

Marriage, Separation and Divorce of Words Use and Misuse of Hyphens and Spaces

The presence or absence of hyphens or spaces affects the meaning of words significantly. This holds true for many languages. While some spell checker programs, such as *TALŌ's, handle hyphens and spaces correctly, many programs don't. This article shows examples taken from several languages that illustrate the importance of correctly used or correctly omitted hyphens and spaces. Correctly handled words are shown in *italics* while incorrectly handled words appear in **bold**.

In this section examples taken from various languages are shown to illustrate the use and misuse of hyphens and spaces.

German

In his article "Als ich noch der **Klasse Sprecher** war" (If I was still the class representative)¹, Bastian Sick addresses the misuse of hyphens in German compounds. He analyses errors and although he provides some examples taken from advertisements in papers, these errors also appear frequently in all kinds of articles. Sick intentionally used the incorrect orthography in the title of his article; the correct orthography is *Klassensprecher*.

The German word *Bienenstich* (bee's sting) is written incorrectly in many newspapers and magazines². The word often shows up as **Bienen-Stich** or **Biene-Stich** and is not recognized as being incorrect by most spell checkers. A **Nazi-Jäger** is a Nazi and a hunter and not a *Nazijäger* (in English: a Nazi hunter) who is someone who hunts Nazis. A nazi criminal is a *Naziverbrecher* and not a **Nazi-Verbrecher**.

If a hyphen results in an inadmissible change of meaning the spell checker should flag the word as being an error. Many commonly used spell checkers, however, accept such errors. Another example appeared in a German advertisement: **Mädchen oder Knaben Slip € 3,99** (girl's or boy's underwear € 3,99). It means that one can get a girl for € 3,99 or boy's underwear for € 3,99. Of course, what is meant is: *Mädchen- oder Knaben-Slip € 3,99*. See also footnote ².

Dutch

Priester-dichter (a priest who is also a poet) is a compound (plural *priesters-dichters*) which should be written with a hyphen. The plural **priesterdichters** is not acceptable. The hyphen in **amateur-historicus**, however, should be omitted and *amateurhistoricus* (an amateur historian) is the correct orthography.

Many spell checkers may accept hyphens in words such as **amazone-kleed** or **Ave-Maria** while the correct spelling is: *amazonenkleed* (the dress of the mythological women) and *Ave Maria* (Hail Mary) in accordance with the official orthography of 2005.

Of course, there are words in which hyphens do have to be used, for instance in acronyms like *AMRO-bank* (not **AMRO bank**) and *Benelux-hof* (not **Beneluxhof**).

Danish

In Danish a stove is a *primusapparat* but a "motive power" is a *primus motor* and not a **primusmotor** or a **primus-motor**. Likewise, hyphens are out of place in words like **skatte-ministerium** (department of the treasury), **snack-bar**, **Sovjet-unionen**, or **spektral-analyse**. In the correct Danish orthography these hyphens are omitted.

*TALŌ is the Germanic root of our words *tell*, *tale*, and *tally*.

It also is the mark of linguistic software which has its roots in neurobiology and human factors.

¹ "Als ich noch der **Klasse Sprecher** war", Zwiebelfisch, Der Spiegel, 2 Aug. 2006, zwiebelfisch@spiegel.de. Please note the deliberately chosen speller error in **Bold**.

² "<http://www.deppenleerzeichen.de>".

Swedish

In words such as **Ordbok-artiklar, beslut-protokoll, samhäll-partiet, näringsfrihet-ombudsmannen** a genitive-s should be used instead of a hyphen. The correct spelling is *ordboksartiklar* (dictionary entry), *beslutsprotokoll* (verdict protocol), *samhällspartiet* (citizens party), *näringsfrihetsombudsmannen* (freedom of trade ombudsman). A non-smoker, however, is an *icke-rökare*. Similar examples can be found in all of the other Nordic languages.

