1 Introduction

> But it was in the verbal conjugation that the Ablaut found its peculiar home, and there it took formal and methodical possession. (Earle 1892: §124)

1.1 The past tense – a descriptive approach

PAST is the most frequently marked verbal category by far (e.g. according to Sampson 2002, based on figures from the British National Corpus), accounting for around 25 per cent of all verb forms in contemporary spoken British English. In comparison, the two next categories, negation or modals, both only account for roughly 12 per cent of verb forms, the perfect for around 8 per cent, and the progressive for under 6 per cent. The passive finally is at best marginal with a text probability of under 1 per cent.

Past tense formation in English appears to be a very simple matter. Nevertheless – or perhaps because of this simplicity – great theoretical significance has been attached to an analysis of the past tense because it is used as the prime example in a long-standing debate in morphological theory (more on which in Chapter 2).

Putting it in simple descriptive terms (although no description is of course theory-free, or truly pre-theoretical), the majority of English verbs today have past tense forms that consist of the present tense stem plus <-ed>.[1] <-ed>, the weak past tense marker, is exactly parallel to the weak past tense in all other Germanic languages and is indeed one of the characterizing features of Germanic. English here is no exception. There are several theories, each deficient in its own terms, of how this common dental suffix evolved with the specific past tense meaning – among them the '*tun* theory' and the '*-tó-* theory'[2] – but a consensus cannot as yet be presented. Although it is probably generally true that, from an Indo-European perspective, the

[1] In contrast to most reference grammars, I disregard variation in spelling here, although I will refer to the graphemes for simplicity's sake.
[2] For a short overview, see West (2001: 53).

weak verbs are the more recent innovation,[3] inside the weak verb class there appear to be different layers: some weak verbs are very old and can be traced to Indo-European roots (and thus constitute rather untypical weak verbs), whereas the majority are probably younger.[4]

Today, for the weak past tense forms, in English we have three regular allomorphs: /əd/ or /ɪd/ after the two alveolar stops /t/ and /d/, /t/ after all other voiceless sounds, and /d/ after all other voiced sounds. This case of phonologically determined allomorphy is perfectly regular and equally productive. The rarer verbs in particular, as well as neologisms and loan words, are weak today. The number of paradigms of weak verbs is very large (because of possible new coinages probably infinite), so that a high type frequency is here coupled with a low token frequency.

A small number of verbs in contemporary (standard) English – Quirk et al. list '250 or so' (Quirk et al. 1985: 104), Huddleston and Pullum have exactly 176 (Huddleston and Pullum 2002: 1608–9), although other linguists name considerably fewer – are irregular and have retained strong past tense forms. This group has been gradually decreasing in number, as strong verbs have changed verb classes and become weak verbs since Old English times (see in particular Krygier 1994 for a detailed analysis through the centuries until Early Modern English). Nevertheless, strong verb forms are still highly visible in present-day English because the frequent verbs in particular have retained their strong forms. Indeed, some text counts put the figure for strong verbs in running text as high as 70 to 75 per cent.[5] For strong verbs, then, low type frequency is coupled with a very high token frequency.[6]

Incidentally, Quirk et al.'s classification seems to be the most inclusive. For them, all verbs that are not regular are irregular. While regular verbs can be defined positively, irregular verbs simply constitute 'the rest' (a rather heterogeneous category that will be discussed further below). Perhaps for this reason, the terms *strong verb* and *weak verb* do not appear in Quirk et al. (1985). Other authors, especially those arguing from a historical point of view, are more discriminatory. Stockwell and Minkova, for example, quoting

[3] As opposed to the strong verbs, which can be shown to re-use the old aorist; for a recent treatment in terms of exaptation, see Lass (1990).

[4] The Newcastle Weak Verb Project aims to shed light on this layering (see West 2001). First studies for Old High German suggest that about 70 per cent of weak verbs are neologisms, 18 per cent are West Germanic, 10 per cent are Germanic and around 2 per cent could be pre-Germanic (West 2001: 54). Figures for Old English were not available at the time of writing.

[5] E.g. in transcripts of parental speech, see Pinker (1999: 227). Based on Sampson's CHRISTINE corpus, a subcorpus of the British National Corpus (BNC), Dahl (2004: 300–1) quotes even more striking figures. Of all verb forms, regular verbs only make up around 9 per cent of all tokens. If one disregards *be*, *have* as well as modals, regular verbs still make up only around 24 per cent of all lexical verb tokens, figures very similar to Pinker's.

