The Use of *EDD Online 4.0* –
A Short Guide
Second edition

Manfred Markus

*English Department, University of Innsbruck, Austria*

© Manfred Markus

Innsbruck, November 2022
Note

This Guide is a “survival kit” that focuses on the most essential search routines of EDD Online 4.0. For a more detailed analysis of the internet platform and an in-depth discussion of its research potential, use the monograph Markus 2021. Only this Guide, however, is based on version 4.0 of the interface. The main modifications and improvements in this version will be mentioned and explained on the following 30-plus pages.

Acknowledgement

I am most obliged to the Austrian Science Fund for granting four successive projects between 2006 and 2022, as well as to the University of Innsbruck for their administrative support. The version of EDD Online here presented (4.0.) is mainly the result of the last three EDD-related projects (2011-16, 2017-18 and 2021-22), but they would not have been successful without the preparatory work during the first phase (2006-10). I would, therefore, like to express my sincere thanks to all the team members over the last sixteen years: Prof. Alexander Onysko, Raphael Unterweger, Dr. Reinhard Heuberger, Mag. Christof Praxmarer, Mag. Christian Peer, Dr. Emil Chamson, Stefan Giuliani, Mag. Christian Stenico, Daniela Jänsch, Anna-Maria Waldner, Mag.a Regina Seiwald, Dr. Thomas Burch (U of Trier), Dr. Hans-Werner Bartz (U of Trier), Prof. Werner Wegstein (U Würzburg) and Dr. Joseph Wang. For EDD Online 4.0, carried out over the last twelve months, I had the strong support of the previous programmers Mag. Joachim Masser and Martin Köll, as well as my philological helper Mag.a Andrea Krapf. For aspects of promoting the EDD Online project, Dr. Robert Spindler joined our team in 2022.
1. Introduction: Basic search for strings in headwords

What can you search for on the interface?

Figure 1 shows us the basic ("simple") version of the interface.

![Diagram of the interface](image)

**Figure 1**: Query for headword with activated sub-filter humour (in filter usage labels)

Clicking "headword" (top left), or rather, leaving this default mode as it is, means working in the "simple" mode, which allows for the basic retrieval of any headword (*) or a specific headword to be typed into the search box (top left). The result is an alphabetical list of headwords in accordance with any of the filters or sub-filters, optionally activated on the right and documented by the search protocol (top right). For example, the string humour has been triggered by opting for one of the eight display filters (top right). These filters can be activated in isolation or combination, with the protocol always giving evidence of the possibly complex arrangement of filters.

The general structure of the interface is as follows: As Figure 1 shows, the vertical dividing line keeps the “retrieval window” on the left apart from the “entry window” on the right. The light green zones above the two windows contain the “search parameters” (on the left) and the “filters” (on the right). Some of the filters have “sub-filters”; for example, the filter dialect areas has three sub-filters: county, region, and nation. The dark bar below the filters keeps three further tools at the users’ disposal: map, image, and context. They will be discussed in due course below.
The options of combining parameters (such as headword) on the left and filters (or sub-filters) on the right are restricted according to rational or linguistic criteria of compatibility. As we were searching in Figure 1 for a pragmatic usage label (humour), it would not make much sense to be simultaneously looking for sources because they refer to different parts of entries than the usage labels. The sources mainly refer to the quotations provided by the EDD to give evidence. To mark incompatible search criteria (both filters and parameters), we have made unacceptable buttons decidedly more visible than earlier by adding a red semi-frame.

2. Retrieval-window (simple mode)

Figure 1 also demonstrates that the options in the left half of the interface are the following (beginning top left):

2.1. Search

In the search box, users may put in strings (irrespective of capitalisation), either as such (e.g. house means immanent truncation on either side) or explicitly truncated at the beginning or end of a string (*house or house*). If truncation is to be excluded, the search string has to be given in double quotation marks ("house"). If users want to search for no particular string, but for all strings fulfilling the demands of activated filters, they just type in an asterisk (*).

2.2. Last result (within search box)

This button allows piggy-back queries, i.e. searches within the sub-set of the findings of a previous search, thus encouraging searches of greater complexity or "second-thought" searches.

Note that the last result button only allows second-thought queries within the set of headwords (or rather: their entries) previously found, so that a change of the parameter (for example, to compounds) only functions within that set frame. On the activation of new filters, users should, in their reference back to the entries previously found, type in an asterisk * in the search box, thus making sure that all these entries of the previous search are now at their disposal. If they by mistake leave the search string of the first search in the search box, the last-result search will lose the matches of the first search.

