CHAPTER 1

Introduction

In view of the large amount of lexicological research in present-day English linguistics, surprisingly few studies are devoted to the spelling of English compounds (cf. 1.1.1). This is particularly striking if one considers the pervasiveness of the phenomenon, since any compound in any text necessarily requires the selection of one spelling variant (usually either as an uninterrupted chain of letters or with an intervening hyphen or space). Most of the time, language users seem to select the spelling automatically without the decision process reaching the level of conscious awareness, and conscious reflection is frequently inconclusive in those cases to which attention is drawn. The spelling of English compounds is a strikingly common difficulty and concerns everybody who produces texts in English: native speakers as well as learners of the language. As a consequence, a number of general ideas about the spelling of English compounds exist and can commonly be found in the literature. Most of these ideas can be subsumed under three general statements:

1. **The spelling of English compounds follows no rules.**

This view is certainly the most widespread one, and it is frequently expressed as a complaint: since no principles seem to underlie the spelling of English compounds, they are considered an important source of error and confusion. As a consequence, language users are frequently advised to consult an up-to-date high-quality dictionary when doubting about particular compounds' spelling. The following quotations are representative of numerous similar passages:

- "Of all the questions which arise in an editorial office, one of the most common has to do with compound words. Should it be written taxpayer, tax-payer, or tax payer? Solid, hyphenated, or open?" (Strumpf and Douglas 1988: 52).
- "There often seems to be little logic or consistency in the hyphenation of words in English. The reason for this is that there are few 'rules' as

such – hyphenation is often more a question of style and common sense, rather than principle" (Cullen 1999: 51).

- "Because of the variety of standard practice, the choice among these styles for a given compound represents one of the most common and vexing of all style issues that writers encounter" (Merriam-Webster 2001: 99).
- "Nowhere (or should that be 'no where'?) is English more chaotic than in its seemingly arbitrary spelling of compound words and phrases" (Wilbers 1997).
- "The chaos prevailing among writers or printers or both regarding the use of hyphens is discreditable to English education" (Fowler 1926: 243). "[U]sage is so variable as to be better named caprice" (Fowler 1921: 7).
- "One of the most common spelling issues involve [*sic*] compound words . . . While consulting a dictionary is usually the best answer, some of these words are not necessarily found in dictionaries, and in a few cases, even if they are the spellings vary. There is also no definitive collection of rules regarding the hyphenation of compound words" (Reiser 2007).

All these statements suggest that there are no general principles offering support in the spelling of English compounds. However, it seems that most language users write the majority of compounds without previously checking their spelling in reference works. Advanced spellers in particular tend to have fairly strong intuitions about how to spell – and sometimes how not to spell – certain compounds. Since this intuition must have some kind of basis, the present study sets out to determine what makes spellers choose one variant over another. The underlying assumption here is that compound spelling is governed by certain principles, even though they can be expected to be numerous and far from obvious – otherwise they would have been discovered and made generally known long ago. At the same time, the spelling of compounds is not entirely fixed. Bauer (2003: 134) illustrates the randomness of English compound spelling by listing the spelling variants of *girl friend* in different dictionaries: *girlfriend* in *Hamlyn's Encyclopedic World Dictionary, girl-friend* in *The Concise Oxford Dictionary* (7th edition) and *girl friend* in *Webster's Third New International Dictionary*. In addition, the most common way of spelling a particular compound may change over time – which provides a direct link to the second general statement:

2. The spelling of individual English compounds usually develops from open via hyphenated to solid spelling.

The second most important general idea about the spelling of English compounds is that they usually start off their life with the constituents separated by a space (*girl friend*), then go through a hyphenated stage (*girl-friend*) and finally finish as a solid, uninterrupted sequence of letters (*girlfriend*). Therefore any theory that attempts to discover principles underlying the spelling of English compounds needs to accommodate diachronic developments. Again, various references to the idea can be found in the literature, particularly in grammars and style guides:

- "Most compounds graduate, so to speak, from separation, through hyphenation, to integration; and everyone is entitled to his own opinion on the present status of any one of them. Thus there is still, after centuries of use, no agreement on the correctness of, say, *common sense, common-sense* or *commonsense; good will, good-will*, or *goodwill*" (Carey 1957: 24).
- "With compounds, the constituents are often written as separate words when the collocation seems relatively unestablished: . . . As the sequence of items becomes more established, it may be hyphenated (especially in BrE) as an intermediate stage before being written solid" (Quirk et al. 1985: 1537).
- "As we have seen, compounds start as separate words, then acquire a hyphen, and end as continuous (or unbroken or solid) words" (Partridge 1953: 138).
- "**Compounds** are sometimes said to progress from being spaced as separate words, to being hyphened, and then set solid, but the pattern is far from universal. In American English they may skip the hyphened stage . . . ; and some, especially longer ones like *daylight-saving*, may never progress beyond the hyphenated stage (in British English, or spaced, in American), however well established they are" (Peters 2004: 119).

The last quotation ends with a statement representative of the third common view regarding English compound spelling:

3. British English uses more hyphens in the spelling of compounds than American English.

- "In American English they [= compounds] may skip the hyphened stage" (Peters 2004: 119).

- "American authors tend to use fewer hyphens than the British do" (Butcher 1992: 154).
- "In AmE, hyphenation is less common than in BrE, and instead we find the items open or solid (more usually, the latter) where BrE may use a hyphen" (Quirk et al. 1985: 1569).

It would thus seem important to take geographical considerations into account as well. The present study focuses on compound spelling in British English and draws comparisons with American English.

To conclude, the spelling of English compounds varies from both a synchronic and a diachronic perspective. The present study attempts to determine whether this variation is item-specific or whether it follows more general principles.

1.1 The State of the Art in English Compound Spelling

The spelling of English compounds is treated in extremely heterogeneous sources. The following sections provide a brief overview of its coverage in the linguistic literature (cf. 1.1.1) and the literature consulted by language users seeking advice on the spelling of particular compounds: style guides for prescriptive rules (cf. 1.1.2), grammars for descriptive rules (cf. 1.1.4) and dictionaries for item-specific information (cf. 1.1.3). Furthermore, spellers may consult a corpus (1.1.6) or use spellcheckers when typing text in word processing programs (cf. 1.1.5).

1.1.1 Linguistic Studies

Linguistic accounts of English compound spelling most frequently discuss spelling as a possible criterion of compounding, e.g. in treatments of word formation such as Plag (2003) or Bauer (1983). Empirical studies of the phenomenon are rare but existent.

In the pre-computerised age, in the earliest widely known study of English compound spelling, Morton Ball (1939: 46–52) assembles compounds whose spelling has changed in six editions of Webster dictionaries. A second list (Morton Ball 1939: 54–59) compares dictionaries by different publishers and "illustrates very clearly the chaotic inconsistency of general practice" (Morton Ball 1939: 43). Morton Ball's (1939: 43) conclusion that the development "has been illogical and in absolute conflict with current usage" in many cases is indicative of the relatively prescriptive style of the book. The author does not intend a quantitative description of compound

spelling, but aims to offer stylistic guidance. It is unclear to what extent the lists are representative of the dictionaries as a whole and how the data – which are not analysed statistically – were selected.

The most important predecessor of the present study is Sepp (2006). To avoid "the gray area of the adjective+noun combinations in English . . . and obtain a less controversial set of compounds" (Sepp 2006: 22), the author restricts herself to nominal noun+noun compounds (unusually, including pronouns and acronyms), which are automatically extracted from two American English corpora that jointly comprise about 14 million words. Solid compounds are gathered by first searching for strings based on a dictionary word list and then by searching for the remaining string. Sepp focuses on the 707 compound types which occur more than thirty-five times in her corpora (Sepp 2006: 78–86). Forty-five per cent of these vary in their orthographic form: 43 occur in all three spellings, 101 have solid or open variants, 161 are either hyphenated or open, and 13 may be solid or hyphenated. Sepp checks a variety of parameters – among them the number of syllables in the compound, compound stress and double consonants across the internal boundary – and finds no single determining feature, but a cumulative effect: ten lexical features account for 66.5 per cent of the variance in solid compounds, nine features for 40.5 per cent in hyphenated compounds and ten features for 67.7 per cent in open compounds (Sepp 2006: 105–106, 109).