[6] This is an oversimplification. In fact, some of the very frequent verbs are weak (*look*, *ask*, *seem*, *want*, *turn* …), while many strong verbs have a very low token frequency. As a statistical trend, however, this statement holds.

Baugh and Cable (1978), only mention sixty-eight strong verbs (Stockwell and Minkova 2001: 130), i.e. those that form the past tense by vowel gradation, going back to similar processes in Indo-European, plus thirteen that are both strong and weak today; Carstairs-McCarthy occupies some middle ground in claiming that 'in all, 150 or so verbs are irregular in that they do not use the *-ed* suffix' (Carstairs-McCarthy 2002: 40), without, however, supplying a list.

As Quirk et al. do provide a comprehensive list of all strong verbs and their various forms, this will constitute the point of departure for my study, the foil against which any non-standard forms will be compared. However, from their list of 250 verbs I excluded 83 which were either morphologically complex (e.g. *deepfreeze*; the simplex *freeze* is included) or behaved as if they were (e.g. *become*, cf. *come*).[7] These were mostly verbs with the prefixes *a-*, *be-*, *for(e)-*, *mis-*, *out-*, *over-*, *re-*, *un-*, *under-*, *up-* and *with-*. Clearly in most cases the prefixes are not semantically transparent today, and many verbs are thus arguably monomorphemic. For our purposes it is important to note, however, that they behave morphologically *as if* they were derivational forms. To avoid skewing due to frequent prefixation of some bases in the later quantitative comparisons, these seemingly derivational forms were excluded. Incidentally, these exclusions bring Quirk et al.'s list very close to the figure '150 or so' mentioned by Carstairs-McCarthy above, namely to a total of 167.[8] Quirk et al.'s complete list of strong verbs with all exclusions can be found in Appendix 1.

1.2 Terminology: strong–weak vs. irregular–regular

A brief note on terminology: in this book, I will use the terms *strong* and *weak verbs* for the verbs that in more modern terminology (see Quirk et al. 1985; Huddleston and Pullum 2002) are usually called *irregular* and *regular*. In particular, *strong verb* will be used as a cover term not only for verbs that display the characteristic Indo-European vowel gradation, but for any other irregular verb as well. The reason for this choice is twofold. Firstly, we will have to have recourse to the concept of *regularization*, an abstract cognitive

[7] Huddleston and Pullum also stress that 'verbs with complex bases' have 'irregular forms matching those of the simple verb in final position' (2002: 1609), pointing out that 'the inflectional-morphological relationship is thus maintained long after the semantic connection has been lost' (2002: 1610). Aronoff goes further and in fact takes 'the inheritance of irregular morphology from a root or morphological head, even in the absence of compositionality' as proof for a level of analysis 'between morphosyntax and morphophonology', i.e. as morphology in the narrow sense, claiming that 'in each case, the set of irregular forms is obviously not a single lexeme … so their unity must be expressed at a purely morphological level' (Aronoff 1994: 28).

[8] Huddleston and Pullum's list is slightly longer with a total of 176 verbs (2002: 1608–9); in contrast to Quirk et al., they include *bid* twice, and add *bust*, *earn*, *fit*, *gird*, *sneak* and *thrive*, as well as the four modals *can*, *may*, *shall* and *will*. On the other hand, their list does not include *knit*, *shit* or *sweat*.

process that can apply at a number of different linguistic levels. As regularization is not necessarily confined to the process of turning irregular verbs into regular verbs, to avoid utter confusion the terms *irregular* and *regular verbs* will not be used in this book after this introduction. If the following sections and chapters mention *strong verbs*, then, it should be borne in mind that this does not only include strong verbs in the narrower sense, i.e. those verb paradigms displaying Indo-European vowel gradation, but also verbs that Stockwell and Minkova call strong and weak, i.e. any verbs that are not weak verbs.[9]

Secondly, the term *regular* (at least in some frameworks) might presuppose, based on perhaps overzealous etymologizing, that a *rule* (Latin *regula*) is involved in the production of this form. This is a presupposition that I will be trying to avoid. In particular, in Chapter 5 and throughout the book I will be arguing that there can be both *weak* ('regular') verbs that are not created through a rule, and, more importantly, *strong* verbs ('irregular' verbs) that nevertheless follow a rule, or pattern, in their formation.