The last-result button is a very useful tool in cases when a search is likely to produce an unacceptably high number of retrievals or is based on too many simultaneous filters. Users should then proceed step by step. The last-result button can also repeatedly be used and, thus, paves the way to more complex queries. For example, a combination of "horse" and "stop" as defining terms, paired with the filter part of speech (= interjection) and with the filter dialect area (all English counties), provides a list of nine interjections addressed to horses, mostly with the meaning "stop!", and relates them to dialect areas.
2.3. Go (the orange button)
Starts the query and produces a list of up to 20,000 matches, plus the parameters or items of filters the user has asked for. The limit of 20,000 has been necessary to keep our server from “revolting”. However, if more matches are available, they are, in addition to the reproduced ones, also counted (up to 100,000), with the frequency number appearing above the retrieval list. We have implemented this function in view of the probably mainly quantitative, rather than qualitative, interest of users in hyper long lists of findings.

2.4. Clear
Deletes whatever has been searched for or previously retrieved. This deletion includes maps, entry images and whatever else has been previously activated in the filters.

2.5. Simple/advanced
The activated mode appears in bold. As an alternative to simple, advanced opens a new window with possible queries that are more specific than those for the headwords, for example, for definitions (i.e. semantically interesting strings) or for phrases (idioms).

2.6. Headword
This standard parameter opens the basic mode of searching for headwords (or "lemmas"). Note that some of Wright's headwords are not lexemes in the modern linguistic sense, but phonemes (very rarely, particularly at the beginning of the letter A, i.e. when Wright and his team were still experimenting with what to include), bound morphemes, derivations, compounds, combinations, and phrases. Almost all headwords are accompanied by their part of speech, for example, GO, v. – the other headwords that are not are variants accompanied by cross references, as in the entry BED-HOUSE, where the information “see BEAD-HOUSE” is added.

2.7. Full text
This mode is recommended mainly for tentative queries, for example, when users wish to know whether a string or filter keyword exists at all within the Dictionary. This mode only reproduces the search strings without the entries they belong to. It does not allow any combination with any other parameter or with a filter. As with the headwords, this mode allows truncation. However, unlike headword searches, it allows for case-sensitivity.

2.8. Original result
This box (see Figure 1) by default presents the matches of a search in the original order of the Dictionary. With headword retrievals, the “original” order is probably the basic and most important mode. But various other sorting modes are also
available in this box. For example, in searches for compounds the "original" mode is that of the alphabetised headwords, whereas the user will probably prefer an alphabetical order of the compounds themselves, taken out of their headword and entry context. There are further sorting modes that isolate the retrieved "columns" and quantify the tokens of the types found in a query. Moreover, sensible combinations are on offer, for example, column 2 with 3 and column 3 with 2, where "column" always refers to the hierarchy level of the parameters and filters involved in a query. The number and type of the sorting options depends on the complexity of a combined query. Figure 2 demonstrates one of the options available, column 3 with

2, based on a query for the string house as part of compounds, with the filter dialect areas (sub-filter: all English counties) combined:

Figure 2: Column 3 with 2 sorting mode for compounds with house in all English counties, with all sorting options opened for demonstration

2.9. Reverse box

The Reverse box on the right-hand side of the retrieval half of the interface, blocked in the case of Figure 2 owing to the specific sorting mode, generally permits a reverse order mode, with findings arranged alphabetically as seen from their word ends (which may be of interest for suffix or rhyme-word studies).
3. **Entry window**

The *entry window* is what you get on the right-hand side of the interface when you click on a retrieved result in the list of retrievals on the left (as shown in Figure 1 earlier).

3.1. **Survey**

The options of the right half of the interface (above the entry window) are the following:

**search protocol**: This wide box documents all steps of queries, including the advanced query routines and the filters (excluding sub-filters, however). This tool is particularly useful in the case of complex queries, with different filters involved.

The box of the protocol is often too small to show all the parameters and filters selected (e.g. *counties*), but when the user moves the mouse into the protocol box, the complete list of the filters pops up (Figure 3).

![Figure 3: Pop-up window of the search protocol, with all English counties selected](image)

At the right end of the protocol box there is a memory function – an arrow for recalling previous search commands.

The big group of the **search filters** allows for (in the order listed in the interface from left to right): (1) *dialect areas*, (2) *usage labels*, (3) *parts of speech*, (4) *sources*, (5) *phonetics*, (6) *morphemics*, (7) *etymology*, and (7) *time spans*. As mentioned earlier, some of the filters have a small red semi-frame – these are the filters that are **not**
acceptable in the case of a given parameter (in Figure 2: compounds). All the other filters are compatible with both each other and the given parameter.

In the following, we will discuss all the filters, irrespective of their acceptability in special cases, in detail.

