

# Orthographic lexicography and its aspects in online dictionary portals

**Tomislav Stojanov**

Institute of Croatian Language and Linguistics  
Republike Austrije 16, 10000 Zagreb, Croatia  
E-mail: [tstojan@ihjj.hr](mailto:tstojan@ihjj.hr)

**Key words:** orthographic lexicography; European Dictionary Portal; metalexicography

## 1. Introduction

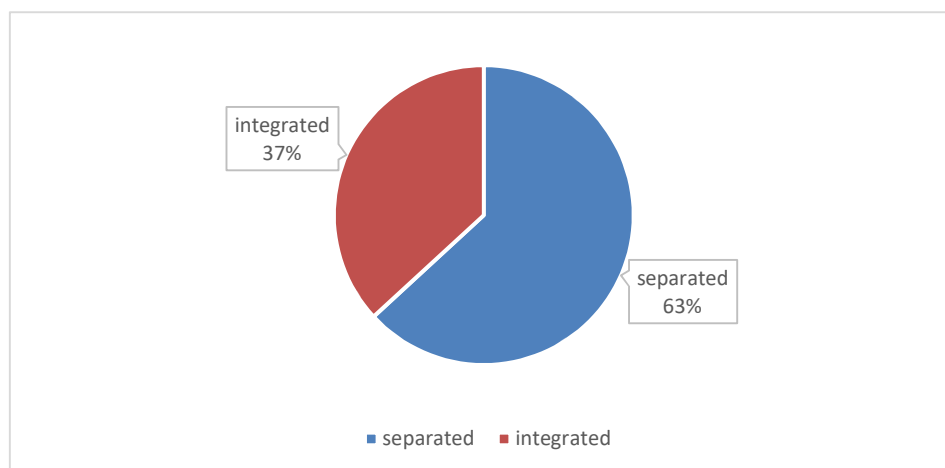
This short paper raises and answers a question related to orthographic lexicography in general and its reference to efforts in making contemporary dictionary portals. As orthographic dictionaries have not yet been researched as a specialized lexicographic variety, part of their metalexicographic description in those European languages that have online normative orthographic dictionaries is presented. Metalexicographic elements that are analyzed were chosen from the perspective of casual and professional users and online dictionary visitors. Regardless of the fact that this is a specific kind of dictionary, as well as of the fact that European orthographic tradition and practice is quite heterogeneous, the belief that the European Dictionary Portal should also include available online orthographic dictionaries is defended. An argument in favor of this could contribute to an awareness of the importance of orthography for online dictionary users, even in those languages whose written form greatly corresponds to the spoken form.

## 2. The scope of orthographic lexicography

Orthographic dictionaries have not been studied to any significant extent, nor has their online dictionary content. They are mentioned in context as a dictionary of certain information type (Svensen 2009: 35), or as “shorthand dictionaries, secretaries’ dictionaries of hard-to-spell words, or crossword puzzle dictionaries” (Burke 2003: 240). The reason they are ignored lies in the very specific, limited, and heterogeneous lexicography of Europe. However, further analysis indicates points of contact supporting the claim that a European orthographic portal (whether as part of the existing European Dictionary Portal or independently) would contribute to the spreading of the concept of European linguistic community.

European lexicography is divided into two major groups based on the criterion of the

relationship between orthography and lexicography. Thirty-eight European languages (national and minority languages) are observed in this research. The first group of languages have a tradition of developing orthographic dictionaries (ODs) separate from general scholarly ones. The second group does not distinguish orthography from general lexical data, and therefore integrates all into one dictionary. The ratio is 24 to 14, meaning that 63% of all observed languages cultivate their orthographic lexicography separate from general dictionaries.