Finally, the data employed here are mainly historical as well as dialectal. While in historical studies it is of course still the case that the terms *strong* and *weak* verbs are used, the situation in dialectology is a little different. Again, works with a strong historical focus tend to avoid the terms *regular* and *irregular* and use *strong* and *weak* instead (despite the title, for example, Cheshire uses 'strong' and 'weak' in her analysis of the English irregular verbs: see Cheshire 1994; see also Miller 2003: 74). When I chart the progress of individual verb forms through history to their dialectal status today, it will be particularly useful to be able to use the same terms, rather than switch from *strong – weak* to *regular – irregular* at some arbitrary point in time (e.g. the change from Middle English to Modern English; the change from historical linguistics to synchronic linguistics; the change from dialectology to sociolinguistics; and what would be the respective dates for these important changes?).

Nevertheless, I am aware of several complications in this choice of terminology. Words that were weak in Old English (like *teach*) would have to be treated as having 'jumped' to the strong verb class, whereas what 'really' happened was of course a series of sound changes that resulted in opacity and, indeed, irregularity for this form.[10] Clearly *taught* is not perceived

[9] Cf. McMahon's terminology, which is similar: 'The Modern English strong verbs … will be defined for present purposes as all those verbs which do not simply add a dental suffix {D} … to mark the past tense, but also, or instead, change the quality of the stem vowel … The term "strong" therefore designates not only historically strong verbs, but also historically weak verbs which now exhibit a vowel mutation in the past tense' (McMahon 2000: 129). In her analysis, it is not clear whether she really wants to exclude paradigms like *hit – hit – hit*.

[10] The Germanic spirant law (or Primärberührung) resulted in the spirantization of /k/ > /x/ before the alveolar in the past tense, but not the present, whereas the vowel change is due to 'reverse vowel gradation' (usually known by its German name of Rückumlaut).

today any longer as containing the regular weak ending <-ed>, and so it would be misleading (for a synchronic analysis) to classify *teach* as a 'weak' verb today.[11] On the other hand, a modern interpretation of 'strong verb' as identical to 'irregular' verb stresses for example the vowel change that takes place between *teach* and *taught*. Although of course it does not go back to an Indo-European ablaut schema, and indeed should not be called ablaut, vowel change between present and past tense stems is still one of the most frequent characteristics in the group of strong verbs (although not all of them, as we shall see in section 3.3).

1.3 Classification of strong verbs

1.3.1 Ablaut series, vowel gradation

Among the strong verbs, several classifications have been proposed. Classically, divisions are historical in nature, but among Germanic scholars it seems widely accepted that 'the English strong verbs are probably the most difficult of any modern West Germanic language to classify in any systematic way' (Durrell 2001: 13), no doubt because English has moved furthest away from its typological relatives German or Dutch in many respects. Typically, for example, verbs are grouped together by the same vowel changes they contain, according to present-day English, Old English, West Germanic, or indeed Indo-European ablaut series (e.g. /ɪ/~/æ/~/ʌ/ *sing – sang – sung; begin – began – begun* vs. /e/~/ɔ/ *bear – bore – borne; tear – tore – torn*, etc.). For present-day English, but clearly based on Old English schemas, this classification typically yields seven verb classes (e.g. Katamba 1993: 102):

Class I:	/aɪ/ /əʊ/ /ɪ/	rise	rose	risen
Class II:	/iː/ /əʊ/ /əʊ/	freeze	froze	frozen
Class III:	/ɪ/ /æ/ /ʌ/	shrink	shrank	shrunk
Class IV:	/eə/ /ɔə/ /ɔː/	bear	bore	borne
Class V:	/ɪ/ /eɪ/ /ɪ/	give	gave	given
Class VI:	/əʊ/ /uː/ /əʊ/	know	knew	known
Class VII:	/æ/ /ʊ/ /ʊ/	stand	stood	stood

The problem with this classification according to ablaut series is that it accounts for only a minority of strong verbs today, even though it is specifically written for present-day English, not historical stages of the language. Katamba's classification, for example, can only include 49 strong verbs – that is less than 30 per cent of the 167 strong verbs today. It neglects many vowel series of verbs that were strong in Old English and have remained so until

[11] Some synchronic descriptions resort to classifying these verbs as 'partial suppletion'; see Aronoff and Fudeman (2005: 168–9), as almost the complete stem /tiːtʃ/ is 'replaced' in /tɔːt/ (with the exception of the initial consonant).

today (e.g. *choose – chose – chosen* or *take – took – taken* or *break – broke – broken*). A classification according to vowel series in general also cannot account for paradigms that have three identical forms, because here the vowel can be quite different from verb to verb (e.g. *cast – cast – cast* vs. *hit – hit – hit* vs. *put – put – put* vs. *cost – cost – cost* vs. *shed – shed – shed*) – nevertheless it would be desirable to capture their intuitive similarity by classifying them together in one class.