3.2. Dialect area

"Area" is the cover term for "county", "region" and "nation". The counties, such as Yorkshire (Yks.), are arranged nation-wise, with England followed by (the whole of) Ireland, Scotland, Wales, Australia, Canada, and the USA. The counties listed also include individually named sub-sections, in particular, cities, and, occasionally, rivers (to be activated by ticking part), and also fuzzy references, such as "some parts of x" or "west of x in y". Moreover, some of the precise references to counties are specified by directions of the compass: north, southeast, etc. In the headline of the dialect filter we have, therefore, provided four sub-filters: prec(ise), part, fuzzy and directions (see Figure 4).

Figure 4: Interface with activated dialect filter "Yorkshire" and the activated sub-filter "directions"

Figure 4, in line with the activated sub-filter, demonstrates the retrieval of only those Yorkshire references that are specified by direction (e/w. etc.). The default mode for the use of the four sub-filters, however, is to have them all included. If users want to
limit their query to the precise data of dialect attribution, they have to keep prec marked and cancel the other three options.

The relationship between the three area types (county, region and nation) and the four modes of precision/sub-filters (prec, fuzzy, part, and directions) is such that queries of all areas theoretically allow a combination with all four precision modes. However, as a matter of fact, region and nation mainly produce precise or fuzzy results, but no outputs for both part and directions.

The second area-group, regions, such as w.Cy. (for 'West Country'), is generally structured in the same way as that of the counties, with England, Ireland etc. on top. But the third area group, nations (such as USA), is much shorter than the county lists and, therefore, simply arranged alphabetically, without any sub-classification. Note, however, that after the thirteen “nations”, from Australia to the West Indies, another group, Colonial, had to be offered for occasional references in the EDD to the West Indies and other 19th-century colonies of the UK. The West Indies are also listed under regions as the only sub-division of Colonial.

All three types of areas – counties, regions and nations – can be combined with each other, by either "AND" or "OR" in the Boolean sense. This is allowed through the respective headline of the entry window (see Figure 5), which shows the result of a query for headword in both Ireland and the US (Boolean AND).

Figure 5: Combination of dialect areas (USA AND Ireland) for any headwords
Note that the OR- versus AND-option generally refers to different levels at the same time: (1) the sub-filters (nation/region/county) in relation to each other; (2) the keywords within these sub-filters, no matter which sub-section of the sub-filters they belong to, so that Yorkshire can be combined with Edinburgh, but also with any region or non-UK nation. The only limitation is that OR (like AND) is always valid for all activated features alike. A combination of the type (Ireland)\textbf{ AND (USA OR Amer.)} is not possible in this mode, but can be achieved with the help of the last-result button (mentioned earlier, 2.2.). \textit{Qua} filter, however, \textit{dialect areas} – in searches for headwords – can be combined with any other filter, such as \textit{parts of speech or usage labels}, where the implicit logic for the filters is always Boolean AND, whereas the sub-filters and keywords can be related by either OR or by AND.

In addition to the Boolean options OR and AND, we have also implemented ONLY. This option helps to retrieve the areal specificity of certain lexemes or features. The button allows for more than one keyword at a time, but the retrieval list will keep the area-specific results apart. The ONLY option is an excellent tool for creating valid county glossaries.

Users not at all interested in specific dialect areas, but, for example, in figurative language use or in flora, may wish to cover all dialects at a time. For such users the headline button select all can be activated, both on the general level and on the specific level of either nations or regions or counties.

Dialect areas concerned by a particular query can be visualised by the map-button (on the right of the entry window). This triggers off either a map of the UK or of the world, depending on which areas have been found in the retrieval window. When users click on a headword, all the areas listed under that headword are shown on the map. If, however, users have searched for, say, \textit{compounds} and also selected \textit{dialect area}, the entry window will show three columns, the headwords in the left column, the compounds in the middle, and the dialect areas in the right column. In this case the map will selectively show all the dialect areas attributed to the \textbf{compound} of interest. Note that the map of dialect distribution will only be produced according to the quantifications of the column 2 counted sorting mode.

In addition to the normal outline maps of the counties and of other regional subdivisions, a physical map and a hybrid map are available (for the UK). The zoom-box, which, together with the outline map, is demonstrated by Figure 6, is a smart-tag device that pops up according to the cursor’s position on the map. It also provides both a county’s abbreviation and its full name.
3.3. **Usage labels**

The search filter *usage labels* reflects the EDD's abundant application of usage information. Such labels are frequent not only in occurrence (as tokens), but also in type: *frequency, reliability, semantics, pragmatics, phonology, prosody, morphology, and syntax* (see Figure 7).
Figure 7: The sub-filters of the filter usage labels, with the keywords of the sub-filter frequency opened.

