

Stanimir Rakić

University of Belgrade, Serbia

Some Further Observations on the Spelling of English Compounds

Summary

This paper explores the main factors which determine the spelling of English N+N compounds. On the basis of a corpus extracted from LDCE (2000) and LDCE (2003), I discuss the following factors, which may have an influence on the spelling of English N+N compounds: the number of syllables, the morphological structure of the first constituent of compounds and the nature of phonotactic transition at the morpheme boundary. The analysis shows that the first two factors exert major influences on the spelling of compounds, while the influence of the third factor varies depending on the number of syllables. In the examined corpus, very few compounds with more than four syllables are spelled solid. The majority of two-syllable compounds are spelled solid under all circumstances, while only 74 four-syllable compounds are spelled solid. The highest percentage of two-syllable compounds are spelled open (49.7%) if the first constituent ending in a consonant builds a consonant cluster at the morpheme boundary. The majority of three-syllable compounds are spelled open unless the first constituent ends in a schwa. The proposed analysis of the extracted corpus shows varied influence of different factors and enables us to establish their partial ranking.

Key words: English compounds, spelling, syllables, morphological structure, phonotactic transitions, factors ranking.

Nekaj novih opažanj glede pisave angleških sestavljenk

Povzetek

Članek raziskuje glavne dejavnike, ki določajo pisavo angleških sestavljenk, tvorjenih iz dveh samostalnikov. Na korpusu sestavljenk iz LDCE (2000) in LDCE (2003) analiziram naslednje dejavnike, ki morda vplivajo na pisavo angleških sestavljenk, tvorjenih iz dveh samostalnikov: število zlogov, morfološko strukturo prvega samostalnika v sestavljenki in naravo fonotaktičnega prehoda na zlogotvorni meji. Analiza je pokazala, da prva dva dejavnika najbolj vplivata na pisavo sestavljenk, medtem ko je vpliv tretjega dejavnika odvisen od števila zlogov. V analiziranem korpusu je zelo malo sestavljenk daljših od štirih zlogov, ki se pišejo skupaj. Večina dvozložnih sestavljenk se piše skupaj v vseh preverjenih pogojih, medtem ko se samo 74 štirizložnih sestavljenk piše skupaj. Največji odstotek dvozložnih sestavljenk (49,7%) se piše ločeno če se prvi del konča na soglasnik, ki tvori soglasniški sklop z začetnim soglasnikom drugega dela sestavljenke. Večina trizložnih sestavljenk se piše ločeno, razen še se prvi del konča na polglasnik. Analiza korpusa kaže različno vplivnost omenjenih dejavnikov in nam omogoča njihovo delno rangiranje.

Ključne besede: angleške sestavljenke, pisava, zlogi, morfološka struktura, fonotaktični prehod, faktorji rangiranja

Some Further Observations on the Spelling of English Compounds

1. Introduction

It is very well known that English compounds may be written in three different ways even in the case of having the same accent (Jespersen 1942, 136). Marchand (1969, 21) also noted “the complete lack of uniformity” in the spelling of English compounds. Quirk et al (1985) made the same observation, noting that some compounds may be written in all three ways: “solid”, “hyphenated” or “open”.¹ Nonetheless they concluded that “there is a progression from “open” to “solid” spelling “as a given compound becomes established”. Bauer (1998, 69) also notices that “English orthography is extremely inconsistent in dealing with noun + noun collocations” so that a single item may be written in three ways (e.g. girl friend, girl-friend and girlfriend). He guesses that orthography might reflect some linguistic intuitions, although it is difficult to pin them down. For example, it might be that longer words are written separately, independently of the stress pattern, while short words are more likely to be written together, but Bauer does not believe that such a statement may be a valuable linguistic generalization. One can, however, note that in one respect English orthography is consistent: the suffixes are assumed to be written solid. In this respect the English orthography seems to reflect the intuitions of native speakers, and I will try to build on this observation.

The spelling of compounds is usually mentioned only in discussing the definition of compounds, but is rarely studied for its own sake. In this paper I intend to do exactly this – to examine the possible regularities in the spelling of noun – noun compounds as they occur in the corpus extracted from the LDCE (2000) and LDCE (2003). I attempt to show that there are some important regularities in the spelling of English compounds which can, however, be described only as statistical tendencies. Here we shall limit our attention to compounds spelled solid or open; those spelled with a hyphen are a rather special case since this orthographic sign seems to be partly subject to special rules. In addition, within the LDCE one finds considerably fewer compounds spelled with a hyphen than those spelled open or solid so that the mentioned limitation will not substantially affect our investigation (see Table 2).

In this paper I try to isolate some factors which influence the way in which English noun – noun compounds are spelled. My observations are based on the corpus of 6020 compounds extracted from the LDCE (2000) and the LDCE (2003), but I do not believe that this fact can impair them as the spoken language is usually simpler than the written one.² It is therefore not plausible that there are in the spoken language compound structures which are not represented in the corpus

¹ These terms seem to be convenient labels for denoting different ways of spelling compounds and we will use them throughout this paper.

² Miller (2006) notes that spontaneous speech is “subject to the limitations of short-term memory”. In such speech “phrases contain fewer words and clauses contain fewer phrases” than in planned writing. He also notes that similar pattern holds for compound nouns. Besides, LDCE (2003) itself is strongly based on spoken language, so that “coverage of the spoken language is second to none” (LDCE 2003, Introduction, XI). One of the consequences of such an orientation is that polymorphous compounds are not at all numerous in our corpus.

extracted from the LDCE (2000) and LDCE (2003). Three main factors seem to influence the spelling of compounds. One of them, the number of syllables, has already informally been identified as the length of compounds. My analysis confirms that observation, but I go a step or two further arguing that there are two other factors which influence the spelling of compounds: the morphological complexity of the constituents and the kind of phonotactic transitions at the morpheme boundary.