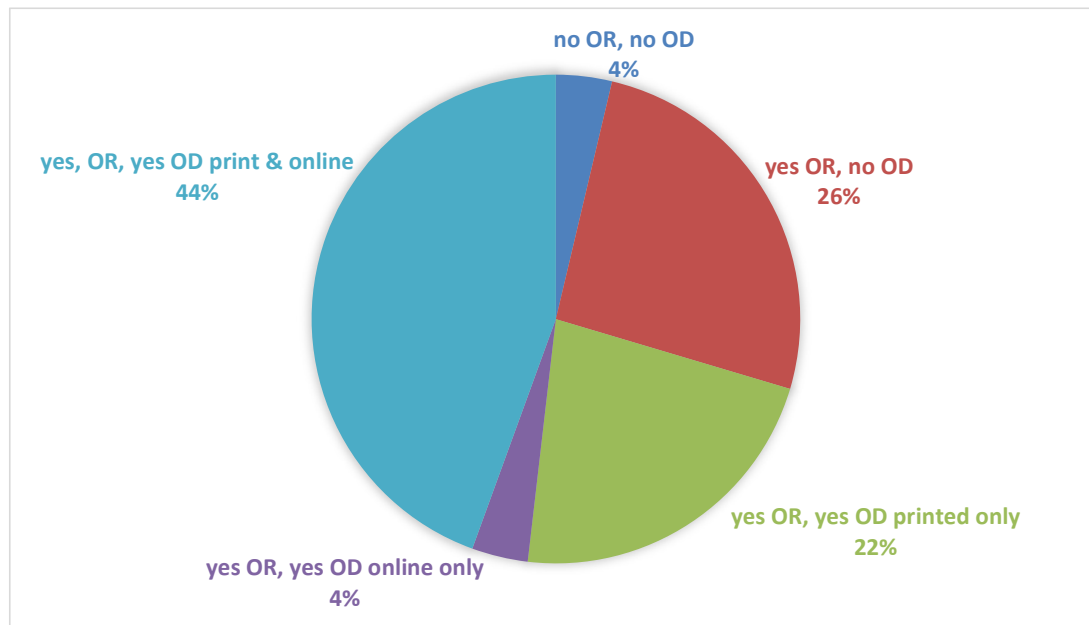


Picture 1. Separate or integrated orthographic lexicography in 38 European languages

### 3. Orthography data and the European Dictionary Portal

A more granular division of these 38 languages, which includes the additional criterion of the media in which ODs have been created, follows (the criterion of *de facto* or *de jure* regulatory body is omitted here):

1. Languages without official orthographic rules (ORs) and with official ODs: English, Welsh, Scots Gaelic.
2. Languages with official ORs and without ODs separate from a general, scholarly dictionary: Cornish, Estonian, Finnish, French, Friulian, Galician, Icelandic, Italian, Maltese, Norwegian, and Spanish.
3. Languages with official ORs and official ODs that are available only in print: Albanian, Bosnian, Bulgarian, Catalan, Hungarian, Lithuanian, and Serbian.
4. Languages with official ORs and official ODs that are available only online: Basque, Frisian, and Slovenian.
5. Languages with official ORs and official ODs that are available both in print and online: Belarus, Croatian, Czech, Danish, Dutch, German, Polish, Portuguese, Romanian, Russian, Slovak, and Swedish.



Picture 2. Distribution of orthographic dictionaries in European languages (excluding minority languages)

The great majority of languages in question (82 %) consider the dictionary a result of orthographic lexicography. Three languages only (14 %) consider them word lists (Dutch, Frisian and Scots Gaelic), while one language (4 %) considers them glossaries (German).

#### 4. Orthographic lexicography online

Nine criteria were established to compare the manner in which orthographic lexicography is cultivated: (i) whether a language has an online orthographic dictionary, (ii) whether it is publicly accessible and free of charge, (iii) whether dictionary content is searchable, (iv) whether the search engine supports advanced search operators, (v) whether search prediction technology (suggestive search) is enabled, (vi) whether results are interactive (i.e. [hyper]linked and/or hovered), (vii) whether it is possible to navigate through the dictionary, (viii) how rich the search result typography is, and (ix) whether the portal is multilingual.

