

Basic verbs in European languages

ÅKE VIBERG

UPPSALA UNIVERSITY

Submitted: 24/08/2025 Revised version: 05/02/2026

Accepted: 13/04/2026 Published: 23/07/2026



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

Abstract

This article presents a corpus-based approach toward a lexical typology of verbs in European languages, based on a comparison of the meanings of the most frequent verbs classified into semantic fields. The comparison is made in two steps. Among the twenty most frequent verbs, there is a set of grammatical verbs that top the list (with a few exceptions): ‘be’, ‘have’ and modals meaning ‘can’ and ‘must’. Among the lexical verbs, there are nine nuclear verbs with meanings that are distributed between the most basic verbal semantic fields (indicated within parentheses): ‘go’, ‘come’ (Motion); ‘give’, ‘take’ (Possession), ‘make’ (Production), ‘say’ (Verbal communication), ‘see’ (Perception), ‘know’ (Cognition), ‘want’ (Desire). Comparisons are made in two ways based on the rank order of the verbs according to descending frequency and according to a frequency measure, based on a logarithmic scale. As a second step, the comparison is extended to the 100 most frequent verbs. The result is compared to the inventories of simple verbs in some non-European languages with a very restricted number of such verbs. The frequency-based comparisons are complemented with the result of corpus-based contrastive studies of selected aspects of the semantic structure of the semantic fields to which the nuclear verbs belong.

Keywords: lexical typology; semantic fields; verbs; basic vocabulary; word frequencies; European languages; corpus-based contrastive study

1. Introduction

The general aim of this article is to find a model for comparing the basic structure of whole vocabularies in order to provide a framework for contrastive and typological

investigations of the semantic organization of lexical structure from a holistic perspective. The study focuses on verbs. In principle, the approach is corpus-based combining frequency counts of large corpora with corpus-based (or usage-based) studies of the basic words that have been identified and the semantic fields to which they belong. The use of frequency is motivated by the central role it plays for language processing and learning (Divjak 2019).

European languages have an open class of verbs which tends to contain something in the range of 10,000 well-established verbs (lexemes). Despite the vast number of words that are potentially available for describing events of various types, it turns out that a small number of verbs are used on most occasions. In Swedish, for instance, the twenty most frequent verbs cover 49% of all the occurrences of verbs in running text, the 50 most frequent verbs cover 60% and the 100 most frequent verbs 68% (Viberg 2006a: 105). Furthermore, there is a striking similarity across languages with respect to the meaning of these verbs. This was shown in Viberg (1993) by comparing the basic meanings of the 20 most frequent verbs in 11 European languages. There are nine lexical meanings that tend to be realized as one of the twenty most frequent lexical verbs in all European languages. These meanings are distributed between the most basic verbal semantic fields (indicated within parentheses): ‘go’, ‘come’ (Motion); ‘give’, ‘take’ (Possession), ‘make’ (Production), ‘say’ (Verbal communication), ‘see’ (Perception), ‘know’ (Cognition), ‘want’ (Desire). For the sake of simplicity, these verb meanings (and only these) will be referred to as “nuclear verbs”. Even when a verb with this meaning does not appear among the 20 most frequent verbs, it is still counted as a nuclear verb but as an exception. Some non-European languages lack direct equivalents of some of the nuclear verbs and the closest correspondences are analytical constructions (see, for example, Section 7). The appearance of ‘be’ and ‘have’ and modal verbs among the most frequent verbs is an areal-specific characteristic of European languages.

The nuclear verbs tend to have many hyponyms, to have characteristic and elaborate patterns of polysemy and to have a strong tendency to develop grammatical meanings. As a set, they cover most of the different types of argument structures characteristic of an individual language and in several respects serve as syntactic prototypes. They also play a central role in first and second language acquisition and in lexical regression such as attrition and certain types of language loss (Viberg 1993, 2002).

Frequency will be used to identify the most basic verbs in an individual language. In a second step, I try to answer the question to what extent basic verb meanings are

shared. Basic verbs that are shared with only a few languages are identified as language-specific. It would be possible to simply talk about the most frequent verbs, but it is convenient to refer somewhat more loosely to basic verbs. Basicness is not an absolute concept. A large part of this paper will deal with the 20 most frequent verbs, but reference will also be made to the 100 most frequent verbs, which are basic but to a lower degree. The cut-off point is in principle arbitrary. What is essential is to try to find the most informative cut-off point to answer a specific research question and then to stick to this cut-off point throughout the discussion. What is referred to as basic verbs has sometimes been referred to as core verbs or nuclear verbs in the literature. To avoid confusion, it is important to underline that the term “nuclear verb” in this paper will be used exclusively to refer to the specific set of nine verbs, which are basic in European languages and tend to belong to the 20 most frequent verbs.

2. Testing the generalization

At the time when the earlier study (Viberg 1993) was carried out, most frequency counts were based on corpora containing around one million words. As will be shown below, the general conclusions turn out to hold for the much larger corpora available today such as the British National Corpus (BNC) containing around 100 million words taken from a wide range of spoken and written genres. In the following, the conclusions from the earlier study will be tested against data from corpora that are available today. What is presented should be regarded as an exploratory study, since there is a need for further development of the methodology and more detailed analysis of each of the verbs that have been identified as basic. To my knowledge, however, there is no contrastive-typological study of basic vocabulary following the same kind of approach.

2.1. Language sample and data sources

The present study is restricted to European languages. The major source of data consists of the frequency lists or dictionaries in Table 1. (To simplify reference, all sources will be referred to as “lists”.) The lists are based on several million words from reasonably balanced corpora with a few exceptions. In principle, there are at least two examples of each of the Germanic, Romance, Slavic and Uralic languages. In addition, there is Greek and Welsh. Lithuanian is included as an example of a Baltic language

despite the restricted size of the corpus (1 million words) and the dominance of formal registers (scientific 32%, fiction 12%, official documents 10%, periodicals 46%). The only major groups that are missing are Albanian and Basque. Swedish and Finnish are included because data from my earlier contrastive studies based on these languages will be used to discuss problematic semantic issues, even though the Swedish corpus contains only one million words and the Finnish corpus only is based on News. Code in Table 1 refers to the abbreviated forms of the names of the languages that will be used in several of the tables to save space. Information is also given about the classification in Glottolog (Hammarström et al. 2026). In the column “Language”, the glottocode appears under the conventional English name of the language (see <https://glottolog.org/glottolog/language>; accessed 2026-07-06).

Code	Language	Size Million words	Composition	Reference
INDO-EUROPEAN				
Germanic				
ENG	English stan1293	100	Balanced Spoken & Written	BNC (Kilgarriff 1998)
GER	German stan1295	20	Balanced Spoken & Written	Jones & Tschirner (2020) 2nd edn
SWE	Swedish swed1254	1	Balanced Written	The Stockholm-Umeå Corpus (SUC 1997)
Romance				
FRE	French stan1290	23	Balanced Spoken & Written	Lonsdale & Le Bras (2009)
SPA	Spanish stan1288	20	Spoken, Fiction, Non- fiction	Davies (2006)
Greek				
GRE	Greek mode1248	47	Written language	Hellenic National Corpus (data obtained from Maria Gavrilidou December 2009 from the version that existed at that time)
Slavic				
CZE	Czech czec1258	20	Spoken, Fiction, Non- fiction, Newspaper	Čermák & Křen (2011)
RUS	Russian russ1263	150	Internet corpus	Sharoff et al. (2013)

(continued)

Code	Language	Size Million words	Composition	Reference
INDO-EUROPEAN				
Baltic				
LIT	Lithuanian lith1251	1	Written language (only 12% fiction)	Dažninis rašytinės lietuvių kalbos žodynas (Andrius Utka 2009)
Celtic				
WEL	Welsh wels1247	11	Spoken and written	Knight et al. (2020) https://doi.org/10.17035/d
URALIC				
Finnic				
EST	Estonian esto1258	15	Fiction, Press, Science (5 mill words each)	The Frequency List of Estonian Literary Language
FIN	Finnish finn1318	44	Press ("printed news")	Wiktionary:Frequency_lists/Finnish https://en.wiktionary.org/wiki/ (Accessed 2026-07-06)
Hungaric				
HUN	Hungarian hung1274	777	Fiction, Non-fiction, Spoken	Hungarian Gigaword Corpus (Oravecz et al. 2014)

Table 1. Language sample and data sources.

2.2. Methodological problems

The frequency lists are developed according to different criteria in many respects, but with few exceptions data are accepted as they are given in the sources without any attempt to compensate for such discrepancies.

2.2.1. Variation across genres/registers

Although the frequency lists are based on “balanced” corpora, it is problematic to decide which genres should be distinguished and what the proportions should be between various sub-corpora. It should also be remarked that most of the discussion in this paper will be restricted to the standard languages. Another restriction is that spoken language is not included in all corpora. This is particularly problematic when the result is compared with typological studies, which typically rely on spoken data. To some extent this can be compensated for by using data from subtitle corpora, which

are available for all European languages, although this has not been accounted for in the present study. This issue is discussed in Viberg (2020).

2.2.2. *Raw frequency and dispersion*

The frequency of a word varies in different types of text. How widely a word is spread across the documents in a corpus can be measured with various measures of dispersion. Even distribution of a word gives a high score, whereas a low score indicates that a word appears only within a restricted range of texts. The French list (Lonsdale & Le Bras 2009) shows the raw frequency of each word, whereas the rank order is based on a combination of the raw frequency and a measure of dispersion called deviation of proportions (DP) (Gries 2008). Ranking that takes dispersion into consideration is more accurate but is available only for some languages.

2.2.3. *Lemmatization*

The lexical items consist of single words but derived and compound words are sometimes treated differently in the lists. Morphological complexity forms a continuum: Simple root > Derived word > Compound word > Lexicalized phrase. Kopecka (2006) shows how derivation forms a continuum with respect to French motion verbs (e.g., *dé-rouler* ‘roll out’ vs. *sortir* ‘go = OUT’). On compound verbs, see Viberg (2017a) and on lexicalized phrases or idioms as a continuum, see Langlotz (2006).

In general, the identification of lemmas in the frequency lists has been followed. In some cases, the frequency lists treat derived forms in a way that reduces comparability. A clear case is the negation in Lithuanian which is represented as a prefix of the verb. Such forms are treated as lemmas in the frequency list, but in this paper the frequencies of the positive and the negative forms have been collapsed and are represented by the positive form of the verb. This appears to be unproblematic in most cases. In Russian and Czech, the frequencies of the perfective and imperfective forms of an aspectual verb pair are indicated separately but have been collapsed. This is not unproblematic but appears to improve comparability. Rank order on the other hand is indicated for the most frequent member of the pair.

2.2.4. Glossing

The identification of the meanings has been based primarily on the glosses given in the frequency dictionaries. For some languages, the source has been a frequency list that does not provide any glosses, and in that case the meaning has been identified with a bilingual dictionary (X to English). When several glosses have been provided, one meaning has been identified as the basic meaning. A further resource that has been used is the Word Alignment Database (WAD) included as a tool in OPUS (Tiedeman 2012). This tool takes a word form in one language as input and shows the word forms that appear in the translations according to a formula. This tool has been used frequently in this study, but it would take up too much space to account for all these searches. See Appendix 1 for an example. The use of translation to identify correspondences across languages in contrastive studies is discussed in Viberg (to appear).

2.3. Basing the comparison on frequency rank order

Table 2 shows the rank of the nine nuclear verbs together with BE, HAVE, CAN and MUST. Ranks higher than 20, which represent exceptions, are shaded.

	BE	HAVE	CAN	MUST	MAKE	GO	COME	GIVE	TAKE	SAY	SEE	KNOW	WANT
ENG	1	2	5	23	9	7	16	17	12	6	10	11	21
GER	1	2	4	6	9	12	11	7	26	5	14	13	8
SWE	1	2	3	16	9	11	6	15	13	8	10	19	14
FRE	1	2	3	7	4	10	18	9	8	6	14	12	11
SPA	1(3)	4(2)	6	15	5	8	26	10	21/30	7	9	11	12
GRE	1	2	4	8	7	18	27	9	13	3	12	34	11
CZE	1	2	3	5	14	6	15	10	19	9	12	4	8
RUS	1	---	2	ADJ	13	11	19	10	33	3	15	6	7
LIT	1	3	2	4	22	36	57	64	18	9	28	7	6
WEL	1	---	9	N	3	4	5	10	25	8	7	13	37
EST	1	---	4	(3)	6	8	5	12	11	7	14	9	15
FIN	1	---	4	?	6	20	5	12	11	8	27	30	13
HUN	1	---	(4)	2	6	22	23	10	7	5	9	4	18

Table 2. The ranking of shared basic verbs in 13 European languages.

Except for Lithuanian, at most two nuclear verbs have a higher rank than 20 and the deviation is not very high in any case. Another exception is that *MUST* is not realized as a verb but as a noun in Welsh and as an adjective in Russian. These peculiarities will be discussed below in Sections 3-9.

2.4. Basing the comparison on levels of the natural logarithm (LN)

An alternative to rank order as a basis for comparison is to base the comparison directly on the frequencies. First the raw frequencies are normalized and specified as frequencies per million words (pm). However, the problem remains to find natural cut-off points. One solution is to use a logarithmic scale and to that end the frequencies of the basic verbs are expressed as values of the natural logarithm (with the Latin acronym LN) in Table 3. (The base of LN is approximately equal to 2.718.) At the bottom of the table the result is summed up by an indication of the median and the highest and the lowest value for each verb.

	BE	HAVE	CAN	MUST	MAKE	GO	COME	GIVE	TAKE	SAY	SEE	KNOW	WANT
ENG	10.7	9.5	8.4	6.6	7.7	7.8	7.3	7.2	7.5	8.1	7.6	7.5	6.9
GER	10.1	9.4	8.6	7.9	7.6	7.4	7.5	7.7	6.4	8.0	7.2	7.2	7.6
SWE	10.1	9.5	8.9	7.1	7.9	7.8	8.1	7.2	7.7	8.0	7.9	6.9	7.3
FRE	10.0	9.8	8.1	7.6	8.4	7.7	6.8	6.8	7.0	8.0	7.2	7.3	7.4
SPA	9.8	8.6	8.3	7.0	8.3	8.2	6.7	7.6	6.5	8.3	7.3	7.5	7.3
GRE	9.4	9.1	7.9	7.6	7.7	6.6	6.4	7.4	6.9	8.1	6.9	6.3	7.1
CZE	10.3	8.6	7.7	7.1	6.9	7.1	6.3	7.0	6.2	7.7	6.6	7.4	7.0
RUS	9.4	---	8.2	ADJ	7.4	7.2	6.5	6.9	6.5	8.2	7.1	7.4	7.1
LIT	9.9	8.4	8.7	7.6	7.3	6.9	6.0	5.9	6.9	7.7	7.0	7.4	7.4
WEL	11.0	---	7.7	N	8.5	8.0	8.0	7.6	6.5	7.7	7.8	6.9	6.2
EST	10.7	---	8.3	(8.3)	8.0	7.8	8.1	7.5	7.5	7.8	7.3	7.6	7.3
FIN	11.0	---	8.4	?	8.3	7.0	8.4	7.4	7.4	8.0	6.9	6.8	7.3
HUN	9.3	---	(7.9)	8.1	7.3	6.6	6.5	6.9	7.1	7.7	7.0	7.9	6.8
<i>Med</i>	10.1	---	8.3	(7.1)	7.7	7.4	6.8	7.2	6.9	8.0	7.2	7.4	7.3
<i>High</i>	11.0	9.8	8.9	8.1	8.5	8.2	8.4	7.7	7.7	8.3	7.9	7.9	7.6
<i>Low</i>	9.3	8.4	7.7	6.6	6.9	6.6	6.0	5.9	6.2	7.7	6.6	6.3	6.2

Table 3. The LNs of shared basic verbs in 13 European languages.

BE has an LN around 10 (Median: 10.1. Lowest: 9.3 – Highest 11.0) followed by HAVE around 9 in the languages where the verb is established (8.4 – 9.8). CAN reaches above 8.0 in most cases (7.7 – 8.9), but MUST is more variable and has a somewhat lower value and is not realized as a verb in some languages. The rest of the verbs in Table 3 make up the nuclear verbs that tend to have an LN above 7 but below 8.0, which corresponds to a frequency between 1,100 and 3,000 per million words (pm). Thus, the use of LN shows how the most frequent verbs tend to appear within specific levels.

2.5. A brief look at the 100 most frequent verbs

The discussion will account systematically for all the nuclear verbs that tend to appear among the 20 most frequent verbs. In addition, several verbs with a rank between 21 and 100 will be commented on below when they belong to the same field as one of the nuclear verbs. Table 4 compares the LNs at various frequency ranks. The median LN for the verb that has rank 100 is 5.4, which is still relatively high and corresponds to a frequency close to 220 pm.

	BE	HAVE	Rank 5	Rank 10	Rank 20	Rank 50	Rank 100
English	10.7	9.5	8.4	7.6	6.9	5.9	5.4
German	10.1	9.5	8.0	7.6	6.6	6.0	5.2
Swedish	10.1	9.5	8.4	7.9	6.9	6.1	5.4
French	10.0	9.8	6.8	7.7	6.3	5.4	4.7
Spanish	9.8	9.2	8.3	7.7	7.0	5.8	5.5
Greek	9.4	9.1	7.8	7.2	6.6	6.0	5.5
Czech	10.3	8.6	7.1	7.0	5.7	5.0	4.5
Russian	9.4	---	8.2	6.9	6.5	5.9	5.4
Lithuanian	9.9	8.4	7.6	6.9	6.7	6.0	5.6
Welsh	11.0	---	8.0	7.6	6.5	6.1	5.4
Estonian	10.7	---	8.1	7.6	7.0	6.1	5.6
Finnish	11.0	---	8.4	7.4	7.1	6.5	6.0
Hungarian	9.3	---	7.7	6.9	6.6	5.8	5.2
Median	10.1	(9.4)	8.1	7.6	6.6	6.0	5.4
Highest	11.0	9.8	8.4	7.9	7.1	6.5	6.0
Lowest	9.3	8.4	7.1	6.9	5.7	5.0	4.5

Table 4. A comparison of the LNs at various frequency ranks.

The disposition of the rest of the paper is as follows. Section 3-9 will be devoted to a discussion of the set of shared basic verbs of European languages, organised in terms of semantic fields. This is followed by a couple of more general sections. The semantics of the nuclear verbs will be primarily discussed based on my own corpus-based contrastive studies. Other work will also be referred to but often rather briefly for reasons of space. More references are given in my earlier studies, which are based primarily on the English-Swedish Parallel corpus (Altenberg & Aijmer 2000) and on the Multilingual Parallel Corpus (MPC), which consists of extracts from 22 Swedish novels and their translations into English, German, French, Spanish and Finnish. Data from further languages will be taken from OPUS, in particular Subtitles and to a lesser extent from Europarl. Several methodological problems are associated with the use of translations. For my view on this subject, see Viberg (2016a).

3. BE and HAVE

Table 5 shows the frequency (pm) and the LN of the verbs meaning BE and HAVE in the sample of European languages.

