

Title: Ordering of capital letters and small letters in English dictionaries
Source: Michael Everson
Date: 1999-06-17

Ireland made the following comments with its negative ballot on FCD 14651:

3. *Annex D states:*

3. *The third decomposition breaks ties for quasi-homographs different only because upper-case and lower-case characters are used. This time, the tradition is well established in English and German dictionaries, where lower case always precedes upper case in homographs, while the tradition is not well established in French dictionaries, which generally use only accented capital letters for common word entries. In known French dictionaries where upper and lower case letters are mixed, the capitals generally come first, but this is not an established and stated rule, because there are numerous exceptions.*

This is, as we have said many times to SC22/WG20, incorrect. Lower case does not precede upper case in English. The concise Oxford dictionary of current English, cited in the JTC1 and CEN directives as a standard for the English language, consistently gives, in its 8th edition (1990) and its 9th edition (1998) the following:

<i>August (month)</i>	<i>May (month)</i>
<i>august (venerable)</i>	<i>may (be able)</i>
<i>March (month)</i>	<i>Polish (of Poland)</i>
<i>march (tread)</i>	<i>polish (shine)</i>
<i>Mass (ritual)</i>	
<i>mass (heap)</i>	

So for a Common Template it is advisable to use English and German traditions, if one wants to group the largest possible number of languages together.

This rationale is therefore unacceptable, as it is untrue. The reason the Common Template has smalls before capitals (which we do not prefer) is because that is what is specified in the Unicode template. This text must be revised.

Let's note here by the way that in Denmark, upper case comes before lower case, a different but well established rule. This is a second fact calling for adaptability in the model used in this standard.

This same rule is used for the English language.

Example: to have the following order: "august", "August", numbers could be assigned indicating respectively "l", "u", where "l" means lower case and "u" upper case.

This example is not sufficient. The actual syntax for ordering smalls before caps which appears in the Common Template should be repeated here, along with the actual syntax for ordering caps before smalls.

The Disposition of Comments prepared by the editor rejects these Irish comments:

Unaccepted. The statement made by Ireland is based on a specific dictionary, not on a universal English language well-documented practice. Furthermore the dictionary quoted by Ireland is one of two suggested dictionaries in ISO Directives, which is by no means presented as a standard in ISO. Both the complete edition of the Oxford English Dictionary (the most complete English dictionary) and Webster dictionaries do the opposite of what Ireland states to be a standard practice for English. SC22/WG20 has then chosen to have a template that reflects more harmonized practices among languages (German, English average -- also no rigorous practice on this in French). That said the informative annex text will nevertheless be nuanced as a compromise to say in spirit that there is no firm existing practice on this in English.

I do not think that Disposition of Comments will satisfy the Irish objection to the informative Annex, though we will have to look at the text proposed. In order to provide to SC22/WG20 with actual evidence for English “well-documented practice” I have examined 52 dictionaries, monolingual and bilingual, published in several countries, which I happen to have in my own collection. The results show that the editor’s assertion is simply incorrect. While it can be said that the ordering practice of casing is not *universal* in English dictionaries, there is nevertheless a very strong tendency to order capital letters before small letters. Of the 52 dictionaries consulted, 29 of them (55%) of them order capital letters before small letters. Only 5 of them (10%) order small letters before capital letters. The remaining 18 dictionaries (35%) have both orders, indicating most likely no specific editorial decision regarding case ordering or other parameters unrelated to deterministic ordering. Looking at the figures in another way, one can say that 90% of them do contain capital-before-small ordering, and 45% of them do contain small-before-capital ordering. The “English average” referred to by the editor is definitely *not* small letters before capital letters. Quite the opposite is true, in fact. The situation is not neutral as it appears from the editor’s assertion that there is “no firm existing practice on this in English”.

English dictionaries which order capital letters before small letters: 29.