The results were graded with 0 (absence of results), 1 (partial support), or 2 (full support). Higher scores mean a friendlier and richer interface.

|                    | online | access | searchability | operators | prediction | interaction | navigation | typography | multilingual | Wikipedia | score     |
|--------------------|--------|--------|---------------|-----------|------------|-------------|------------|------------|--------------|-----------|-----------|
| <i>Basque</i>      | 2      | 2      | 1             | 0         | 0          | 0           | 1          | 1          | 0            | 1         | <b>8</b>  |
| <i>Belarus</i>     | 2      | 2      | 2             | 1         | 0          | 0           | 1          | 1          | 0            | 1         | <b>10</b> |
| <i>Bosnian</i>     | 0      | 0      | 0             | 0         | 0          | 0           | 0          | 1          | 0            | 0         | <b>1</b>  |
| <i>Catalan</i>     | 0      | 0      | 0             | 0         | 0          | 0           | 0          | 1          | 0            | 0         | <b>1</b>  |
| <i>Croatian</i>    | 2      | 2      | 2             | 1         | 2          | 0           | 0          | 1          | 0            | 1         | <b>10</b> |
| <i>Czech</i>       | 2      | 2      | 0             | 0         | 2          | 0           | 1          | 2          | 0            | 1         | <b>10</b> |
| <i>Danish</i>      | 2      | 2      | 2             | 2         | 1          | 1           | 1          | 2          | 0            | 1         | <b>14</b> |
| <i>Dutch</i>       | 2      | 2      | 2             | 0         | 1          | 1           | 1          | 1          | 0            | 1         | <b>11</b> |
| <i>Frisian</i>     | 2      | 2      | 0             | 0         | 1          | 0           | 1          | 1          | 0            | 0         | <b>7</b>  |
| <i>German</i>      | 2      | 2      | 2             | 1         | 0          | 0           | 1          | 1          | 0            | 1         | <b>10</b> |
| <i>Hungarian</i>   | 0      | 0      | 0             | 0         | 0          | 0           | 0          | 1          | 0            | 1         | <b>2</b>  |
| <i>Lithuanian</i>  | 0      | 0      | 0             | 0         | 0          | 0           | 0          | 1          | 0            | 0         | <b>1</b>  |
| <i>Montenegrin</i> | 0      | 0      | 0             | 0         | 0          | 0           | 0          | 1          | 0            | 0         | <b>1</b>  |
| <i>Polish</i>      | 2      | 2      | 1             | 0         | 1          | 0           | 1          | 2          | 0            | 0         | <b>9</b>  |
| <i>Portuguese</i>  | 2      | 2      | 2             | 0         | 0          | 0           | 1          | 1          | 0            | 1         | <b>9</b>  |
| <i>Romanian</i>    | 2      | 2      | 2             | 0         | 1          | 0           | 1          | 1          | 0            | 1         | <b>10</b> |
| <i>Russian</i>     | 2      | 2      | 0             | 0         | 1          | 0           | 1          | 1          | 0            | 1         | <b>8</b>  |
| <i>Slovak</i>      | 2      | 2      | 1             | 2         | 0          | 0           | 1          | 2          | 0            | 0         | <b>10</b> |
| <i>Slovenian</i>   | 2      | 2      | 2             | 2         | 0          | 0           | 0          | 2          | 0            | 2         | <b>12</b> |
| <i>Swedish</i>     | 2      | 2      | 1             | 0         | 0          | 0           | 1          | 1          | 0            | 1         | <b>8</b>  |

Table 1. Language grade scores for online orthographic lexicography.

In the following, languages with online ODs are evaluated based on their user-friendliness.

The added sections – list of labels, glossary of terms and description of usage – have been given a grade of 1.

If there were orthography rules attached to the dictionary and if they were searchable, the grade of 2 has been given. Grade of 1 means if only one of the conditions were met. The same is with PDFs: rules published and downloadable in PDF – one point, rules with the dictionary content – 2 points. The update grade means that editors have listed all the improvements since the previous edition.