Language	BE			HAVE		
	<i>form</i>	<i>LN</i>	<i>pm</i>	<i>form</i>	<i>LN</i>	<i>pm</i>
English	<i>be</i>	10.65	42396	<i>have</i>	9.53	13756
German	<i>sein</i>	10.08	23949	<i>haben</i>	9.43	12452
Swedish	<i>vara</i>	10.09	24094	<i>ha</i>	9.53	13826
French	<i>être</i>	10.02	22372	<i>avoir</i>	9.78	17610
Spanish	<i>ser</i>	9.84	18710	<i>tener</i>	8.58	5332
	<i>estar</i>	8.58	5321	<i>haber</i>	9.20	9848
Greek	<i>ίμε</i>	9.44	12539	<i>έxo</i>	9.08	8782
Czech	<i>být</i>	10.27	28850	<i>mit</i>	8.57	5271
Russian	<i>byt'</i>	9.36	11588	<i>(imet')</i>	6.96	1056
Lithuanian	<i>(ne)būti</i>	9.86	19187	<i>turėti</i>	8.41	4502
Welsh	<i>bod</i>	11.10	66137	WITH-possessive	---	---
Estonian	<i>olema</i>	10.66	42653	LOCATIVE	---	---
Finnish	<i>olla</i>	11.03	61736	LOCATIVE	---	---
Hungarian	<i>van</i>	9.33	11308	DATIVE	---	---

Table 5. BE and HAVE.¹

¹ In Lithuanian, the frequency of the positive form has been collapsed in Table 5 and the following tables with the frequency of the negative form of the verb, which is indicated with the prefix *ne-* and

BE-verbs have been studied extensively from a crosslinguistic perspective, at least since the six volumes edited by Verhaar (1967-1972). The seminal typological work is Stassen (1997), which is based on data from 210 languages, followed recently by Bertinetto et al. (2026a,b). In a similar way, HAVE, which is not realized as a simple verb in many languages, represents a type of predicative possession described in Stassen (2009), see also Heine (1997: 45–142), and, for Existentials, Sarda & Lena (2023) and Viberg (2010: 129–133)². The following discussion will be restricted to European languages.

3.1. BE- and HAVE-languages

BE has rank 1 in all European languages. In Spanish, there are two verbs, *ser* and *estar*, but *ser* can be regarded as the major BE-verb. As can be seen, the most important dividing line between European languages turns out to be whether there is a verb meaning HAVE or not. This motivates a special term HAVE-language for the languages that have such a verb (Isačenko 1974).

The European languages that are not HAVE-languages appear in the Eastern and Western parts of Europe and consist of Uralic languages such as Finnish, Estonian and Hungarian and of Celtic languages such as Welsh, Irish and Breton. The Baltic and Slavic languages are HAVE-languages, except Russian and Latvian, whereas Ukrainian and Belorussian are in a stage of transition according to Isačenko (1974). Somewhat unexpectedly, a verb glossed ‘have’ (*imet’*) is found among the twenty most frequent Russian verbs, but this verb is primarily used together with abstract nouns to express Abstract possession in phrases modelled on Western European languages: *imet’ talant* ‘have the talent’, *imet’ mesto* ‘take place’ (compare French *avoir lieu*).

According to Stassen (2013a), “a core area for the Have-Possessive is western and central Europe; however, we also find quite a few instances of this option in Africa and in the Americas”.

represented as a separate word in the frequency list. In Welsh, the frequency given for the infinitive *bod* has been collapsed with that of the 3SG form *mae* (2596), which is indicated separately in the frequency list, since only this form is used in the Existential and Possessive functions.

² For reasons of space, I will only be able to give reference to relevant works without being able to discuss them properly, which is unsatisfactory. I have also included rather many references to my own work where I have motivated my own analysis, since I’m not able to do that in this paper. I hope to be able to return to the topics of some of the individual sections in separate papers.

3.2. Contrastive patterns of *BE* and *HAVE*

The range of functions covered by *BE* and *HAVE* constitutes an important component of a contrastive/typological description of basic verbs in European languages. To give a simple overview of the basic functions in European languages, English examples are used in Table 6. (The range of grammatical functions varies across languages.)

Basic functions	Examples
Identity	The tall fellow over there <i>is</i> my brother.
Description:	
Nominal description	Peter <i>is</i> a teacher.
Adjectival description	Peter <i>is</i> tall.
Location	Peter <i>is</i> in the kitchen.
Existence	There <i>is</i> a fridge in the kitchen.
Possession	Peter <i>has</i> a car.
Grammatical uses	Examples
Cleft	It <i>is</i> Peter who <i>is</i> in the kitchen.
Progressive	Peter <i>is</i> snoring.
Passive	Peter <i>was</i> killed.
Perfect	Peter <i>has</i> arrived.
Modal: Obligation	Peter <i>has</i> to go now.

Table 6. Major functions of *be* and *have* in English.

The Identity construction is used to state the identity of two referents, which may have been treated as two independent referents in earlier discourse. In this construction, the NP in the complement is always referential, whereas it is non-referential in the Nominal description. Stassen (1997) refers to Adjectival description as Property predicate and Nominal description as Class-inclusion predicate.

In European languages, a predicate nominal in general requires a copula and the copula is always a verb. This situation is not unique to Europe, but outside Europe copulas can be realized as another type of word, most frequently a personal or demonstrative pronoun. Outside Europe, there are also languages where a Zero copula is the norm. A concentration of such languages appears in Australia and New Guinea according to Stassen (2013b), who also found across a world-wide sample that a Zero

copula was possible at least in some context in 175 languages but impossible in 211 languages.

The Grammatical uses listed in Table 6 occur in various combinations across European languages. For example, in a quantitative review, Čermáková et al. (2023) found a large difference with respect to the proportion of auxiliary uses of the BE verb in Czech (47%), English (26%) and Norwegian (6%). The study was based on fiction. The large proportion in Czech primarily depends on the use of the copula to form the past tense. The use as a progressive auxiliary is prominent in English but is missing in the other two languages.

The functions of BE and HAVE form a continuum that will be referred to as the Stative scale. Table 7 shows the basic contrasts in a selection of European languages exemplifying various possibilities.

	English	Swedish	French	Spanish	Greek	Russian		Finnish
						PRS	-PRS	
Identity	be	vara	être	ser	ίμε	Ø	byt'	olla
Description								
Stable								
Unstable								
Location				Postural				
Existence		finnas		haber				
Possession	have	ha	avoir	tener	έxo			
Abstract						imet'		

Table 7. The stative scale. Basic contrasts in the use of BE and HAVE in a selection of European languages.

In English, *be* covers all functions except Possession. Swedish is an example of a HAVE-language where BE has a restricted use in constructions that express Location and Existence. French and Greek are examples of languages where HAVE has extended into Existence. In French, *être* ‘be’ can be used to refer to Location, but a wide range of other options are frequent (Viberg 2013a: 153). In Spanish, *haber* has the second rank but is used as an auxiliary and in Existential constructions. Another verb *tener* with rank 4 functions as a verb of possession. Spanish also has a special locative copula

estar that also can be used as a descriptive copula but primarily to refer to unstable properties. (This is a complex and much debated issue, see Camacho 2012, Leonetti et al. 2015). Russian is the only European language that makes extended use of a Zero-copula in the present tense. It is also a BE-language with an emergent HAVE (*imet'*) restricted to Abstract possession. Finnish is a consistent BE-language, where the copula is used in all functions.

BE plays an exceptionally prominent role in the clausal structure of Welsh (see King 2016). In this language, verbs in general mark tense periphrastically with *bod* 'be' except for the preterite (past perfective). Thus, the construction in (1) corresponds both to the simple and the progressive present form in English.

(1) Welsh (King 2016: 169)

<i>mae</i>	<i>hi</i>	<i>'n</i>	<i>prynu</i>
be.PRS.3SG	she	CM	buy[VN]

'She buys. / She is buying.'

Since Welsh is a VSO-language, the copula appears initially. A Complement Marker (CM) *yn* ('*n* after a vowel) is required in front of the complement. The verb in periphrastic constructions formally has the properties of a noun and is referred to as Verbal Noun (VN). The copula also has special forms to indicate whether a clause is affirmative, interrogative or negative. This means that the copula carries very much grammatical information that appears in the initial position.

There is a trade-off between the use of the copula and the realization of concepts as a certain word class. Adjectival description, where an adjective appears in predicative position, requires the use of a copula in European languages. Outside Europe, there are languages where property concepts corresponding to adjectives in European languages are primarily realized as verbs. Stassen (2013c) refers to these as verbal property concepts and to adjectives as non-verbal property concepts. Europe, in which non-verbal property concepts dominate, belongs to a mega-area comprising also Central Asia and Siberia, India the Middle East and northern Africa, whereas verbal encoding predominates in several areas such as Africa south of the Sahara, East and Southeast Africa and North America. There are also languages such as Hausa (haus1257; Afro-Asiatic, Chadic), where property-concepts are primarily realized as nouns and appear in a possessive construction. In European language, the part-of-speech of property concepts can vary, for example, the favored realization of certain

types of Physiological Conditions or Emotions: *I'm hungry, I'm thirsty, I'm afraid*; French: *j'ai faim, j'ai soif, j'ai peur*; Greek: *pináo, ðipsáo, fováme*; German: *Ich bin hungrig / Ich habe Hunger, Ich bin durstig / Ich habe Durst, Ich habe Angst*. The large differences between European languages with respect to the frequency of the copula depend on differences in the lexicalization of concepts as a certain word class, which remains to be studied in detail, and on the variable grammatical uses of the copula, which is well described in the grammars of individual languages but needs to be systematically quantified to make contrastive comparisons.

4. Modals: CAN and MUST

In the introduction to the comprehensive reference work “Modals in the languages of Europe”, Hansen & de Haan (2009: 3) present the following working definition of modals: “word-like elements which are polyfunctional in the sense that they express no less than two types of modality”. The expression “word-like” element is motivated by the fact that modals appear on a continuum from fully lexical items to inflections. The part-of-speech also varies, but for the purposes of the present paper words that are regarded as verbs in the frequency dictionaries are regarded as modal verbs³ if they fulfill the semantic criteria, even if they may be defective in some respect.

Three types of modality are specifically mentioned by Hansen & de Haan (2009): Dynamic modality basically refers to the ability (*I can read*) or internal needs (*I must eat*) of the subject to do something (internal dynamic modality) or to possibilities or necessities due to the external circumstances (external dynamic modality). Compare: *I can (know how to) play tennis* (internal), *but I cannot play today* (external), *since the tennis court is closed*. Deontic modality typically involves permission or obligation of the subject to do something (*You may / must leave now*) or more generally indicates the degree of moral desirability of the situation referred to, whereas Epistemic modality refers to the estimation (typically by the Speaker) of the degree to which the described states of affairs apply in the world along the scale: certainty – probability –

³ In English, the modals have been included even though they are not counted as verbs in the source (Kilgarriff 1998). The reason for that is that the closest correspondences in the other European languages are counted as verbs. The frequency is based on the combined frequency of the different forms, which are listed as separate lexemes in the source: *will/would, can/could, may/might* and *shall/should*.

possibility (*He must be tired / I may be wrong*). Within each of these types, a contrast can be made between Possibility and Necessity, which forms the basis of the influential typology presented in van der Auwera & Plungian (1998). For a discussion of various views on modality, see Nuyts (2014).

The frequency data used for the present study make it possible in most European languages to identify one verb centered around Possibility (symbolized CAN) and one verb centered around Necessity (symbolized MUST) that is the major realization of the concept. Table 8 shows the frequency per million words and the corresponding value of the natural logarithm for the two types of modals in 13 European languages, which makes it possible to compare the frequency ranges of the major modals with those of the nuclear verbs.

Language	CAN			MUST		
	<i>form</i>	<i>LN</i>	<i>pm</i>	<i>form</i>	<i>LN</i>	<i>pm</i>
English	<i>can</i> (including <i>could</i>)	8.38	4345	<i>must</i>	6.58	721
German	<i>können</i>	8.58	5304	<i>müssen</i>	7.93	2787
Swedish	<i>kunna</i>	8.89	7265	<i>måste</i>	7.13	1251
French	<i>pouvoir</i>	8.13	3395	<i>devoir</i>	7.61	2021
Spanish	<i>poder</i>	8.25	3837	<i>deber</i>	7.01	1112
Greek	<i>boró</i>	7.94	2807	<i>prépi</i>	7.61	2010
Czech	<i>moc</i>	7.67	2149	<i>muset</i>	7.11	1229
Russian	<i>(s)moč'</i>	8.19	3594	<i>(dolžen)</i>	(6.99)	1091
Lithuanian	<i>(ne)galėti</i>	8.74	6234	<i>reikėti</i>	7.61	2015
Welsh	<i>gallu</i>	7.73	2267	<i>(rhaid)</i>	NOUN	
Estonian	<i>võima</i>	8.33	4156	<i>pidama</i>	(8.34)	(4183)
Finnish	<i>voida</i>	8.41	4479	<i>täytyä</i>	?	
Hungarian	<i>(tud 'know,can')</i>	7.90	2688	<i>kell</i>	8.06	3156

Table 8. The major modal verbs expressing Possibility and Necessity.⁴

⁴ Russian *smoč'* pf. 'become possible' 353/5.87 is included; *moč'* + *smoč'* 3,594/8.19 *dolžen* 'must' is an adjective; Welsh *rhaid* is a noun meaning 'necessity' (King 2016: 264); Estonian *pidama* represents two homonymous verbs meaning 'must' and 'hold' (Raag 2014: 317), which both appear to be frequent; The Finnish cognate *pitää* 'hold' can also express obligation. Hungarian *tud* means both 'know' and 'can'.

The identification of meanings in Table 8 is primarily based on the glosses in the frequency dictionaries. The detailed descriptions in Hansen & de Haan (2009) have also been consulted, although it has not been possible to account for this in detail. A further resource that has been used is the Word Alignment Database (WAD) included as a tool in OPUS (Tiedeman 2012). The analysis is representative also for analyses carried out on the nuclear verbs.

Even though MUST has been used to symbolize the major modal verb that expresses Necessity, *must* is rivalled by alternative expressions in English. As can be observed in Table 8, *must* has a lower frequency than its closest correspondences in the other European languages. This is the result of an ongoing change. Aijmer (2019) shows that *must* is the main proponent for epistemic modality in present-day English but is losing ground to *have to* and other semi-modals in the area of deontic modality and this development is related to a cultural change towards the use of less authoritative and face-threatening language that appears to have parallels in some other European languages. Variable expression of MUST is characteristic of further languages. In Finnish, a frequent alternative to express MUST is *täytyy* (3SG.PRES of *täytyä*), which is used with a subject in the genitive: *Minun täytyy lähteä* ‘I (lit. ‘my’) must leave’. In addition to *täytyy*, *pitää* ‘hold’ is a frequent translation of *must* in Subtitles. In Europarl, however, *on* ‘be.PRS.3SG’ dominates completely in a special construction with a passive present participle as in (2).

(2) Finnish (Europarl)

<i>taistelu on</i>	<i>käy-tä-vä</i>	<i>pääosin</i>	<i>kansainvälise-llä</i>	<i>taso-lla</i>
battle is	employ-PASS-PRS.PTCP	mainly	international-AD	level-AD

‘This battle must, in the main, be fought at international level.’

To sum up, CAN is realized as one of the 3-6 most frequent verbs in the majority of European languages and its frequency expressed as LN typically reaches above 8.0, which is superseded only by BE and HAVE. MUST is also expressed as a frequent verb and its LN usually reaches above 7.0, which is the cut-off point for the 20 most frequent verbs, but the picture is more complex than for CAN. One reason for that is that constructions with HAVE (or in Finnish BE) are frequent alternatives to a modal verb in several languages especially in certain text types.

5. MAKE and DO

Both *make* and *do* belong to the most frequent verbs in English, but in most European languages this verb pair has one major correspondence, which is symbolized by MAKE in Table 9. Dutch has been included, since it is one of the few European standard languages that have a do-verb. For a wide range of functions of DO in various varieties of Germanic languages, see Tieken-Boon van Ostade et al. (1998). See also Juvonen (2016) on the use of MAKE/DO as light verbs in pidgin lexica.

Language	MAKE				DO			
	<i>form</i>	<i>Rank</i>	<i>LN</i>	<i>Pm</i>	<i>form</i>	<i>LN</i>	<i>Rank</i>	<i>pm</i>
English	<i>make</i>	9	7.68	2173	<i>do</i>	8.63	4	5596
German	<i>machen</i>	9	7.59	1971	<i>tun</i>	6.62	21	752
Swedish	<i>göra</i>	9	7.89	2669				
French	<i>faire</i>	4	8.37	4330				
Spanish	<i>hacer</i>	5	8.31	4081				
Greek	<i>káno</i>	7	7.65	2101				
Czech	<i>dělat/udělat</i>	14	6.90	997				
Russian	<i>delat'/sdelat'</i>	13	7.40	1640				
Lithuanian*	<i>(pa)daryti</i>	22	7.25	1409				
Welsh	<i>gwneud</i>	3	8.46	4738				
Estonian	<i>tegema</i>	6	8.04	3114				
Finnish	<i>tehdä</i>	6	8.26	3896				
Hungarian**	<i>tesz</i>	6	7.31	1490				

*Lithuanian *sudaryti* 'form; conclude': 1266 pm

**Hungarian *tesz* 'make, do; put'

Table 9. MAKE and DO in European languages.

Gilquin & Viberg (2009) and Viberg (2006b) distinguish 6 main functions of *do* and *make* displayed in Table 10, which will be taken as a point-of-departure for the following brief survey. The function Pro-Verb includes two main subtypes.

Pro-verb covers what Van Valin & LaPolla (1997) refer to as a *Generalized Action Verb*, which is used when the nature of the event is unknown as in *I wonder what he did*. This function can be regarded as basic for languages that have a specific do-verb. Pro-verb also covers the anaphoric use as in: *He left and she **did** too*. Dutch *doen* is used as a pro-verb in both these uses, but this does not apply to German *machen*, which is

used as a generalized action verb: *Was macht er?* ‘What is he doing?’ The number of do-languages is too small to form general rules. Another case that would be worth looking at is the alternation between *lave* and *gøre* in Danish. The use of English *do* in interrogative and negative sentences stands for a large proportion of the number of tokens. It is unique at least among European languages and will not be commented on further (on *do* as a placeholder, see Hammarberg & Viberg 1977: 150–151).

Main functions	Examples
Production	She <i>made</i> her own wedding dress.
Transformation	It was this movie which <i>made</i> him a star.
Causative-ADJ	What <i>makes</i> you so happy?
Causative-VP	I like him because he <i>makes</i> me laugh.
Support verb	I can’t <i>make</i> a decision just yet.
Pro-verb:	
Generalized Action Verb	What did they <i>do</i> ?
Anaphoric verb	He left and she <i>did</i> too.

Table 10. Main functions of DO/MAKE.

For *make*, Production (or Creation), which refers to causing something to come into existence, can be regarded as the basic function and all the verbs symbolized as MAKE in Table 9 can be used to express this meaning. MAKE can be regarded as the nuclear verb of the semantic field, which also contains more specialized verbs such as *sew*, *build*, *manufacture* and *produce*. This would be worth a more extensive discussion since some of the other more abstract and grammatical meanings are more frequent (see Viberg 2006b for Swedish). The decisive factor is that production is a meaning from which the other meanings can be derived by extension in the most natural way. Transformation, which follows Production in Table 10, can be used with a construction typical of a verb of Production: *The director managed to make a star* [Product] *out of this unexperienced teenager* [Material] and the causative meanings Causative-VP and Causative-ADJ are more abstract extensions with a modified argument structure. In examples such as *make a decision* and *do harm*, the action is specified in the noun and *do* and *make* serve as support verbs. Both DO and MAKE have this function also in other languages, but this function will not be discussed further since the usage patterns vary a lot across languages with respect to which

abstract nouns are used. Table 11 shows the distribution of possible functions. (Data on English, Dutch, Swedish and French are based on Gilquin & Viberg 2009).

	ENG		DUT		SWE	FRE	GRE	RUS	FIN	
Production	make		maken		göra	faire	káno	delat'	tehdä	
Transformation						rendre				
Causative-ADJ				doen	få	faire				
Causative-VP										
Support verb	do	maken		göra				saada		
Pro-verb									tehdä	

Table 11 Possible functions of DO/MAKE in selected European languages.