In a recent work of mine based on the same corpus (Rakić 2008), I have shown that compounds with morphologically complex constituents tend to be spelled open. I have in particular demonstrated that those compounds in which suffixes take an inner position (e.g. *absentee ballot*, *aircraftwoman*, *amusement arcade*, *apartment block*, *blotting-paper* ...) are more often spelled open than those in which suffixes take an outer position (e.g. *adult education*, *aid worker*, *air conditioning*, *air freshener*, *air-hostess*, *baby-minder*...). If we denote the former group of compounds with the formula (Stem+Suffix)+Stem, and the latter with the formula Stem+(Stem+Suffix), we can illustrate the frequencies within the different ways of spelling in the following Table 1.

Compound structure	Sum	Open	solid	hyphenated
Stem.+ (Stem+Suffix)	879=100%	547 = 62.23%	237 = 26.96%	95 = 10.81%
(Stem+Suffix) + Stem	733=100%	650 = 88.68%	82=11.19%	11 = 1.50%

Table 1.

The significant difference of these percentages shows that the morphological structure of constituents influences the way in which compounds are written. In particular, we have shown that the fact that suffixes are attached to the first constituent rather than to the second considerably influences the spelling of compounds.

In this paper I continue to examine the effects of morphological complexity in the spelling of compounds. Here I examine how the fact that the first constituent of compounds is a suffixal derivative affects the spelling of compounds which have a different number of syllables. In addition, I examine how the phonotactic transitions at the morpheme boundary affect the spelling of compounds. I distinguish two cases: the first constituent may end in a schwa, or it may end in a consonant which makes up consonant clusters at the morpheme boundary. The distribution of some suffixes in three-syllabic compounds suggests that such a distinction is relevant for the spelling of compounds. The analysis in Section 2 below suggests that different phonotactic transitions may have different effects on the spelling of compounds. In this way I can expand the analysis to compounds in which the constituents are morphologically simple. I can, therefore, claim that there is also some regularity in the spelling of compounds whose constituents are morphologically simple, although these regularities may only be presented as statistical tendencies.

My approach has been influenced by Hay's analysis on the parsability of complex words. She found that the parsability of complex words depends on two main factors: the probability of phonotactic transitions at the morpheme boundaries, and the relative frequency of the basis in respect to that of the derived word. The affixes which at the morpheme boundaries build phonemic combinations

which are unusual in monomorphemic words are easier to recognize and separate from the basis, and similar effects have a relative frequency of the basis in respect to the derived word. If the basis is independently more frequent than the derived word, the greater is the probability that the complexity of the derived word will be noticed as well as the particular meanings of its parts. It follows from these assumptions that derived words may have a different degree of complexity, depending on the kinds of affixes and the frequency of the constituent parts. Some affixes are easier to separate and recognize than the others – for example the suffixes which begin with consonants are usually easier to identify than the suffixes which begin with vowels, although the interference of the frequency makes it impossible to establish one-way dependence.

Hay (2000) did not consider a possible application of her theory to the orthography of English compounds, and it is not obvious that such an application is at all possible. It is easy to see that the relative frequency of constituents can hardly have an impact on the spelling of compounds in English. Following the lead of Hay's theory, one might assume that the compound is spelled solid if the compound is more frequent than its constituent. But it turns out that such an assumption is false in most cases. For example, in Francis and Kučera (1982) we can find that the frequency for the noun *ash* is 17, and for the noun *tray* 21, but for the compound *ashtray* just 1, and similar examples abound, so that it is difficult to find even a single counterexample; a quick check in a Google inquiry only confirms such a conclusion. Therefore, the clause of relative frequency is not applicable to the spelling of compounds. But when we consider phonotactic transitions, some regularity in the compound spelling seems to emerge. In order to keep different factors separated, I will compare the effects of phonotactic transitions on the spelling of compounds in which first constituents are not suffixal derivatives.³ It turns out that phonotactic transitions have some influence on the spelling of compounds, although the effects of this influence greatly depend on the number of syllables.

The paper is organized as follows. In Section 2, I discuss the factors affecting the spelling of compounds. In sections 3, 4 and 5, I discuss respectively the spelling of two-, three- and four-syllable compounds. Finally, in Section 6 the main observations are discussed and summarized.

2. The factors affecting the spelling of compounds

A basic fact about the spelling of compounds is that it heavily depends on the number of syllables. This is shown in Table 2.

	2 syll.	3 syll.	4 syll.	5 syll.	6 syll.	7 syll.	8 syll.	9 syll.	10 syll.
Solid	1448	558	74	4	3	0	0	0	0
Open	840	1488	892	320	112	38	15	2	1
Hyphen	65	109	37	8	4	1	1	0	0
Sum	2353	2155	1003	332	119	39	16	2	1

Table 2.

It is obvious that the number of compounds quickly decreases as the number of syllables rises. The number of compounds written solid decreases even faster, so that not one compound with

³ If the first constituent is not a suffixal derivative, it can still be a prefixal derivative or compound; there are, in fact, in our corpus 3.73% of such compounds. For this reason, we cannot use the adjective "monomorphemic" instead of the expression "not a suffixal derivative".

more than six syllables is written solid. Table 2 therefore shows clearly that the greater number of syllables goes hand in hand with the open spelling of compounds.

I shall here examine the compounds which can be written solid or open, and these occur in greater numbers only with two-, three- and four-syllable compounds. There are very few compounds with the number of syllables greater than 4 which are spelled solid, so that all of them can be given in (1):

- (1a) deliveryman, assemblywoman, upperclasswoman, dipsomaniac
- (1b) radiotherapy, aromatherapy, laboratoryman.