Mondorf's (2009, 2000) descriptive linguistic treatment of the phenomenon is the by-product of a study on comparison in adjectives. As a consequence, Mondorf's research is also confined to a single part of speech. She finds that about 85 per cent of adjectival compounds are hyphenated (Mondorf 2009: 378) and agrees with Sepp (2006) that the spelling of English compounds is presumably based on a large number of interacting factors.

Plag et al. (2008) consider the spelling of English noun+noun compounds regarding the effect on stress placement. Based on Google frequencies (which treat hyphens as spaces; cf. Plag et al. 2008: 776), they find "that compounds written in one word . . . are more frequently left-stressed than compounds written in two words" (Plag et al. 2008: 778).

Rakić's (2009: 60) study also investigates only noun+noun compounds, but his material comes from a dictionary rather than a corpus: 5,270 compounds were "excerpted from" the *Longman Dictionary of Contemporary English* (LDOCE), but by what method is not elucidated. It also remains unclear why his modern British English material from a single source – whose spelling could also represent a house style – is linked

to American English frequency data from the 1960s, more specifically the Brown Corpus (Rakić 2009: 72). Rakić mainly focuses on morphological structure and finds that morphological complexity and compound length favour open spelling.

The most recent large-scale study of English compound spelling is Kuperman and Bertram (2013). Like most of its predecessors, it is limited to noun+noun compounds, but additionally excludes compounds with a final constituent consisting of verb+ *-ing*, such as *house+warming* (Kuperman and Bertram 2013: 944). Kuperman and Bertram (2013) use two corpora: the regionally indeterminate 2008 Wikipedia corpus (with 1.2 billion tokens) and a diachronic corpus of American English containing 1.8 million documents from the *New York Times* between 1986 and 2007 (Kuperman and Bertram 2013: 943–945), which may be strongly influenced by house style. Their detailed study – which restricts itself to compounds occurring in alternative variants (Kuperman and Bertram 2013: 941) – takes a multitude of factors into account in order to "identify orthographic, phonological, semantic and distributional explanatory factors" for spelling preferences (Kuperman and Bertram 2013: 943).

The remaining empirical linguistic research related to the spelling of English compounds can generally be subsumed under the heading of psycholinguistic or cognitive studies. These investigate various aspects regarding the processing of open, hyphenated and solid spelling such as the following:

- **Is it favourable to insert spaces into normally solid English compounds?** According to research by Juhasz, Inhoff and Rayner (2005), conserving the open spelling of compounds usually spelled that way leads to fast results for first fixations on the compound and lexical decision tasks. The insertion of spaces into normally solid compounds (e.g. *softball* > *soft ball*) speeds up first fixations on the compound and lexical decision. However, when refixations on the compounds are considered, spelling a solid compound open significantly disrupts processing, particularly if it consists of adj+n (rather than n+n) – presumably, because that is frequently accompanied by a change in semantics, as in the example *soft+ball*. The results indicate that open spelling facilitates access to the constituent lexemes, whereas solid spelling benefits the specification of full compound meaning, because "the direct look-up of the whole compound" is "able to begin in parallel to accessing the first constituent of the compound" (Juhasz, Inhoff and Rayner 2005: 314). All this is in line with the more general

assumption that new compounds, whose meaning can often be deduced from their parts, tend to be spelled open, whereas solid spelling is frequently found in compounds with a more specialised, lexicalised meaning (cf. 5.12.4).