Schultze-Berndt (2008) in a typological study of do-verbs⁵ based on a sample of nine very diverse non-European languages identifies further functions, some of which are characteristic also of European languages. The defining feature for a do-verb is that it can function as a Generalized Action Verb in the sense of Van Valin & LaPolla (1997). Like the European languages with a combined meaning DO/MAKE, languages like Kalam (kala1397; Nuclear Trans New Guinea, Madang) and Samoan (samo1305; Austronesian, Malayo-Polynesian) have a do-verb that also functions as a verb of Creation (Production). In some languages, do-verbs also function as productive causative markers, for example Ewe (ewee1241; Atlantic-Congo, Volta-Congo), but the presence of the creation sense seems to be a precondition for that use. (Dutch *doen* is an exception.) Crosslinguistically, a common use of do-verbs is to introduce quotations (Schultze-Berndt 2008: 191–192). European languages primarily rely on a specific verb SAY (see Section 8), but *faire* can be used to introduce direct reports in French, even though this use is not very frequent (Gilquin & Viberg 2009: 76). It is, however, very frequent in some regional varieties of spoken Italian (Francesca di Garbo, personal communication). A related function, which is widely attested crosslinguistically, is the use of a do-verb as a verbalizer with sound-symbolic elements. In several languages, the do-verb functions as an equivalent of ‘feel’. In

⁵ Schultze-Berndt’s term do-verb covers also the verbs referred to as MAKE in this paper. Her choice of term is obviously motivated by the focus on the use as generalized action verb and appears to be compatible with the analysis in the present paper with respect to the phenomena that are discussed in both papers.

Hausa, for example, *yi* ‘do’ is used to express emotions such as ‘be angry’, ‘be happy’ and ‘feel regret’.

6. Motion: GO, COME and the variable PUT

There are two basic types of motion verbs. *Come* and *go* refer to subject motion, where the moving entity functions as a subject, whereas verbs such as *put*, *pull* and *carry* refer to object motion (also referred to as caused motion). As shown in Table 12, verbs meaning GO and COME appear among the 20 most frequent verbs in all the studied languages. Among the object-motion verbs, the verbs of putting tend to be most frequent and come close to being nuclear verbs. For that reason, they will be discussed at some length in Section 6.3.

Lang.	GO			COME			PUT		
	<i>form</i>	<i>LN</i>	<i>pm</i>	<i>form</i>	<i>LN</i>	<i>pm</i>	<i>form</i>	<i>LN</i>	<i>pm</i>
ENG	<i>go</i>	7.82	2495	<i>come</i>	7.33	1519	<i>put</i>	6.55	700
GER	<i>gehen</i>	7.41	1648	<i>kommen</i>	7.52	1837	POSTURAL		
SWE	<i>gå</i>	7.81	2476	<i>komma</i>	8.12	3348	POSTURAL		
FRE	<i>aller</i>	7.69	2194	<i>venir</i>	6.76	860	<i>mettre</i>	6.75	855
SPA	<i>ir</i>	8.17	3518	<i>venir</i>	6.68	799	<i>poner</i>	6.92	1017
GRE	<i>páo</i>	6.61	741	<i>érxume</i>	6.43	622	<i>θέτο</i>	6.47	647
							<i>vázo</i>	5.13	169
CZE	<i>jít</i>	7.07	1175	<i>přijít</i>	6.26	522	GIVE		
RUS	<i>idti/pojti</i>	7.19	1323	<i>prixodit’</i>	6.53	686	POSTURAL		
LITH	<i>eiti</i>	6.92	1011	<i>ateiti</i>	5.95	385	<i>dėti</i>		
WEL	<i>mynd</i>	8.22	3726	<i>dod</i>	8.03	3072	GIVE		
EST	<i>minema</i>	7.75	2317	<i>tulema</i>	8.09	3264	<i>panema</i>	6.98	1073
FIN	<i>mennä</i>	7.02	1118	<i>tulla</i>	8.38	4371	<i>panna</i>		
HUN	<i>megy</i>	6.56	704	<i>jön</i>	6.48	653	<i>(tesz)*</i>	7.31	1490

*Hungarian *tesz* is glossed ‘do, make; put’ and represented in Table 9.

Table 12. GO, COME and PUT in European languages.

6.1. Come and go

COME relates the motion to a place that can be implicit: the Deictic Center (DC). In the prototypical case, DC is the place of the speaker at the time of utterance: *Come (here)!* This serves as a defining feature. DC can be transferred step by step, but here

languages differ. DC can be transferred to the place of the speaker at reference time, to the place of the hearer, or even to the central character in a third person narrative. GO is in principle negatively defined and covers the cases not covered by COME in individual languages. Fillmore (1997) is the classical discussion of *come* and *go* in English. Ricca (1993) studied the contrast between COME and GO in 20 European languages. Lewandowski (2014: 46) proposed a universal hierarchy of extensions from the prototype with four steps, which is a good starting point for the discussion, since it emphasizes the great variation with respect to the identification of DC across languages. In Portuguese (step 1), COME is restricted to the speaker's location at speaking time but cannot be extended to speaker's location at reference time, whereas the latter is possible in Spanish (step 2). English (step 3) in addition allows the extension to the addressee's location, whereas in East and West Slavic languages (step 4), COME can refer to any Goal. The situation in languages representing the fourth step is studied more closely in Martinková & Janebová (2024), who look at the use of venitive verbs in Czech, where the closest correspondences of *come* as in other East and West Slavic languages are verbs with the perfectivizing prefix *při-*. This prefix productively combines with basically imperfective manner-of-motion verbs such as Czech *přijít* [*při-walk*], *přiběhnout* [*při-run*], *přijet* [*při-drive*] and *přilétet* [*při-fly*]. These verbs are not manner neutral like English *come*. Lithuanian is another language, where COME is not a simple verb but a morphologically marked version of GO: *eiti* 'go' – *ateiti* 'come'. Interestingly, German has a manner neutral verb *kommen* but also deictic particles *her-* (roughly 'to DC') and *hin-* ('from DC'), which can be combined with manner-of-motion verbs (see Meex 2020, for a more exact description). There are also great differences even between closely related languages such as English and Swedish with respect to the range of extended meanings. With respect to grammatical meanings, Swedish uses *komma* as a future marker, whereas English uses *go* (*be going to*). In addition, Swedish *gå* 'go' is used to express a specific modal meaning. There are also many abstract lexical meanings that contrast across the two languages (Viberg 1999a, 2003).

Even if the universality of the distinction between COME and GO was questioned already in a thought-provoking article by Wilkins & Hill (1995) and is clearly not an absolute universal, it represents a wide-spread lexical contrast. Based on a study of the Gospel According to Mark in 100 languages, Wälchli & Cysouw (2012) found that a distinction between COME and GO was made in most languages. This system is not always stable. DeLancey (1985) has shown how Tibeto-Burman languages change in a cyclical fashion. In essence, at a certain stage 'come' develops into a bound affix

which can be combined with other motion verbs, among them ‘go’. At a later stage, ‘come’ can arise anew through the merger of this affix with ‘go’. There are also languages that make further distinctions and use special verbs to indicate whether DC serves as the home base or not. In Diuxi Mixtec spoken in Oaxaca, Mexico (Kuiper & Merrifield 1975), there is a set of motion verbs which contrast with respect to both deictic distinctions and whether a location is the homebase (Base) of the subject of the verb, for example *ndisi* ‘move to Base toward DC’ and *vasi* ‘move to non-Base toward DC’. Deixis in motion verbs is dealt with systematically and in an innovative way in Matsumoto (2025).

6.2. Basic contrasts between European languages: Assisted and Indeterminate motion

There is a geographical belt in Europe, where motion in a vehicle is obligatorily marked in the verb (Viberg 2013b,c). Continental Scandinavian languages together with German and the Baltic and the East and West Slavic languages form a block and make a lexical distinction between moving on foot and travelling in a vehicle and do not have a general motion verb corresponding to English *go* (general ‘go’) but rather a verb which can be symbolized as ‘go = walk’. In Swedish, for example, the verb *åka* must be used in a sentence such as *Barnen åker buss till skolan* ‘The children go by bus to school’. *Barnen går till skolan* can only mean that they go by foot.

In East and West Slavic languages, there is also a lexical contrast between determinate and indeterminate motion verbs (Jakobson 1971, Sussex & Cubberley 2006: 482–483). Determinate verbs describe motion towards a Goal or Endpoint, whereas indeterminate motion verbs refer to aimless motion without a specific goal (‘go around’), habitual repetition and ability to walk, run etc. Not all motion verbs contrast in this way and the number of pairs vary in Slavic languages between 9 and 20.

6.3. Other basic motion verbs

PUT appears as one of the 20 most frequent verbs in several European languages, but Germanic and Slavic languages tend to make an obligatory differentiation between two or three verbs of putting based on postural verbs (SIT/STAND/LIE). As discussed in Viberg (2015a), the number of contrasts between the postural verbs have been reduced and in some languages one such verb has developed towards a general put verb. In addition, there is a general trend towards the development of a general verb

of putting with another basic meaning. In some languages, DO, LET or GIVE are used to fulfill that function. Lenz et al. (2020) describe the use of German *tun* ‘do’ as a general put verb in non-standard varieties across the whole German speaking area. This study also shows how *dát* ‘give’ in Czech is the most frequent put verb in the spoken language and the second most frequent (after *položit* ‘lay’) in the literary language. From a typological perspective, verbs of putting have been studied in Kopecka & Narashimhan (2012).

There are no basic motion verbs other than COME and GO that appear among the 100 most frequent verbs in all 13 languages. There are, however, some that appear in most of them. A verb glossed LEAVE appears in 11 languages and 10 of the 13 languages have a verb glossed FOLLOW, which refers to co-motion. The basic meaning of FOLLOW is ‘move behind and in the same direction as some other mover’ even though the meaning may be wider in various ways. In Swedish, *följa* corresponds both to FOLLOW and to ACCOMPANY ‘move together with’. The two meanings can be kept apart by adding a particle: *följa efter* (‘after’) ‘follow’, *följa med* (‘with’) ‘accompany’.

With respect to possible language-specific features, it is interesting to look for possible reflections of Talmy’s (1985) typology of motion events in the lists of the 100 most frequent verbs. Verbs referring to motion into and out of a container (ENTER/EXIT) are found in this frequency range in the Verb-framed languages that typically encode Path in the verb (rank within parentheses): French *entrer* (88), Spanish *entrar* (41) and Greek *béno* (72), whereas French *sortir* (77), Spanish *salir* (28) and Greek *vjéno* (100) express EXIT. Among European languages, the use of frequent verbs meaning ENTER and EXIT is a characteristic feature of Verb-framed languages. The situation is less clear with respect to verbs referring to motion UP/DOWN. Besides verbs glossed ASCEND and DESCEND, which are characteristic of Verb-framed languages, certain such verbs such as FALL appear among the 100 most frequent verbs in 4 languages (ENG, GER, HUN, SPA) and in Swedish *falla* has rank 103. Verbs covering the concepts of RISE, RAISE and/or LIFT appear among the 100 most frequent verbs in 6 languages (FRE, SPA, EST, FIN, LIT, WEL).

7. Possession: GIVE and TAKE versus TAKE and BRING

Traditionally, TAKE is regarded as the opposite of GIVE, but in studies of English, *take* and *bring* have also been regarded as transitive or causative counterparts of *go* and

come, since deixis is also involved in the contrast (Fillmore 1997, Levin 1993). Both perspectives are valid.

7.1. GIVE and TAKE

The data for the nuclear verbs GIVE and TAKE are shown in Table 13.

Language	GIVE			TAKE		
	<i>form</i>	<i>LN</i>	<i>pm</i>	<i>form</i>	<i>LN</i>	<i>pm</i>
English	<i>give</i>	7.18	1314	<i>take</i>	7.49	1792
German	<i>geben</i>	7.71	2240	<i>nehmen</i>	6.43	618
Swedish	<i>ge</i>	7.24	1399	<i>ta</i>	7.69	2189
French	<i>donner</i>	6.75	851	<i>prendre</i>	6.96	1058
Spanish	<i>dar</i>	7.61	2012	<i>tomar</i>	6.45	632
Greek	<i>díno</i>	7.43	1689	<i>pérno</i>	6.86	950
Czech	<i>dát/dávat</i>	7.01	1111	<i>vzít/brát</i>	6.18	483
Russian	<i>davat’/ dat’</i>	6.86	956	<i>vzjat’/brat’</i>	6.54	694
Lithuanian	<i>duoti</i>	5.88	357	<i>(pa)imti</i>	6.90	991
Welsh	<i>rhoi</i>	7.57	1942	<i>cymryd</i>	6.46	638
Estonian	<i>andma</i>	7.52	1851	<i>võtma</i>	7.54	1884
Finnish	<i>antaa</i>	7.35	1550	<i>ottaa</i>	7.37	1581
Hungarian	<i>ad</i>	6.94	1033	<i>vesz</i>	7.05	1156

Table 13. The frequencies of GIVE and TAKE.

There is an extensive literature on GIVE from a crosslinguistic perspective (Newman 1998, Nolan et al. 2015, Bouveret 2021). The seminal work is Newman (1996), who presents an extensive analysis of the basic meaning of English *give* and its rich semantic network from a cognitive linguistic perspective in combination with a broad typological survey. Across languages, literal GIVE tends to extend to more abstract fields, in particular Interpersonal communication (*give advice to someone*), and to Emergence (of Sound, for example: *give a yell*). Newman also discusses the development of grammatical uses such as Enablement and Causation as well as Recipient and Benefactive markers.

The status of GIVE as a nuclear verb appears to represent a strong universal tendency, but it is clearly not an absolute universal. In some languages, such as Sango (sang1328; Atlantic-Congo, Volta-Congo), GIVE and TAKE are the same verb and the semantic distinction between them is signaled by the construction (Thornell 2002).

Neither is GIVE always a simple verb. Reesink (2013) describes Papuan languages where GIVE is expressed with a combination of serial verbs and Rice (1998) shows how Chipewyan (chip1261; Athabaskan-Eyak-Tlingit, Athabaskan) completely lacks specialized verbs of possession and uses physical action verbs describing handling and manipulation as the closest equivalents of the verbs of possession. In Japanese, GIVE is expressed with a set of verbs that are differentiated with respect to the social status and interpersonal relations between speaker and hearer and the giver and the receiver (Loveday 1986: Chapter 3).

Viberg (2010, 2025) shows how TAKE as a verb of Possession can be analyzed as a goal-directed physical action verb and how an important concrete use of TAKE can be described as extending the arm and grasping an object as the first part of a chain of actions in order to gain control i.e. possession of the object, see (3). Ownership is often not an issue in this use, only physical possession.

		VERB
(3)	And he took the coarse chunks of bread, one in each hand, and ate.	<i>take</i>
FRE	et prit les morceaux de pain noir,	(source: MPC) <i>prendre</i>

The verbs used in this example are the ones identified as TAKE in Table 13. In addition to this concrete use, TAKE has many abstract extended uses.

7.2. TAKE and BRING

In many languages, another verb than TAKE is preferred (or required) when the verb refers to caused motion of the grammatical object combined with the co-motion of the Agent as in (4). The meaning of these verbs will be referred to as TAKE^{CM} (CM: Co-Motion).

		VERB
(4)	She took me to see this movie.	<i>take</i>
FRE	Elle m' a emmené voir le film.	<i>emmener</i>
	she me has taken ^{CM} see the film	(source: OPUS Subtitles)

In English, *take* and *bring* are regarded as opposites, but in languages that use separate verbs to express TAKE and TAKE^{CM}, it is the verb with the latter meaning that is the opposite of BRING. TAKE^{CM} refers to motion from DC and contrasts with BRING, which refers to motion to DC. As is the case with COME, the rules for assigning DC varies across languages. BRING is represented by a frequent special verb in all the studied

languages except Welsh and, in actual practice also Swedish, in which *bringa* has rank 933 and a frequency of only 14 pm. BRING is characteristic of European languages, even if it is not as frequent as any of the nuclear verbs.

In French, there are two pairs of verbs that correspond to TAKE^{CM} and the same applies to BRING. The pair *emporter/emmener* refer to motion from a deictic center (DC) and correspond to TAKE^{CM}, whereas *apporter/amener* refer to motion to DC and correspond to BRING. In principle, the first verb in each pair is used with inanimate objects and the second with human objects. The verbs are derived versions of *porter* ‘carry’ and *mener* ‘lead’. Verbs meaning CARRY and LEAD also describe co-motion but do not refer to DC. The contrasts between TAKE, TAKE^{CM} and BRING are summed up in a selection of representative European languages in Table 14.

Language	TAKE ^{CM}			BRING		
	<i>form</i>	<i>LN</i>	<i>pm</i>	<i>form</i>	<i>LN</i>	<i>pm</i>
English	<i>take</i>	See Table 13		<i>bring</i>	6.08	439
Swedish	<i>ta</i>			<i>(bringa)</i>	2.64	14
French	<i>emporter</i>	4.01	55	<i>apporter</i>	5.06	158
	<i>emmener</i>	3.76	43	<i>amener</i>	4.44	85
Spanish	<i>llevar</i>	6.75	853	<i>traer</i>	5.52	250
Greek	<i>páo</i> ‘go; cause to go’	See GO		<i>féro</i>	6.01	408
Welsh	<i>mynd âh</i> ‘go with’			<i>dod âh</i> ‘come with’		
Finnish	<i>viedä</i>	6.53	687	<i>tuoda</i>	6.73	841

Table 14. The frequency of TAKE^{CM} and BRING in some European languages.

In Greek, *páo* ‘go’ can be used as a co-motion verb and is a frequent translation of *take*, when it refers to co-motion as in (5).

(5) Greek (OPUS Subtitles)

θα σε πάο σπίτι με το aftokínito μου
 FUT you.ACC go.1SG house with DEF car my
 ‘I’ll **take** you home in my car.’

In Swedish, *bringa* has a very low frequency and is used primarily with abstract objects: *bringa lycka* ‘bring luck’, *bringa klarhet* ‘throw light on’, *bringa ordning*, ‘bring order’. When *bring* has a concrete meaning, a frequent translation is *komma med* ‘come with’ but also *ta med* ‘take with’ and *ha med* ‘have with’ are frequent. (The situation is similar in Norwegian, see Ebeling 2017). In Welsh, both co-motion verbs are rendered

with more analytical expressions: TAKE^{CM} is expressed with *mynd âh* ‘go with’ and BRING with *dod âh* ‘come with’ King (2016: 469, 471). In Czech, Russian and Lithuanian, TAKE^{CM} has three major correspondences (see, for example, Naughton & von Kunes 2021: 214 for Czech.). BRING is expressed by adding a prefix to these verbs. TAKE and BRING in European languages are treated more in detail in Viberg 2025 ms.) There is an interesting typological study by Margetts et al. (2022) of caused accompanied motion (my TAKE^{CM} and BRING) based on data from eight languages from the Americas and from the Pacific area. The study shows how caused accompanied motion is described in a more analytical way, for example by transitivity COME and GO by adding causative or applicative markers (see also Margetts et al. 2022).

7.3. *Summing up Verbs of Possession*

GIVE serves as a nuclear verb in European languages and tends to be basic even in languages outside Europe, but this is only a tendential universal. As a verb of possession and as an antonym of GIVE, TAKE has a similar status although its use in non-European languages is not as well documented as that of GIVE. The use of TAKE as a co-motion verb (TAKE^{CM}) and as an antonym of BRING should be regarded as a language-specific characteristic of languages such as English and Swedish. In these languages TAKE can be extended to cover TAKE^{CM}, whereas this meaning is expressed with other verbs than TAKE in most European languages. BRING is characteristic of European languages except some like Swedish, Norwegian and Welsh, where this meaning is expressed analytically by COME + WITH. Verbs glossed RECEIVE and GET, which both appear among the 100 most frequent verbs, will be commented on in Section 11.3. A couple of other possession verbs tend to be frequent but do not appear among the 100 most frequent verbs in all 13 languages. PAY appears within this range in 9 languages and BUY in 7.

8. Verbal Communication: SAY and TELL

Table 15 shows the frequency of SAY, which is the most frequent verbal communication verb in European languages.

A very large inventory of Verbal Communication Verbs (VCVs) has been described for several European languages. Ballmer & Brennenstuhl (1981) account for 4,800 English speech act verbs (including multiword expressions) and Harras et al. (2004,

2007) devote two volumes to German VCVs. Wierzbicka's (1987) analysis of English VCVs is particularly relevant for the present study, since the primitive SAY appears in the definitions of all the other VCVs, which can be interpreted as support for the idea that SAY is a nuclear verb.