In (1a) are the compounds with 5 syllables, in (1b) with 6. No compounds with 7, 8, 9 or 10 syllables are written solid in our corpus. I assume that this fact also has to do with processing – it is more difficult to process words with a greater number of syllables if the morpheme boundaries are not overtly indicated. My task is thereby greatly simplified because we are able to analyze only the spelling of compounds with 2, 3 and 4 syllables.

One of our premises is that the morphological structure of the constituents has a considerable effect on the spelling of compounds. Two-syllable compounds can contain only the suffix *-s* in the first or second constituent. There are 47 compounds containing *-s* in the first constituent, 39 of them are written solid and 8 open. The following compounds are spelled solid:

- (2) bridesmaid, bandsman, batsman, clansman, capslock, cockscomb, craftsman, cranesbill, draftsman, draughtsman, fieldsman, groomsman, groundsman, guardsman, helmsman, herdsman, huntsman, jobsworth, kinsfolk, kinsman, lambswool, linesman, marksman, nutscase, oarsman, pointsman, salesgirl, salesman, sportscast, sportsman, sportswear, statesman, steersman, swordsman, townsfolk, tribesman, tradesman, woodsman, yachtsman.

The form *man* can be regarded as a combining form (a semi-suffix).⁴ If we take away the examples with *man*, we are left with only 12 examples:

- (3) bridesmaid, capslock, cockscomb, cranesbill, jobsworth, kinsfolk, lambswool, nutscase, salesgirl, sportscast, sportswear, townsfolk.

There are also 8 compounds which contain the suffix *s* and are spelled open:

- (4) brains trust, sales slip, sales tax, sports car, sports coat, sports day, sports bra, trials bike.

⁴ The problems of combining forms (semi-suffixes) have been discussed in Marchand (1969), Allen (1978), Giegerich (1985); these forms are supposed to appear also as independent words. In literature different lists of such forms have been given. The semi-affixes (or affixoids) have been recently analysed by Booij (2005) from the position of the theory of grammaticalization and construction grammar. The rise of semi-affixes is a typical process of grammaticalization in that content words are becoming grammatical morphemes. When they are used as affixes their meaning is reduced, but they can still be used independently with “a greater range of meaning”. According to this interpretation, the words derived with semi-affixes are special constructions which take the intermediate positions between compounds and derivatives. The use of the term ‘combining forms’ is justified in the context of my analysis because it points out to the particular character of the compounds in (2), in which some lexical units repeatedly appear as a second constituent. I follow COED (2004), in which the forms *man*, *woman*, *person*, *ware* and *work* have been classified in this way. I add to this list the form *way*, which, surprisingly, in COED has been labelled as a full suffix. For similar treatment of combining forms see *Collins Cobuild English Guide 2. Word Formation* (1993).

None of the combinations with *man* occur in (4), although it occurs in some compounds without *-s* (e.g. *hit man*, *stunt man*, *front man*). The combining forms (semi-suffixes) seem to be more ready to combine with the bases ending in consonant clusters than lexical free words. In general, *-s* clusters do not seem to disfavour solid spelling: in spite of the fact that we have eliminated the examples with combining forms, the number of solid spelling remains greater than the number of open spelling. This may be a particular property of the phoneme /s/: It is known that at the end of consonant clusters /s/ is extrasyllabic, it does not count as a member of coda (s. Giegerich 1993, 149, Ewen and Hulst 2001, 136); One can expect that the effect of consonant clusters in the spelling of compounds with 3 and 4 syllables may be different.

Two-syllabic compounds can contain only constituents derived by *-s*. Other suffixes can appear only in three- and four-syllable compounds, to which we now turn. The suffixes in three-syllable compounds spelled solid are given in Table 3, and those in three-syllable compounds spelled open are given in Table 4 and Table 5. The signs *suf1* and *suf2* refer to the suffixes added to the first and second constituents of compounds respectively.

	s	er	age	ice	craft	iers	ing	ant	ion	ess
suf1	15	10	2	1	1	1	0	0	0	0
suf2	10	128	0	0	0	0	39	2	1	3

Table 3. Suffixes in three-syllable compounds written solid.

The number of suffixes occurring in the compounds spelled open is much greater than that occurring in the compounds spelled solid so that the former must be given in two tables. This bare fact shows the impact of the morphological structure on the spelling of compounds – the compounds with morphologically complex constituents are more often spelled open than solid. In particular, there are a number of suffixes occurring only once in the first constituent of compounds spelled open. We represent them in Table 5.

	er	ing	s	age	ty	ion	y	ure	ings	al	dom	ie	ice	ist	ness	eer
suf1	35	205	8	15	8	8	5	3	3	3	2	2	2	2	2	2
suf2	81	43	44	2	0	11	1	5	0	0	0	0	4	1	3	0

Table 4. Suffixes in three-syllable compounds written open.

	ance	ee	ent	ery	ity	ment	um	or	th
suf1	1	1	1	1	1	1	1	1	1
suf2	0	0	0	0	0	2	0	5	0

Table 5. Suffixes in three-syllable compounds written open.

Table 3 shows that there are few suffixes ending in a consonant which apply to the first constituents in the compounds written solid. Let us look at these compounds:

(5) *carriageway*, *passageway*, *aircraftman*, *frontiersman*, *serviceman*.

The segments *way* and *man* can be considered as combining forms (semi-suffixes), and this means that we are left only with the suffixes *-s* and *-er*, which are the only suffixes appearing in a greater number of forms in Table 3. But these suffixes have a quite different impact on the character of the morpheme boundary: the suffix *-s* can build consonant clusters at the morpheme boundary, while the suffix *-er* cannot. One can notice that in Table 3 the suffix *-er* appears more often with the second constituent in solid spelling, and in Table 4 occurs in a much greater number with the first constituent in open spelling. And in general, there are 10 compounds in which the first constituent contains a suffix ending in a schwa and are spelled solid and 39 compounds with the same type of first constituent spelled open. Does that mean that compounds in which first constituents end in a schwa are more likely to be spelled open than solid? Not necessarily, this difference may also be the consequence of the fact that in Table 3 the first constituent is a suffixal derivative, for which we already know that it is an important factor favouring open spelling. Nonetheless, it seems to be a natural move to examine whether a vocalic ending of the first constituent favours open or solid spelling while assuming that it is not a part of suffixal derivation.