August (month)	May (month)
august (venerable)	may (be able)
March (month)	Polish (of Poland)
march (tread)	polish (shine)

- Airlie, Maree, ed. 1996. *Collins Gem Russian dictionary: Russian-English English-Russian*. Glasgow: HarperCollins. ISBN 0-00-458652-2 (omits august)
- Akdikmen, Resuhi. 1992. *Langenscheidt's pocket Turkish dictionary: English-Turkish Turkish-English*. Maspeth: Langenscheidt. ISBN 0-245-60405-7
- Atkins, Beryl T., et al. eds. 1987. *Robert-Collins dictionnaire français-anglais anglais-français = Collins-Robert French-English English-French dictionary*. Glasgow: HarperCollins; Paris: Dictionnaires le Robert. ISBN 0-00-433451-5, ISBN 2-85036-088-0 (omits May)
- Axelksen, Jens. 1979. *Engelsk-dansk ordbog*. 10. udgave. København: Gyldendal. ISBN 87-01-33451-4
- Clark, M., & O. Thyen. 1998. *The concise Oxford-Duden German dictionary*. 2nd edition. Oxford: Oxford University Press. ISBN 0-19-864230-X
- Delaporte, Raymond. 1995. *Elementary Breton-English & English-Breton dictionary = Geriadurig brezhoneg-saozneg ha saozneg-brezhoneg*. Lesneven: Mouladurioù Hor Yezh. ISBN 2-86863-078-2 (omits august, Polish)
- Gonzalez, Mike, ed. 1993. *Collins Gem Spanish dictionary: Spanish-English English-Spanish*. 3rd edition. Glasgow: HarperCollins. ISBN 0-00-470048-1 (omits august)
- Goris, Richard, & Yukimi Okubo. 1995. *Collins-Shubun English-Japanese dictionary*. [s.l.]: HarperCollins & Shubun International. ISBN 0-00-470823-7 (omits august)
- Hawkins, Joyce M. 1996. *The Oxford school dictionary*. Oxford: Oxford University Press. ISBN 0-19-910377-1 (God/god, Miss/miss; omits August, March, May, Polish)
- Laučka, A., et al., eds. 1992. *Anglų-lietuvių kalbų žodynas = English-Lithuanian dictionary*. Vilnius: Mosklas.
- Levițchi, Leon, & Andrei Bantaș. 1971. *Dicționar englez-român*. București: Editura Științifică.
- Lewis, Henry. 1960. *Collins-Spurrell Welsh dictionary*. London and Glasgow: Collins. ISBN 0-00-433402-7 (omits Polish)
- Love, Catherine E., & Michela Clari, eds. 1989. *The Collins paperback Italian dictionary: Italian-English English-Italian*. London and Glasgow: HarperCollins. ISBN 0-00-433334-9
- Knezovič, Martin. 1993. *Kapesní anglicko-český česko-anglický slovník*. [s.l.]: Nakladatelství KPS. (omits august, Polish)
- Komac, Daša. 1996. *Angleško-slovenski slovensko-angleški slovar = English-Slovene Slovene-English dictionary*. Ljubljana: Cankarjeva založba. ISBN 86-361-0820-9 (omits Polish)
- Macleod, Iseabail, Ruth Martin, & Pauline Cairns. 1995 (1992). *The pocket Scots dictionary*. Edinburgh: Chambers. ISBN 0-550-11854-3 (not modern English, but: Mairch/mairch)
- Mac Mathúna, Séamus, & Ailbhe Ó Corráin, eds. 1995. *Collins Gem Irish dictionary: Irish-English English-Irish*. Glasgow: HarperCollins. ISBN 0-00-470753-2 (omits august)