To keep the lists of options, for example, of the markers of pragmatic relevance, in moderate length, we have normalised the terms and abbreviations of the Dictionary so that sometimes half a dozen or more strings (such as emphasis, emph., emphatic, emphatically, emphasise, stressed, stress, highly stressed etc.) were reduced to just one (emphatic). This principle was generally applied whenever necessary. For example, the list of semantic features presented to the users of our interface is also a normalised one, in line with modern linguistic terminology.

3.4. **Parts of speech**

In this menu, a list of eighteen word classes, from adjective to verb, is available. "Hybrids" of word classes, however, have been split: in "verbal noun" and "adverbiaal adjective", for example, noun and adjective have been classified as word classes, but the specifying attributes have been marked by the tag grammar and can be traced via the usage labels in the sub-filter syntax.

To illustrate, Figure 8 shows the surprisingly long list of interjections (745 matches) after a search for them without any further specification.
3.5. Sources

Four types of sources have been kept apart, generally in line with Wright's own practice (Figure 9).
Figure 9: Sub-filters of *sources*, with the first source reference opened and highlighted (whereas the background is faded)

The four options of sources are *dialectal (selection)*, *literary (selection)*, *dial. and lit. (complete)*, and *unprinted*. The third (*complete*) option is difficult to access due to its enormous length and also its generally non-transparent abbreviations. It is a merger of the first two lists, the dialectal and the literary one, but it also lists many other titles, names and abbreviations for which the attribution to either of the two groups would have been questionable and was difficult or impossible for us to achieve. Users may wish to identify authors or titles from this comprehensive list, or they may consider the selected lists, which are about a fifth or less of the whole, as a "survival kit" for retrieving dialectal secondary titles, on the one hand, and "literature", on the other.

Both the dialectal and the literary selection of titles have been considerably revised for EDD 4.0. The dialectal list includes all authors and/or titles quoted by Wright at least two hundred times. With the literary sources, the same principle of selection has prevailed, but we have also included a few titles that seemed of interest owing to the author’s or work’s reputation. The *select-all* option, now available for all subtypes of sources, has newly been implemented to allow for comparative source studies. Given the large number of source titles, particularly in the *complete* list, users of the *select-all* options should reckon with longer times of retrieval.

The fourth type of source references is *unprinted*. This is a complete list of the full names of various helpers who contributed to Wright’s project in the form of letters or slips sent to him or by word of mouth. Users can select individual names of contributors that we have been able to identify, or they can summarily opt for the unidentified abbreviations of contributors/correspondents.

As Figure 9 also shows, the printed titles can be smart-tagged, thus presenting the full title of a book or several titles of books of the listed author. This transparency of what a name or title stands for was also implemented for all titles of the complete list, but, unlike the selections, the keywords on that list were not regularised or normalised. They were edited up to a point (the blue colour stands for our additions or considerable changes/corrections to Wright’s bibliographical information in his references), but they essentially come in the shape of the first and usually most common abbreviation used in the *EDD*. Users interested in all these – often inconsistent and sometimes non-transparent – abbreviations and what we decided they stand for may open the Excel-table offered at the very top of our interface (*EDD References*). They can find two tables there, one for printed and the other for unprinted sources. The extremely long table of printed sources, at its end, includes the source abbreviations used in the Supplement; they are added to our Excel table as a separate block in green fonts. They are new abbreviations standing for new sources, many of them from the early 1900s.
Despite its limits, the present bibliography is much more comprehensive and transparent than the book version, and given the great share of titles filled in by us (about a third of all titles), we can boast of having considerably improved the quality of the EDD bibliography by our editorial work (though there is still space for further source identification).

3.6. **Phonetics**

Wright's phonetic transcription is very similar to the IPA transcription (International Phonetic Association), which was to become standard in the 20th century, the main difference being that word accent is marked in the EDD by a high dot after the stressed syllable of a word, rather than by a stroke before it. Special symbols of transcription, as well as diacritics, can be produced by users with the help of the keyboard which opens when users click on the filter *phonetic*. The order of the special characters and diacritics on this keyboard is simple: the keyboard first lists the vowels from *a* and its "derivatives" (such as the ash-ligature), then the “variants” of *e, i, o* and *u*; in a final line, it provides consonants and diacritics, such as the raised dot.

The keyboard is also meant to be used for the production of special characters in whatever context, i.e. outside the search for phonetic transcriptions, for example, in full-text queries. As Figure 10 shows, this is the mode offered first ("keyboard use only"). It is, in fact, the default mode.