Another peculiarity in Table 3 and Table 4 is a curious distribution of the suffix *-ing*. This otherwise very productive suffix appears only in the second constituent in solid spelling, but it massively occurs in the first constituent in open spelling. The disproportion of 0:205 is so conspicuous that it is not likely that we can ascribe it only to the tendency of the compounds in which the first constituents are suffixal derivatives to be spelled open. This state of affairs provides us with another hint: the compounds N+N seem to be spelled open rather than solid when the first constituent ends in a consonant cluster. In Table 3 we find the suffix *-craft*, which ends in a consonant cluster, but it occurs only in the compound *aircraftman* in which the second constituent *man* can be understood as a combining form (semi-suffix).

Let us now look at some simple examples in which the first constituent builds consonant clusters at the morpheme boundary. Some examples with the noun *child* as the first constituent are presented in (6):

- (6a) childbirth, childcare,
- (6b) childbearing, childminder, child support,
- (6c) child molester, child benefit, child prodigy.

In (6a) the two-syllable compounds are spelled solid, while there is a vacillation with the three-syllable compounds (6b); in (6c) the four-syllable compounds are spelled open. The number of syllables must therefore be taken account of in the discussion of compound spellings – we can only compare compounds with the same number of syllables.⁵

We can now start examining the hypothesis that besides the number of syllables three other factors influence the spelling of compounds: consonant clusters at the morpheme boundary, the schwa at the end of the first constituent, and, particularly, whether or not the first constituent is a suffixal derivative. In order to estimate the strength of these conditions independently, we have

⁵ As consonant clusters at the morpheme boundary, we count only three-member clusters in the case when the first constituent ends in a consonant, and the second constituent begins with a consonant. The cluster CC.C can be dubbed a coda-cluster, and C.CC an onset cluster. The clusters in (7) are the examples of coda-clusters.

to distinguish the following subsets of compounds:

- (7a) the compounds in which the first constituents are derived by suffixes,
- (7b) the compounds in which the first constituents are not suffixal derivatives,
- (7c) the compounds in which the first constituents are not suffixal derivatives and end in a schwa,
- (7d) the compounds in which the first constituents are not suffixal derivatives and make up a consonant cluster at the morphemic boundary,
- (7e) the compounds in which the first constituents are not suffixal derivatives and do not fulfil any of the previous conditions, except (7b).

For each of these subsets we have to determine the percentage of compounds spelled solid and open. The subset (7e) is intended to show us what the “neutral” case is, i.e. what percent of compounds are spelled in either way when no special conditions on morphological structure and on phonotactic transitions obtain. All conditions (7a) – (7e) could have a different impact on the spelling of compounds depending on the number of syllables they consist of. We must, therefore, examine how these conditions are fulfilled in compounds with different numbers of syllables. We turn first to two-syllable compounds, and then to three- and four-syllable compounds.

3. Two-syllable compounds

In two-syllable compounds condition (7c) cannot be fulfilled; we can therefore examine only conditions (7a), (7b), (7d) and (7e). We start with condition (7a), and then examine conditions (7b), (7d) and (7e).

3.1 We first examine condition (7a), which refers to the compounds in which the first constituents are suffixal derivations. The counting above (see (3) and (4)) has shown that 12 two-syllable compounds in which the first constituent is derived by *-s* are spelled solid, and 8 open. We get the relation (10) with rounded percentages:

$$(8) \quad 12:8=60\%:40\%.$$

This means that six in ten compounds with *-s* added to the first constituents are spelled solid.

3.2 Next, we turn to condition (7b), which refers to the compounds in which the first constituents are not suffixal derivatives. There are 1448 two-syllable compounds spelled solid, and 840 spelled open. From 1448 we first subtract 12 compounds in which the first constituents are suffixal derivatives, and then 113 compounds containing the combining forms *man* (53), *work/works* (34), *way/ways* (26) and *ware* (3). Subtracting this number, we get 1323 compounds spelled solid. In open spelling, only in 8 compounds are the first constituents suffixal derivatives. The subtraction gives 832 compounds in open spelling. The corresponding percentages are shown in (9).

$$(9) \quad 1323:832=61.39\%:38.61\%.$$

This means that the addition of the suffix *-s* slightly influences the spelling of two-syllable compounds. Only 1.39% more compounds without *-s* in the first constituent are spelled solid than those with *-s*. This also suggests that *-s* in two-syllable compounds does not function like a linking element.

3.3 Now we examine condition (7d). This condition refers to the first constituents which are not suffixal derivations, and form consonant clusters at the morpheme boundary. We can distinguish the compounds with onset clusters and with coda clusters. There are 222 compounds spelled solid with consonant clusters at the morpheme boundary, the examples of those with onset clusters are shown in (10a), and those with coda clusters in (10b).

(10a) bloodstain, breadcrumbs, buckskin, catchphrase, cockcrow, dogsled, drumstick, flagstaff, footbridge, beansprout;

(10b) breastbone, calfskin, chestnut, clingfilm, driftnet, dustpan, fieldmouse, foxtrot, goldbrick, benchmark, campfire, wristband.

On the other hand, 220 compounds with consonant clusters at the morpheme boundary are spelled open (e.g. *brain drain, bank book, blood group, box lunch, brand name, change purse, crown price, change purse, death squad, dump truck, dust bowl*). We can now calculate the percentages of compounds spelled solid and open:

(11a) $222:220=50.23\%:49.7\%$.

(11b) shows that two-syllable compounds with consonant clusters at the morpheme boundary are slightly more often spelled solid than open.

