

A corpus-based learning list of irregular verbs in English

Eva Grabowski and Dieter Mindt
Freie Universität Berlin

Abstract: For learners of English, irregular verbs represent one of the most difficult aspects of the language. They are normally presented in the form of alphabetical lists. Alphabetical lists do not, however, take into account the actual occurrence of these verbs. The article presents a new corpus-based learning list, in which the verbs are ranked in their order of frequency in authentic English. The new list ensures that the learner has encountered the most important verbs, no matter when the learning process ends. For the author of teaching materials and the teacher, the new list supplies an empirical basis for the selection and gradation of irregular verbs in language courses.¹

1. Problem and goal

Irregular verbs haunt learners of English from the beginning to the end of their studies. Full mastery of irregular verbs seems to be very rarely achieved.

There are very few verbless sentences in English and irregular verbs belong to the core of English. In all types of texts, forms of irregular verbs outnumber those of regular verbs.

The learning of irregular verbs is normally based on the well-known alphabetical lists which range from *ABIDE* to *WRITE* (Quirk *et al.* 1985: 115ff.), from *ARISE* to *WRITE* (Collins *COBUILD English Grammar* 1990: 451) or from *BE* to *WRITE* (Ungerer *et al.* 1989: Iff.). All these lists differ in the number of verbs.

Unfortunately, the grammars make no mention of the basis on which these lists were compiled. The lists are probably based on a comparison of previous lists and the experience and/or intuition of the author(s) of a grammar.

Alphabetical lists of irregular verbs are appropriate for looking up

verb forms. The order of verbs in alphabetical lists is, however, of no avail in the learning of these verbs, because their order does not reflect their importance in language use.

This article aims at establishing a new learning list of irregular verbs, a rank list in which the verbs are ordered according to their frequency in English. Learners who use this list will be able to learn the most important verbs first and the least important ones last. If learning is discontinued, the learner can be sure to have encountered the most important verbs irrespective of when the learning process is broken off.

For authors of textbooks or other teaching materials the list presents new empirical foundations for decisions on the selection and gradation of irregular verbs in a course.

Two machine-readable standard corpora of English, the BROWN Corpus of American English and the LOB Corpus of British English (cf. Aijmer and Altenberg 1991: 315-318), were selected to form the basis of this learning list. For both corpora the tagged versions were used, where every word is accompanied by a tag which indicates its word class. The study was performed with UNIX tools, especially AWK (Aho, Kernighan and Weinberger 1988). The results were checked against a more recent corpus: about 7.3 million words of the *Longman/Lancaster English Language Corpus* (Longman Dictionaries Division), henceforth LOLAC.

2. Data collection

The first step was to retrieve all verb forms in BROWN and LOB and to lemmatize them. The figures in the following tables are total numbers for the verbs and verb forms in BROWN and LOB. Lemmatization was accompanied by a number of standardizations.

Standardizations

- a) of spelling variants:
e.g. *re-writes* was lemmatized to *REWRITE*; *built-up* to *BUILD*, *running-on* to *RUN*.
- b) of compounds and derivations:
Compounds or derivations were assigned to their base form when there is no change in the basic meaning of the compound or derivation, e.g.:

<i>AWAKE</i>	>	<i>WAKE</i>	<i>OUTSPREAD</i>	>	<i>SPREAD</i>
<i>BEFALL</i>	>	<i>FALL</i>	<i>OVERWRITE</i>	>	<i>WRITE</i>
<i>FORBEAR</i>	>	<i>BEAR</i>	<i>PRECUT</i>	>	<i>CUT</i>
<i>INLAY</i>	>	<i>LAY</i>	<i>RETELL</i>	>	<i>TELL</i>
<i>MISREAD</i>	>	<i>READ</i>	<i>UNDERFEED</i>	>	<i>FEED</i>
<i>OFFSET</i>	>	<i>SET</i>	<i>UNFREEZE</i>	>	<i>FREEZE</i>
<i>housebroken</i>	>	<i>BREAK</i>			
<i>half-understood</i>	>	<i>UNDERSTAND</i>			
<i>floodlit</i>	>	<i>LIGHT</i>			

Defective verbs

Three defective verbs (verbs which do not display the full set of forms) were found: *bereft*; *clad*, *cladding*; *wrought*. They were treated as individual lexical forms and were not assigned to the learning list of irregular verbs.

These procedures resulted in a list of 179 verbs with irregular forms.