Language	SAY	LN	pm
	<i>form</i>		
English	<i>say</i>	8.11	3335
German	<i>sagen</i>	7.99	2954
Swedish	<i>säga</i>	7.96	2868
French	<i>dire</i>	7.97	2898
Spanish	<i>decir</i>	8.29	3967
Greek	<i>léyo</i>	8.07	3183
Czech	<i>řict/řikat</i>	7.72	2249
Russian	<i>skazat'/govorit'</i>	8.16	3503
Lithuanian	<i>(pa)sakyti</i>	7.72	2260
Welsh	<i>dweud</i>	7.74	2308
Estonian	<i>üttelema</i>	7.83	2523
Finnish	<i>sanoa</i>	7.98	2913
Hungarian	<i>mond</i>	7.72	2262

Table 15. The nuclear Verbal communication verb SAY.

8.1. The nuclear verb SAY

SAY appears to have a correspondence in most (possibly all) languages outside Europe, although SAY also has a tendency to acquire a wide range of more general meanings. Goddard (2001: 31–32) in a survey of lexico-semantic universals states: “In many languages there is a unique lexical form meaning ‘say’, as in English. But there are languages in which the word for ‘say’ can also express the meanings ‘do’ (as in Samoan) or ‘want’ (as in Kalam and Mangaaba-Mbula). Another common polysemy in which ‘say’ is involved concerns ‘making sounds’ of various kinds.” In Papuan languages, SAY according to Foley (1986) is the verb most commonly used in periphrastic causative constructions and it also tends to be used in desiderative constructions. Another geographical area where SAY has developed many secondary meanings is Ethiopia, where one of the most characteristic secondary meanings is

intention (Ferguson 1976: 71). In addition, SAY has a strong tendency to grammaticalize into a quotative or a more general complementizer (Lord 1993: Chapter 7, Heine & Kuteva 2002).

8.2. Other basic verbs of verbal communication

SAY and TELL. English *tell* has rank 22 and the frequency 772 pm (LN 6.65). Its frequency is only slightly lower than what is characteristic of a nuclear verb. *Tell* must, however, be regarded as a language-specific basic verb, since the closest correspondences in the other European languages tend to have a much lower frequency and the most frequent VCVs, which are identified as SAY in this study, are glossed ‘say, tell’ in some of the frequency lists. As discussed in a corpus-based contrastive study of *tell* and its Swedish correspondences (Viberg 2017b), *tell* can be used as an information verb and report a statement. In that function, *tell* is often translated into Swedish by *säga* ‘say’. The verbs *si* ‘say’ in Norwegian and *dire* ‘say’ in French are also frequent correspondences of *tell* (Egan 2021). A distinctive characteristic of *tell* as an information verb is that the Addressee is explicitly mentioned in most cases (as much as 90% in Viberg 2017b: 64). The semantically closest equivalent to *tell* can usually serve as a verb that reports a narrative, which typically refers to a complex set of events that may be fictional. These correspondences have different ranges of uses in addition to the prototypical ‘tell a story’ (see Viberg 2017c).

SAY and SPEAK. In English, the verb *say* typically requires explicit reference to the Message (*He said that he would leave / He said so*) but can optionally refer also to the Addressee and the Topic: *He [Speaker] said something [Message] about it [Topic] to Alice [Addressee]*. Speech activity verbs refer to the act of communicating verbally without referring to the message. They can refer simply to the activity as in *He was talking* but can optionally include reference to the Addressee and the Topic: *He [Speaker] talked to Alice [Addressee] about the wedding [Topic]*. European languages usually have one or sometimes like English (*speak, talk*) two speech activity verbs that belong to the 100 most frequent verbs. At least one speech activity verb appears among the 100 most frequent verbs in 12 of the 13 languages in the sample. The exception is Russian in which the situation is more complicated. SAY is expressed by two suppletive aspect forms *skazat’* pf and *govorit’* impf., but in particular the imperfective form is used also as a speech activity verb.

WRITE. A verb meaning ‘write’ appears among the 100 most frequent verbs in all 13 languages. In English, this verb can be used both as a reportative verb *He wrote that he would come* and as an activity verb: *He has been writing all day. He wrote a few words [Message] about it [Topic] to me [Addressee] in a letter [Medium]*.

ASK and ANSWER are also among the most frequent VCVs. In English, *ask* can be used both in the sense ‘ask-inquire’ (*He asked what I was doing*) and in the sense ‘ask-request’ (*He asked me to open the window*). All 13 languages have a verb that can mean ask-inquire among the 100 most frequent verbs (including some cases in which this verb can also mean ‘ask-request’ such as English *ask* and French *demander*). ANSWER appears among the 100 most frequent verbs in 7 languages. The polysemy of *ask* and *answer* is described in Rudzka-Ostyn (1989, 1995) and contrastively compared with the Swedish correspondences in Viberg (2016b).

9. Mental: SEE, KNOW and WANT

Mental is a macro-field that comprises several general semantic fields: Perception, Cognition, Emotion and Desire.

9.1. Perception

Verbs of perception can refer to intentional Activities such as *look* and *listen* or to non-agentive Experiences such as *see* and *hear*. Only Experiences will be discussed here. (See Viberg 2019 for the general structure of the semantic field of perception verbs.) Table 16 shows the frequencies of the verbs meaning SEE.

SEE is the most frequent perception verb in all European languages and with a few exceptions has an LN above 7, which is characteristic of the nuclear verbs, whereas HEAR and FEEL have a lower frequency. LN is usually above 6 for HEAR but not above 7. FEEL is more variable and ranges between 5.4 – 7.0 (214 – 1067 pm). The rank order between the latter two verbs varies across European languages. The major reason for this appears to be that the verbs identified as FEEL in some languages have very frequent cognitive meanings (see Section 9.2.1). Verbs referring to Experiences of TASTE and SMELL have a much lower frequency.

SEE appears to be the most basic perception verb also in non-European languages (Viberg 1983a). See also Evans & Wilkins (2000), San Roque et al. (2015) and Norcliffe & Majid (2024a,b). The status of SEE as a nuclear verb thus represents a very strong tendency across the world’s languages but is not an absolute universal. As discussed in

Viberg (2001), there are languages where SEE is not a simple verb as in Ngarinjin (non-Pama-Nyungan, Northern Kimberley, Australia).

Language	SEE <i>form</i>	<i>LN</i>	<i>Pm</i>
English	<i>see</i>	7.6	1917
German	<i>sehen</i>	7.2	1357
Swedish	<i>se</i>	7.9	2592
French	<i>voir</i>	7.2	1344
Spanish	<i>ver</i>	7.3	2293
Greek	<i>vlépo</i>	6.9	1021
Czech	<i>(u)vidět</i>	6.6	712
Russian	<i>(u)videt'</i>	7.1	1227
Lithuanian	<i>(pa)matyti</i>	7.0	1039
Welsh	<i>gweld</i>	7.8	2460
Estonian	<i>nägema</i>	7.3	1542
Finnish	<i>nähdä</i>	6.9	1014
Hungarian	<i>lát</i>	7.0	1052

Table 16. The frequencies of SEE across European languages.

There are also languages that do not have any verb clearly tied to a specific sense modality. A well-documented case of this type is the Papuan language Kalam (Pawley 2020) from which example (6) is taken.

(6) Kalam (Pawley 2020: 290)

mñi nŋ-sp-an?

now perceive-PROG-2SG

‘Now do you notice/see/hear/smell/feel/understand (it)?’

9.2. Cognition

Cognition covers Knowledge, Belief, Understanding, Memory, Judgment, Desire, Intention, Problem solving, Planning and Resolution. (The concept Meaning is regarded as part of Verbal communication.) Verbs of thinking and other cognitive verbs have been discussed from a typological perspective in Fortescue (2001) and in Palmer et al. (2003), see in particular Goddard (2003) in that thematic issue.

9.2.1. KNOW

Etymology is particularly informative with respect to the verbs of knowing in European languages. According to Buck (1949: KNOW), the English verb *know* covers two meanings that were originally expressed by two different verbs and still is in many Indo-European languages. The first meaning is ‘know as a fact’ and verbs with that meaning are shown under KNOW in Table 17. The other meaning according to Buck is ‘be acquainted with (a person or thing)’. The meaning of these verbs, which are represented as KNOW2 in Table 17, is variable and it is difficult to pin down one basic meaning shared across all languages by these verbs.

Language	KNOW			KNOW2		
	<i>form</i>	<i>LN</i>	<i>pm</i>	<i>form</i>	<i>LN</i>	<i>pm</i>
English	<i>know</i>	7.5	1855	---		
German	<i>wissen</i>	7.2	1358	<i>kennen</i>	6.0	400
Swedish	<i>veta</i>	6.9	1032	<i>känna</i> ‘feel’	7.0	1067
French	<i>savoir</i>	7.3	1423	<i>connaître</i>	6.1	461
Spanish	<i>saber</i>	7.5	1855	<i>conocer</i>	6.4	613
Greek	<i>kséro</i>	6.3	547	<i>γνóριζο</i>	6.1	448
Czech	<i>vědět</i>	7.4	1691	<i>znát (se)</i>	5.4	220
Russian	<i>znat’</i>	7.4	1579	<i>znat’</i>	= KNOW	
Lithuanian	<i>(ne)žinoti</i>	7.37	1589	<i>pažinti</i>		
Welsh	<i>gwybod</i>	6.9	961	<i>adnabod</i>		
Estonian	<i>teadma</i>	7.6	2058	<i>tundma</i> ‘feel’	6.5	997
Finnish	<i>tietää</i>	6.8	873	<i>tuntea</i> ‘feel’	6.5	639
Hungarian	<i>tud</i> ‘know, can’	7.9	2688	<i>ismer</i>	5.8	332

Table 17. Verbs of knowledge in European languages.

Latin *sapere* ‘taste; have sense, be wise’ is the source of KNOW in the Romance languages except Romanian: French *savoir*, Spanish *saber* and Italian *sapere*; Romanian *ști* (from Latin *scire*). In most Germanic and Slavic languages, KNOW can be traced back to the Indo-European root **weid-*, perfective **woida* (have seen > know) and this represents the most well-known example of the extension from SEE to KNOW (see Sweetser 1990). Examples are Swedish *veta*, German *wissen*, Czech *vědět* and Polish *wiedzieć*. Another source is IE **gēnə-/gēnō-*. This is the source of English *know* and Russian *znat’* but also of a second verb (KNOW2) in several languages such as French *connaître* and Spanish *conocer*.

and Greek *ynorízo*. The same root has also developed into Gothic *kunnan*, which has developed further into a modal verb. The German verb *kennen* and Swedish *känna* are causative versions (Gothic *kan+jan*), that represent KNOW2. In the Fenno-Ugrian languages, Hungarian *tud* ‘know, can’ is historically related to Estonian *tundma* and Finnish *tuntea* (Collinder 1955: 63), which both cover ‘feel’ as well as KNOW2. The semantic distinction in Greek between *kséro* KNOW and *ynorízo* KNOW2 is unclear. Both verbs can refer to ‘know as a fact’. Variation across registers appears to be involved. *Kséro* is the dominant translation of *know* in Subtitles, whereas *ynorízo* dominates in Europarl based on WAD.⁶

Table 18 shows how some representative meanings of *know* are distributed in a selection of European languages (see also Viberg 2005, 2015b).

Meaning	Example	SWE	SPA	CZE	RUS	LIT
KNOW A FACT	Mary knows that Harry left	<i>veta</i>	<i>saber</i>	<i>vědět</i>	<i>znat’</i>	<i>(ne)žinoti</i>
KNOW ABOUT A FACT	Mary didn’t know of/about the decision	<i>känna</i> <i>till</i>	<i>conocer</i>	<i>znát</i>	<i>znat’</i>	?
KNOW A PERSON	Mary knows Harry	<i>känna</i>	<i>conocer</i>	<i>znát</i>	<i>znat’</i>	<i>pažinti</i>
RECOGNITION	Mary recognized Harry	<i>känna</i> <i>igen</i>	<i>reconocer</i>	<i>poznat</i>	<i>uznavat’</i>	<i>(ne)atpažinti</i>

Table 18. Distributions of meanings across KNOW and KNOW2.

Like English, Russian primarily uses one verb *znat’* to cover both KNOW and KNOW2, whereas the other languages use two verbs. A frequent meaning of Spanish *conocer*, which is not shown, is ‘meet’.

9.2.2. Verbs of thinking

The verb *think* equals the nuclear verbs in English in terms of its general meaning and its high frequency (R 15, 1539 pm, LN 7.3). It is, however, problematic to regard THINK as a nuclear verb characteristic of European languages in general because English *think*

⁶ Subtitles Δεν ξέρω πώς βρέθηκε στην εφημερίδα. <den *kséro* pós vréthike stin efemerída>
‘I don’t know how it got in the paper.’

EuroParl Όλοι γνωρίζουμε ότι η αειφόρος ανάπτυξη είναι μια διαδικασία γνώσης.
<óli *ynorízume* óti i aifóros anáptiksi íne mia diadikasía gnósis>
‘We all know that sustainable development is a learning process.’

appears to be much more general than the closest correspondences in many other European languages. A verb roughly glossed BELIEVE tends to come closer in frequency to THINK than in English, and in Swedish, *think* has three frequent translations (Viberg 2005: 142–143. See also the experimental study by Frommherz 2022).

The picture is complicated also by the fact that the verbs of thinking tend to develop special discourse and interactional functions, when they are used with first person present reference. Such uses can reach a very high frequency in some languages – especially in various types of spoken discourse. Aijmer (2015: 204) discusses the use of English *I think* as a discourse particle with very flexible meaning. Fetzer & Johansson (2012) compare the use with first person reference of the English verbs *think* and *believe* with French *penser* and *croire* in argumentative discourse. Parenthetical uses are most frequent with *I think* in English, whereas such uses are most frequent with *je crois* in French. Pragmatically based uses also are at play in Spanish and motivate why *creer* is more frequent than *pensar*, which is categorized as THINK. In Spanish according to Mulder (2017), *creo* ‘I believe’ stands out as marker of evidentiality and epistemic modality, whereas *pienso* ‘I think’ more often appears in its primary meaning and refers to the act of thinking or reasoning or is used to express intentions or plans.

9.3. *Desire*

Desire refers to verbs such as *want*, *desire*, *hope* and *will*. The nuclear verb in present-day English is *want* even if *will* can still have a similar meaning besides its use as future and modal auxiliary. The verb *want* can refer to a person who wants an object (*I want a beer*), someone who wants to do something (*I want to leave*) or someone who wants another person to do something (*I want you to leave*). *Want* can also refer to someone wanting other types of events to occur (*I want it to stop snowing*). European languages contrast with respect to which types of constructions are used to express such functions. The major correspondences of *want* are shown in Table 19.

In European languages, the situation where someone wants to do something is in general expressed with the construction NP₁ + VP^{Infinitive} (see (7)). In this construction, the subject is coreferential with the Wanter and left implicit in the complement. The only exception is Greek and some other Balkan languages, where the coreferential subject is marked by a person ending on the verb in the clausal complement (SBJV: subjunctive particle; *ði* [ðɛi] 3SG simple subjunctive, a perfective form).

(7) Greek (OPUS Subtitles)

kápios *théli* *na* *me* *ði*
 someone want.PRS.3SG SUBJUNCTIVE me see.PRF.3SG
 ‘Oh, a man just wants to see me.’

Haspelmath (2013) in a typological survey based on 283 languages shows that the complement subject is implicit in the majority of the languages (51%) and that an explicit subject is found in 25% of the languages. The rest signal the desiderative meaning with an affix or a particle. In some languages, a verb is used that also means ‘say’ and/or ‘think’. The typological study by Harkins (1995) based on 50 diverse languages gives several examples of that.

Language	WANT	LN	pm
	<i>form</i>		
English	<i>want</i>	6.85	943
German	<i>wollen</i>	7.61	2017
Swedish	<i>vilja</i>	7.34	1536
French	<i>vouloir</i>	7.37	1586
Spanish	<i>querer</i>	7.27	1435
Greek	<i>thélo</i>	7.12	1236
Czech	<i>chtít</i>	7.02	1120
Russian	<i>(za)xotet’</i>	7.07	1174
Lithuanian	<i>(ne)norėti</i>	7.40	1641
Welsh	<i>eisiau</i>	6.23	509
Estonian	<i>tahtma</i>	7.26	1427
Finnish	<i>halua</i>	7.32	1508
Hungarian	<i>akar</i>	6.77	871

Table 19. WANT in European languages.

9.4. Other basic verbs of cognition

Two more verbs in addition to the ones discussed above tend to appear among the 100 most frequent verbs. A verb meaning DECIDE can be found in 11 of the languages and a verb meaning UNDERSTAND in 10. Memory is an important part of cognition, but REMEMBER and FORGET appear among the 100 most frequent verbs only in a few of the languages.

10. How (un)usual are European languages? Universal and language-specific aspects of the inventories of the nuclear verbs in European languages

The characterization of basic verbs presented in this paper is intended to serve as a reference point for studies aiming to produce descriptions of basic verbs in the world's languages. To what extent is it possible to find similarities and to what extent is it possible to identify languages that represent completely different types? More specifically, to what extent can equivalents to the nuclear verbs be found among the most frequent verbs in non-European languages and is it possible to identify other lexical verbs that hold a similar position?

The nuclear verbs do not represent any absolute universals in the sense that a direct equivalent can be found in all languages. In each of Sections 5-9, examples have been given of languages that lack a simple verb that expresses the corresponding meaning. It appears, however, that the nuclear verbs have frequent equivalents in many languages. Their status as universal tendencies appears to be an interesting working hypothesis that is worth testing. Viberg (2012) looked at data on the 20 most frequent verbs in some non-European languages with a special focus on nuclear verbs. As can be observed in Table 20, the nuclear verbs tend to be basic also in these languages.

Motion	Possession	Production	Verbal	Perception	Cognition	Desire
GO (9)	GIVE (6)	MAKE (9)	SAY (9)	SEE (9)	KNOW (8)	WANT (7)
COME (8)	TAKE (3)					
	GIVE/TAKE (1)					

Languages: Kurdish (Kurmanji), Turkish, Mongolian, Megrelian, Arabic, Chinese, Sango, Swahili, Lao.

Table 20. The nuclear verbs in 9 non-European languages.

The result supports the working hypothesis, but the test is based on a very restricted database and frequency data from a representative sample of the world's languages will probably not be available in the recent future. A more promising approach would be to look at descriptions of languages with a restricted set of simple verbs. The most well-known are north Australian languages (more exactly languages in north-western and north-central Australia), see Dixon (1980: 280–281). An example that shows how such a language can look is presented in Table 21 based on the sketch of Nungali (nung1291; Mirndi, Yirram) in Hoddinot & Kofod (1976).

General	Motion	Motion	Possession	Physical	Production
Dynamic	Subject	Object		Contact	
<i>-nand(u)-</i>	<i>-Nga-</i>	<i>-ra-</i>	<i>-Nana-</i>	<i>-NaNga-</i>	<i>-malin-ma-</i>
‘be/stay’	‘go’	‘put’	‘give’	‘hit’	‘make’
	<i>-ruma-</i>	<i>-djama-</i>	<i>-ndja-</i>	<i>-maNu-</i>	
	‘come’	‘bring’	‘take’	‘hit/kill’	
	<i>-da-</i>	<i>-mili-</i>	<i>-yal-gara-</i>		
	‘fall’	‘gather’	‘send’		
	<i>-wada-gara-</i>	<i>-Ngu-</i>			
	‘follow’	‘throw’*			
Verbal	Perception	Disconnection	Ingestion	Cooking	Other Fields
Communication					
<i>-Ngu-</i>	<i>-Nawu-</i>	<i>-wiri-</i>	<i>-ndji-</i>	<i>-riga-</i>	<i>-yada-</i>
‘say (throw) ’	‘see’	‘bite’	‘eat’	‘cook/burn’	‘spear’
	<i>-mala-Nawu-</i>				<i>-nama-</i>
	‘hear’				‘dance’

*-Ngu- ‘say/throw’ is listed twice in accordance with its two meanings

Table 21. The 22 simple verbs in Nungali (Hoddinott & Kofod 1976).