Other verbs used to be weak in Old English times, but today have become irregular through devoicing (e.g. *spill – spillt – spillt* or *bend – bent – bent*). A second class of weak Old English verbs (particularly the weak class III verbs) today are still differentiated by their consonants, while the vowel has remained the same (e.g. *make – made – made* or *have – had – had*). A third group of weak Old English verbs have become strong through the regular process of Middle English open syllable lengthening (MEOSL), so that the Great Vowel Shift operated on different forms of the same paradigm differently. These regular phonological processes have resulted in markedly irregular paradigms with vowel changes as well as sometimes an added suffix (e.g. *mean – meant – meant* or *bite – bit – bitten*) that should be included in a present-day classification like Katamba's above. Finally, some verbs that were weak in Old English have undergone both vowel and consonant changes such as the Germanic spirant law, turning /g/ or /k/ into /x/ before the past tense alveolar stop (but not in the present tense), and deleting any preceding nasal; vowel changes even in Old English were due to *Rückumlaut* (or 'reverse vowel gradation'); with the subsequent deletion of /x/ in the majority of verbs this again results in present-day verb paradigms with a clear vowel change, but with very different present tense forms (e.g. *buy – bought – bought* with present tense /aɪ/; *teach – taught – taught* with present tense /iː/; or *catch – caught – caught* with present tense /æ/). As we have seen above, synchronically these verbs are today classified by many as 'partial suppletion' (see Aronoff and Fudeman 2005: 168–9) because their past tense forms are so radically different from their bases. Clearly, the intuitive similarity between these past tense forms is poorly accounted for in the form of vowel series.

Finally, it is no great help to start from the seven Old English strong verb classes either, as some verbs switched verb classes, many became weak, a large number simply fell into disuse, and of course some verbs entered the system after Old English times (for Old English verb classes, see Cassidy and Ringler 1971; and of course Krygier 1994. For sound change, see Campbell 1959).

1.3.2 Dental suffix

A second classificatory criterion generally applied is the presence or absence of a (dental) suffix in the past tense and (or) the past participle – the

advantage is that this criterion can also be applied to all weak verbs, in addition to those strong verbs (former weak verbs) like *dream – dreamt – dreamt* which do have a suffix; some authors also include a nasal suffix here and would therefore classify *shake – shook – shaken* as belonging to this special group. As the examples already show, this criterion cuts across the first one of ablaut series, as dental or nasal suffixation may go hand in hand with vowel alternation (but need not do so). Clearly, however, this criterion on its own does little to structure the group of strong verbs, as almost half of them – around 47 per cent – have either a dental or a nasal suffix; if employed, this criterion probably always has to be combined with other criteria to result in a workable classification.

1.3.3 Abstract formal identity

A third, more interesting criterion characterizing verb paradigms today is the formal identity or non-identity of forms, and this is the one that will be chiefly applied in this book. Quirk et al. for example – if only in passing – distinguish five patterns of paradigms[12] (Quirk et al. 1985: 103), as do Nielsen (1985) and Hansen and Nielsen (1986: 181) in some more detail: (a) all forms are the same (e.g. *cut – cut – cut*); (b) only past tense and past participle are identical (e.g. *meet – met – met*); (c) infinitive and past tense are identical (e.g. *beat – beat – beaten*); (d) infinitive and past participle are identical (e.g. *come – came – come*); and (e) all three forms are different (e.g. *speak – spoke – spoken*) (Quirk et al. 1985: 103).[13] These are the five patterns that are logically possible, and as the examples already show, all five (one one-form pattern, three two-form patterns, and one three-form pattern) are actually attested in English.

Diagrammatically, the five logically possible patterns are displayed in Figure 1.1.

Quirk et al. list these possibilities without further qualification (Quirk et al. 1985: 103). It has to be stressed, however, that these five possibilities are by no means equivalent functionally (and they are also not equally distributed

[12] I use the term paradigm to refer to what has traditionally been known as the principal parts of the verb, i.e. present tense stem – past tense stem – past participle.

