulary acquisition. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 195-211). Clevedon: Multilingual Matters.

Thorne, S. L., & Hellermann, J. (2015). Sociocultural approaches to expert–novice relationships in second language interaction. In N. Markee (Ed.), *The handbook of classroom discourse and interaction* (pp. 281-297). New York: Wiley Blackwell.

Tschichold, C. (2012). French vocabulary in *Encore Tricolore*: do pupils have a chance? *The Language Learning Journal, 40*(1), 7-19.

Webb, S. (Ed.) (2020). *The Routledge handbook of vocabulary studies*. Routledge.

Zhang, P. (2022). How does repetition affect vocabulary learning through listening to the teacher’s explicit instruction? The moderating role of listening proficiency and preexisting vocabulary knowledge. *Language Teaching Research, Online First,* 13621688221140521.

Zhang, P., & Graham, S. (2020). Vocabulary learning through listening: Comparing L2 explanations, teacher codeswitching, contrastive focus-on-form and incidental learning. *Language Teaching Research, 24*(6), 765-784.

Table 1: Sample lemma frequencies in *FLELex* texts at different CEFR levels (From François et al., 2014)



Table 2: *FLELex*/Beacco lemmas by CEFR level

|  |  |
| --- | --- |
| CEFR level | No. lemmas matched to level |
| A1 | 1,247 |
| A2 | 678 |
| B1 | 1,751 |
| B2 | 5,087 |
| C1 | 3,164 |
| C2 | 2,313 |

Table 3: Distribution of lexical tokens in teacher speech by selected MOR syntactic category (interjections excluded)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Syntactic category** | **Tokens** | **% of total** | **Syntactic category** | **Tokens** | **% of total** |
| Adjectives | 1804 | 4.1 | Numerals | 3820 | 8.6 |
| Adverbs | 2377 | 5.4 | Prepositions | 1437 | 3.2 |
| Auxiliary *être* [to be]\* | 29 | 0.1 | Subject pronouns | 3503 | 7.9 |
| Auxiliary *avoir* [to have] | 290 | 0.7 | Object pronouns | 532 | 1.2 |
| Conjunctions | 1487 | 3.4 | Verb *être* [to be] | 357 | 0.8 |
| Determiners | 5050 | 11.4 | Verb *avoir* [to have] | 385 | 0.9 |
| Nouns | 8070 | 18.2 | Other lexical verbs | 4332 | 9.8 |
| Negators (excl. *non*) | 610 | 1.4 | Interrogative words | 226 | 0.5 |

\*The MOR programme distinguishes between use of *avoir* and *être* as main and auxiliary verbs, but this is not the case in *LLB*

Table 4: Distribution of selected word classes in example lessons (tokens)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Less-on** | **Adjectives** | **Adverbs** | **Conjunct-ions** | **Nouns** | **Numerals** | **Preposit-ions** | **Lexical verbs** |
|  | No. | **%** | No. | **%** | No. | **%** | No. | **%** | No. | **%** | No. | **%** | No. | **%** |
| 1 | 62 | **6.1** | 90 | **8.8** | 69 | **6.7** | 116 | **11.6** | 7 | **0.7** | 36 | **3.5** | 384 | **37.6** |
| 10 | 8 | **0.8** | 39 | **3.9** | 39 | **3.9** | 262 | **26.3** | 140 | **14.1** | 24 | **2.4** | 148 | **14.9** |
| 20 | 20 | **2.1** | 40 | **4.2** | 76 | **8.0** | 419 | **44.3** | 95 | **10.1** | 4 | **0.0** | 110 | **11.6** |
| 30 | 15 | **1.6** | 41 | **4.4** | 46 | **4.9** | 253 | **26.9** | 52 | **5.5** | 65 | **6.9** | 132 | **14.0** |

Table 5: The lexis of selected topics: Non-*LLB* items in teacher speech (types)