Seven of the nuclear verbs have parallels according to the glosses: GO, COME, GIVE, TAKE, MAKE, SAY and SEE. HIT and EAT represent verbs that tend to have a high frequency in several non-European languages.

There is a systematic overview by McGregor (2002) of Australian languages with restricted sets of simple verbs according to which such languages have a very small inventory of inflecting verbs (from half a dozen up to around one hundred). These are typically combined with large sets of particles or other types of adjuncts to form compound verbs. Among the inflecting verbs, there are fourteen basic verbal meanings that tend to appear across languages with small sets of inflecting verbs. The lexical meanings of these verbs when they are used as simple verbs tend to belong to seven semantic domains characterized in the following way: “speech (‘say, do’); stance (‘sit, be’, ‘stand’); motion (‘go’, ‘fall’, and ‘become’); induced motion (‘carry’, ‘throw’, ‘put’); acquisition (‘catch, get’, ‘give’); violence (‘hit’, ‘poke’); and perception (‘see’)” (McGregor 2002: 153). This list includes approximate parallels to four nuclear verbs: SAY, GO, GIVE and SEE. HIT is a potential nuclear verb in the sense that it serves as the core within the field of physical contact verbs and has rich pattern of polysemy in many languages including grammatical uses (Viberg 2016c), but it does not appear among the 20 most frequent verbs in any European language. EAT has a wide range

of meanings in many languages including grammatical uses in some languages (Newman 2009).

A small, closed set of simple verbs is also characteristic of several Papuan languages. Foley (1986: 127) refers to “the comparative dearth of verb-stems” in many Papuan languages. This is in general compensated for in one of two ways, either by verb serialization, which is prominent among Sepik languages (e.g., Yimas, Alamlak, Barai), or by adjunct constructions, i.e. compound verbs consisting of nominal + verb-stem, which are prominent among highlands languages (e.g., Hua, Enga and Kewa). The most reduced number of verbs is found in the two closely related Kobon (kobo1249; Nuclear Trans New Guinea, Madang) (Davies 1981) and Kalam spoken in the Kaironk valley. Kalam has been described in quite some detail by Pawley (e.g., 1993), who has also written a survey article comparing Australian and Papuan languages with closed verb classes (Pawley 2006). There are only around 130 verb roots in Kalam, and among these the ten most frequent ones make up almost 80% of the verb roots in running text. Kalam can form complex verbs by adding adjuncts to the simple verbs as in north Australian languages but in addition uses verb serialization to a certain extent. The verbs that are most productively used to form complex verbs are glossed SAY, HAPPEN/MAKE, GO, COME and STABILIZE (PUT/FORM) (Pawley 2006: 1–6). Pawley also compares the meaning of the most frequent verbs in English with the most frequent verbs in Kalam and in Jaminjung, a north Australian language, whose system of simple and complex verbs has been thoroughly described in Schultze-Berndt (2000). Pawley’s basic conclusion is that the three languages are “not so different after all”, which lends some support to the claim that the meaning of the European basic verbs reflect universal tendencies. One general characteristic is that the languages with closed verb classes can have certain verb roots that cover the meaning of several of the European nuclear verbs. To what extent these “generalized nuclear verbs”, as they may be called, are polysemous and have separate senses corresponding to the various nuclear verbs or whether they have a more general meaning appears to be difficult to determine. An example of this is Kalam *ɲɲ*-‘perceive’, which (often as part of a complex verb) covers SEE in addition to other verbs of perception and also cognitive verbs such as KNOW and THINK. These meanings are closely related also in languages with open verb classes.

Many Indo-Iranian languages also have relatively small classes of simple verbs. Liljegren (2010: 55) shows how Palula (phal1254; Indo-European, Indo-Aryan) spoken in northern Pakistan places itself between the European languages and Kalam with respect to its reliance on its most frequent verbs. In Palula, the five most frequent

verbs account for 51.5% of the total textual occurrences of verbs and the 20 most frequent verbs account for close to 80%. Five of these correspond to nuclear verbs: MAKE, GO, COME, GIVE, SAY; In addition, HIT and EAT appear among the 20 most frequent verbs. There is a relationship between the size of the set of simple verbs and the morpho-syntactic structure of the language. Palula uses compound verbs consisting of V preceded by Noun or another Non-Verb Component: NVC + V to a greater extent than European languages, whereas Kalam relies extensively on both serial and compound verbs.

Persian according to Miller & Aghajanian-Stewart (2017) has around 150 simple verbs with any high frequency and according to Windfuhr & Perry (2014), around 18 of these are used as light verbs to form compound verbs together with a noun or other NVC, for example *goš dadæn* ‘listen’ (lit. ‘ear give’) and *fekr kærðæn* ‘think’ (lit. ‘idea make’). All nine nuclear verbs of European languages appear among the 20 most frequent verbs as can be observed in Table 22. In addition, BE is the most frequent verb but HAVE does not have rank 2 but rank 4, because *ʃodæn* ‘become’ (rank 2) and *kærðæn* ‘make’ (rank 3) are so frequently exploited to form compound verbs. Note that HIT (*zæðæn*, rank 18) also appears among the 20 most frequent verbs.

Concrete actions				Verbal communication
<i>Motion</i>	<i>Possession</i>	<i>Production</i>	<i>Physical contact</i>	
11 <i>ræftæn</i> ‘go’	4 <i>daftæn</i> ‘have’	3 <i>kærðæn</i> ‘make’	18 <i>zæðæn</i> ‘hit’	5 <i>goftæn</i> ‘say’
13 <i>amæðæn</i> ‘come’	10 <i>dadæn</i> ‘give’			
12 <i>residæn</i> ‘arrive’	8 <i>gereftæn</i> ‘take’			
20 <i>gozæftæn</i> ‘put, leave’	16 <i>aværdæn</i> ‘bring’			
Mental			Grammatical	
<i>Perception</i>	<i>Cognition</i>	<i>Desire</i>	<i>Modal</i>	<i>Dynamic</i>
15 <i>didæn</i> ‘see’	14 <i>danestæn</i> ‘know’	6 <i>xastæn</i> ‘want’	7 <i>tævanestæn</i> ‘can’	1 <i>budæn</i> ‘be’
17 <i>nemudæn</i> ‘show, do’			9 <i>bajestæn</i> ‘must’	2 <i>ʃodæn</i> ‘become’
18 <i>jaftæn</i> ‘find’				

Table 22. The most frequent verbs in Persian (based on Miller & Aghajanian-Stewart 2017).

The ranking is based on the raw frequencies.

Family (2014: 25) in her monograph on Persian light verbs lists 14 such verbs in descending order of frequency: *kærdæn* to do, *zædæn* to hit, *dadæn* to give, *gereftæn* to get, *daftæn* to have, *amædæn* to come, *aværdæn* to bring, *xordæn* to eat, *kešidæn* to pull, *šodæn* to become, *bordæn* to take, *ræftæn* to go, *ændaxtæn* to throw and *oftadæn* to fall. Among these, there are 5 nuclear verbs: MAKE ('do'), GO, COME, GIVE and TAKE. In addition, the two verbs HIT and EAT that have extensive patterns of polysemy in many non-European languages are used as light verbs. The verb *xordæn* 'eat' has a high frequency (rank 28, 476 pm, LN 6.2).

Turkish (nucl1301; Turkic, Common Turkic) is particularly interesting since it represents a highly agglutinating language. The display of the 20 most frequent verbs in Table 23 is based on the part of speech index in Aksan et al. (2017: 307).

Concrete actions				Verbal communication
<i>Motion</i>	<i>Possession</i>	<i>Production</i>	<i>Ingestion</i>	
6 <i>gel</i> 'come'	4 <i>al</i> 'take, get'	2 <i>et</i> 'do, make'	20 <i>ye</i> 'eat'	5 <i>de</i> 'say'
9 <i>çık</i> 'go out'	7 <i>ver</i> 'give'	3 <i>yap</i> 'do, make'		18 <i>söyle</i> 'say, tell'
11 <i>git</i> 'go'				
14 <i>geç</i> 'pass'				
Mental			Grammatical	
<i>Perception</i>	<i>Cognition</i>	<i>Desire</i>	<i>Modal</i>	<i>Dynamic</i>
8 <i>gör</i> 'see'	15 <i>bil</i> 'know'	13 <i>iste</i> 'want'		1 <i>ol</i> 'be, become'
10 <i>bul</i> 'find'	16 <i>anla</i> 'understand'			
19 <i>bak</i> 'look at'				
Other: 12 <i>çalış</i> 'work'; 17 <i>kal</i> 'stay, remain'				

Table 23. The 20 most frequent verbs in Turkish (based on Aksan et al. 2017).

Table 23 shows the stems of the verbs. (Turkish dictionaries traditionally use the infinitive marked with the suffix *-mek/-mak* as a citation form.) Because of that, unresolved ambiguities arise making some lemmas homographic. Among the verbs in Table 23, this only affects *et*, which is represented as “*et* v, n (v) to do, to make (n)

meat; flesh” in the frequency index (Aksan et al. 2017: 8). The nuclear verbs appear among the 20 most frequent verbs also in Göz (2003), which is based on a corpus with around one million words and shows the verbs lemmatized in the infinitive form. In this list, *etmek* does not appear among the 20 most frequent verbs.

All the nuclear verbs have clear parallels except that there are two verbs glossed ‘do, make’. It is worth noticing that Turkish belongs to the languages where EAT appears among the most frequent verbs. The most frequent grammatical verbs in European languages do not have direct correspondences. BE is special (see Lees 1972).

HAVE and the Modals are missing from the list of the most frequent verbs. HAVE is expressed with the existential predicate *var* ‘exist’ and its negative counterpart *yok*. Modals are described in Johansson (2009). Possibility is expressed with the suffix *-(y)A-bil* derived from the verb *bilmek* ‘to know’ preceded by the converb (CVB) *-(y)A-: gel-e-bil-* (come-CVB-know) ‘to be able to come’. This suffix can be interpreted as a morphological correspondence to CAN. Necessity presents a more complex picture and can be expressed with the two adjectives *gerik* or *lazım* (the latter of Arabic origin).

11. The 100 most frequent verbs in European languages

This section will briefly discuss what meanings are characteristic of verbs in European languages with a rank between 1 and 100. The discussion has accounted systematically for all the nuclear verbs that tend to appear among the 20 most frequent verbs. In addition, several verbs with a rank between 21 and 100 have been commented on when they belong to the same field as one of the nuclear verbs. As a background to the following discussion, the 100 most frequent verbs in English are shown in Table 24. The frequency ranking is based on Kilgariff (1998).

Grammatical verbs					
<i>Dynamic system</i>			<i>Phasal</i>	<i>Modal</i>	<i>Attempt/Success</i>
State	Change	Cause			
1 <i>be</i>	26 <i>become</i>	49 <i>help</i>	43 <i>begin</i>	2 <i>will</i>	34 <i>try</i>
	61 <i>happen</i>	67 <i>let</i>	46 <i>start</i>	5 <i>can</i>	
	74 <i>change</i>		66 <i>continue</i>	12 <i>shall</i>	
			85 <i>stop</i>	14 <i>may</i>	
				23 <i>must</i>	
				30 <i>need</i>	

(continued)

Concrete actions				
Posture	Motion		Possession	Creation
	Subject	Object		
59 stand	7 go	24 put	2 have	9 make
64 sit	16 come	41 turn	8 get	68 produce
100 lie	28 leave	42 bring	13 take	93 build
	40 follow	48 set	17 give	
	47 run	58 lead	37 keep	
	51 move	62 carry	38 provide	
	81 fall	86 send	52 pay	
	92 return		76 lose	
	99 reach		84 buy	
			91 receive	
			94 spend	
Mental verbs				
Metalinguistic	Verbal Communication	Perception	Cognition	Desire
27 mean	6 say	10 see	11 know	21 want
	22 tell	18 look	15 think	
	33 ask	20 find	56 believe	
	36 call	31 feel	72 consider	
	45 write	34 show	78 expect	
	63 talk	53 hear	79 remember	
	70 require	65 appear	87 decide	
	71 suggest	73 read	89 understand	
	75 offer		98 learn	
	82 speak			
	95 describe			
	96 agree			
Other fields: 4 do, 19 use, 25 work, 39 hold, 50 play, 54 meet, 55 include, 57 allow, 60 live, 69 involve, 77 add, 80 remain, 83 open, 88 win, 90 develop, 97 increase.				

Table 24. The 100 most frequent verbs in English classified into semantic fields based on Kilgarriff (1998).

11.1. Grammatical verbs

This paper has focused on verbs with a lexical meaning except BE and the modals CAN and MUST. Verbs referring to “Concrete actions” and “Mental verbs” have already been discussed in Sections 5-9. For that reason, the verbs referred to broadly as “grammatical verbs” will be commented on separately in this section, see Table 25.

In this table and Table 27, the number of languages in which a verb with the indicated meaning can be found among the 100 most frequent verbs is given within parentheses. Meanings which usually appear among the 20 most frequent verbs are shown in bold face. Lexical verbs with frequent grammaticalized meanings do not appear in the table, since information about the frequency of various uses of these verbs is available only to a limited extent (see discussion below).

Grammatical verbs					
<i>Dynamic system</i>			<i>Phasal</i>	<i>Modal</i>	<i>Attempt/Success</i>
State	Change	Cause			
BE (13)	BECOME (9)	LET (9)	BEGIN (13)	CAN (13)	TRY (11)
	CHANGE (10)	HELP (9)	CONTINUE (9)	MUST (11)	SUCCEED (6)
	HAPPEN (13)		END (10)		

Table 25. Verbs with a grammatical meaning among the 100 most frequent verbs.

The group labelled “Dynamic system” refer to State, Change (or Inchoative) and Cause. Table 26 shows how BECOME, which appears among the 100 most frequent verbs in 9 of the languages, can be related to BE (dealt with in Section 4) and to MAKE (Section 5).

Dynamic	Verb	Property	Class inclusion
State	<i>be</i>	Eve is happy	Alphie is a big star.
Change	<i>become</i>	Eve became happy	Alphie became a big star.
Cause	<i>make</i>	Bob made Eve happy	The famous director made Alphie a big star

Table 26. A sketch of the dynamic system.

BE is realized as a verb in all European languages but BECOME is realized by several different verbs with other basic meanings in some languages, for example Spanish (*ponerse* ‘put + REFL’, *hacerse* ‘make + REFL’, *volverse* ‘turn + REFL’). In Finnish, the major realization of BECOME is *tulla* ‘come’, but several other options exist. Tentatively HAPPEN is also regarded as part of the dynamic system together with CHANGE.

As mentioned, the grammatical meanings are in many cases expressed with extended uses of lexical verbs. The examples of “Cause” in Table 26 correspond to Causative-Adj and Transformation in Table 10, which illustrates uses of MAKE. A verb with this core meaning is used as a periphrastic causative in several languages. The reason why this verb does not appear in Table 25 is that only grammatical verbs

without a basic lexical meaning are shown. This is a consequence of the decision to assign only one basic meaning to each verb. The principle to stick to one meaning for each verb is done to get a manageable overview of lexical structure. The grammatical meanings of lexical verbs are accounted for both in the description of the pattern of polysemy of each verb and in the description of the systems of grammatical meaning briefly discussed here. It is an open question whether some verbs which have both a frequent lexical meaning and a frequent grammatical meaning should be treated as two separate verbs. At present, only information about the overall frequency of such verbs is available for many languages.

The function Causative-VP in Table 10 as in *He made me laugh*, which is usually referred to as the periphrastic or analytic causative, is expressed with several different verbs across languages. A variety of finer semantic distinctions can be expressed in this way but there is a major contrast between make-type and let-type causation according to Levshina (2015), who has analyzed such structures in 18 European languages. The former type is expressed by MAKE in English, the Romance languages and Greek but with GET in Finnish and Swedish. Observe that *doen*, not *maken*, can be used to express this type of causative in Dutch (see Table 11). A verb glossed ‘let’ and in some languages ‘let, leave’, which is symbolized LET, appears in 9 languages, where it can be used as a verb of causation. Tentatively, this meaning is regarded as the basic one, but LEAVE might be more appropriate for some languages. In Swedish, there is a language-specific combination of the meanings LET and SOUND in the verb *låta* (Viberg 2009). A verb with the basic meaning GIVE fulfils that function in Russian, Finnish and Greek. A verb meaning ALLOW appears in as many as 11 of the languages, but it is left open for the time being whether the basic meaning is causative or belongs the field of Verbal communication verbs (‘say that something is permitted’). According to Levshina (2015: 514), “Letting” is the prototype of European causatives, whereas make-causatives are less salient. One reason for this is that they compete with lexical causatives and in some languages also with morphological causatives as in Finnish and Hungarian. HELP represents a special type of causative meaning and a verb with that meaning is found among Top 100 in 9 languages.

Phasal verbs represent another field of grammatical verbs that tend to be frequent in European languages. All languages have at least one verb representing each of the meanings BEGIN and STOP (or FINISH, END) among the Top 100 and CONTINUE is found in this range in most of the languages. The most basic *Modals* are described in Section 4. English has unusually many very frequent modal verbs but that is rather a language-specific feature. *Attempt/Success* refers to the concepts Attempt, Success and Failure. TRY is the most frequent verb of this type and appears among the top 100 in 11 languages. Finnish is particularly rich in such verbs (Flint 1980).

The limits set up when looking at the 20 or 100 most frequent verbs are heuristic but in principle arbitrary. Based on the first limit, it could be established that the nuclear verbs could be found within this limit with only a restricted number of exceptions and that most of the exceptions had a rank only slightly above 20. When looking at the 100 most frequent verbs, no systematic attempt has been made to find out whether a certain verb could be found outside this range, which means that verbs that deviate only slightly are missed. Further studies would have to include more verbs, say with a rank up to 200, to find the most informative cut-off point.

11.2. Lexical verbs

Table 27 sums up the set of lexical verbs discussed in Sections 3-9 together with a few more language-specific verbs that will be commented on below.

Concrete actions					
<i>Posture</i>	<i>Motion</i>		<i>Possession</i>	<i>Creation</i>	<i>Life cycle</i>
	Subject	Object			
STAND (8)	GO (13)	LEAD (11)	HAVE (8)	MAKE (13)	LIVE (13)
SIT (7)	COME (13)	CARRY (10)	TAKE (13)	CREATE (11)	
LIE (5)	LEAVE (11)		GIVE (13)		
	FOLLOW (10)		BRING (9)		
			PAY (9)		
			SELL (7)		
			GET (9)		
			RECEIVE (9)		
Mental verbs					
<i>Metalinguistic</i>	<i>Verbal communication</i>		<i>Perception</i>	<i>Cognition</i>	<i>Desire</i>
MEAN (7)	SAY (13)		SEE (13)	KNOW (13)	WANT (13)
	SPEAK (12)		HEAR (12)	THINK (12)	
	WRITE (13)		SHOW (12)	BELIEVE (12)	
	ASK (13)		FEEL (12)	DECIDE (11)	
	ANSWER (7)		SEEM (13)	UNDERSTAND (10)	
	TELL/narr. (8)				
Other verbs: USE (12), WORK (12)					

Table 27. Lexical verbs that appear among the 100 most frequent verbs

The field “Life cycle” has been added, which contains LIVE in the sense ‘be alive’ together with verbs such as GROW and DIE. Among “Other verbs”, USE might have been considered as a grammatical verb categorized as Instrumental, because of the alternation *Peter **used** a crowbar to open the window / Peter opened the window **with** a crow bar*. A verb glossed WORK appears among the 100 most frequent verbs in 12 languages. The sense ‘do a job’ appears to be basic but that requires further study.