The search in Figure 10 was triggered by the activation of the now-blue button on the right bottom. This search function only includes signs or strings that are marked in the text as phonetic transcriptions, that is, by square brackets. Note that “full text” also appears as a main parameter in contrast to headword (top left in Figure 10); searches on this “channel” include every single sign or string in a retrieval, in our case the Greek letter χ, irrespective of its possibly non-phonetic function.
3.7. Morphemic

This filter permits the opening of a limited list of common bound suffixes and prefixes, such as *-ing and *be-. Since word compositions with at least one bound morpheme are, by definition, derivations, the morphemic query automatically refers, on the one hand, to the parameter of derivations (in the expanded mode), and, on the other, to headwords, which are included in this query because entries themselves are occasionally derivations. In other words, morphemic is simply an icon implemented for the sake of convenience. You select a prefix or suffix, which is then automatically copied into the query box and applied on both headwords and derivations. Needless to say, users may also search for bound morphemes of their choice that are not listed in the morphemic selection. But then they have to search twice: once for headwords and then for derivations.

The selective lists of affixes have been limited to morphemes relevant to word formation. Grammatical morphemes, such as the third person singular *-s, have been excluded.

The example of a morphemic filter in Figure 11 provides all strings ending with *able and attested to all English counties (132 items).

![Figure 11: *able as a morphemic query feature, with beginning of retrieval list and the sample Unkennable opened](image)

3.8. Etymology

In the "comments" of the entries (see 4.5), the EDD keeps referring to etymological roots. The etymology filter not only often provides the main earlier languages that Late Modern English dialect is based on, such as Old English and Norwegian, but allows access to many cognates, such as from Gothic and Low German (cf. Chamson 2014). In a query, the filter provides the languages searched for, the etyma/words of these languages and, as usual, the headwords involved. Figure 12 shows the 185 matches of a search for Dutch etymologies, combined with the filter of all county dialects of England. The map – I have selected the relative per mille option – shows the different degrees of Dutch influence in the English counties.
Figure 12: Search for dialect words in English counties affiliated with Dutch etymology (incl. Old and Middle Dutch and Flemish), plus distribution map with regard to counties

The map reveals the relatively strong impact of Dutch on dialect in Kent and along the North Sea coast up to Northumberland and along the Scottish border. It is based on the county-specific figures visible in the retrieval window after they have gone through a process of being normalised in relation to the sum total of references to the counties concerned, irrespective of the additional reference to the Dutch/Flemish etymological background.

When areas outside the UK are concerned, a world map pops up instead of the more usual map of the British Isles. Note that there are also two alternative options of maps in addition to the "political" (i.e. outline) map, a physical map and a hybrid one. Whichever option is preferred, one can always click on an area-specific abbreviation, for example, Yks., to see on the map where it is, what the abbreviation stands for and what the numerical basis of the attributed colour is.

The etymological filter also allows searches for specific etyma, including possibly morphemic strings as well as graphemes, such as $<\text{æ}>$ and $<\text{ȝ}>$, with both strings and languages provided. The sorting routine offers the option of getting the findings in a language-specific order so that all Old English etyma, for example, are found side by side.
The strings searched for have to be typed into the general query box. Figure 13 illustrates this option with the string <æ> by showing an extract of the screen.

**Figure 13:** Words with ash-ligature arranged language-wise (filter etymology)

### 3.9. Time spans

The *EDD* does not offer any direct information on the time of usage of a certain form or word, but only the dates of the sources provided for evidence (in the case of written sources). These source references, of course, do not equally cover all periods of the English language, given that the *EDD*, unlike the *OED*, is not primarily a historical dictionary. Nevertheless, the generally large number of source references allows for valid conclusions concerning the time of usage.

Our pool of dates is only one of years; months and other more precise information had to be ignored. Circumscriptive time references, however, of the type "in the first quarter of the 19th century", have been normalised and transformed into clear numbers of years or of spans of years (in this case, to 1801-1825).

Users can activate either of two modes: *time span* or *exact or truncated year* (see Figure 14).
Figure 14: The two modes of time reference

In the first mode, users may fill in two years to mark the span of time they are interested in, such as 1840 to 1849, and thus produce a list of the headwords whose time of popularity, i.e. the phase of publication dates, overlaps at least at one point in time with the searched time span. Entries are excluded in retrievals if their spans of time do not overlap with the spans searched for.

The query result of mode 1 is illustrated in Figure 15, with the entry-specific spans of time in the retrieval window on the left and the first entry opened.

Figure 15: Search for time span (= mode 1), with example of entry ABRICOCK
The second mode of the time filter (see again Figure 14) is meant for temporal close-up studies. Here, searches for time do not refer to entries as a whole (as in mode 1), but to their sub-sections, for example, *compounds* or *variants*. This higher degree of referential precision is, of course, an advantage. However, if users type in the year number "1840", only exact hits are retrieved, unfairly excluding cases with sources from, for example, 1839 and 1841. Therefore, this pinning down of dates to a single year also has a clear disadvantage. It could, however, be of interest, for example, for users focusing on a particular year, such as the last years covered by Wright, 1903 and 1904. Moreover, truncation is possible: "190*" means "1900" to "1905" (there is no later year). By the same token, a search for “18*” in this mode includes the whole of the 19th century and retrieves the exact years concerned.