Verbs with irregular and regular forms

This list of 179 irregular verbs contains both verbs with irregular forms only and verbs with irregular as well as regular forms (*learnt/learned*). Verbs whose regular forms occur more frequently than their irregular forms were not included in the learning list, provided that the irregular forms can be recognized as variants of the regular ones both in their spoken and their written forms, e.g. *LEARN* (cf. Table 1).

Table 1: Verb forms of *LEARN*

	Absolute frequency	Past tense form	Absolute frequency	Past participle form
Regular forms	80	<i>learned</i>	95	<i>learned</i>
Irregular forms	22	<i>learnt</i>	17	<i>learnt</i>

There are 20 verbs with more regular than irregular forms (cf. Table 2). The figures for past tense and past participle forms are summarized in Table 2.

Table 2: Verbs with more regular than irregular forms

Absolute frequency	Regular forms	Absolute frequency	Irregular forms
16	<i>blessed</i>	3	<i>blest</i>
54	<i>burned</i>	14	<i>burnt</i>
2	<i>cleaved</i>	1	<i>cleft</i>
34	<i>dreamed</i>	2	<i>dreamt</i>
64	<i>fitted</i>	12	<i>fit</i>
4	<i>forfeited</i>	1	<i>forfeit</i>
14	<i>heaved</i>	2	<i>hove</i>
5	<i>hoisted</i>	1	<i>hoist</i>
9	<i>knitted</i>	6	<i>knit</i>
71	<i>leaned</i>	8	<i>leant</i>
29	<i>leaped</i>	7	<i>leapt</i>
175	<i>learned</i>	39	<i>learnt</i>
72	<i>proved</i>	14	<i>proven</i>
6	<i>shaved</i>	2	<i>shaven</i>
19	<i>smelled</i>	7	<i>smelt</i>
7	<i>sneaked</i>	1	<i>snuck</i>
4	<i>spilled</i>	2	<i>spilt</i>
9	<i>spoiled</i>	3	<i>spoilt</i>
112	<i>stayed</i>	1	<i>staid</i>
8	<i>wedded</i>	2	<i>wed</i>

For *SPELL* there are 7 instances of *spelled* and 8 instances of *spelt*. The figures for LOLAC are: 188 instances of *spelled* vs 13 instances of *spelt*. The verb *SPELL* was therefore excluded from the learning list.

In some cases it is not so easy to recognize the irregular forms as variants of the regular ones: *hove* < *HEAVE*, *snuck* < *SNEAK*. These forms are so infrequent (between 1 and 2 cases in BROWN and LOB altogether; LOLAC: *hove* 1, *heaved* 36; *snuck* 4, *sneaked* 11) that most learners will very rarely come across them. They were also excluded from the learning list.

The forms *dreamt* < *DREAM* and *leapt* < *LEAP* can easily be recognized as variants of the regular forms in writing, but less easily in speaking. In both cases the regular forms far outnumber the irregular forms (LOLAC: *dreamed* 60, *dreamt* 14 ; *leaped* 188, *leapt* 42). The verbs *DREAM* and *LEAP* were also excluded from the learning list.

Some grammars (e.g. Quirk *et al.* 1985: 115ff.) do not distinguish the verbs *CLEAVE* ('to split or cause to split') and *CLEAVE* ('to cling or adhere'), which have different etymologies. Historically and synchronically *CLEAVE* ('to cling or adhere') is a regular verb (*cleave, cleaved*). For the irregular verb *CLEAVE* ('to split or cause to split') there are 4 instances (*cleaving, cleft*) in more than 9 million words, mostly in archaic contexts (e.g.: *Rock of ages, cleft for me, let me hide myself in thee*). Because of its archaic character, we decided not to include *CLEAVE* ('to split or cause to split') in the learning list.

For those who are not ready to follow these decisions, the following changes would have to be made to the learning list: *DREAM* would have to be added after position 93, *LEAP* after position 105, *HEAVE* after 133, *SNEAK* after 144 and *CLEAVE* after 153. For *LEAP*, *DREAM*, *HEAVE*, and *SNEAK* it would be appropriate to mention that the regular forms prevail.²

For all other verbs with both irregular and regular forms, the number of regular forms is lower than or equal to the number of irregular forms (cf. Tables 3 and 4).