When orthography portals have a convenient URL address for direct queries, it was rated as 1.

The CSS grade stands for the quality of modern web design according to the criterion of responsiveness.

The App grade shows whether there is a software designed specifically for mobile devices in the *Apple App Store* and *Google Play* application repositories.

|                   | <i>labels</i> | <i>terms</i> | <i>usage</i> | <i>rules</i> | <i>PDFs</i> | <i>updates</i> | <i>URL</i> | <i>CSS</i> | <i>app</i> | <i>score</i> |
|-------------------|---------------|--------------|--------------|--------------|-------------|----------------|------------|------------|------------|--------------|
| <i>Basque</i>     | 1             | 0            | 0            | 1            | 2           | 0              | 1          | 1          | 0          | <b>6</b>     |
| <i>Belarus</i>    | 1             | 0            | 0            | 0            | 0           | 0              | 1          | 0          | 0          | <b>2</b>     |
| <i>Croatian</i>   | 0             | 1            | 1            | 1            | 0           | 0              | 0          | 1          | 0          | <b>4</b>     |
| <i>Czech</i>      | 1             | 0            | 1            | 2            | 0           | 0              | 0          | 0          | 0          | <b>4</b>     |
| <i>Danish</i>     | 1             | 0            | 1            | 2            | 0           | 1              | 1          | 0          | 0          | <b>6</b>     |
| <i>Dutch</i>      | 0             | 1            | 1            | 2            | 0           | 1              | 1          | 0          | 0          | <b>6</b>     |
| <i>Frisian</i>    | 0             | 0            | 0            | 2            | 0           | 0              | 1          | 0          | 0          | <b>3</b>     |
| <i>German</i>     | 1             | 0            | 0            | 2            | 2           | 0              | 1          | 0          | 0          | <b>6</b>     |
| <i>Polish</i>     | 0             | 0            | 0            | 0            | 0           | 0              | 0          | 1          | 0          | <b>1</b>     |
| <i>Portuguese</i> | 0             | 0            | 0            | 2            | 0           | 0              | 0          | 0          | 1          | <b>3</b>     |
| <i>Romanian</i>   | 0             | 0            | 0            | 0            | 0           | 0              | 1          | 1          | 0          | <b>2</b>     |
| <i>Russian</i>    | 0             | 0            | 0            | 0            | 0           | 0              | 1          | 0          | 0          | <b>1</b>     |
| <i>Slovak</i>     | 0             | 0            | 0            | 0            | 0           | 0              | 1          | 1          | 0          | <b>2</b>     |
| <i>Slovenian</i>  | 1             | 0            | 1            | 1            | 1           | 1              | 1          | 0          | 0          | <b>6</b>     |
| <i>Swedish</i>    | 0             | 0            | 0            | 0            | 0           | 0              | 0          | 0          | 1          | <b>1</b>     |

Table 2. Language grade score of metalexigraphic explanations

## 5. Conclusion

The great distribution of European languages regarding orthographic dictionaries represents the basis for further metalexigraphic and sociolinguistic study, which could possibly have implications for the European Dictionary Portal concept and structure, which is a joint research and technological goal of Working Group 1 (“Integrated Interface to European Dictionary Content”) in the e-COST Action IS1305 “European Network of e-Lexicography”.

The research of relationships between orthographic lexicography and the user experience confirmed the assumption on the high level of heterogeneity among online orthographic dictionaries, which represents an attempt to contribute to the goal of finding the best user experience.

What is the relationship between the European Dictionary Portal and the displayed results? First, 15 of the languages presented have well-developed online orthographic lexicography that could be added to the content in the European Dictionary Portal relatively easily.

Second, orthographic data are highly normative linguistic data that imply that the websites on which they are published are heavily visited and referenced. Data on visitation and the profile of visitors to the Croatian Orthographic Dictionary ([www.pravopis.hr](http://www.pravopis.hr)) support this.

Third, orthographic heterogeneity confirms two great deficiencies in European orthographic lexicography: (i) the lack of integrative factors that would raise writing above the national level, and (ii) the lack of a metalexigraphic standard for lexicographic norms.