[13] Quirk et al. go on to use a mixture of all three criteria (presence/absence of suffix, identity/non-identity only of past tense and past participle, and vowel identity across all three forms). This mixture results in a very detailed classification, nevertheless again with seven main classes and many subclasses. They do not, however, justify their use of only employing identity of past and past participle as a criterion. Huddleston and Pullum in contrast use four criteria: (1) secondary *–ed* formation, (2) vowel alternation, (3) participle <-en>, (4) 'other formations', where these four are not mutually exclusive (Huddleston and Pullum 2002: 1600–8). In their Student Grammar, by contrast, they have reduced irregular verb classes to just two: those where simple past and past participle are identical (with eight subtypes, including a 'miscellaneous' class), and those where simple past and past participle are not identical (with six subtypes, again including a 'miscellaneous' one) (Huddleston and Pullum 2005: 274–7).

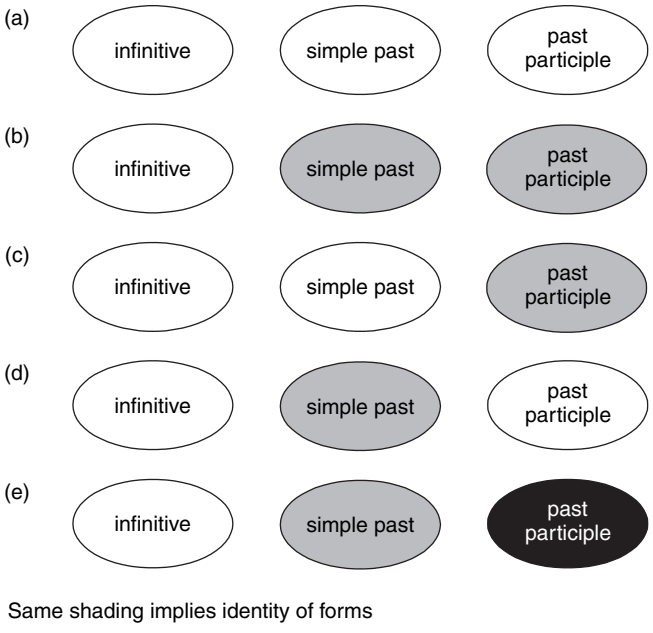

Figure 1.1 Formal identity of forms

across the English vocabulary, as Nielsen 1985 points out, and as will become apparent shortly).

The first type of verbs they mention – equality of all forms – is clearly not optimal in functional terms. Verbs without any morphological tense distinctions certainly have moved furthest on their way towards the 'isolate word'. For them, temporal distinctions can be recovered by the context only. Nevertheless, twenty-four verbs of Quirk et al.'s list fall into this class (i.e. around 14 per cent of all strong verb paradigms listed there – certainly a sizeable subgroup). However – not surprisingly, considering the less than optimally functional nature of this class in the system – for many verbs weak alternatives are recorded (e.g. *rid – rid – rid* but also *ridded*; *bet – bet – bet* but also *betted*), and thus this subclass seems at present to be diminishing. (That historically this pattern has been quite attractive is stressed by Bauer 1997.)

The second group of verbs (e.g. *say – said – said; find – found – found*) – despite having identical past tense and past participle forms – is not dysfunctional at all. Any tense contrasts that might involve the past tense forms and the past participle must also involve further auxiliaries, so that the tenses can always be unambiguously decoded, even if the form of the lexical verb is identical. In particular, the past participle is used for the perfect (obligatorily with forms of HAVE) as well as for the passive (obligatorily with forms of BE); cf. *I found* vs. *I have found/I was found* or *he said* vs.

he had said/it was said. Although this pattern is not the prototypical pattern of strong verbs, half of all strong verbs do pattern like this (81 in Quirk et al.'s reduced list of 167, or over 48 per cent); this is indeed the largest group of strong verbs. More importantly, despite of course forming the past tense by a different process, all weak verbs also follow this abstract pattern. One can therefore say that this type constitutes the prototypical weak verb pattern. As we shall see, in non-standard dialect systems this pattern acts as a powerful attractor for a range of strong verbs, and the weak verb pattern receives additional support in the system from the large subgroup of strong verbs that already pattern alike. This pattern also seems attractive from a cross-linguistic perspective. Durrell, for example, notes for Dutch that here more strong verbs have been retained and indeed more verbs have entered the strong verb classes than in other West Germanic languages, and that these stable strong verbs 'all … have the same vowel in the preterite and the past participle. This levelling seems to have simplified the paradigms and stabilized them, facilitating analogical levelling towards these classes' (Durrell 2001: 13).