- Nance, R. Morton. 1978 (1952-1958). *An English-Cornish and Cornish-English dictionary*. [Penzance]: Cornish Language Board. (omits august, Polish)
- Petti, Vincent, & Inger Hesslin Rider. 1994. *Norstedts engelsk-svenska ordbok*. Stockholm: Norstedts Förlag. ISBN 91-1-935472-X
- Pring, J. T. 1992 (1982). *The Oxford dictionary of Modern Greek: Greek-English and English-Greek*. Oxford: Clarendon Press. ISBN 0-19-864137-0 (omits Polish)
- Rebora, Piero. 1967. *Cassell's Italian dictionary: Italian-English English-Italian*. New York: Funk & Wagnalls. ISBN 0-308-10020-4.
- [s.n.] 1990 (1989). *The Collins pocket reference German dictionary: German-English English-German*. London and Glasgow: Collins. ISBN 0-00-433258-X (omits august)
- Stokkeland, Jostein. 1996. *Engelsk-nynorsk nynorsk-engelsk ordbok*. Oslo: Det Norske Samlaget. ISBN 82-521-3953-1
- Thompson, Della, ed. 1990. *The concise Oxford dictionary of current English*. 8th edition. Oxford: Clarendon Press. ISBN 0-19-861200-1
- Thompson, Della, ed. 1995. *The concise Oxford dictionary of current English*. 9th edition. Oxford: Clarendon Press. ISBN 0-19-861319-9
- Thorne, David A. 1992. *Collins Gem Welsh dictionary: Welsh-English English-Welsh*. Glasgow: HarperCollins. ISBN 0-00-470199-2 (omits Polish)
- Ushkevich, Alexander, & Alexandra Zezulin. 1992. *Byelorussian-English English-Byelorussian dictionary with complete phonetics*. New York: Hippocrene Books. ISBN 0-87052-114-4 (omits august, march, Polish)
- Williams, Nicholas J. A. [In press]. *English-Cornish dictionary*. Dublin: Everson Gunn Teoranta; Redruth: Agan Tavas. ISBN 1-899082-03-4
- Мюллер, В. К. (V. K. Müller). 1977. *Англо-русский словарь = English-Russian dictionary*. Москва: Русский язык.

English dictionaries which order small letters before capital letters: 5.

august (venerable)	may (be able)
August (month)	May (month)
march (tread)	polish (shine)
March (month)	Polish (of Poland)

- Kenyon, John Samuel, & Thomas Albert Knott. 1953. *A pronouncing dictionary of American English*. Springfield: G. & C. Merriam. SBN 87779-047-7
- Morris, William, ed. 1981. *The American Heritage dictionary of the English language*. Boston: Houghton Mifflin. ISBN 0-395-20360-0
- Takebayashi, S., Y. Kojima, & N. Higashi. 1995. *Kenkyusha College Lighthouse English-Japanese dictionary*. [s.l.]: Kenkyusha. ISBN 4-7674-1526-8
- Torikashvili, John J. 1992. *Hippocrene concise dictionary: Georgian-English English-Georgian dictionary*. New York: Hippocrene Books. ISBN 0-87052-121-7 (omits august, may, Polish)
- Uzicanin, Nikolina. 1996. *Hippocrene compact dictionary: Bosnian-English English-Bosnian*. New York: Hippocrene Books. ISBN 0-7818-0499-X (omits august, Polish)

English dictionaries which have mixed capital/small ordering usage: 18.

- Aulestia, Gorka, & Linda White. 1992. *Basque-English English-Basque dictionary*. Reno: University of Nevada Press. ISBN 0-87417-178-4 (march/March/march, May/may; omits august, Polish)
- Castillo, Carlos, & Otto F. Bond. 1987. *The University of Chicago Spanish dictionary*. Chicago & London: University of Chicago Press. ISBN 0-226-10402-8 (march/March, may/May, Polish/polish; omits august)
- Cowie, A. P. , & A. Evison. 1986. *Concise English-Chinese Chinese-English dictionary*. Beijing: The Commercial Press; Hong Kong: Oxford University Press. ISBN 0-19-584048-8 (March/march, may/May; omits Polish)
- de Bhaldraithe, Tomás. 1959. *English-Irish dictionary*. Baile Átha Cliath: Oifig an tSoláthair. (august/August, March/march, may/May, polish/Polish)
- Evans, H. Meurig, & W. O. Thomas. 1983. *Y geiriadur mawr: the complete Welsh-English English-Welsh dictionary*. Llandybïe: Christopher Davies; Llandysul: Gwasg Gomer. ISBN 0-7154-0543-8, ISBN 0-85088-462-4 (august/August, March/march, May/may, polish/Polish)
- Girard, Denis. 1971. *The new Cassell's French dictionary: French-English English-French*. New York: Funk & Wagnalls. (august/August, March/march, May/may, Polish/polish)
- Griffiths, Bruce, & Dafydd Glyn Jones. 1995. *Geiriadur yr Academi = The Welsh Academy English-Welsh dictionary*. Caerdydd: Gwasg Prifysgol Cymru. ISBN 0-7083-1186-5 (august/August, March/march, may/May, polish/Polish)