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Total items in T talk** | **Items found in LLB** | **T talk items not found in LLB (types, selected examples)** |
| Food & drink | 60 | 22 | *Ananas* [pineapple], *banana* [banana], *carotte* [carrot], *casse-croute* [snack], *cerise* [cherry], *fraise* [strawberry], *framboise* [raspberry], *frites* [chips], *hamburger* [hamburger]… |
| Animals | 30 | 19 | *Dinosaure, hamster, insecte, lapin* [rabbit], *lion, papillon* [butterfly], *poussin* [chick], *serpent* [snake], *tortue* [tortoise], *vache* [cow] *…* |
| Classroom routines & crafts | 96 | 16 | *Autocollant* [sticker], *challenge* [challenge], *colle* [glue], *colorier* [to colour], *stylo* [pen], *(danse) solo, taille-crayon* [pencil sharpener], *tampon* [stamp], *tricher* [to cheat]… |
| Body details | 27 | 8 | *Bague* [ring], *bracelet* [bracelet], *menton* [chin], *moustache* [moustache], *ongle* [nail], *sourcil* [eyebrow], *t-shirt* [t-shirt], *tatouage* [tattoo]… |
| Colours | 15 | 8 | *Brun* [brown], *châtain* [chestnut], *marron* [brown], *noisette* [hazel], *roux* [red/ginger], *violet* [violet]… |
| Emotions | 29 | 8 | *Barbant* [boring], bisou [kiss], *dégoûtant* [disgusting], *délicieux* [delicious], *laid* [ugly], *mignon* [cute], *moche* [ugly]… |

Table 6: Animal names in teacher talk, *FLELex* and *LLB*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Item** | **Frequency in T talk** | ***FLELex* A1-A2** | ***FLELex* B1-C2** | ***LLB* 1-2k** | ***LLB* 3-5k** |
| *chat* [cat] | 158 | √ |  |  | √ |
| *chien* [dog] | 124 | √ |  | √ |  |
| *poisson* [fish] | 124 | √ |  |  |  |
| *lapin* [rabbit] | 101 | √ |  |   |   |
| *chenille* [caterpillar] | 96 |  | √ |   |   |
| *oiseau* [bird] | 92 | √ |  |  | √ |
| *hamster* [hamster] | 77 |  | √ |  |  |
| *souris* [mouse] | 60 |  | √ |  | √ |
| *cheval* [horse] | 52 | √ |  |  | √ |
| *animal* [animal] | 38 | √ |  | √ |  |
| *tortue* [tortoise] | 36 |   |   |   |  |
| *papillon* [butterfly] | 35 |  | √ |   |  |
| *cochon d’inde* [guinea pig] | 29 |   |   |   |  |
| *cochon* [pig] | 15 |  | √ |  | √ |
| *poussin* [chick] | 14 |  | √ |   |  |
| *rat* [rat] | 14 |  | √ |  | √ |
| *serpent* [snake] | 14 | √ |  |   |   |
| *éléphant* [elephant] | 11 | √ |  |   |   |
| *lion* [lion] | 11 |  | √ |   |   |
| *girafe* [giraffe] | 9 |   |   |   |   |
| *mouton* [sheep] | 3 | √ |  |  | √ |
| *vache* [cow] | 3 |  | √ |  | √ |
| *araignée* [spider] | 2 |  |  |  |  |
| *crocodile* [crocodile] | 2 |  | √ |  |  |
| *gorille* [gorilla] | 2 |  |  |  |  |
| *canard* [duck] | 1 |  | √ |  |  |
| *dinosaur* [dinosaur] | 1 |  | √ |  |  |
| *guêpe* [wasp] | 1 |  | √ |  |  |
| *insect* [insect] | 1 |  | √ |  |  |
| *poulet* [chicken] | 1 | √ |  |  | √ |

Figure 1: Distribution of selected ‘animals’ vocabulary in teacher talk (tokens per lesson)

Figure 2: Distribution of selected conjunctions in teacher talk (tokens per lesson)

Figure 3: Distribution of teacher lexis by *LLB* frequency bands (word types)



Figure 4: Distribution of teacher lexis by *FLELex* CEFR levels (word types)