3.4 Let us now look at condition (7e), which defines the “neutral” case when no special conditions examined so far are imposed. There are 1448 compounds spelled solid, and 840 spelled open. We assume that consonant clusters affect the spelling of compounds, as well as suffixes added to the first constituents. We have, therefore, to subtract the numbers of compounds in which these special conditions materialize. 113 combining forms *man* (53), *way* (24), *work* (33) and *ware* (3) occur almost all in the compounds spelled solid.⁶ We get, therefore, $1448-12-113-222=1101$ for the compounds spelled solid, and $840-8-220=612$ for the compounds spelled open.

We therefore get 1101 compounds spelled solid, and 612 compounds spelled open which do not satisfy any special conditions. The corresponding percentages are given in (12):

(12) $1101:612=64.27\%:35.73\%$.

Thereby we get that 64.27% of two-syllable compounds are spelled solid, and 35.73% spelled open. This proportion characterizes the “neutral” case for two-syllable compounds.

3.5 The results of the analysis of two-syllable compounds may now be presented in Table 6:

	solid	open
no special conditions	64.27%	35.73%
first constituents expended by the suffix –s	60%	40%
first constituents are not suffixal derivation	61.39%	38.61%
first constituents make consonant clusters	50.23%	49.77%

Table 6. The effects of different conditions on the spelling of two-syllable compounds.

The first row shows a neutral case – this relation is specific for two-syllable compounds, since for

⁶ Only 4 examples with one-syllabic combining forms *man* and *work* are spelled open (*hit man, stunt man, front man, grunt work*) spelled open.

different numbers of syllables we expect to get different proportions. We assume that consonant clusters increase the open spelling of compounds, and *-s* suffix clusters do it considerably less than the clusters made by the constituents of the compounds. The fact that *s* – clusters favour open spelling to a lesser degree than other consonant clusters may be explained by the often made assumption that *s* at the end of consonant clusters is extrasyllabic in English.

4. Three-syllable compounds

4.1 We first examine condition (7a), which refers to the first constituents derived by suffixes. We can simply count the number of compounds with the suffixes in Table 2 for solid spelling, and in Table 4 and 5 for open spelling. There are 30 compounds spelled solid with the first constituents derived by suffixes, but (5) shows that 5 of them have combining forms as second constituents. We are therefore left with 25 compounds in solid spelling, 15 with *-s* and 10 with *-er*:

(13a) *craftswoman, kinswoman, oarswoman, salesperson, saleswoman, sportswoman, thanksgiving, townspeople, yachtswoman, backwoodsman, tribeswoman, tradespeople, sportsperson, sportswriter, codswallop.*

(13b) *trollerman, fisherman, pokerwork, chatterbox, chatterline, checkerboard, clapperboard, halterneck, rubberneck, boilerplate.*

When we eliminate the compounds with combining forms in (13), we are left with 12 examples:

(14) *thanksgiving, townspeople, tradespeople, sportswriter, codswallop, chatterbox, chatterline, checkerboard, clapperboard, halterneck, rubberneck, boilerplate.*

Summing up the numbers for *suf1* in Table 4 and Table 5 we see that 314 compounds spelled open have the first constituent derived by suffixes. Thereby we get the proportion (15):

(15) $12:314=3.68\%:96.32\%$.

(15) shows how much the morphological structure of the first constituent influences the spelling of compounds.

4.2 Now we shall examine the compounds in which the first constituents are not suffixal derivatives. There are 558 three-syllable compounds spelled solid, but 12 of them have the first constituent derived by some suffixes. Subtraction gives us the result of 546 compounds spelled solid. Again, we have to subtract 65 combining forms built with *man* (28), *woman* (14), *person* (3), *ware* (3), *work* (10) and *way* (7).⁷ The number of compounds with the monomorphemic first constituent is therefore reduced to 481.

⁷ Here is the list of these words: *aircraftman, airwoman, alleyway, anchorman, backwoodsman, basketwork, bodywork, bogeyman, bridleway, businessman, carriageway, cattleman, chairperson, charwoman, clergyman, congressman, councilman, countryman, craftswoman, dairyman, donkeywork, entryway, ferryman, fisherman, frontiersman, highwayman, horsewoman, ironware, ironwork, journeyman, jurywoman, kinswoman, linkwoman, longshoreman, lumberman, masterwork, merchantman, metalwork, midshipman, minuteman, motorway, muscleman, needlework, oarswoman, paperwork, passageway, patrolman, pokerwork, policeman, repairman, saleswoman, salesperson, serviceman, signalman, silverware, spacewoman, sportsperson, sportswoman, tableware, trawlerman, tribeswoman, waterway, waterworks, weatherman, yachtswoman.*

There are 1488 three syllable compounds spelled open. When we subtract 314 compounds in which the first constituent is derived by some suffixes, we arrive at the number of 1174. Therefore, the proportion (16):

$$(16) 481:1174=29.06\%:70.94\%.$$

This proportion shows that less than three in ten compounds are spelled solid under the given condition.

4.3 Now we shall examine the compounds in which the first constituents are not suffixal derivatives and end in a schwa. Counting shows that there are 101 solid compounds which satisfy these conditions, and 120 open compounds with the same property. We get the proportion (17):

$$(17) 101:120=45.70\%:54.30\%.$$

(17) shows that more compounds are spelled open than solid if the first constituents are not suffixal derivatives and end in a schwa in three-syllable compounds.

4.4 Now we shall examine the compounds in which the first constituents are not suffixal derivatives and make up consonant clusters at the morphemic boundary. Counting shows that there are 60 such compounds in solid spelling and 210 in open spelling. We can calculate the proportion (18):

$$(18) 60:210=22.22\%:77.78\%.$$

This proportion shows that more than two compounds in ten is spelled solid under the given conditions.