- **Is it favourable to insert hyphens into normally solid compounds?** Bertram et al. (2011: 537) find an identical effect to that described earlier for spaces: the insertion of a hyphen into usually solid compounds benefits early processes and disrupts later processes. However, Bertram et al.'s (2011) results are based on triconstituent compounds, which are longer and structurally more complex than the compounds in Juhasz et al. (2005), and based on compounds from two languages other than English (Dutch and Finnish). While some results can be transferred between languages, others – particularly those referring to differences in orthographic systems – may not be so easily applied to other languages, since the expectations of readers in a language that obligatorily concatenates all compounds will presumably differ from those of readers in a language that permits more variation in the spelling.
- **Is it favourable to remove spaces from normally open English compounds?** When Juhasz et al. (2005) also examined compounds from the reverse perspective, their subjects did not find it very disturbing if usually open compounds were spelled solid – in any case less so than if solid compounds were spelled open. Furthermore, readers' eyes land further towards the centre of solid compounds, which reduces the tendency towards refixations.
- **Is the spelling of English compounds influenced by the spelling of morpho-semantically related compounds?** De Jong et al. (2002) answer this question in the affirmative after considering both the position family size and the position family frequency (cf. 5.10.3) of English compounds. They find that a large number of solid and hyphenated family members leads to faster responses from participants. The number of family members with open spelling, by contrast, seems to be irrelevant or even inhibitory, since the words with few open family members were responded to faster than those with many.
- **Are there any differences in the processing of open, hyphenated and solid compounds?** In the same study, de Jong et al. (2002) conclude that English open compounds are not part of the morphological families of simplex words and that the space between the constituents of open compounds renders their processing more comparable to that of simplex words, since it is also influenced by position family size rather than position family frequency. This outcome suggests that open

compounds may be represented differently from solid and hyphenated compounds at a central level.

- **What is the best visual cue for compound constituent boundaries?** Research for German by Inhoff, Radach and Heller (2000: 45) suggests that spacing is a particularly good visual cue to indicate constituent boundaries.

The collective view emerging from the psycholinguistic research data seems to be that open, hyphenated and solid spellings each have advantages and disadvantages for language processing. The experimental studies can only consider a relatively limited number of research items, and there is a strong focus on noun compounds only. Furthermore, the theoretical background of the phenomenon (such as the issues of norm, variation and language change) is only considered to a limited extent. Most importantly, all the studies only investigate the spelling of English compounds from a receptive perspective and do not take production into account, although there are important differences: thus readers may accept any out of two or even more spelling variants without comprehension problems (and possibly even without awareness of potential variation), but production necessarily requires the selection of the single spelling variant deemed most suitable, so that minor differences in acceptability have larger repercussions on compound spelling compared to reading.

1.1.2 Style Guides

The majority of books dealing with the spelling of English compounds are style guides treating it as one among many other issues, e.g. Bailey's (1979) *English Punctuation in Brief*, Carey's (1958) *Mind the Stop* and McDermott's (1990) *Punctuation for Now*, all of which focus on punctuation. Listing all style guides on the market is almost impossible, since many of the large publishing companies and newspapers have their own usage guides, e.g. *Webster's New World Guide to Punctuation* (Strumpf and Douglas 1988) or *Chambers Guide to Punctuation* (Cullen 1999). Style guides (which are also called *usage guides*) provide easily accessible information on constructions[1] of unclear usage (which is more difficult to retrieve from descriptive reference grammars; cf. Busse and Schröder 2010b: 99). Since style guides are almost exclusively prescriptive, and rarely indicate

[1] The term *construction*, understood in the sense of "a pairing of form with meaning/use" (Goldberg 1996: 68), is used here to refer to linguistic entities without the necessity of specifying whether these are categorised as compounds or phrases or as situated on a gradient.

their sources, they will merely be considered as the subject of investigation here, e.g. with regard to the criteria proposed for compound spelling (cf. Chapter 5) and with the aim of empirically testing the spelling rules advocated therein.

