THE HYPER USAGE GUIDE OF ENGLISH DATABASE

USER MANUAL

Robin Straaijer
# Table of Contents

Table of contents ................................................................. i
Acknowledgements .............................................................. iii
Foreword ................................................................................. v

1. Introduction ........................................................................ 1
   Bridging the Unbridgeable ................................................... 1
   Usage guides and usage problems .................................. 1
   Research questions ......................................................... 2
   The HUGE database ....................................................... 3
   Applications ........................................................................ 3

2. Construction and data ....................................................... 5
   Varieties of English ............................................................ 5
   Usage guides ....................................................................... 6
   Usage problems .................................................................... 7
   Entries ................................................................................. 8
   Secondary references ....................................................... 8
   Persons ............................................................................... 9
   Types of access .................................................................... 9

3. Searching the database .................................................... 11
   Home screen, login and search module ............................. 11
   Main search menu ............................................................. 11
   Constructing queries ....................................................... 12
   Using operators ............................................................... 12
   Running queries ............................................................. 13
   Displaying additional search results ............................... 14
   Show buttons ...................................................................... 15
   Displaying entry text ....................................................... 16
   Exporting search results ................................................ 17

4. Example searches .............................................................. 18
   Example search 1: Entries ................................................ 18
   Example search 2: Persons and the operator AND ............ 20
   Example search 3: References and the operator OR .......... 23
   Example search 4: usage problems ................................ 26
   Example search 5: usage guides ..................................... 28
   Example search 6: entries ................................................ 32

References .............................................................................. 35
ACKNOWLEDGEMENTS

The Hyper Usage Guide of English (HUGE) database was funded by the Netherlands Organisation for Scientific Research (NWO), under the name A database of usage guides and usage problems (project number 360-70-441). It is part of a larger project called Bridging the Unbridgeable: linguists, prescriptivists and the general public at Leiden University (NWO project number 360-70-440).

The following people were involved in the creation of the HUGE database:

Ingrid Tieken-Boon van Ostade          project director Bridging the Unbridgeable: linguists, prescriptivists and the general public
Robin Straaijer                      lead researcher A database of usage guides and usage problems, design and admin HUGE database
Coen Zimmerman                      design and development HUGE database
Cynthia Lange                        assistant data entry HUGE database
Inge Otto                            assistant data entry HUGE database
Emily Maas                           intern data entry HUGE database & proofreading
Vera Willems                         assistant data entry pilot database
Rob Goedemans                        ICT consultant, design and development pilot database
Sander Gellaerts                     legal consultant
FOREWORD

Welcome to the Hyper Usage Guide of English database!

The project Bridging the Unbridgeable: linguists, prescriptivists and the general public, of which the HUGE database is a part, was started on 1 September 2011 at the Leiden University Centre for Linguistics in The Netherlands. On 27 June 2014, the HUGE database was officially launched at a symposium at the English Faculty of the University of Cambridge, which was organised for the occasion.

This is the instruction manual for users of the Hyper Usage Guide of English (HUGE) database. A user name and a password are required to gain access to the database. If you do not have a user name and a password, please use the form on the project website http://bridgingtheunbridgeable.com to request access to the database. The first part of this manual describes the content of the database, and how it was built, including an explanation of the choices that were made. The second part is meant as an illustration of how to use the database. It contains instructions on how to construct queries and example searches.

Your feedback on the database is very welcome, as is any you may have on this manual. There is a contact form on the website which can be used to send corrections or ask questions.

Robin Straaijer
Leiden, February 2015
1. INTRODUCTION

Bridging the Unbridgeable

In 2011, the research project Bridging the Unbridgeable: linguists, prescriptivists and the general public was started at the Leiden University Centre for Linguistics to investigate English usage in a general sense, and the genre of the usage guide in particular, from a sociolinguistic perspective. The following abstract from the project proposal articulates its goals.

Usage guides are a controversial topic among linguists because of their function to present a norm of correctness to whoever wishes to consult them. Linguistics as a discipline, however, is concerned with describing rather than prescribing usage. Nevertheless, usage guides are extremely popular with the general public, and even increasingly so despite centuries of prescriptivism. This project seeks to close the gap between the three main players in the field of prescriptivism: the linguists themselves, the prescriptivists (as writers of usage guides) and those who depend upon such manuals. It will do so by compiling a database of usage guides and usage problems that will be of use to all concerned, including the direct participants in the project.

—Tieken-Boon van Ostade 2010a

As the last sentence indicates, a database of usage guides and usage problems was seen as an important tool in the stated goal.