Obviously, less compounds are spelled solid under the given conditions than in 4.2.

4.5 We now turn to condition (7e) and examine the compounds whose first constituents are not suffixal derivatives and do not fulfil any other special conditions. According to the calculation in 4.2 there are 481 compounds spelled solid in which the first constituents are not suffixal derivations and do not contain combining forms. From that number we have still to subtract 101 compounds whose first constituents end in a schwa and 60 compounds which make up consonant clusters at the morpheme boundary. We finally arrive at the required number of 320 compounds.

From 4.2 we have 1181 compounds spelled open in which the first constituents are not suffixal derivatives. When we subtract the number of compounds whose first constituent ends in a schwa or forms consonant clusters at the morpheme boundary, we get 851 compounds spelled open which do not fulfil any special conditions on their first constituents. The proportion of the resulting numbers should reflect the “neutral” case characteristic for three-syllable compounds:

$$(19) 320:851=27.33\%:72.67\%.$$

These percentages should correspond to the probability of spelling a three-syllable compound solid or open if we do not impose any special morphological or phonotactic conditions on the first constituent.

4.6 We can now represent the results of our examination of the spelling of three-syllable compounds in Table 7.

	solid	open
no special conditions	27.33%	72.67%
first constituents are suffixal derivations	3.68%	96.32%
first constituents are not suffixal derivations	29.06%	70.94%
first constituents end in a schwa	45.70%	54.30%
first constituents make up clusters	22.22%	77.78%

Table 7. The effects of different conditions on the spelling of three-syllable compounds.

From the first row we see that more than about one compound in four is spelled solid, if no special conditions on the first constituents are imposed. We can infer that the fact that the first constituent is a suffixal derivation considerably increases the number of compounds spelled open since $96.45\% > 72.67\%$. The condition that the first constituent ends in a schwa provides an interesting result regarding the proportion when no special conditions apply. This condition favours solid spelling because the greater percentage of compounds is spelled solid (45.21%) than in case when no special condition applies (27.33%). It is also conspicuous that under the condition that the first constituents of compounds are not suffixal derivations, more compounds are spelled solid than in the “neutral” case ($29.06\% > 27.33\%$). This “deviation” probably follows from the fact that condition (7b) does not exclude condition (7c) which apparently favours solid spelling. The most important difference is however between the first and the second row of the table: if the first constituent is a suffixal derivation only 3.68% of the compounds satisfying this condition are spelled solid, and this percentage falls far behind 27.33%, which represents the neutral case. It is evident that the mere fact that the first constituents are derivatives changes the outcome. The suffix *-s* in two-syllable compounds does not have the same impact.

5. Four-syllable compounds

Let us now look at the four-syllable compounds. There are only 74 compound spelled solid, and 892 compounds spelled open. The numbers of suffixes which can be added to the first constituents of compounds spelled solid and open are quite different: 9 suffixes can be added to the first constituents of compounds spelled solid, and 217 to the first constituents of compounds spelled open. All these suffixes are represented in Table 8, Table 9 and Table 10. The huge difference in the number of suffixes shows that the compounds with morphologically complex first constituents are much more readily spelled open than solid. The morphological complexity of constituents is obviously an important factor influencing the spelling of compounds.

	er	ance	craft	ery	ia	ice	ion	y	ing	s
suf1	2	1	1	1	1	1	1	1	0	0
suf2	17	0	0	0	0	0	0	0	2	1

Table 8. Suffixes in four-syllable compounds written solid.

	ing	er	ion	ment	ance	ty	age	ness	y	ity	s	ure	ation	ette	ery
suf1	95	29	12	10	8	7	6	6	6	4	4	3	3	3	3
suf2	40	107	25	10	6	1	2	0	2	1	31	3	0	1	4

Table 9. Suffixes in four-syllable compounds written open.

	ee	ice	ie	ist	al	eer	ent	hood	ings	ry	shire	ess	ions
suf1	2	2	2	2	1	1	1	1	1	1	1	0	0
suf2	0	3	0	1	5	1	3	0	0	1	0	1	3

Table 10. Suffixes in four-syllable compounds written open.

5.1 There are 9 compounds in which the first constituents are suffixal derivatives. All these compounds are presented in (20):

- (20) ambulanceman, aircraft, woman, washerwoman, whippersnapper, nurseryman, militiaman, servicewoman, companionway, assemblyman.⁸

If we eliminate the compounds with combining forms we are left only with one compound: *whippersnapper*.⁹ In open spelling, there are 224 compounds in which the first constituents are derived by suffixes. The suffixes ending in consonant clusters like *-ing*, *-ment*, *-ance* and *-ings* are well represented in open spelling, but scarcely appear in solid spelling, and, if any, only with combining forms (e.g. *ambulanceman*, *aircraftwoman*). The proportion (21) shows that an overwhelming majority of compounds with first constituent derived by suffixes are spelled open.

- (21) 1:224=0.44%:99.56%.

5.2 We now examine condition (7b), which requires that the first constituent of compounds is not a suffixal derivation. There are 74 compounds spelled solid, but in 9 of them the first constituents are derived by suffixes, and in one case the linking element *-o* appears between the constituents. Subtraction yields 64 compounds spelled solid. We have still to subtract the compounds with combining forms. There are 24 compounds including such forms:

- (22) aircraftwoman, ambulanceman, anchorperson, anchorwoman, cavalryman, clergywoman, companionway, infantryman, liveryman, militiaman, needlewoman, nurseryman, policewoman, servicewoman, underclassman, washerwoman, weatherperson, countrywoman, assemblyman, congresswoman, councilwoman, upperclassman, cameraman, clergywoman.

However, 8 such forms have already been subtracted above. The subtraction of the remaining 16 makes for 48 compounds in which the first constituents are suffixal derivatives.