- Grzebienowski, Tadeusz. 1963. *Mały słownik angielsko-polski = A concise English-Polish dictionary*. Warszawa: Wiedza Powszechna. (August/august, March/march, may/May, polish/Polish)
- Grzebienowski, Tadeusz. [s.d.]. *Langenscheidt's pocket Polish dictionary: English-Polish Polish-English*. Maspeth: Langenscheidt. ISBN 0-88729-109-0. (August/august, March/march, may/May, polish/Polish)
- Hall, John R. Clark. 1962. *A concise Anglo-Saxon dictionary*. Cambridge: Cambridge University Press. (not modern English, but: angel/Angel, Bryten/bryten, Mirce/mirce)
- Hornby, A. S. 1981 (1974). *Oxford advanced learner's dictionary of current English*. Warszawa: Oxford University Press. ISBN 83-01-02448-8 (august/August, March/march, may/May, polish/Polish)
- Kidd, D. A. 1996. *Collins Gem Latin dictionary: Latin-English English-Latin*. 2nd edition. Glasgow: HarperCollins. ISBN 0-00-470048-1 (august/August, March/march, May/may; omits Polish)
- MacLennan, Malcolm. 1925. *A pronouncing and etymological dictionary of the Gaelic language: Gaelic-English English-Gaelic*. Edinburgh: John Grant. (August/august, March/march, may/May, omits Polish)
- Nöjd, Ruben, Astrid Tornberg, & Margareta Ångström. 1979. *McKay's modern English-Swedish and Swedish-English dictionary*. New York: David McKay Company. ISBN 0-679-10079-2 (august/August, March/march, May/may, Polish/polish)
- Samuelian, Thomas J. 1993. *Armenian dictionary in transliteration*. New York: Hippocrene Books. ISBN 0-7818-0207-5 (Mach/march, may/May; omits august, Polish)
- Silvet, J. 1991. *Inglise-eesti sõnaraamat = An English-Estonian dictionary*. 2 vols. 5th ed. Tallinn: Valgus. ISBN 5-440-01310-5 (august/August, March/march, may/May, polish/Polish)
- Tchomakov, Ivan. 1997. *Hippocrene compact dictionary: Bulgarian-English English-Bulgarian dictionary*. New York: Hippocrene Books. ISBN 0-7818-0535-X (omits august, march, polish but has March/march and miss/Miss)
- Wells, J. C. 1989 (1969). *Concise Esperanto and English dictionary*. Sevenoaks: Hodder and Stoughton. ISBN 0-340-27576-6 (august/August, March/march, may/May, Polish/polish)

I wish to reiterate the Irish position here. Ireland said on the ballot that one of our requirements was that:

3. The assertion that small letters ordered before capital letters is the normal practice for the English language is not made and is removed from informative annex D.

The editor has proposed to reject the Irish comment:

Given the actual practice of English dictionaries which is not monolithic, one can say that there is no standard practice in English with regard to the lexicographic order of case. If anything there are more dictionaries following German practice in English on this issue than otherwise. However the text in question will be nuanced, not indicating as before that English and German practice are the same.

On the contrary, I have shown here that, if anything, there are more dictionaries which do *not* follow German practice than otherwise. The text in informative annex D is incorrect and misleading, implying that small-before-capital ordering is, and therefore should be, the norm for English.

I have put the question to the editors at Merriam-Webster and at Oxford. They each have explicit policies. It is the editorial policy of Merriam-Webster to order small letters before capital letters.