Table 3: Verbs with more irregular than regular forms

Lemma	Irregular: regular forms	Past tense		Past participle	
		irregular	regular	irregular	regular
<i>BID</i>	7:1	2 <i>bade</i> 1 <i>bid</i>	1 <i>bided</i>	3 <i>bid</i> 1 <i>bidden</i>	
<i>COST</i>	31:1	20 <i>cost</i>		11 <i>cost</i>	1 <i>costed</i>
<i>CUT</i>	193:1	46 <i>cut</i>	1 <i>upper-</i> <i>cuted</i>	146 <i>cut</i> 1 <i>precut</i>	
<i>KNEEL</i>	17:2	16 <i>knelt</i>	2 <i>kneeled</i>	1 <i>knelt</i>	
<i>KNOW</i>	1296:2	738 <i>knew</i>	1 <i>know'd</i>	558 <i>known</i>	1 <i>knowed</i>
<i>LIGHT</i>	50:38	31 <i>lit</i>	15 <i>lighted</i>	19 <i>lit</i>	23 <i>lighted</i>
<i>QUIT</i>	4:2	3 <i>quit</i>		1 <i>quit</i>	2 <i>quitted</i>
<i>RING</i>	43:2	39 <i>rang</i>		4 <i>rung</i>	2 <i>ringed</i>
<i>SAY</i>	4023:1	3561 <i>said</i>	1 <i>sayed</i>	462 <i>said</i>	
<i>SEW</i>	4:1			4 <i>sewn</i>	1 <i>sewed</i>
<i>SHOW</i>	396:5			396 <i>shown</i>	5 <i>showed</i>
<i>SPEED</i>	12:5	12 <i>sped</i>	3 <i>speeded</i>		2 <i>speeded</i>
<i>STRING</i>	9:1	1 <i>strung</i>		8 <i>strung</i>	1 <i>stringed</i>
<i>STRIVE</i>	5:2	5 <i>strove</i>	2 <i>strived</i>		
<i>THROW</i>	190:2	100 <i>threw</i>	2 <i>throwed</i>	90 <i>thrown</i>	
<i>UNDER-</i> <i>STAND</i>	130:1	43 <i>under-</i> <i>stood</i>		87 <i>under-</i> <i>stood</i>	1 <i>under-</i> <i>standed</i>
<i>WAKE</i>	49:2	44 <i>woke</i>	1 <i>waked</i>	5 <i>woken</i>	1 <i>waked</i>
<i>WET</i>	5:2	3 <i>wet</i>	1 <i>wetted</i>	2 <i>wet</i>	1 <i>wetted</i>
<i>WIND</i>	21:2	14 <i>wound</i>		7 <i>wound</i>	2 <i>winded</i>

Table 4: Verbs with equal numbers of irregular and regular forms

Lemma	Past tense	Past participle form
<i>HEW</i>	–	1 <i>hewn</i> 1 <i>hewed</i>
<i>SAW</i>	–	1 <i>sawn</i> 1 <i>sawed</i>

The verbs of Table 3 were included in the learning list, as well as those of Table 4, because in LOLAC the irregular forms are more frequent than the regular forms for *HEW* and *SAW*.

For the verbs *BESEECH*, *BET*, *DWELL*, *RID*, *SHEAR*, *SHINE*, *SHOE*, *SOW*, *SWELL*, which are sometimes claimed to have both irregular and regular forms, we found irregular forms only. These verbs were added to the learning list.

There are grammars which list the following irregular verbs: *ABIDE*, *BUST*, *CHIDE*, *DIVE*, *MOW*, *PLEAD*, *SHRED*, *SWEAT*, *THRIVE*. For these verbs no irregular forms were found. Consequently, they were not included in the learning list.

After these decisions the learning list consists of 158 irregular verbs.

The verbs *SLAY* and *SLINK* are found in most lists of irregular verbs but there are no occurrences of past forms or past participle forms in either *BROWN* or *LOB*. In LOLAC, however, there are 16 instances of *slew* and 22 instances of *slain* as well as 22 instances of *slunk* (representing both the past tense and the past participle form) and one instance of *slinked* (past tense). The two verbs were accordingly included in the learning list.

3. The learning list and its interpretation

The final list of irregular verbs contains 160 verbs. These 160 verbs represent 192,868 verb forms (tokens). Table 5 gives a comparison of regular and irregular verbs and verb forms.