The corresponding numbers for compounds spelled open are 892 and 678. We get the proportion (23):

- (23) 48:678=6.61%:93.39%.

⁸ In this list the example *speedometer* has not been included. In *speedometer* *-o-* could be classified as a linking element, and, therefore, has a quite different function than suffixes.

⁹ There are five compounds with forms *man* and *woman* spelled open: *family man*, *lollipop man*, *medicine woman*, *career woman*, and *cleaning woman*. In open spelling these forms always bear secondary accents.

5.3 Now we apply condition (9c), which refers to the compounds in which the first constituents are not suffixal derivation and end in a schwa. The counting shows that 10 compounds spelled solid satisfy this condition:

(24) butterfingers, elderberry, motherfucker, motorcycle, paperhanger, quarterfinal, quartermaster, rumourmonger, watercolour, watermelon.

116 compounds spelled open also fulfil the same conditions. Here are some examples:

(25) banner headline, chamber commerce, danger money, feather bedding, fibre optics, flower people, future perfect, labour exchange, water cannon, etc.

Now we can calculate the proportion of the compounds spelled solid and open:

(26) $10:116=7.94\%:92.06\%$.

This means that almost eight compounds in hundred are spelled solid under the given conditions.

5.4 Now we shall look at condition (9d) which requires that the first constituents which are not suffixal derivatives build consonant clusters at the morphemic boundary. Only 2 compounds spelled solid satisfy this condition:

(27) passionflower, toilettraining.

Both examples in (27) represent onset clusters.

The number of compounds spelled open which satisfy condition (7d) is much greater – 76. Some examples of these clusters are shown in (24):

(28) bank holiday, batch processing, carpet sweeper, cement mixer, distance learning, field glasses, film festival, back-seat driver, blood transfusion, inkjet printer, remand centre, etc.

Both onset and coda clusters are possible in open spelling. We calculate the proportion of compounds spelled solid and open in the same way as in earlier cases:

(29) $2:76=2.56\%:97.44\%$.

The proportion (29) shows that less than three compounds in a hundred are spelled solid under the given conditions.

5.5 We see that four-syllable compounds are spelled open in a much greater proportion than compounds with a lower number of syllables. To estimate this relation in a more exact way we shall now calculate the “neutral” case as defined by condition (7e). It is sufficient to subtract from the number of compounds in which the first constituents are not suffixal derivatives the numbers established in 4.2. and 4.3. Therefore, we get $48-10-2=36$. For the compounds spelled open we get $678-116-76=486$.

(30) $36:486=6.90\%:93.10\%$.

5.6 We can now represent the results of our analysis in Table 11.

	solid	open
no special conditions	6.90%	93.10%
first constituents are suffixal derivatives	0.44%	99.56%
first constituents are not suffixal derivations	6.61%	93.39%
first constituents end in a schwa	7.94%	92.06%
first constituents make up consonant clusters	2.56%	97.44%

Table 11. The effects of different conditions on the spelling of four-syllable compound.

We see again that suffixal derivation of the first constituents greatly favours open spelling over solid spelling since we have $0.44\% < 6.90\%$ and $99.56\% > 93.10\%$. The condition that the first constituents end in a schwa favours solid spelling of compounds because the percentage 7.94% is greater than the percentage 7.13% when this special condition does not apply. This confirms that the analysed conditions produce the similar results across different numbers of syllables, though the corresponding percentages may be different.

6. Discussion and conclusions

I have examined three factors which could influence the spelling of compounds: the number of syllables, whether or not the first constituent of compounds is a suffixal derivation and the nature of phonotactic transitions at the morpheme boundary. The nature of phonotactic transition at the morpheme boundary is further presented with two mutually excluded conditions: the first constituents may end in a schwa or it may build a consonant cluster at the morpheme boundary.

The number of syllables has already been noted as a factor which may influence the spelling of compounds, although informally identified as the “length” of compounds. The statistics of our corpus provide more precise information in this respect. Only two-syllable compounds are more often spelled solid than open; if the number of syllables is greater than two, the number of compounds spelled solid drops rapidly, so that no compound with more than six syllables is spelled solid in our corpus.

The number of syllables is also a factor that is always present: every compound must have a particular number of syllables. If the other factors are not present, we get a “neutral” case characteristic for a particular number of syllables. These “neutral” cases are presented in Table 12:

	2 syllables	3 syllables	4 syllables
solid spelling	64.27%	27.33%	6.90%
open spelling	35.73%	72.67%	93.10%

Table 12. The spelling of compounds with different number of suffixes - “neutral” cases.

When the first constituent is not a suffixal derivation, and no special phonotactic transitions at the morpheme boundary apply, only two-syllable compounds are more often spelled solid than open. If compounds have more than two syllables, the number of compounds spelled solid drops sharply.

If the first constituent is a suffixal derivative, only two-syllable compounds are more often spelled solid than open. Few three- and four-syllable compounds are in that case spelled solid. The percentages may be represented in Table 13.

	2 syllables	3 syllables	4 syllables
solid spelling	60%	3.68%	0.44%
open spelling	40%	96.32%	99.56%

Table 13. *The spelling of compounds in which the first constituent is a suffixal derivative.*

The first constituent in two-syllable compounds can be extended only by *-s*. This *-s* usually can be interpreted as the plural or as the genitive case of the first constituent, and it is often not clear whether it belongs to the underlying forms (e.g. *sales, spots*), the ambiguity which in dictionary notes surfaces with the remarks “usually” or “often” in plural. These are the circumstances which favour solid spelling. Besides, *s* at the end of consonant clusters may be interpreted as extrasyllabic. Nonetheless, in three-syllable compounds, more compounds with *-s* are spelled open than solid, and all four-syllable compounds with *-s* are spelled open.¹⁰

The spelling of compounds when the first constituent is not a suffixal derivative should show a considerable difference in respect to the spelling when the first constituent is a suffixal derivative. In fact, such a difference is obvious for three- and four-syllable compounds.