11.3. Language-specific and variable features

The lists of the 100 most frequent verbs can also be used to identify language-specific features in European languages. Verbs that have a high frequency only in one language or in a restricted group of languages are language-specific. One example of that is DO discussed in Section 5. English *tell* discussed in Section 8 should also be regarded as language-specific. Another type of language-specific characteristic is the lack of a verb that is frequent in most of the other European languages such as the lack of a frequent verb meaning BRING in Swedish and Welsh as described in Section 7. Some features are variable without being language-specific since they are characteristic of certain groups of languages. An example is *Postural verbs*, which have a high frequency in Germanic and Slavic languages, in which all the three verbs SIT, STAND and LIE appear among the 100 most frequent verbs, whereas other European languages only occasionally have a postural verb in this frequency range. One important reason for this is the frequent use of the Postural verbs to indicate the location of physical objects in the Slavic and Germanic languages, except in English where location most frequently is indicated with BE and *sit*, *stand* and *lie* have a lower frequency than their correspondences in Swedish and German, where postural verbs are among the 50 most frequent verbs (see Viberg 2013a). Among verbs of Motion, ENTER and EXIT, which are characteristic of V-framed languages, have already been discussed in Section 6.4. Among the verbs of Possession, verbs meaning GET (in the sense ‘receive’) have a particularly high frequency in Swedish (*få*), Finnish (*saada*) and Estonian (*saama*), where these verbs have rank 5 or lower, see the articles in Lenz & Rawoens (2012). The verbs glossed RECEIVE have a much lower frequency even if they appear among the 100 most frequent verbs. Language-specific and other variable features that are characteristic of the 100 most frequent verbs are displayed in Table 28.

Semantic field	Variable features
Production	DO vs. MAKE
Spatial	
<i>Location</i>	BE vs. Postural location verbs
<i>Motion</i>	Assisted vs. Non-assisted motion The deictic component of COME A general verb PUT vs. other alternatives Path verbs
Possession	HAVE vs. BE A frequent, polysemous GET TAKE vs. TAKE ^{CM} TAKE vs. BRING GIVE → LET
Verbal	SAY vs. TELL ASK = REQUEST vs. two separate verbs
Perception	
<i>Experiences</i>	Separate verbs in all sense modalities vs. FEEL extended to various extents
Cognition	The extension of KNOW Contrasts between verbs of thinking

Table 28. Some important variable features in European languages.

12. Why are nuclear and other basic verbs important?

Viberg (1993) claimed that nuclear verbs tend to have many hyponyms, to have characteristic and elaborate patterns of polysemy and to have a strong tendency to develop grammatical meanings. Grammaticalization is an aspect of ordinary language change. Nuclear verbs also play a central role in lexical progression and regression, i.e. in developing and in reduced lexical systems. My own interest in research on basic vocabulary arose when I studied the acquisition of Swedish as a second language by children in preschool and early primary school. It turned out that frequency of occurrence was an important factor for the acquisition of verbs. At early stages, several of the nuclear verbs were relied on heavily in oral production. Such verbs were overused in terms of frequency in comparison to L1 children of the same age and had a strong tendency to be semantically overextended (Viberg 1993, 1999b). Together

the nuclear verbs cover major distinctions in the semantic space of actions in a systematic way. Learners can make themselves understood to a certain degree even if the reference is less precise than when more specific verbs are used by native speakers. As a set, nuclear verbs cover most of the different types of argument structures characteristic of an individual language and in several respects serve as syntactic prototypes (Viberg 1993, 2002).

More recently, several studies have appeared that point to the importance of (what I refer to as) nuclear verbs in first language acquisition. Maouene et al. (2008, 2016) related the early development of the verb lexicon to embodiment and looked at the relationship between early verbs, body parts and syntactic frames (argument structure) in corpora of naturalistic speech by two groups of children 20 and 28 months old and of their mothers' speech in interactions with their children. With growing age, there was a clear development from relatively many mouth verbs (e.g., *bite*, *blow*, *drink*), to more hand verbs (e.g., *put*, *take*, *write*), to verbs less closely associated with the body. What is especially relevant for the study of basic verbs is the relationships that were found between syntactic frames and the most frequent verbs appearing in a specific frame as shown in Table 29.

28-month-olds and mothers (both groups)			20-month-olds
<i>Syntactic frame</i>	<i>Body region</i>	<i>Most frequent verb(s)</i>	<i>Body region</i>
V	HEAD	<i>go, look</i>	ARM & HEAD
V Locative	LEG	<i>go</i>	LEG
V NP	ARM	<i>get, eat</i>	HEAD & ARM
V NP Locative	ARM	<i>put</i>	ARM
V NP NP	ARM	<i>give</i>	---
V Sentence	HEAD	<i>say, think</i>	---

Table 29. The relationships between syntactic frames and their most frequent verbs in child language (freely based on Maouene et al., 2016).

The result was similar for 28-month-olds and mothers of both age groups, whereas 20-month-olds lacked examples of the two most complex syntactic frames. The general conclusion was that syntactic development was driven both by high frequency and embodiment, which was shown to play an independent role for the type of less frequent verbs that appeared in the frames. Three nuclear verbs are found among the

most frequent verbs in Table 29; *go*, *give*, *say* and perhaps also *think* are nuclear verbs. High frequency plays a role also for some typologically variable verbs such as *put*, which is regarded as a language-specific basic verb. What is specific about nuclear verbs is that they have a high frequency in all European languages and tentatively also in many other languages.

Davis & Landau (2021) present a detailed study how perception and mental (my “cognitive”) verbs interact with sentence frames in children’s language development. There is a special focus on embedded clauses (cf. V + Sentence, above), since such clauses are key to referencing mental states for both verb classes. The study is based on a large sample of child-produced and parent-produced utterances (age range 2;0-4;11). The most frequent verbs are shown in Table 30.

Perception	N	Mental	N	Other
<i>see</i>	2175	<i>know</i>	1216	296
<i>look</i>	1190	<i>think</i>	386	
<i>watch</i>	297	<i>forget</i>	84	
<i>hear</i>	170	<i>guess</i>	70	
TOTAL			5884	

Table 30. The most frequent child-produced perception and mental verbs
(based on Davis & Landau 2021: 35).

The nuclear verbs *see* and *know* dominate in their respective fields, even though some frequent language-specific verbs such as *watch* are also used from early on, but not as frequently. Contrary to what has often been expected, mental verbs turned out to lead the development of embedded clauses, because the children produced embedded clauses for mental verbs more often, when they started to use such clauses more regularly around age three. Perception verbs at this age were frequently used with less complex sentence frames and less abstract meaning.

13. Brief summary and Conclusion

This article can partly be regarded as a replication of an earlier study (Viberg 1993) based on corpora of a restricted size (around 1 million words). Further support has been provided for the identification of nine nuclear verbs and a couple of grammatical verbs as the most basic lexical verbs in European languages, see Table 31.

a. Reflecting universal tendencies: The nuclear verbs						
<i>Motion</i>	<i>Possession</i>	<i>Production</i>	<i>Verbal com.</i>	<i>Perception</i>	<i>Cognition</i>	<i>Desire</i>
GO	GIVE	MAKE	SAY	SEE	KNOW	WANT
COME	TAKE					
b. Dynamic system		c. Areally specific basic verbs		d. Subareally specific basic verb		
<i>Copula</i>		<i>Modal verbs</i>		<i>Possession: state</i>		
BE		CAN		HAVE		
		MUST				

Table 31. Basic verbs in European languages.

In most of the 13 tested European languages, the verbs in Table 31 are among the 20 most frequent verbs and have a frequency above 1100 pm (LN 7.0). The only clear exception is HAVE, which is completely lacking in some languages. There are also a few languages where MUST is not expressed by a verb. It has only been possible to sketch the analysis of individual semantic fields the basic verbs belong to, but more extended discussion can be found in the references to earlier contrastively oriented work by the author and to work by other scholars in Sections 3-9. The next step, which has only been sketched, is to investigate to what extent there are parallels in non-European languages.

The extension of the study to the 100 most frequent verbs points to possible further generalizations. The approach also makes it possible to identify frequent but language-specific basic verbs. Such verbs make up a significant part of the typological lexical profile of an individual language (Viberg 2006a) and are interesting among other things because they tend to represent serious problems for second language learners.

In non-European languages, the concepts associated with the nuclear verbs and other basic verbs may be realized as frequent and basic expressions that are not verbs or are not simple verbs but multiwords or morphologically complex verbs. From a theoretical point of view, it would be interesting to base the comparison on word stems or on morphemes (cf. Section 10 on Turkish). For polysynthetic languages, it appears to be necessary to look at morphemes. Compound words and lexicalized phrases represent another problem but do not in general appear among the 100 most frequent verbs in European languages. One example of a different kind of language is Persian, which was briefly discussed for comparison in Section 10. Several of the characteristic compound verbs are treated as separate entries, for example *fekr kærðæn* ‘think’ and *pejda kærðæn* ‘find’ (*kærðæn* ‘do/make’). This is obviously motivated by

the fact that there are very few simple verbs in Persian. This could motivate to base the comparison directly on concepts such as THINK or FIND. Partly, this has already been done in the present study, but I must leave it to further studies to refine such theoretical issues. (I made an early attempt by looking at the way the sense modality hierarchy of perception verbs is reflected in morphological complexity in Viberg 1983b.) There is no denying that cultural differences lead to certain concepts being more established in a way that is reflected in the distinctions expressed by words belonging to a certain semantic field, but a hypothesis worth testing is that certain concepts are so basic that they tend to be expressed by very frequent expressions in most or all languages. Since European languages have a large number of simple verbs, it is natural that basic events are expressed by frequent simple verbs, but the analysis must be adjusted to accommodate languages where compound verbs are the norm (except for a restricted number of simple verbs) or languages with different types of morphological structure than European languages such as polysynthetic languages.

Acknowledgements

I wish to thank the two anonymous reviewers and Francesca Di Garbo for insightful comments and Nick Ellis, Jonė Grigaliūnienė and Maria Koptjevskaja-Tamm for help with various aspects of the analysis of Welsh, Lithuanian and Russian, respectively. Remaining shortcomings are my own responsibility.

Abbreviations

1 = 1 st person	CM = complement marker	PROG = progressive
2 = 2 nd person	DEF = definite	PRS = present
3 = 3 rd person	FUT = future	PTCP = participle
ACC = accusative	PASS = passive	SG = singular
AD = adessive	PRF = perfect	VN = verbal noun

References

- Aijmer, Karin. 2015. Pragmatic markers. In Karin Aijmer & Christoph Rühlemann (eds.), *Corpus pragmatics. A handbook*, 195–218. Cambridge: Cambridge University Press.

- Aijmer, Karin. 2019. What is happening to *must* in Present-day English? A contrastive perspective on a declining modal auxiliary. In Harumi Sawada, Yoshio Nitta & Masa-aki Yamanashi (eds.), *Bamen to shutaisei/shukansei* [Situation and subjectivity], 387–402. Tokyo: Hituzi Syobo.
- Aksan, Yeşim, Mustafa Aksan, Ümit Mersinli & Umut Ufuk Demirhan. 2017. *A frequency dictionary of Turkish*. London: Routledge.
- Altenberg, Bengt & Karin Aijmer. 2000. The English-Swedish Parallel Corpus: A resource for contrastive research and translation studies. In Christian Mair & Marianne Hundt (eds.), *Corpus linguistics and linguistic theory*, 15–23. Amsterdam: Rodopi.
- Ballmer, Thomas & Waltraud Brennenstuhl. 1981. *Speech act classification. A study in the lexical analysis of English speech activity verbs*. Berlin: Springer-Verlag.
- Bertinetto, Pier Marco, Luca Ciucci & Denis Creissels. 2026a. *Non-verbal predication in the world's languages: A typological survey. Volume 1: Eurasia, North America, South America*. Berlin: De Gruyter Mouton.
- Bertinetto, Pier Marco, Luca Ciucci & Denis Creissels. 2026b. *Non-verbal predication in the world's languages: A typological survey. Volume 2: Africa, Austronesia, Papunesia, Australia*. Berlin: De Gruyter Mouton.
- Bouveret, Myriam (ed.). 2021. *Give constructions across languages* [Constructional Approaches to Language 29]. Amsterdam: Benjamins. <https://doi.org/10.1075/cal.29>.
- Buck, Carl Darling. 1949. *A dictionary of selected synonyms in the principal Indo-European languages*. Chicago & London: The University of Chicago Press.
- Camacho, José. 2012. *Ser* and *estar*: The individual/stage-level distinction and aspectual predication. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The Handbook of Hispanic Linguistics*, 453–475. Oxford: John Wiley & Sons.
- Čermáková, Anna, Jarle Ebeling & Signe Oksefjell Ebeling. 2023. 'Be' verbs in a contrastive perspective: The case of BÝT, BE and VÆRE. *Nordic Journal of English Studies* 22(2). 27–58.
- Collinder, Björn. 1955. *Fenno-Ugric vocabulary*. Stockholm: Almqvist & Wiksell.
- Davies, John. 1981. *Kobon* [Lingua Descriptive Studies]. Amsterdam: North-Holland.
- Davis, E. Emory & Barbara Landau. 2021. Seeing and Believing: The relationship between perception and mental verbs in acquisition. *Language Learning and Development* 17(1). 26–47. <https://doi.org/10.1080/15475441.2020.1862660>.
- DeLancey, Scott. 1985. The analysis-synthesis-lexis cycle in Tibeto-Burman: A case study in motivated change. In John Haiman (ed.), *Iconicity in syntax*, 367–89. Amsterdam: Benjamins.

- Divjak, Dagmar. 2019. *Frequency in language: Memory, attention and learning*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. 1980. *The languages of Australia*. Cambridge: Cambridge University Press.
- Ebeling, Signe Oksefjell. 2017. Bringing home the bacon! A contrastive study of the cognates *bring/bringe* in English and Norwegian. *Kalbotyra* 70. 104–126. <https://doi.org/10.15388/Klbt.2017.11193>.
- Egan, Thomas. 2021. TELLING in English, Norwegian and French: A three-way contrast. *BLS* 11(1). 47–68.
- Evans, Nicholas & David Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76(3). 546–592.
- Family, Neiloufar. 2014. *Semantic spaces of Persian light verbs. A constructionist account* [Brill's Studies in South and Southwest Asian Languages Series 6]. Leiden: Brill.
- Ferguson, Charles. 1976. The Ethiopian language area. In M. Lionel Bender, J. Donald Bowen, Robert L. Cooper & Charles A. Ferguson (eds.), *Language in Ethiopia*, 63–76. Oxford: Oxford University Press.
- Fetzer, Anita & Marjut Johansson. 2012. Cognitive verbs in context. A contrastive analysis of English and French argumentative discourse. In Stefania Marzo, Kris Heylen & Gert de Sutter (eds.), *Corpus studies in contrastive linguistics*, 89–115. Amsterdam: Benjamins. <https://doi.org/10.1075/bct.43.05fet>.
- Fillmore, Charles J. 1997. *Lectures on deixis* [CSLI lecture notes 65]. Stanford: CSLI Publications.
- Flint, Aili. 1980. *Semantic structure in the Finnish lexicon: verbs of possibility and sufficiency* [Suomalaisen kirjallisuuden seuran toimituksia 0355–1768; 360]. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- Fortescue, Michael. 2001. Thoughts about thought. *Cognitive Linguistics* 12(1). 15–45.
- Frommherz, Yannick. 2022. Thinking things in German versus Swedish. A cross-linguistic comparison of verbs of thinking in two genetically close languages. *Studia Linguistica* 76(2). 464–506.
- Gilquin, Gaëtanelle & Åke Viberg. 2009. How to make do with one verb: a corpus-based contrastive study of DO and MAKE. *Revue française de linguistique appliquée* 14(1). 67–82.
- Goddard, Cliff. 2001. Lexico-semantic universals: A critical overview. *Linguistic Typology* 5. 1–65.

- Goddard, Cliff. 2003. Thinking across languages and cultures: six dimensions of variation. *Cognitive Linguistics* 14(2/3), 109–140 Special issue edited by Gary B. Palmer, Cliff Goddard & Penny Lee.
- Göz, İlyas. 2003. *Yazılı türkçenin kelime sıklığı sözlüğü* [A Frequency Dictionary of Written Turkish]. Ankara: Türk Dil Kurumu.
- Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13(4). 403–437.
- Hammarberg, Björn & Åke Viberg. 1977. The place-holder constraint, language typology, and the teaching of Swedish to immigrants. *Studia Linguistica* 31. 106–163.
- Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2026. Glottolog 5.3. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at: <http://glottolog.org> (Accessed 2026-07-06).
- Hansen, Björn & Ferdinand de Haan (eds.). 2009. *Modals in the languages of Europe*. Berlin: Mouton de Gruyter.
- Harkins, Jean. 1995. Desire in language and thought: A study in cross-cultural semantics. Canberra: Australian National University. (Doctoral Dissertation.) <https://doi.org/10.25911/5d7638e7c1b34>.
- Harras, Gisela, Edeltraud Winkler, Sabine Erb & Kristel Proost. 2004. *Handbuch deutscher Kommunikationsverben 1. Wörterbuch*. Berlin: de Gruyter.
- Harras, Gisela, Kristel Proost & Edeltraud Winkler. 2007. *Handbuch deutscher Kommunikationsverben 2. Lexikalische Strukturen*. Berlin: de Gruyter.
- Haspelmath, Martin. 2013. ‘Want’ complement subjects. In Matthew Dryer & Martin Haspelmath (eds.), *WALS Online*. Available online at: <http://wals.info/chapter/124> (Accessed 2026-07-06)
- Heine, Bernd. 1997. *Possession. Cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Hoddinott, William G. & Frances M. Kofod. 1976. Djamindjungan. In R. M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 698–704. New Jersey: Humanities Press.
- Isačenko, Alexander V. 1974. On ‘have’ and ‘be’ languages. In Michael S. Flier (ed.), *Slavic forum: Essays in linguistics and literature*, 43–77. The Hague: Mouton.
- Jakobson, Roman. 1971. *Selected Writings*, vol. 2. The Hague: Mouton.

- Johansson, Lars. 2009. Modals in Turkic. In Björn Hansen & Ferdinand de Haan (eds.), *Modals in the languages of Europe*, 487–510. Berlin: Mouton de Gruyter.
- Juvonen, Päivi. 2016. Making do with minimal lexica. Light verb constructions with MAKE/DO in pidgin lexica. In Päivi Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts*, 223–248. Berlin: de Gruyter Mouton.
- Kilgariff, Adam. 1998. BNC lists. <https://www.kilgariff.co.uk/bnc-readme.html> (Accessed 2026-07-06).
- King, Gareth. 2016. *Modern Welsh. A comprehensive grammar*. London: Routledge.
- Kopecka, Anetta. 2006. The semantic structure of motion verbs in French: Typological perspectives. In Maya Hickman & Stéphanie Robert (eds.), *Space in languages. Linguistic systems and cognitive categories*, 83–101. Amsterdam: Benjamins.
- Kopecka, Anetta & Bhuvana Narasimhan (eds.). 2012. *Events of Putting and Taking*. Amsterdam: Benjamins.
- Kuiper, Alberta & William R. Merrifield. 1975. Diuxi Mixtec verbs of motion and arrival. *International Journal of American Linguistics* 41(1). 32–45.
- Langlotz, Andreas. 2006. *Idiomatic creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam: Benjamins.
- Lees, Robert B. 1972. The Turkish copula. In John W. M. Verhaar (ed.), *The Verb 'Be' and its synonyms. Philosophical and grammatical studies*, Part 5, 64–73. Dordrecht: Reidel Publishing Company.
- Lenz, Alexandra, Fabian Fleißner, Agnes Kim & Stefan Michael Newerkla. 2020. GIVE as a German PUT verb – Analyses on Slavic-German language contact. *Journal of Linguistic Geography* 8(2). 67–81.
- Lenz, Alexandra N. & Gudrun Rawoens (eds.). 2012. The art of getting: GET verbs in European languages from a synchronic and diachronic point of view. Special issue of *Linguistics* 50(6).
- Leonetti, Manuel, Silvia Gumiel-Molina & Isabel Pérez-Jiménez. 2015. *Ser and estar: Outstanding questions*. In Isabel Pérez-Jiménez, Manuel Leonetti & Silvia Gumiel-Molina (eds.), *New perspectives on the study of ser and estar*, 1–20. Amsterdam: Benjamins.
- Levin, Beth. 1993. *English verb classes and alternations*. Chicago: The University of Chicago Press.
- Levshina, Natalia. 2015. European analytical causatives as a comparative concept: Evidence from a parallel corpus of film subtitles. *Folia Linguistica* 49(2). 487–520. <https://doi.org/10.1515/flin-2015-0017>.