### 4. Retrieval window (general tools and advanced mode)

#### 4.1. General mode of highlighting and use of colour

In the retrieval window, on the left of the interface, the entries are always automatically added (in capital letters). The strings of the query appear in boldface. The clicked and, thus, opened element (headword, or whatever) is marked by a vertical stroke to the left of that element. In the entry window (on the right) the elements searched for are highlighted. Here the main units of the entry come in different colours: the headwords in blue, the parts of speech in red, etc. The text is presented without the line-breaking of words by hyphens and other layout details that are liable to "confuse" the computer, but with the (frequent) complemented units of complex lexemes that were abbreviated in the original by hyphens (and are now marked by a light grey).

An alternative mode of presentation is that of the original images of the *EDD* entries, as found and reproduced from the paper version, in black and white. Users, including my own team, can, thus, always check the correctness of our editing. Figure 16 shows us this basic marking policy for PENT-HOUSE, one of the 53 findings in a query for *house*. 
Figure 16: Text of the entry for PENTHOUSE with image of the EDD paper version

Figure 16 is still based on a query within the simple mode for headwords. Boldfacing and highlighting are, however, always the same, no matter what a query refers to.

4.2. Survey: retrieval window

With a click on advanced eight optional parameters are provided. All these have a significant share in the EDD’s content, but they are somewhat hidden within entries, unlike in most modern dictionaries, where many of these types, for example compounds, are lemmatised.

The sub-menu of the advanced button is given in Figure 17.
As one can see in Figure 17, the advanced mode allows for access to eight parameters.

Before we start discussing these in detail, we should raise the general question as to which parameters can be combined with each other and how they can be combined with filters. The answer is given by the red frames surrounding unacceptable options. Figure 18 demonstrates the principle.

![Figure 18: Marking acceptable and unacceptable combinations of parameters and filters](image)

Figure 18 demonstrates that the search for compounds also allows for a simultaneous search for derivations and combinations but excludes pairings with all other parameters. As to the filters, only five of them can be combined with the parameters just focused on. While the inclusion of phrases would also have made sense here from the philological point of view, our programmers objected in this case (and a few other cases) because the structure of phrases was incompatible with the XQuery routine applied for compounds.  

4.3. Definitions

The first parameter, definitions, restricts a query to those sections of entries that topicalize the meaning of lemmas or word formations in them. This mode offers itself as a basis for studies of semantic fields. For example, users may search for girl, at the same time activating the dialect-button (for Scottish counties). They will then get 99 strings girl in the definition block, plus the dialects attributed to them (see Figure 19).

---

1 X-Query was our search language for writing our search commands.
Unlike in Version 3.0 of *EDD Online*, there is now a direct relationship between retrieved search strings and dialect areas, which is made transparent to the user by the addition of the *numerus currens* figures in the list of retrievals: the only occurrences of a search string retrieved are those that share the same counted section of an entry with the requested dialect areas. In the case of BIRD, this applies to the abbreviation for Ayrshire (Ayr.) in section 1 of the entry. As regards any of the other filters, they can also be added within the general limits, i.e. usage labels, sources, and etymology can likewise be opted for with terms of definitions.

The parameter *definitions* has also fundamentally been revised so that it now allows searches not only for strings, but also for complete texts, i.e. of whole definitions. Users may try out this additional query routine; however, when searching for * in definitions, they should avoid overburdening the system by activating filters or by limiting the search in the first place. For example, they may retrieve the entry for BIRD (as a headword) and then, with the help of the *last-result* button, search for * in definitions by adding all possible dialect areas. The result, particularly in the column-2 counted mode, will allow them to draw interesting conclusions as to the semantic productivity of the word *bird* in dialect.

4.4. **Citations**

*Citations* triggers searches for strings of texts quoted by Wright from sources. The syntax in these blocks of *citations* is the most unpredictable part of the entries. This
comes as no surprise because the "style sheet" of the often-abbreviated references variably depends on the authors quoted. The citations are, thus, a mixture of illustrative and explanatory text. Only the first type can provide evidence of authentic dialect text.

Two tools are provided for users to split up the two types of text. One is the option of a kwic-concordance routine, which provides the immediate neighbourhood of a string in the text (kwic = "key word in context"). The concordance can be activated with a click on citations followed by another click on with concordance in the pop-up window. The default mode is without concordance because the production of the concordance takes extra time.