Group (c), although at first glance perhaps a little similar, is really quite different. Here, the identity lies between infinitive and the simple past. In contrast to the prototypical weak verb pattern above, the simple present – employing the base form – and the simple past are never further distinguished by auxiliaries; present tense and past tense are after all the only purely morphological (i.e. inflectional) tenses of English (indeed, of the Germanic languages). Similar to those patterns that have identical forms everywhere, therefore, the context is the only source for clues about the temporal reference. Only one formal difference exists between present tense and past tense, namely in the third person singular. Here the present tense regularly has the suffix *–s*, whereas the past tense does not; cf. *I beat* (present? past?) vs. *she beats me* (present)/*she beat me* (past). For spoken language, in particular, the importance of this criterion should not be underestimated, as much discourse is in fact in the third person singular.[14] Again not surprisingly, this type does not contain too many verbs (in Quirk et al.'s list, *beat* is

[14] The figures from FRED are as follows:

Pronoun	Occurrence	% of total	Pronoun	Occurrence	% of total
I	61,458	23.4%	we	27,240	10.4%
he	29,733	11.3%	you	54,163	20.6%
she	9,418	3.6%	they	38,608	14.7%
it	41,776	15.9%			
			Total	262,396	

In other words, even in FRED – heavily biased towards first person narratives – the third person singular accounts for around a third of all pronouns. In addition, of course, all singular noun phrases are in the third person singular.

in fact the only verb, and thus accounts for only around 0.6 per cent of all strong verb types).

Quirk et al.'s fourth pattern – with identity of base form and past participle (e.g. *come – came – come; run – ran – run*) – along the same lines of argument is not particularly non-functional, at least not for the expression of tenses, as there is no possible area of confusion between the simple present (*I come/she comes, you run/we run*), and any perfect form (*I have come, she has come; you have run, we have run*); the important morphological distinction between simple present and simple past is maintained for this verb type. Nevertheless, this pattern is also very much a minority pattern, accounting basically for only the two verbs *come* and *run* (and a number of derivational forms which, as detailed above, have been excluded from these calculations) together making up just over 1 per cent of all strong verbs. Low type frequency is here obscured by extremely high token frequency, with *come* and *run* being some of the most frequent words in general.[15] This is no doubt the reason that this pattern appears intuitively quite common. Although it cannot really be called non-functional, there is a very strong trend in non-standard systems to 'level' the morphologically distinct past tense forms of both *come* and *run*, resulting in three identical forms. A detailed analysis of past tense *come* and *run* (in Chapter 6) aims to shed more light on this phenomenon.

The final pattern – three distinct forms for base form, past tense and past participle, e.g. *sing – sang – sung; eat – ate – eaten; fall – fell – fallen* – results in a maximally distinct three-way paradigm and constitutes the prototypical strong verb pattern. In Quirk et al.'s list, 59 out of 167 or around 35 per cent – a little more than a third – of all verbs conform to this pattern. Not surprisingly, the Old English ablaut series have survived in this pattern especially. Although it is certainly not dysfunctional in any way, the three-way contrast is redundant. In particular, a formal distinction between past tense and past participle is not necessary to assign tenses unambiguously (from the viewpoint of the listener), and perhaps for this reason many non-standard systems tend to 'level' the simple past–past participle contrast for these verbs – at least and especially for one particular subgroup, like *sing – sang – sung*, or *drink – drank – drunk*, namely to *sing – **sung** – sung* or *drink – **drunk** – drunk*. In other words, these verbs become more like prototypical weak verbs, the most frequent group, in particular like a subgroup of these, provisionally designated *Bybee verbs*. These are verbs like *cling – clung – clung*, *win – won – won* or *stick – stuck – stuck* and they have in common a certain phonological shape (to be detailed in Chapter 5), in particular a past tense form in /ʌ/. These

[15] In Francis and Kučera's adjusted frequency list for the Brown corpus (American English), *come* has rank 60, *become* rank 99 and *run* rank 204 (Francis and Kučera 1982: 465–7). This means that *come* is the 60th most frequent word in the corpus, *become* the 99th most frequent, and *run* the 204th most frequent. In fact, *come* is the 11th most frequent verb after *be*, *have*, *do*, *will*, *say*, *make*, *can*, *could*, *go* and *take*.