- Lewandowski, Wojciech. 2014. Deictic verbs: typology, thinking for speaking and SLA. *SKY Journal of Linguistics* 27. 43–65.
- Liljegren, Henrik. 2010. Where have all the verbs gone? On verb stretching and semi-words in Indo-Aryan Palula. *Himalayan Linguistics* 9(1). 50–79.
- Lonsdale, Deryle & Yvon Le Bras. 2009. *A frequency dictionary of French*. London: Routledge.
- Lord, Carol. 1993. *Historical change in serial verb constructions* [Typological Studies in Language 26]. Amsterdam: Benjamins. <https://doi.org/10.1075/tsl.26>.
- Loveday, Leo. 1986. *Explorations in Japanese Sociolinguistics*. Amsterdam: Benjamins.
- Margetts, Anna, Katharina Haude, Nikolaus P. Himmelmann, Dagmar Jung, Sonja Riesberg, Stefan Schnell, Frank Seifart, Harriet Sheppard & Claudia Wegener. 2022. Cross-linguistic patterns in the lexicalisation of *bring* and *take*. *Studies in Language* 46(4). 934–993. <https://doi.org/10.1075/sl.19088.mar>.
- Margetts, Anna, Sonja Riesberg & Birgit Hellwig (eds.) (2022). *Caused accompanied motion: Bringing and taking events in a cross-linguistic perspective* [Typological Studies in Language 134]. Amsterdam: Benjamins.
- Maouene, Josita, Shohei Hidaka, & Linda B. Smith. 2008. Body parts and early-learned verbs. *Cognitive Science* 32. 1200–1216.
- Maouene, Josita, Nitya Sethuraman, Mounir M. Maouene & Sango Otieno. 2016. Contingencies between verbs, body parts, and argument structures in maternal and child speech: a corpus study. *Language and Cognition* 8. 237–282. <https://doi.org/10.1017/langcog.2014.48>.
- Martinková, Michaela & Markéta Janebová. 2024. English and Czech venitive Verbs in contrast: Deictic or not?. *Nordic Journal of English Studies* 23(2). 34–60.
- Matsumoto, Yo (ed.). 2025. *Motion event descriptions across languages. Volume 1: Case studies of linguistic representations of motion*. Berlin: De Gruyter Mouton.
- McGregor, William B. 2002. *Verb classification in Australian languages*. Berlin: Mouton de Gruyter.
- Meex, Birgitta. 2020. Patterns of path encoding in German. In Yo Matsumoto & Kazuhiro Kawachi (eds.), *Broader perspectives on motion event descriptions*, 63–104. Amsterdam: Benjamins.
- Miller, Corey & Karineh Aghajanian-Stewart. 2017. *A Frequency dictionary of Persian*. London: Routledge.
- Mulder, Gijs. 2017. When feeling is thinking. A lexical-semantic analysis of evidential and epistemic predicates in Spanish. In Juana I. Marín Arrese et al. (eds.),

- Evidentiality Revisited: Cognitive grammar, functional and discourse-pragmatic perspectives*, 105–121. Amsterdam: Benjamins.
- Naughton, James & Karen von Kunes. 2021. *Czech. An essential grammar*. 2nd edn. London: Routledge.
- Newman, John. 1996. *Give. A cognitive linguistic study*. Berlin: Mouton de Gruyter.
- Newman, John (ed.). 1998. *The linguistics of giving*. Amsterdam: Benjamins.
- Newman, John (ed.). 2009. *The linguistics of eating and drinking*. Amsterdam: Benjamins.
- Nolan, Brian, Gudrun Rawoens & Elke Diedrichsen (eds.). 2015. *Causation, permission, and transfer. Argument realisation in GET, TAKE, PUT, GIVE and LET verbs* [Studies in Language Companion Series, 167]. Amsterdam: Benjamins. <https://doi.org/10.1075/slcs.167>.
- Norcliffe, Elisabeth & Asifa Majid. 2024a. Verbs of perception: A quantitative typological study. *Language* 100(1). 81–123.
- Norcliffe, Elisabeth & Asifa Majid. 2024b. Word formation patterns in the perception domain: a typological study of cross-modal semantic associations. *Linguistic Typology* 28(3). 419–459.
- Nuyts, Jan. 2014. Analyses of the modal Meanings. In Jan Nuyts & Johan van der Auwera (eds.), *The Oxford Handbook of Modality and Mood*, 31–49. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199591435.001.0001>.
- Palmer, Gary B., Cliff Goddard & Penny Lee (eds.). 2003. Talking about thinking across languages. Special issue of *Cognitive Linguistics* 14(2/3).
- Pawley, Andrew. 1993. A language which defies description by ordinary means. In William Foley (ed.), *The role of theory in language description*, 87–129. Berlin: Mouton de Gruyter.
- Pawley, Andrew. 2006. Where have all the verbs gone? Remarks on the organisation of languages with closed verb classes. (Paper presented at the 11th Biennial Rice University Linguistics Symposium, 16-18 March 2006.) http://www.ruf.rice.edu/~lingsymp/Pawley_paper.pdf (Accessed 2026-07-06).
- Pawley, Andrew. 2020. The depiction of sensing events in English and Kalam. In Helen Bromhead & Zhengdao Ye (eds.), *Meaning, life and culture: In conversation with Anna Wierzbicka*. ANU Press. Canberra: Australia. <https://doi.org/10.22459/MLC.2020.20>.
- Raag, Raimo. 2014. *Estniska från grunden* [Finsk-ugriska läromedel FU 25]. Uppsala: Uppsala University.

- Reesink, Ger. 2013. Expressing the GIVE event in Papuan languages: A preliminary survey. *Linguistic Typology* 17(2). 217–266.
- Ricca, Davide. 1993. *I verbi deittici di movimento in Europa: una ricerca interlinguistica* [Deictic Motion Verbs in Europe: A Cross-Linguistic Study]. Florence: La Nuova Italia Editrice.
- Rice, Sally. 1998. Giving and taking in Chipewyan: the semantics of THING-marking classificatory verbs. In John Newman (ed.), *The linguistics of giving*, 97–134. Amsterdam: Benjamins.
- Rudzka-Ostyn, Brygida. 1989. Prototypes, schemas and cross-category correspondences: The case of *ask*. *Linguistics* 27. 613–661.
- Rudzka-Ostyn, Brygida. 1995. Metaphor, schema, invariance: The case of verbs of answering. In Louis Goossens, Paul Pauwels, Brygida Rudzka-Ostyn, Anne-Marie Simon-Vandenberghe & Johan Vanparys (eds.), *By Word of Mouth*, 205–243. Amsterdam: Benjamins.
- San Roque, Lila, Kobin H. Kendrick, Elisabeth Norcliffe, Penelope Brown, Rebecca Defina, Mark Dingemanse, Tyko Dirksmeyer, Nicholas J. Enfield, Simeon Floyd, Jeremy Hammond, Giovanni Rossi, Sylvia Tufvesson, Saskia van Putten & Asifa Majid. 2015. Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics* 26(1). 31–60.
- Sarda, Laure & Ludovica Lena (eds.). 2023. *Existential constructions across languages. Forms, meanings and functions*. Amsterdam: Benjamins.
- Schultze-Berndt, Eva. 2000. *Simple and complex verbs in Jamingung*. Nijmegen: MPI Series in Psycholinguistics.
- Schultze-Berndt, Eva. 2008. What do “do” verbs do? The semantic diversity of generalised action verbs. In Elisabeth Verhoeven, Stavros Skopeteas, Yong-Min Shin, Yoko Nishina & Johannes Helmbrecht (eds.), *Studies on grammaticalization*, 185–208. Berlin: De Gruyter.
- Stassen, Leon. 1997. *Intransitive predication*. Oxford: Clarendon Press.
- Stassen, Leon. 2009. *Predicative possession*. Oxford: Oxford University Press.
- Stassen, Leon. 2013a. Predicative Possession. In Matthew S. Dryer & Martin Haspelmath (eds.), *WALS Online*. Available online at: <http://wals.info/chapter/117> (Accessed 2026-07-06).
- Stassen, Leon. 2013b. Zero Copula for Predicate Nominals. In Matthew S. Dryer & Martin Haspelmath (eds.), *WALS Online*. Available online at: <http://wals.info/chapter/120> (Accessed 2026-07-06).

- Stassen, Leon. 2013c. Predicative Adjectives. In Matthew S. Dryer & Martin Haspelmath (eds.), *WALS Online*. Available online at: <http://wals.info/chapter/118> (Accessed 2026-07-06).
- Sussex, Roland & Paul Cubberley. 2006. *The Slavic languages*. Cambridge: Cambridge University Press.
- Sweetser, Eve. 1990. *From etymology to pragmatics*. Cambridge: Cambridge University Press.
- Talmy, Leonard. 1985. Lexicalization patterns: semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description. III. Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.
- Thornell, Christina. 2002. The uses of the Sango verb *mû* ‘transfer’. *Afrika und Übersee* 85. 267–299.
- Tiedemann, Jörg. 2012. Parallel data, tools and interfaces in OPUS. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 2214–2218. Istanbul: European Language Resources Association (ELRA).
- Tieken-Boon van Ostade, Ingrid, Marijke van der Wal & Arjan van Leuvensteijn (eds.) 1998. *DO in English, Dutch and German. History and present-day variation*. Amsterdam: Stichting Neerlandistiek & Nodus Publikationen.
- van der Auwera, Johan & Vladimir A. Plungian 1998. Modality’s semantic map. *Linguistic Typology* 2(1). 79–124.
- Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning, and function*. Cambridge: Cambridge University Press.
- Verhaar, John W. M. (ed.). 1967-1972. *The Verb ‘Be’ and its synonyms. Philosophical and grammatical studies*, Part 1-6 [Foundations of Language Supplementary Series, vols. 1, 6, 8, 9, 14 & 16.]. Dordrecht: Reidel Publishing Company.
- Viberg, Åke. 1983a. The verbs of perception: a typological study. *Linguistics* 21(1). 123–162.
- Viberg, Åke. 1983b. A universal lexicalization hierarchy for the verbs of perception. In Fred Karlsson (ed.), *Papers from the 7th Scandinavian Conference of General Linguistics*, 260–275. Helsinki: Department of General Linguistics, University of Helsinki.
- Viberg, Åke. 1993. Crosslinguistic perspectives on lexical organization and lexical progression. In Kenneth Hyldenstam & Åke Viberg (eds.), *Progression and regression in language*, 340–385. Cambridge: Cambridge University Press.

- Viberg, Åke. 1999a. The polysemous cognates Swedish *gå* and English *go*. Universal and language-specific characteristics. *Languages in Contrast* 2(1). 87–113.
- Viberg, Åke. 1999b. Lexical development in a second language. In Guus Extra & Ludo Verhoeven (eds.), *Bilingualism and migration*, 165–185. Berlin: Mouton de Gruyter.
- Viberg, Åke. 2001. The verbs of perception. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals. An international handbook*, 1294–1309. Berlin: De Gruyter.
- Viberg, Åke. 2002. Basic verbs in lexical progression and regression. In Petra Burmeister, Thorsten Piske & Andreas Rohde, A. (eds.), *An integrated view of language development. Papers in honor of Henning Wode*, 109–134. Trier: Wissenschaftlicher Verlag.
- Viberg, Åke. 2003. The polysemy of the Swedish verb *komma* ‘come’: A view from translation corpora. In Kasia M. Jaszczolt & Ken Turner (eds.), *Meaning through language contrast*, vol. 2, 75–105. Amsterdam: Benjamins.
- Viberg, Åke. 2005. The lexical typological profile of Swedish mental verbs. *Languages in Contrast* 5(1). 121–157.
- Viberg, Åke. 2006a. Towards a lexical profile of the Swedish verb lexicon. *Sprachtypologie und Universalienforschung* 59(1): 103–129. Thematic issue on: The Typological Profile of Swedish.
- Viberg, Åke. 2006b. What one verb can do: The Swedish verb *göra* in a crosslinguistic perspective. In Mickael Suominen et al. (eds.), *A Man of Measure: Festschrift in Honour of Fred Karlsson on his 60th Birthday. Special Supplement to SKY Journal of Linguistics* 19. Turku, The Linguistic Association of Finland, 243–257.
- Viberg, Åke. 2009. The meaning patterns of Swedish *låta* ‘sound’; ‘let’. A corpus-based contrastive study. In Petra Bernardini, Verner Egerland & Jonas Granfeldt (eds.), *Mélanges plurilingues offerts à Suzanne Schlyter à l’occasion de son 65ème anniversaire*. [Études Romanes de Lund 85], 455–480. Centre for Languages and Literature, Lund University.
- Viberg, Åke. 2010. Basic verbs of possession. In Maarten Lemmens (ed.), *Unison in multiplicity: Cognitive and typological perspectives on grammar and lexis* [CogniTextes 4]. <http://cognitextes.revues.org/308> (Accessed 2026-07-06).
- Viberg, Åke. 2012. Basic verbs in typological perspective. In Marleen Van Peteghem, Peter Lauwers, Els Tobbak, Annemie Demol & Laurence De Wilde (eds.), *Le verbe en verve. Réflexions sur la syntaxe et la sémantique verbales*, 255–273. Ghent: Academia Press.

- Viberg, Åke. 2013a. Posture verbs. A multilingual contrastive study. *Languages in Contrast* 13(2). 139–169.
- Viberg, Åke. 2013b. Seeing the lexical profile of Swedish through multilingual corpora. The case of Swedish *åka* and other vehicle Verbs. In Karin Aijmer & Bengt Altenberg (eds.), *Advances in corpus-based contrastive linguistics. Studies in honour of Stig Johansson*, 25–56. Amsterdam: Benjamins.
- Viberg, Åke. 2013c. Ord om konsten att förflytta sig: Svenskans gå, åka och köra i tvärspråkligt perspektiv. In Björn Lindblom (ed.), *Text, tal & tecken. Några perspektiv inom språkforskningen* [KVHAA Konferenser 83], 116–138. Stockholm: Kungl. Vitterhets Historie och Antikvitets Akademien.
- Viberg, Åke. 2015a. Contrasts in construction and semantic composition: Crosslinguistic perspectives on the verbs of Putting in English and Swedish. In Signe Oksefjell Ebeling & Hilde Hasselgård (eds.), *Cross-Linguistic Perspectives on Verb Constructions*, 222–253. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Viberg, Åke. 2015b. Sensation, Perception and Cognition. Swedish in a typological-contrastive perspective. *Functions of Language* 22(1). 96–131.
- Viberg, Åke. 2016a. What happens in translation? A comparison of original and translated texts containing verbs meaning SIT, STAND and LIE in the English-Swedish Parallel Corpus (ESPC). *Nordic Journal of English Linguistics (NJES)* 15(3), 102–148. Available online at: <https://publicera.kb.se/njes/article/view/26824/21895> (Accessed 2026-07-06).
- Viberg, Åke. 2016b. Asking and answering: A contrastive study of English and Swedish basic communication verbs. *Linguistics Beyond and Within* 2. 180–212. Available online at: <https://czasopisma.kul.pl/index.php/LingBaW/article/view/5646> (Accessed 2026-07-06).
- Viberg, Åke. 2016c. Polysemy in action: The Swedish verb *slå* ‘hit, strike, beat’ in a crosslinguistic perspective. In Päivi Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts*. [Cognitive Linguistics Research 58], 177–222. Berlin & Boston: de Gruyter Mouton.
- Viberg, Åke. 2017a. Contrasts in morphology. The case of UP/DOWN and IN/OUT as bound morphemes in Swedish and their English correspondents. In Markéta Janebová, Ekaterina Lapshinova-Koltunski & Michaela Martinková (eds.), *Contrasting English and other languages through corpora*, 32–74. Cambridge: Cambridge Scholars.

- Viberg, Åke. 2017b. SAYING, TALKING and TELLING. Basic verbal communication verbs in Swedish and English. In Thomas Egan & Hildegunn Dirdal (eds.), *Cross-linguistic correspondences: From lexis to genre*, 37–74. Amsterdam: Benjamins
- Viberg, Åke. 2017c. Finding a model for contrastive lexical semantics: A look at verbal communication verbs. *Linguistics Beyond and Within* 3. 195–215. Available online at: <https://czasopisma.kul.pl/index.php/LingBaW/article/view/5659> (Accessed 2026-07-06).
- Viberg, Åke. 2019. Phenomenon-based perception verbs in Swedish from a typological and contrastive perspective. In Éric Gilbert (ed.), *Perception, perceptibilité et objet perçu. Approches inter-langues* [Syntaxe & Sémantique 20], 17–48. Caen: Presses Universitaires de Caen.
- Viberg, Åke. 2020. Contrasting semantic fields across languages. (Using the verbs of cutting and breaking as an example). In Renata Enghels, Bart Defrancq & Marlies Jansegers (eds.), *New approaches to contrastive linguistics. Empirical and methodological challenges*. [Trends in Linguistics 336], 265–312. Berlin: de Gruyter Mouton.
- Viberg, Åke. 2025. What kind of verb is TAKE? The English and Swedish verbs *take/ta* and their correspondences in other European languages. Unpublished manuscript.
- Viberg, Åke (to appear). Word meaning in comparative and contrastive linguistics. In Dirk Geeraerts & Dylan Glynn (eds.), *The Cambridge handbook of lexical semantics*. Cambridge: Cambridge University Press.
- Wälchli, Bernhard & Michael Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3). 671–710.
- Wierzbicka, Anna. 1987. *English speech act verbs. A semantic dictionary*. New York: Academic Press.
- Wilkins, David P. & Deborah Hill. 1995. When “go” means “come”: Questioning the basicness of basic motion verbs. *Cognitive Linguistics* 6(2/3). 209–259.
- Windfuhr, Gernot & John R. Perry. 2010. Persian and Tajik. In Gernot Windfuhr (ed.), *The Iranian languages*, 416–544. Abingdon: Taylor & Francis Group.

Frequency dictionaries/lists

<i>Language</i>	<i>Size</i> <i>Million words</i>	<i>Reference</i>
Czech	20	Čermák, František & Michal Křen. 2011. <i>A frequency dictionary of Czech</i> . New York: Routledge.

English	100	Kilgariff, Adam. 1998. <i>BNC lists</i> . Available at: https://www.kilgariff.co.uk/bnc-readme.html (Accessed 2026-07-06).
Estonian	15	The Frequency List of Estonian Literary Language https://www.cl.ut.ee/ressursid/sagedused1/ (Accessed 2026-07-06).
Finnish	43	Wiktionary:Frequency lists/Finnish/Press data https://en.wiktionary.org/wiki/Wiktionary:Frequency_lists/Finnish/Press_data (Accessed 2026-07-06).
French	23	Lonsdale, Deryle & Yvon Le Bras. 2009. <i>A frequency dictionary of French</i> . London: Routledge.
German	20	Jones, Randall L. & Erwin Tschirner. 2020, 2 ^{ed} . <i>A frequency dictionary of German</i> . London: Routledge.
Greek	47	Hellenic National Corpus: http://hnc.ilsp.gr/index.php?current_page=main&lang=en (Accessed 2026-07-06; data obtained from Maria Gavrilidou December 2009 from the version that existed at that time)
Hungarian	777	Oravecz, Csaba, Tamás Váradi & Bálint Sass. 2014. The Hungarian Gigaword Corpus. <i>Proceedings of LREC 2014</i> . http://hnc.nytud.hu (Accessed 2026-07-06).
Lithuanian	1	Andrius Utkā, Sudarė. 2009. <i>Dažninių rašytinės lietuvių kalbos žodynas: 1 milijono žodžių morfologiškai anotuoto teksto pagrindu</i> . Kaunas: Vytauto Didžiojo universiteto leidykla. https://sitti.vdu.lt/wp-content/uploads/2019/06/Dazninis_zodynas.pdf (Accessed 2026-07-06).
Persian	10	Miller, Corey & Karineh Aghajanian-Stewart. 2017. <i>A Frequency Dictionary of Persian</i> . London: Routledge.
Russian	150	Sharoff, Serge, Elena Manskaya & James Wilson. 2013. <i>A frequency dictionary of Russian</i> . London: Routledge.
Spanish	20	Davies, Mark. 2006. <i>A frequency dictionary of Spanish</i> . London: Routledge.
Swedish	1	SUC 1997. <i>SUC 1.0. Stockholm Umeå Corpus</i> . Produced by Dept. of Linguistics, Umeå University and Dept. of Linguistics, Stockholm University. CD-Rom.