Our policy is to put the common nouns or other generic words first, followed by proper nouns or adjectives like "Polish." This would not appear to me to be based on historical order, which is the general principle for homograph ordering. For example, in Merriam-Webster's Collegiate Dictionary, Tenth Edition, "March" comes after the regular "march" entries even though it is older.

I can only speculate on the origins of a policy that was begun long ago, to wit: The dictionary always gives priority to the lexicon as opposed to facts in the encyclopedic vein, and so common nouns and other ordinary words appear to get first billing.

My own advice as an editor as well as a lexicographer is to be consistent above all, and if I had a preference it would be to put the lowercase words first as we do.

*Thomas Pitoniak, Ph.D.
Assistant Editor
Merriam-Webster Inc.*

However, it is the editorial policy of Oxford to order capital letters before small letters.

The issue of how one should sort homographic words seems to be dealt with differently depending upon the dictionary in question. For example, the OED ignores any capitalization, and sorts firstly on part of speech (the order being noun, pronoun, adjective, verb, etc.), with each part of speech usually dealt with in a separate entry. Then, once that order is established, the sorting is done on a chronological basis, with the earliest coming first. As the OED is so large in its scope, the issue is further complicated because it has uncommon senses ('march' has 6 noun entries and 2 verb entries), so I have given a gloss beside the headword for you. These are simplified somewhat and are intended only to differentiate each from the other. So, to take your example for 'march', the OED has:

march (noun, meaning 'wild celery')
March (noun, meaning 'third month of the year')
march (noun, meaning 'some kind of boundary')
march (noun, meaning 'the track of an otter')
march (noun, meaning 'an act of moving forward; a procession, etc')
march (noun, meaning 'part of a loom')
march (verb, meaning 'to border')
march (verb, meaning 'to walk in a military manner')

And for 'polish' it has:

polish (noun, meaning 'the act of making something smooth, etc')
Polish (adj and noun, meaning 'of or relating to Poland; the Polish language')
polish (verb, meaning 'to make something smooth and shiny').

The New Oxford Dictionary of English and the Concise Oxford English Dictionary sort in a different way to the OED, but in the same way as each other, giving the capitalized headword first. These dictionaries do not give separate entries for different parts of speech:

March (noun)
march (verb and noun, meaning 'to walk in a military manner'; the act of doing this, a procession)
march (noun and verb, meaning 'a frontier; to border')

And for 'polish':

Polish (adjective and noun, meaning 'of or relating to Poland; the Polish language')
polish (verb and noun, meaning 'to make smooth and shiny'; a substance for making something smooth and shiny).

So there seems to be a tendency in what we might call 'desk dictionaries' to sort upper-case headwords first (a quick glance at other dictionaries confirms this). There doesn't appear to be any discernible reason for it, but it now seems to be an established procedure.

*Fiona McPherson
Senior Assistant Editor OED.*

I think it is clear that a deterministic algorithm is applied in the case of the "desk dictionaries", whatever the origin of the original decision. (Arguments for such a decision are given at <http://www.indigo.ie/egt/standards/capsmall.html>) For the big *Oxford English Dictionary* the explicit editorial policy is based on historical principles, that is, editorial

choice, and is not deterministic.

Ireland is adamant that the text in FCD2 is both incorrect and misleading. It does not describe the usual ordering for English as evidenced in the inventory given above. It is true that Merriam Webster and Oxford have opposing practices, and as they are the most important dictionary publishers in English it is certainly unacceptable for the text of the FCD to state what it states. Nota bene: Ireland and the UK are preparing national standards on ordering, and it is likely that they will follow Oxford practice in ordering as they do in spelling. In such a case, the text in FCD2 as it stands could be considered prejudicial to our national standards, which of course is unacceptable. The text in the annex is simply wrong. The reason capital letters precede small letters in the template is because that is the way it is in the Unicode algorithm and I don't know who knows why that is, considering that in ASCII and in all 10646 tables capital letters precede small letters.

The German tradition of ordering small letters before capital letters is not the English tradition. More dictionaries in English which have a preference sort capital letters before small letters.

We look forward to seeing what the editor has proposed in FCD3.