Figure 20 demonstrates the functioning of the parameter of citations, with *house* as a search string and the concordance mode activated.

As Figure 20 reveals in the small window of a drop-down menu, additional sorting options are provided, allowing for users to arrange the retrieval list according to the first or second slot, either to the left or to the right. Moreover, users can, of course, opt for any of the filters compatible with generating a concordance, for example, by selecting a certain dialect area. This offers them a clearer picture of dialect keywords in context.

The second tool for creating context, new in EDD Online 4.0, can provide complete quotation texts.
In Figure 21 I have selected Aberdeenshire. The retrieved passages of text are likely to include many Scotticisms from Aberdeenshire. The presentation mode is still that of the original result, with the quotations correlated with the headwords. In order to get a small Aberdeenshire corpus, users can change the sorting mode to column 2 counted; more than 7,000 quotations are then alphabetically arranged.

4.5. Comments

The comments attached by Wright at the end of entries are, from the point of view of dialectology, of secondary importance. They occasionally refer to dialects, particularly to American and other overseas English, but mainly topicalize literature (including Shakespeare, Chaucer and medieval anonymous works), as well as the history of English before the 18th century and etymology. They also often refer to affixes. Moreover, they contain many labels.

Comments generally refer to entries as a whole, so that the more general search for headwords seems, by and large, more advisable, as such a search would cover the whole entry. All relevant information concerning strings in the comments can be retrieved this way.

Searches for complete texts (*) in comments, however, are a different matter. Similar to the built-in function of providing coherent citations (see 4.4), we have now implemented the option of retrieving the complete comments. Combined with any of the four filters that are accessible, for example, Dutch OR Flemish as an etymological
marker, the production of the comments concerned can be seen as a treasure trove – as Figure 22 may illustrate.

4.6. Variants

Most of the variants that have been tagged by us as such are – as in the Oxford English Dictionary – phonological and spelling variants, but some are also generally lexical or semantic "variants", i.e. synonyms or homophones/polysemes. As regards filters, variants can only be combined with dialect areas. The compatibility of other filters in EDD Online 3.0 has been cancelled because their correlation with variants turned out to be misleading: the other filters practically always refer to units in an entry other than the variants.

Figure 23 shows a search for variants beginning with a y-, combined with the dialect filter (all English counties).
My query in Figure 23 is motivated by the role of so-called *j*-insertion (with \(j\) represented by <y>), which I examined in Markus 2011. This dialectal deviation from the English standard, a phenomenon complementary to *h*-dropping, was common practice in the LModE period. Figure 23, after the result’s rearrangement in the *column-2* mode, would show the distribution of *y*-variants, such as *yalhoose* for ‘ale-house’, in English counties.

**4.7. Derivations**

While Wright's concept of "derivations" was surprisingly correct, there have occasionally been contestable cases where the "bound" quality of a prefix or suffix, the prerequisite of a derivation, may be questioned. In the Innsbruck project, we did not want to investigate this issue theoretically, but usually followed Wright, who, in the *EDD*, generally introduces derivations by the introductory marker *hence*. In rare cases, however, *hence* has also been used for the introduction of other types of word formation, for example, “combinations”. In our work this was no big problem as long as the types were clearly kept apart. But occasionally the *EDD* provides different groups of word formation under the heading "hence", or under some other heading.
Such cases of inconsistency took us extra time to disentangle, but our aim has been to keep the different types of word formation apart. Figure 24 demonstrates searches within the parameter *derivations*.

Figure 24: Search for *derivations* ending in -ee

There is an output in Figure 24 of 14 matches. Generally speaking, the parameter *derivations* permits a combination with all filters except *phonetics* and *morphemics*. The *dialect* filter is automatically activated.

Users interested in derivations may, however, make use of the filter *morphemics*, mentioned earlier, which offers a selective list of the most common bound pre- and suffixes. The advantage of this filter is that it does not only find derivations within entries and that it neatly marks them as such (like Gaggee in Figure 24), but it also finds headwords that happen to be derivations. The search for -ee in the filter mode would have provided many more matches (namely 443, vs. the merely fourteen of Figure 24).

Note that in either search routine concerning morphemes, the query is based on the mere string *ee*, no matter whether it is morphemically functional or not. Therefore, findings such as *agree* and *flee* have to be eliminated from the retrieval list.