French

While the hyphen in *assurances-vie* (life insurance) is correct, the hyphen in the following words is not: **assurances-sociales** (should be *assurances sociales* (health insurance)), **Génération-Mitterand** (should be *Génération Mitterand* (Mitterand's generation)), **Génération-perdue** (should be *Génération perdue* (lost generation)), and the Swiss word **Jeûne-fédéral** (should be *Jeûne fédéral*), a religious feast sacred to the love of the homeland (à l'amour de la patrie). And then there is also the dilemma of hyphens or spaces in proper names and geographical names: *Grande Barrière* not **Grande-Barrière**, *Grande-Bretagne* not **Grande Bretagne**, *Nouvelle-Sibérie* not **Nouvelle Sibérie**, *Nouvelle Vague* not **Nouvelle-Vague**. *TALŌ's spell checker will help a writer to pick the right combination.

Portuguese

Hyphens are essential in compounds such as *chapéu-armado* (hammerhead shark). While the plural here is *chapéus-armados* and isn't **chapéu-armados**, it is the other way around in words like *pés-d'agua* (plurality of rain i.e. shower), not **pés-d'aguas**, *pré-digestões* (predigestions), not **prés-digestões**, *mulungus-crista-de-galo* (a big tropical tree), not **mulungus-cristas-de-galo**, *negro-spirituais*, not **negros-spiritual**, *pingos-de-mel* (a kind of big fig, literally drops of honey), not **pingo-de-méis**, *pinhos-de-flandres* (Scottish pines), not **pinhos-de-flandre**. The "acordo ortográfico" adds cases such as *fim de semana* (weekend) that used to be **fim-de-semana**. In spite of being "in accordance" not all differences between Portugal and Brazil have disappeared. Spanish and Italian provide similar, and similarly confusing, examples.

British English and American English

In both versions of the English language hyphens and spaces present a dilemma. For instance: *Taxpayer* is a closed compound and correct in both British English and American English, unlike **tax payer** or **tax-payer**. An **afterdinner cup** is an error too, it should be *after-dinner cup* but if afternoon is used as a noun it has to be *afterdinner*. While in American English *makeup* is correct, it is *make-up* in British English. A *Druid stone* might be found at Stonehenge, but there will be neither a **Druid-stone** nor a **Druidstone**. An altogether different type of word is an eggcorn³, which sounds like acorn. An eggcorn sounds like a legitimate word but isn't. **Soaping-wet** (as in a **soaping-wet T-shirt**) is such a word. It has to be a *sopping-wet T-shirt*. **Soaping-wet** sounds like *sopping-wet*. Two other examples are: **coal-hearted** (*cold-hearted*) and **coal-blooded** (*cold-blooded*). The reason why these odd words are not noticed by many spell checkers is technically related to the odd processing of hyphens and spaces.

"Why" a Hyphen or Space

In a Dutch questionnaire youngsters appear to be more enthusiastic users of English loanwords, while older people try to keep native Dutch words (Telegraaf 4 feb 2009), a trend also noticed by many linguists. The TV, business commercials, and many other new developments use these kinds of words and concepts introduced from North America, including their hyphenated and spaced variants. The (mis)usage of language has increased over time and computer text processing software has not been able to keep up with this language change. The major part of these tools is still based on ideas introduced by developers from North America.

An alternative

As stated above most spelling checking programs cannot deal adequately with hyphens and spaces because they are based on the derivation of words from elementary roots. The results are very often words that do not really exist and, consequently, errors are not recognized, as is apparent from the examples shown in this article. Instead of a derivation of words from elementary roots an appropriate language-specific model is required, i.e. a language model that includes descriptions of characteristics of an individual language. This is the approach taken by *TALŌ⁴, to successfully deal with the "hyphen / no hyphen / space / no space" confusion as demonstrated in this article. An external plug-in, called "Smart Speller" is available. It is the result of a cooperation between Woodwing Software bv and *TALŌ bv.

³ "<http://eggcorns.lascribe.net/>"

⁴ "<http://www.talo.nl/>"