Table 5: Absolute and relative frequency of regular and irregular verbs³

Kind of verbs	Number of different verbs (<i>types</i>)	Percentage of all verbs	Absolute frequency of verb forms (<i>tokens</i>)	Percentage of all verb forms
regular verbs	4,240	96.4	141,403	42.3
irregular verbs	160	3.6	192,868	57.7

Altogether 4,400 different verbs were found, with 160 (3.6%) irregular verbs. Of all verb forms, 192,868 forms are instances of irregular verbs. This is to say that 160 irregular verbs account for about 58% of all verb forms. Regular verbs make up about 42% of all verb forms. This percentage represents 4,240 different verbs. Mastery of irregular verbs is thus an essential aspect of the learning of English.

The final question is whether there are irregular verbs whose exceptional status in English does not make it necessary to include them in the learning list. This is true of the verbs *BE*, *HAVE*, and *DO*. These three verbs do not only occur as main verbs but are very frequently used as components of important grammatical constructions such as the progressive and the passive (I *am* going – he *was* greeted), the perfect and past perfect (she *has* left – they *had* sat), question, negation, emphasis (*do* you know? – he *did* not wait – *do* come in). Moreover, the verbs *BE*, *HAVE*, and *DO* are the most frequent verbs in English. The full range of their forms has to be mastered very early in order to understand and to produce some of the most elementary structures of English. For these reasons these three verbs do not belong to the final learning list of irregular verbs.

The final lists of irregular verbs are displayed in Table 6. The learning list (highlighted, on the left) contains 157 verbs. The list of all irregular verbs including *BE*, *HAVE*, and *DO* is on the right and contains 160 verbs.

The columns in Table 6 give the following information:

- Column 1:* Rank position in learning list.
- Column 2:* Relative frequency (percentage) of learning list.
- Column 3:* Cumulative frequency (percentage) of learning list. For every position the column indicates the percentage of verb forms that has been covered if the learner started at the beginning and has proceeded in the order of the learning list. If learning is discontinued after position 5 (*COME*) the learner has learned 27.3% of all irregular forms with five verbs (*SAY, MAKE, GO, TAKE, COME*). If learning is discontinued after *LEAVE* the first 15 verbs have covered 56% of all irregular forms.
- Column 4:* Lemma/Base form.
- Column 5:* Past tense form.⁴
- Column 6:* Past participle form.
- Column 7:* Absolute frequency in BROWN and LOB.
- Column 8:* Relative frequency (percentage) of all irregular verbs.
- Column 9:* Cumulative frequency (percentage) of all irregular verbs.
- Column 10:* Rank position in list of all irregular verbs.

4. Conclusion

The new learning list of irregular verbs presents the verbs in an order that reflects their importance in English. The list ensures that the most important verbs are learned first and that there is always the maximum yield for the student's learning effort, irrespective of when the learning process is broken off.

For teachers and authors of textbooks and other teaching materials the learning list presents an empirical basis for decisions concerning the selection and gradation of irregular verbs in foreign language courses.

For this important area of English, the new learning list can help to achieve learning objectives faster and more effectively than the traditional alphabetical lists.

Notes

1. This article is an abridged version of Grabowski and Mindt (1994). Errors in the preceding article have been corrected (cf. Note 3).

2. The 7 cases of *leapt* constitute 0.009 % of all irregular forms, which makes it clear that we are dealing with a marginal problem here.
3. The figures presented in this table differ slightly from those given in Table 5 in Grabowski and Mindt (1994: 344). After all errors had been corrected we unfortunately failed to make the necessary changes in Table 5 in Grabowski and Mindt (1994). Thus Table 5 is presented in revised form here.
4. Columns 5 and 6 (past tense form and past participle form) give only the most frequent form. Less frequent variants are not mentioned in the learning list.

References

- Aho, Alfred V., Brian W. Kernighan and Peter J. Weinberger. 1988. *The AWK programming language*. Reading, Mass.: Addison-Wesley.
- Aijmer, Karin and Bengt Altenberg (eds.). 1991. *English corpus linguistics*. London: Longman.
- Collins COBUILD English grammar*. 1990. London & Glasgow: Collins.
- Grabowski, Eva and Dieter Mindt. 1994. Die unregelmäßigen Verben des Englischen: eine Lernliste auf empirischer Grundlage. *Die Neueren Sprachen* 93/4:334-353.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Ungerer, Friedrich, Peter Pasch, Peter Lampater and Rosemary Hellyer-Jones. 1989. *Learning English Grundgrammatik*. Ausgabe für Gymnasien. Stuttgart: Klett.