### 4.8. Compounds

The usual introductory marker in the original text is "Also in comp". The high frequency of this marker is in line with the observation that dialects are extremely creative, even hyper-productive in compounding (cf. Markus 2012). Accordingly, compounds in the *EDD* may be a very promising field of study in dialectology.
One point of warning may be added. Hyphenation is a non-reliable criterion of compounding. After all, Wright was mainly confronted with spoken dialect use and, thus, obviously found it difficult or impossible to decide in individual cases whether a compound had to be hyphenated when written down, or not. His quotations confirm the generally inconsistent use of hyphens in dialect texts. Accordingly, the hyphens Wright has actually used in practically all compounds – very short ones so that they could almost be addressed as points – are ambiguously bi-functional: they always mark the morphemic border between the constituents of compounds, and they additionally suggest hyphens in many cases.

The filters to be combined with compounds are: dialect areas, etymology, usage labels, sources and time spans. Note that the dialect areas, if not specified by the user, are by default added in the retrieval output.

In the example of Figure 25, the search string *man, combined with the filter “all English counties”, has produced 156 matches.

![Figure 25 Search for *man in compounds, filtered by “all English counties”](image)

4.9. **Combinations**

**Combinations** show a lower degree of juncture between their elements than compounds, a difference which probably goes hand in hand with factors such as (a)
young age of the coining concerned, (b) (still) separate spelling of the elements of the combination, and (c) the lack of the typical compound stress pattern, i.e. lack of the initial stress. In addition to these features we may say with Onysko (2010) that "combinations" in Wright's sense tend to be multi-word non-nominal or only partly nominal constructions, usually with orthographic separation. By contrast, "compounds" are nearly always "two-word nominal" constructions with "orthographic unity" (if we ignore the dot-like hyphen), and phrases tend to be constructions of three or more "content and function words [...] in orthographic separation" (2010: 143).

Which of the filters do combinations allow? The answer is, the same as the compounds, i.e. dialect areas, etymology, usage labels, sources, and time spans.

The equally important role of combinations vs. compounds in the EDD can be gathered from the high number of matches in the query shown in Figure 26: 136 occurrences. The search string and the filter (all English counties) were the same as with compounds.

Figure 26: Search for combinations with *man (plus filter all English counties)

Before we leave the parameter of combinations, it should be repeated that this parameter can be combined with compounds and derivations in one query. We have
implemented this facility mainly because Wright's division between the three categories is not always linguistically reliable.

4.10. Phrases

Given Onysko's definition, mentioned above, that "phrases" in the EDD are three or more "content and function words [...] in orthographic separation" (2010: 143), the modern, more strictly syntactic definition of "phrase" must be taken with a grain of salt. Users interested in phrasal verbs, for example, should try to find them under both phrases and combinations. Clearly Wright's prototypical concept of a "phrase" included, apart from similes and proverbs. multi-word idiomatic expressions, (cf Onysko 2010: 143).

Notwithstanding this lack of full clarity in the definition of phrases, I present an example in Figure 27 which illustrates the prolific role of common verbs in the formation of phrases.

Figure 27: Search for phrases with (non-truncated) go

The 306 matches shown in Figure 27 give rise to the suspicion that phrases, in the particular shape of "compositional predicates"², play a more important role than in the English standard.

The parameter phrases allows for the same optional filters as the compounds and combinations, i.e. dialect areas, etymology, usage labels, sources, and time spans. The filter usage labels, for example, would encourage the selection of

figurative/metaphorical phrases. Irrespective of the filters, searching for *phrases* means that users type in a lexical part of the phrases and get the phrases in their complete extension. Needless to add that users can also search for phrases as such, i.e. without typing in a keyword in the search box. Criteria of selection would then come from any of the five filters.

**5. Final remarks**

This guide to the interface of *EDD Online* is intended as a go-to source of help for users dealing with our software’s complex options. After over a year of applying the new platform myself, I am convinced that its value for English dialect studies is enormous and that it is largely free of major mistakes. However, minor bugs cannot entirely be excluded. Future users who eventually discover any or who have questions or recommendations regarding *EDD Online* are invited to contact me or Reinhard Heuberger (U of Innsbruck, English Department). My email address is provided on our interface next to the icon for this guide.

One tool that has not yet been mentioned due to its recent implementation is the availability of *context*; that is, of the surroundings of a given headword in the text of the dictionary. While the parameters and filters allow for all kinds of ordered selections, the original order of entries in the dictionary is not entirely irrelevant and is a frame of reference that users may occasionally wish to come back to.

In addition to reading this guide, users are invited to explore the more sophisticated tools and gadgets by themselves, in particular trying out the *last-result* mode, the sorting options and the quantification illustrated by maps. Hopefully, *EDD Online* will contribute to invigorate the general interest in (English) dialects, focusing not only on individual words, but also on types of words, features and areas (cf. Markus 2018, 2021, 2022). And perhaps dialectology, as an ancient academic discipline, will profit from the Innsbruck interface and change from a rather immobile caterpillar into an attractive butterfly.

**References**


End of file