	2 syllables	3 syllables	4 syllables
solid spelling	61.39%	29.06%	6.61%
open spelling	38.61%	70.94%	93.39%

Table 14. *The spelling of compounds in which the first constituent is not a suffixal derivative.*

The difference is much smaller for two-syllable compounds (61.39%-60%=1.39%), because of the special characteristics of the suffix *-s* in these compounds.

The first constituent which is not a suffixal derivative can end in a schwa only in three- and four-syllable compounds.

	3 syllables	4 syllables
solid spelling	45.70%	7.94%
open spelling	54.30%	92.06%

Table 15. *The spelling of compounds in which the first constituent is not a suffixal derivative and ends in a schwa.*

¹⁰ There are 8 three-syllable compounds spelled open (*drinks machine, honours list, rights issue, sports centre, sports jacket, trades union* and *Brussels sprout*) and 5 solid (*thanksgiving, townspeople, tradespeople, sportswriter, codswallop*). The four-syllable compounds *futures markets, honours degree* and *sports scholarship* are all spelled open.

Both three- and four-syllable compounds are more often spelled open than solid, but the relatively high percentage of three-syllable compounds spelled solid shows that the schwa ending of the first constituent favours to a certain extent solid spelling. The percentage of these compounds spelled solid (45.70%) is higher than the percentage of solid spelling in the “neutral” case (27.33%).

If the first constituent which is not a suffixal derivative and forms a consonant cluster at the morpheme boundary, only two-syllable compounds are spelled solid rather than open.

	2 syllables	3 syllables	4 syllables
solid spelling	50.23%	22.22%	2.56%.
open spelling	49.77%	77.78%	97.44%.

Table 16. The spelling of compounds in which the first constituent is not a suffixal derivative and forms a consonant cluster at the morpheme boundary.

For three- and four-syllable compounds the number of compounds spelled solid is considerably lower than for the compounds in which the first constituent ends in a schwa. The difference between solid spelling and open spelling for two-syllable compounds reaches here the lowest level of 0.26%.

We can conclude that two-syllable compounds are under all circumstances more often spelled solid than open, although the difference is minimal if the first constituent which is not a suffixal derivative makes a consonant clusters at the morpheme boundary (Table 16). Only two-syllable compounds are more often spelled solid than open if the first constituent is a suffixal derivation. This property of two-syllabic compounds is contributed to by the specific properties of the suffix *-s*. However, slightly more two-syllable compounds are spelled solid (61.39% >60%), if the first constituent is not a suffixal derivative. This shows that *-s* at the morpheme boundaries in English compounds does not function as a linking element.

The three- and four-syllable compounds in our corpus are spelled open rather than solid under all conditions. This is especially the case if the first constituents are suffixal derivatives – the percentages of solid spelling are 3.68% and 0.44% respectively for these compounds. The consonant clusters at the morpheme boundary also lower the percentages of these compounds spelled solid – they are lower than in neutral case (22.22% < 27.33% and 2.56% < 6.90%). If the first constituents end in a schwa, theses percentages are, however, higher than in a neutral case (45.70% > 27.33% and. 7.94%>6.90%), although the difference in the case of four-syllable compounds is small (1.04%). These relations indicate the ranking of different conditions influencing the spelling. The number of syllables and the fact that the first constituent is a suffixal derivative exhibit a dominant influence on the spelling of compounds. The fact that the first constituent forms consonant clusters at the morpheme boundary also increases open spelling, while the fact that the first constituent ends in a schwa apparently works in a different direction. The impact of the last two factors is, however, subordinate to that of the other two factors mentioned first.

Bibliography

- Allen, M. 1978. *Morphological Investigations*. PhD diss., University of Connecticut.
- Bauer, L. 1998. When is a sequence of noun + noun a compound in English? *English Language and Linguistics* 2: 65–86.
- Booij, G. 2005. Compounding and derivation: evidence for Construction Morphology. In *Morphology and its demarcations*, ed. W.U. Dressler, F. Rainer, D. Kastovsky and O. Pfeiffer, 109–32. Amsterdam and Philadelphia: John Benjamins.
- COED. 2004. *Concise Oxford English Dictionary*. Eleventh Edition. Oxford: Oxford University Press.
- Collins Cobuild English Guides* (2) – Word Formation. 1991. London: Harper Collins Publishers.
- Ewen, C., and H. van der Hulst. 2001. *The Phonological Structure of Words*. An Introduction. Cambridge: Cambridge University Press.
- Francis, W.N., and H. Kučera. 1982. *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.
- Giegerich, H. 1985. *Metrical Phonology and Phonological Structure*. German and English. Cambridge: Cambridge University Press.
- – –. 1993. *English Phonology. An Introduction*. Cambridge: Cambridge University Press.
- Hay, J. 2003. *Causes and Consequences of Word Structure*. New York and London: Routledge.
- Jespersen, O. 1942. *A Modern English Grammar on Historical Principles, vol. VI: Morphology*. Copenhagen: Ejnar Munksgaard.
- LDCE. 2000. *Longman Dictionary of Contemporary English*. Second Edition. Edinburgh Gate, Harlow and Essex: Longman.
- LDCE. 2003. *Longman Dictionary of Contemporary English*. Third Edition. Edinburgh Gate, Harlow and Essex: Longman.
- Marchand, H. 1969. *The Categories and Types of Present-day English Word-Formation*. Munich: Beck.
- Miller, J. 2006. Spoken and Written English. In *Handbook of English Linguistics*, ed. B. Aarts and A. McMahon, 670–91. Oxford: Blackwell Publishing.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English language*. London, New York: Longman.
- Rakić, S. 2009. Some Observations on the Structure, Type Frequencies and Spelling of English Compounds. *SKASE* 6(1): 59–82.