Since European integration is moving in the direction of examining communal linguistic

policy, the unification of available European dictionary content through the EDP would influence the awareness of communal European lexicographic work and might assist the development of a communal metalexicographic framework or standard orthographic online lexicography. Instead of cooperation between European linguistic regulatory/normative bodies being carried out exclusively in the context of linguistic policy (as has been the case in e.g. the work of EFNIL), EDP could clearly show other possible forms of cooperation.

## 6. Acknowledgements

This research was conducted in April 2015 under the e-COST IS1305 Short Term Scientific Mission at the Fryske Akademy and Instituut voor Nederlandse Lexicologie. The extracted research results shown here represent data as of April 2015.

The author would like to express his gratitude to all employees and members of the Mercator European Research Centre on Multilingualism and Language Learning (<http://www.mercator-research.eu/minority-languages/database-of-experts/>), who provided him with data on European minority languages.

## 7. References

### Books:

Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press

### Book sections:

Burke, S. M. (2003). *The design of online lexicons*. In Piet van Sterkenburg (ed.) *A Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 240–249

Dziemianko, A. (2012). *On the use(fulness) of paper and electronic dictionaries*. In pp. Sylviane Granger and Magali Paquot (eds.) *Electronic Lexicography*. Oxford: Oxford University Press, pp. 319–337

### Websites:

*Dicționarul ortografic, ortoepic și morfologic al limbii române*. Accessed at: [http://www.webdex.ro/online/dictionar\\_ortografic\\_al\\_limbii\\_romane/](http://www.webdex.ro/online/dictionar_ortografic_al_limbii_romane/). (1-30 April 2015)

*Foarkarswurdlist*. Accessed at: <http://taalweb.frl/foarkarswurdlist>. (1-30 April 2015)

*Hiztegi batua*. Accessed at: [www.euskaltzaindia.eus/hiztegiatua](http://www.euskaltzaindia.eus/hiztegiatua). (1-30 April 2015)

*Internetová jazyková příručka*. Accessed at: <http://prirucka.ujc.cas.cz/>. (1-30 April 2015)

*Pravopisni rječnik*. Accessed at: <http://pravopis.hr/rjecnik/>. (1-30 April 2015)

*Retskrivningsordbogen*. Accessed at: <http://prirucka.ujc.cas.cz/>. (1-30 April 2015)

*Slovar*. Accessed at: <http://bos.zrc-sazu.si/sp2001.html>. (1-30 April 2015)

*Slovník*. Accessed at: <http://slovník.azet.sk/pravopis/>. (1-30 April 2015)

*Svenska Akademiens ordlista*. Accessed at:

[http://www.svenskaakademien.se/svenska\\_spraket/svenska\\_akademiers\\_ordlista/saol\\_pa\\_natet/ordlista](http://www.svenskaakademien.se/svenska_spraket/svenska_akademiers_ordlista/saol_pa_natet/ordlista). (1-30 April 2015)

*Vocabulário Ortográfico Comum da Língua Portuguesa*. Accessed at:

<http://voc.cplp.org/>. (1-30 April 2015)

*Wielki słownik ortograficzny*. Accessed at: <http://sjp.pwn.pl/lista>. (1-30 April 2015)

*Woordenlijst Nederlandse Taal*. Accessed at: <http://woordenlijst.org/>. (1-30 April 2015)

*Wörterverzeichnis*. Accessed at:

<http://www.korrekturen.de/regelwerk/woerterverzeichnis/>. (1-30 April 2015)

*Русский орфографический словарь*. Accessed at: <http://goo.gl/1b70IB>. (1-30 April 2015)

*Слоўнік беларускай мовы*. Accessed at: <http://www.slounik.org/sbm>. (1-30 April 2015)

### **Dictionaries:**

OCDSE: *Oxford Collocations Dictionary for Students of English*. (2009). 2nd edition. Oxford: Oxford University Press.

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>