Since most style guides are published by some authority, they often describe the in-house style of a particular publisher or editorial board – e.g. *New Hart's Rules* (Ritter 2005a) for Oxford University Press. Such style guides represent the conventions agreed upon by particular relatively small groups, and they sometimes contradict each other. While they are arbitrary to a certain extent, there seems to be a common core of agreement: some rules, such as the hyphenation of ad hoc compounds, are cited by practically all books in this category – but not by Peters (2004). Her *Cambridge Guide to English Usage* is exceptional by basing its advice on corpora and questionnaires to determine acceptability, but the style guide format forces Peters to omit her statistical basis. In addition, hyphenation is treated only relatively briefly, because the style guide covers many different areas of the English language. An important dilemma of usage guides is that they have to "cover a vast amount of linguistic ground, without necessarily being an expert in all areas" (Busse and Schröder 2010a: 49). As far as hyphenation is concerned, however, the author of the classic English style guide *A Dictionary of Modern English Usage* (1926), Henry Watson Fowler, had published an almost verbatim tract on hyphenation in 1921.

Beside these shorter treatments of English compound spelling, there are two whole books devoted to the topic by Alice Morton Ball, *Compounding in the English Language* (1939) and *The Compounding and Hyphenation of English Words* (1951). The former can be considered a digest style guide discussing the treatment of English compound spelling in an impressive number of different reference works. Morton Ball also sets up her own spelling system for the *Department of State Style Manual* (1939: 67), the predecessor of the US Government Printing Office's *Style Manual* (2008), whose chapter 6 still deals with compounding rules. In Morton Ball's own prescriptive judgement, hers is "the only complete rational system that has ever been formulated up to the present time" (Morton Ball 1939: 66). A disadvantage of her work is that some rules include so many conditions that they actually correspond to three or even more rules, e.g. when she states that a "noun or adjective and a verb, used jointly as a verb; and two verbs, or a verb and a noun, used jointly as a noun" belong in the category of "Words Properly Compounded" (Morton Ball 1951: 7).

Style guides differ in the amount of detail that they offer on compound spelling in English, regarding both the space and the number of examples

devoted to the phenomenon. Thus Peters' (2004) discussion comprises little more than one page (259–260) and is limited to compounds with two constituents, while Merriam-Webster (2001) extends the treatment of compound spelling over more than twelve full pages (98–111), differentiated by part of speech and supported by additional sections on hyphens and hyphenated compounds, drawing on citation files of 15 million examples and using hedges such as *usually, generally, sometimes*. Surprisingly, Bell's (2009) *Rules and Exceptions of English Spelling* does not discuss compound spelling at all, and the prescriptive classic *Eats, Shoots and Leaves: The Zero Tolerance Approach to Punctuation* merely treats hyphenation in a "teeny-weeny, hooked-on, after-thought-y chapter" (Truss 2003: 168), in which the author expresses her regret at disappearing (and also misplaced) hyphens, whereas the question when open and solid spelling are appropriate is not discussed at all. The American classic *The Elements of Style* also limits itself to a single piece of advice in this respect: "Do not use a hyphen between words that can better be written as one word: *water-fowl, waterfowl*. Common sense will aid you in the decision, but a dictionary is more reliable" (Strunk and White 2000: 35).

To sum up, style guides usually offer one or more of the following types of information:

a) prescriptive rules or descriptive principles on how to spell compounds
b) the spelling of a relatively small number of prototypical compounds
c) the spelling of a relatively small number of particularly frequent compounds
d) the spelling of a relatively small number of particularly difficult compounds
e) exceptions to the rules/principles

whereas:

f) the spelling of individual compounds in general

is relegated to dictionaries – even if some style guides, such as Morton Ball (1939, 1951), also offer long lists of compounds and may be considered hybrid in this respect.

1.1.3 Dictionaries

Although dictionaries cannot inform on the spelling of all compounds in a language, they are far more comprehensive than style guides in this respect, which leads Strumpf and Douglas (1988: 49) to reach the (debatable)