Turkish	51	Aksan, Yeşim, Mustafa Aksan, Ümit Mersinli & Umut Ufuk Demirhan. 2017. <i>A Frequency Dictionary of Turkish</i> . London: Routledge.
Welsh	11	Knight, Dawn, Steve Morris, Beth Tovey-Walsh, Tess Fitzpatrick, Laurence Anthony. 2020. <i>Yr Amliadur: Frequency Lists for Contemporary Welsh</i> (Version 1.0.0). Cardiff University. https://doi.org/10.17035/d2020.0120164107

Appendix

The identification of CAN and MUST

In several cases, it is difficult to decide which verbs correspond most closely across languages. To a certain extent there are detailed contrastive studies of the relevant verbs. When this is not the case, a resource that has been used is the Word Alignment Database (WAD) included as a tool in OPUS (Tiedeman 2012). When nothing else is said, data are taken from the Subtitles corpus. All 13 languages except Welsh are represented in OPUS.

This appendix will use CAN and MUST, which are described in Table 8 in Section 4, as an example how OPUS has been used. The modals are particularly suited as a test case since there is a wealth of detailed studies against which the result can be evaluated. The same type of analysis has been carried out on several of the nuclear verbs, when the correspondence has been unclear, but it is not possible to account for all such analyses in detail. I plan to make available appendices or articles about other semantic fields based on OPUS and the many other corpora that now are becoming freely available.

CAN. The Word Alignment Database takes a word form in one language as input and shows the word forms that appear in the translations according to a formula. Table App1:1 shows the output when the English word form *can* is used as input. There is a possibility to choose among source texts. The option “Subtitles 2018” was chosen. The result should be seen as tentative, since it should be compared to corpus-based studies carried out with manual coding like the one by Aijmer (2019) referred to in Section 4.

#	prob	cze		#	prob	ell	>>	#	prob	est	>>
84078	(0.14)	může	 	195178	(0.21)	μπορώ	 	88296	(0.27)	saa	 
				84376	(0.17)	μπορώ να	 	30348	(0.12)	suuda	 
				141662	(0.16)	μπορείς	 	19842	(0.08)	võid	 
				50503	(0.11)	μπορεί να	 	20956	(0.08)	saan	 
				67496	(0.08)	μπορούμε	 	20392	(0.07)	võib	 
#	prob	fin	>>	#	prob	fre		#	prob	ger	
338682	(0.44)	voi	 	622830	(0.44)	peux	 	378566	(0.47)	kann	 
25520	(0.06)	voin	 	318466	(0.26)	peut	 	161446	(0.24)	können	 
22518	(0.05)	voit	 	108756	(0.10)	pouvez	 	103404	(0.18)	kannst	 
20664	(0.05)	pysty	 								
25684	(0.05)	saa	 								
#	prob	hun		#	prob	lit	>>	#	prob	rus	>>
53508	(0.08)	lehet	 	8036	(0.20)	gali	 	274684	(0.38)	могу	 
40018	(0.07)	tud	 	4292	(0.13)	galiu	 	102750	(0.17)	можешь	 
40230	(0.07)	tudok	 	1576	(0.06)	galime	 	96580	(0.14)	может	 
45904	(0.06)	tudom	 	1638	(0.05)	Gali	 	80262	(0.14)	можем	 
				1584	(0.05)	Galiu	 	46258	(0.08)	можете	 
#	prob	spa	>>	#	prob	swe		Translations of <i>could</i>			
529512	(0.30)	puedo	 	555334	(0.63)	kan	 	#	prob	est	
414236	(0.26)	puede	 					18648	(0.15)	võib	 
258850	(0.17)	puedes	 					9886	(0.11)	võiks	 
111688	(0.09)	pueden	 					4788	(0.05)	saaks	 
85698	(0.07)	pueda	 								

Table App1:1. The translations of English *can* in OPUS Word alignment (accessed 2026-07-06).

In Table App1:1, the languages are shown in alphabetical order, but the following comments follow the order in Table 8. (*ell* refers to *elleniká* Greek). German is represented by present tense forms of *können* inflected for person and number. In Swedish, the form *kan* dominates completely. This is the present tense of the Swedish cognate *kunna* ‘to be able’, which has the full set of endings of an ordinary verb. It should be noted that verbs are not inflected for person and number in Swedish. The French and Spanish translations contain forms inflected for person and number of the

verbs *pouvoir* and *poder*, which are ultimately derived from Classical Latin *posse* ‘be capable of’ (present *poteo* ‘I am capable’), see Cornillie et al. (2009), Kronning (2024). Greek is represented by the present tense of *boró* ‘I can’ (the citation form) in several persons (*boró* [μπορώ] 1SG, *borís* [μπορείς] 2SG, *borí* [μπορεί] 3SG, *boríme* [μπορούμε] 1PL). This verb is combined with the subjunctive particle *na* (να), which shows up in two cases in Table App1:1. When the verb has an epistemic meaning, only the 3SG form *borí* is used (Tsangalidis 2009). Present-day Slavic languages have a modal verb related to Old Church Slavonic *mošti* to express Possibility (Besters-Dilger et al. 2009: 178). In Czech, the most frequent translations are inflected forms of Czech *moci* (*můž*e PRS.3SG, *mohou* PRS.3PL) and in Russian of *moč’* (*mogu* [могу] PRS.1SG, *možet* [может] PRS.3SG, *možeš’* [можешь] PRS.2SG, *možem* [можем] PRS.1PL, *mogut* [могут] PRS.3PL). Lithuanian is represented by present tense forms of the verb *galėti* inflected for person (*gali* 2/3SG, *galiu* 1SG, *galime* 1PL). In Welsh, CAN is expressed with the verbs *gallu* and *medru*, which are said to be interchangeable, but the choice is subject to regional variation. The verbs can refer both to ability/physical possibility as well as epistemic possibility, but the construction tends to shift (King 2016: 249–255).

In the list of Estonian translations, three different verbs are represented *võima* ‘can’ (*võid* PRS.2SG, *võib* PRS.3SG); *saama* ‘get; can’ (*saan* PRS.1SG and the form used in the present negative *ei saa*); *suutma* ‘be able’ (*suud* used in present negation: *ei suud*). In the translations of *could*, which are shown only for Estonian, forms of *võima* dominate (*võib* PRS.3SG, *võiks* COND.PRS.3SG, NEG: *ei võiks*), but there is also one form of *saama* (the conditional present combined with negation: *ei saaks*). *Saama* and *võima* are combined with one of the two infinitives, the *da*-infinitive. In Finnish, various inflected forms of *voida* ‘can, to be able’ dominate, but two other verbs also are represented. The most frequent translation in Hungarian is the verb *tud*, whose basic meaning is ‘know’, as in (1). There is no information about frequency of the modal use.

(1) Hungarian (OPUS Subtitles)

- a. *nem tudom, ki maga, barátom, de remélem,*
 not know.PRS.DEF.1SG who you friend.POSS.1SG but hope.PRS.DEF.1SG
tud főzni
 know.2SG.FORMAL cook.INF
 ‘I don’t **know** who you are, friend, but I hope you **can** cook.’ (Internal)

- b. *most* *Paul* *nem tud játszani*
 now Paul NEG know.PRS.3SG play.INF
 ‘Paul **cannot** play now.’ (External)

Possibility can also be expressed with the suffix *-hAt* and with *lehet* (<‘be’ + *-hAt*) ‘be possible’. The latter is a frequent translation of *can* according to Table App1:1 but not as frequent as *tud*.

MUST. Even though *MUST* has been used to symbolize the major modal verb that expresses Necessity, *must* is rivalled by alternative expressions in English. As can be observed in Table 8 in Section 4 that shows the frequency in the frequency dictionaries, English *must* is less frequent than its correspondences in the other languages. This is obviously connected to the ongoing development of alternative expressions of obligation. For that reason, the searches with OPUS Word Alignment Database that are shown in Table App1:2 are based on Swedish *måste*. Like English *must*, *måste* has the same form in present and past tense.

#	prob	cze		#	prob	ell		#	prob	eng	> >
30638	(0.17)	musí	✓✗	62568	(0.27)	πρέπει	✓✗	86450	(0.28)	have to	✓✗
15686	(0.09)	muset	✓✗			να	✓✗	73366	(0.25)	must	✓✗
				11548	(0.09)	πρέπει	✓✗	42300	(0.16)	gotta	✓✗
						να	✓✗	42164	(0.15)	need to	✓✗
				5526	(0.06)	πρέπει	✓✗	19740	(0.08)	got to	✓✗
				5314	(0.06)	πρέπει	✓✗				
#	prob	est		#	prob	fin		#	prob	fre	> >
35156	(0.21)	pean	✓✗	74056	(0.30)	täytyy	✓✗	75400	(0.27)	dois	✓✗
28158	(0.18)	peame	✓✗	61076	(0.22)	pitää	✓✗	48782	(0.18)	doit	✓✗
28540	(0.18)	pead	✓✗					28224	(0.11)	faut	✓✗
18480	(0.12)	peab	✓✗					14944	(0.07)	devez	✓✗
								12510	(0.06)	devons	✓✗
#	prob	ger		#	prob	hun		#	prob	lit	
106348	(0.35)	muss	✓✗	123714	(0.30)	kell	✓✗	3628	(0.11)	reikia	✓✗
77824	(0.28)	müssen	✓✗					3226	(0.10)	turi	✓✗
33374	(0.14)	musst	✓✗					1740	(0.06)	turiu	✓✗
								1304	(0.05)	turime	✓✗

(continued)

#	prob	rus	>>	#	prob	spa		Europ	prob	eng	>>
32972	(0.16)	должен	✓✗	15792	(0.11)	debe	✓✗	69536	(0.42)	must	✓✗
33644	(0.15)	нужно	✓✗	8106	(0.06)	tiene que	✓✗	15266	(0.10)	need	✓✗
26338	(0.14)	должны	✓✗	8096	(0.06)	tengo que	✓✗	12630	(0.10)	must be	✓✗
8992	(0.09)	надо	✓✗					10082	(0.08)	have to	✓✗
7442	(0.08)	должна	✓✗					8378	(0.07)	need to	✓✗

Table App1:2. The translations of Swedish *måste* in OPUS Word alignment (accessed 2026-07-06).

Starting with English this time, we can see that the most frequent translation in Subtitles is *have to*. *Must* is the second most frequent alternative closely followed by *got to*, and *gotta* and *need to* are also relatively frequent. For comparison, the program was also run on texts from Europarl (bottom of right column in Table App1:2). In this text type, *must* was clearly dominant and more frequent than the combined frequencies of *need (to)* and *have to*. These observations are in accordance with the thorough study by Aijmer (2019), which shows that *must* is the main proponent for epistemic modality but is losing ground to *have to* and other semi-modals in the area of deontic modality and this on-going change is related to a cultural change towards the use of less authoritative and face-threatening language that appears to have parallels in some other European languages.

German is represented by inflected forms of *müssen*, which is inflected for both tense and person. French *devoir* and Spanish *deber* have developed from Latin *debere* ‘must’, which originally meant ‘I owe to / I am indebted to’ (Cornillie et al. 2009). Table App 1:2 shows inflected forms of these two verbs, but French also uses the impersonal *faut* and in Spanish, expressions based on *tener* ‘have’ are slightly more frequent than forms of *deber*. However, in Europarl only inflected forms of *deber* appear. Greek uses the impersonal verb *prepi* (πρέπει), which like *boró* ‘can’ is combined with the subjunctive particle *na* (να) and a main verb inflected for person and aspect. The Czech translations of *måste* represent inflected forms of the verb *muset*, which is etymologically related to ‘must’ in several other Slavic languages (Polish *musieć*, Belorussian *music’*) and to *may* and its cognates in Germanic languages. The major translation in Russian is an adjective *dolžen* that takes an infinitive as complement. The adjective is inflected for gender (*dolžen* [должен] m., *dolžna* [должна] f., *dolžno* [должно] n., *dolžny* [должны] pl.). All but the neuter form appear in App1:2. Tense is indicated with the copula (Zero in present tense). A frequent alternative is the impersonal predicate *nužno* (нужно) and the less frequent *nado*

(надо). In Welsh (not included in App1:2), MUST is expressed with a noun *rhaid* ‘need, necessity’, see (2). *Rhaid* can also be used with epistemic meaning but then appears in a different construction (King 2016: 265–267)

(2) Welsh (King 2016: 265)

(mae) *rhaid i ti fynd*
 is need for you go
 ‘You must go.’

Lithuanian uses inflected forms of the verb *turėti* ‘have’ as the major expression and this holds also for Europarl. There is also an impersonal expression *reikia*, which is the third person present form of *reikėti* ‘require’ (< *reika* ‘necessity’) and is combined with a dative subject. Estonian is represented by inflected forms of *pidama*. This verb means both ‘must’ and ‘hold’ and has a similar inflection except for indicative past (see Raag 2014: 317). The Finnish cognate *pitää* ‘hold’ can also express obligation (see below). The only translation of *máste* in Hungarian is *kell* which is counted as a verb but lacks agreement markers. The subject is optionally marked with a possessive suffix on the infinitive as in *Mennem kell* ‘must go’ [go.INF.POSS.1SG + *kell*]. It can also be represented by a noun or pronoun in dative form.

SUMMARY. CAN is realized as one of the 3-6 most frequent verbs in the majority of European languages and its frequency expressed as LN typically reaches above 8.0, which is superseded only by BE and HAVE. MUST is also expressed as a frequent verb and its LN usually reaches above 7.0, which is the cut-off point for the 20 most frequent verbs, but the picture is more complex than for CAN. One reason for that is that constructions with HAVE (or in Finnish BE) are frequent alternatives to a modal verb in several languages especially in certain text types.

Complementing the analysis with concordance data

The analysis can be taken one step further with another tool than WAD: OPUS – Corpus query (CWB), which allows searches in parallel corpora of a word in one language as input and provides concordances of the word as key word in context (KWIC) and translations into several other languages. Searches can be done i.a. in Subtitles and Europarl. Only a few short examples will be given here. This kind of analysis has not been carried out systematically due to the complexity of the semantics of the modals, but the tool has been used as a complement to the literature that has

been consulted. The tool is useful to show the characteristic constructions of the modals, especially when a language has several alternative correspondences, for example of Swedish *måste*. Example (3) shows the major correspondences in the languages which are displayed. All combine the modal verb with a main verb in the infinitive. In Finnish, the subject is in the genitive.

- (3) SWE Du *måste* förbli vaken, min son! (source: OPUS Subtitles)
 ENG You *must* stay awake, my son!
 FRE Tu *dois* rester éveillé.
 SPA *Debes* permanecer despierto, hijo.
 FIN Sinun *täytyy* pysyä hereillä!
 you.SG.GEN must.PRS.3SG stay.1INF awake

Example (4) shows the impersonal construction *il faut* in French and the use of *tener* ‘have’ to express MUST in Spanish, and the Finnish example in (4) shows the use of *pitää* ‘hold’ to express MUST in Finnish. Also in this case, the subject is marked with genitive.

- (4) SWE Vi *måste* förflytta oss snabbt. (source: OPUS Subtitles)
 we must move REFL.1PL fast
 ENG We’ve *gotta* move fast.
 FRE Il *faut* agir vite.
 it must.PRS.3SG act fast
 SPA *Tenemos* *que* ser rápidos.
 have.PRS.1PL that be fast.PL
 FIN Meidän *pitää* iskeä heti.
 we.GEN hold.PRS.3SG strike.1INF fast

In Finnish, there is also a special construction with *olla* ‘be’ combined with a passive present participle as in (5), which does not appear in Table App1:2.

- (5) Finnish (OPUS Subtitles)
 sinun on nai-ta-va Aella
 you.SG.GEN be.PRS.3SG marry-PASS-PRS.PTCP Aella
 ‘You must marry Aella.’

In Europarl, this alternative dominates completely, see (6) for a typical example.

(6) Finnish (Europarl)

taistelu on käy-tä-vä pääosin kansainvälise-llä taso-lla
 battle is employ-PASS-PRS.PTCP mainly international-AD level-AD
 ‘This battle must, in the main, be fought at international level.’

The corpus tool (CWB) can also be used to study the functions with which the modals are used and to test the requirement that modals should express no less than two types of modality. Example (3) above shows how the major modals in the languages that are included are used to express deontic modality (obligation). The following example show how they can also express epistemic modality. In (7), modality is expressed by an adverbial in Finnish, which is a rather frequent option also in the other languages. (Actually, it is difficult to find examples where a modal is found as a translation in all languages.)

(7) [There is a knock.] That **must** be Mulder. (source: OPUS Subtitles)

SWE Det **måste** vara Mulder.
 FRE Ce **doit** être Mulder.
 SPA **Debe** ser Mulder.
 FIN Tuo on **varmaan** Mulder. [That is surely Mulder]

Example (8) shows how the modal *täytyä* can express an epistemic meaning, whereas in this example an adverbial is used in French.

(8) They **must** have spotted you. (source: OPUS Subtitles)

SWE De **måste** ha sett er! [They must have seen you]
 FRE Ils vous ont **sûrement** repéré. [They have surely detected you]
 SPA **Deben** haberte seguido. [They must have followed you]
 FIN Heidän on **täytynyt** nähdä sinut.

Abbreviations

1 = 1 st person	GEN = genitive	PRS = present
1 _{INF} = 1 st infinitive	INF = infinitive	PTCP = participle
2 = 2 nd person	NEG = negation	REFL = reflexive
3 = 3 rd person	PASS = passive	SG = singular
AD = adessive	PL = plural	
DEF = definite conjugation	POSS = possessive	

References

- Aijmer, Karin. 2019. What is happening to *must* in Present-day English? A contrastive perspective on a declining modal auxiliary. In Harumi Sawada, Yoshio Nitta & Masa-aki Yamanashi (eds.), *Bamen to shutaisei/shukansei* [Situation and subjectivity], 387–402. Tokyo: Hituzi Syobo.
- Besters-Dilger, Juliane, Ana Drobnjaković & Björn Hansen. 2009. Modals in the Slavonic languages. In Björn Hansen & Ferdinand de Haan (eds.), *Modals in the languages of Europe*, 167–197. Berlin: Mouton de Gruyter.
- Cornillie, Bert, Walter De Mulder, Tine Van Hecke & Dieter Vermandere. 2009. In Björn Hansen & Ferdinand de Haan (eds.), *Modals in the languages of Europe*, 107–137. Berlin: Mouton de Gruyter.
- King, Gareth. 2016. *Modern Welsh. A comprehensive grammar*. London: Routledge.
- Kronning, Hans. 2024. Les verbes modaux *devoir* et *pouvoir*. In *Encyclopédie Grammaticale du Français, en ligne*. Available online at: <http://encyclogram.fr> (Accessed 2026-07-06).
- Raag, Raimo. 2014. *Estniska från grunden* [Finsk-ugriska läromedel FU 25]. Uppsala: Uppsala University.
- Tiedemann, Jörg. 2012. Parallel data, tools and interfaces in OPUS. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 2214–2218. Istanbul: European Language Resources Association (ELRA).
- Tsangalidis, Anastasios. 2009. Modals in Greek. In Björn Hansen & Ferdinand de Haan (eds.), *Modals in the languages of Europe*, 139–163. Berlin: Mouton de Gruyter.

Electronic sources

OPUS. <https://opus.nlpl.eu/> (Accessed 2026-07-06).

CONTACT

ake.viberg@lingfil.uu.se