The project Bridging the Unbridgeable: linguists, prescriptivists and the general public is subsidised by the Netherlands Organisation for Scientific Research (NWO http://www.nwo.nl), which allocates subsidies on behalf of the Dutch Ministry of Education, Culture and Science (Ministerie van Onderwijs, Cultuur en Wetenschap http://www.rijksoverheid.nl/ministeries/ocw). It is their philosophy that since the tax payers ultimately, though indirectly, fund the generation of scientific knowledge, academics should share that knowledge with the public. Therefore, since the project is publicly funded, the HUGE database should ideally be made freely available to the general public as the very final stage of the database-project. Some of the older usage guides that have been included in the HUGE database have reverted to the public domain, but most of them still have copyright on them. In order to be able to provide public access to the HUGE database, we are negotiating with the holders of the copyrights of the usage guides used.

Usage guides and usage problems

"What are we to make of that neglected genre, the Usage Guide?", Edmund Weiner asked himself more than 25 years ago (Weiner 1988: 171). As of 2010, the usage guide was no longer neglected, with scientific work having been published on usage in general (Burchfield 1991, Algeo 1994, Peters and Young 1997, Finegan 1998, Finegan 2000, Peters 2006) and on specific usage guides or the genre itself (Busse and
Schröder 2010, Tieken-Boon van Ostade 2010b). But the neglect of the usage guide as a genre is understandable, since, as a genre, the usage guide is somewhat of a mixed bag, as Weiner also noted.

In one dimension, the scope of a usage guide is as broad as the English language, covering spelling, punctuation, phonology, morphology, syntax, and lexis, and involving sociolinguistic considerations. But obviously a usage guide does not describe the whole language. It takes for granted the bulk of it, and it assumes that its audience are native speakers or advanced learners. It has nothing to say about some of the central facts of language, while going into great detail about others.

—Weiner 1988: 173

The usage problems, those items of disputed usage that are discussed in usage guides, reflect this. Again, Weiner noted this.

The usage guide addresses itself to a tiny fragment of the language, or rather to a number of tiny fragments, for the subjects it treats are not inherently linked together, as the phenomena explored in a grammar are interconnected by being part of an overall system. What holds the parts together is external function.

—Weiner 1988: 173

However, external function is what language in actual usage is all about, and the study of language – at least that of modern languages, even in a historical context – cannot be considered complete without taking actual usage into account. The ‘sociolinguistic considerations’ mentioned by Weiner are not to be underestimated. For instance, the selection of the topics that are included in a usage guide is almost wholly determined by such considerations, as noted by Peters and Young.

[… ] certain topics have indeed become conventional for usage books, and the commentators note without justification what is “right”/“wrong”, “preferred”/“to be avoided” […] The writer’s value system is foregrounded, with little attempt to correlate judgements with external sources, either primary or secondary

—Peters & Young 1997: 317

Despite existing research on usage, there are still a number of questions about usage problems in British and American English that have yet to be answered. It is expected that at least some of these will be answered in the course of the Bridging the Unbridgeable-project.

*Research questions*

One aspect of usage problems about which we have questions is their individual histories. *When does a certain usage become problematic, or perceived as such, and when does a certain usage stop being (perceived as) problematic or disputed?* In other words, when do usage problems ‘begin’ and ‘end’? *And which usage problems persist?* Another aspect is the discussion of usage problems in usage guides. Questions are: *Does the discussion of specific usage problems change, and if so, in what way? And are there differences in usage advice for different varieties of English?* These questions naturally bring us to the following: *Why do these things happen? And what is the role of usage guides in these processes?*

The main obstacle to begin to answer these questions is that no historical overview of usage advice in English exists. A database of English usage guides and usage problems was conceived as a solution to this problem. This database would have full-text entries of many usage problems. In addition to this, the
database would also include scholarly secondary works dealing with usage guides or specific usage problems.

**The HUGE database**

The answer to these questions lies in the *Hyper Usage Guide of English* database, or HUGE for short. HUGE is hyper in two senses. One, the usage advice in the HUGE database transcends any one particular usage guide or language variety. And two, the information in the HUGE database transcends the usage advice itself. Especially with the incorporation of secondary references, it can operate on a meta level.

As part of an earlier project proposal, there was a MS Access pilot database of historical grammars and usage guides. Wanting to improve upon this pilot database led to the design of a new Access database, which was never realised. Instead, looking forward to a time when an online resource would be wanted, a new relational database was designed and built, which came to be called the HUGE database.

The main strength of the HUGE database is that it combines nearly 250 years worth of usage advice in new ways. Because it is a relational database, every bit of data in the database can in principle be extracted from it with the right query. The way the database is structured allows for the development of new types of queries, such as statistical queries that provide information about the numbers of things in the database. The total amount of data is limited since usage guides are not included in their entirety.

**Applications**

This database aims to benefit linguists, educators, as well as language professionals such as writers and editors, and in its ideal form – and projected final inception – also the general public.

On an academic level, the database is useful for historical linguists and sociolinguists. The database can provide a temporal and social focus for historical and sociolinguistic studies regarding usage. Using the HUGE database, it is possible to track when certain usages become or cease to be problematic, or to be perceived as being so. It is possible to compare usage advice across time, language variety and author. It is also possible to investigate the treatment of usage advice and usage guides in the linguistic literature at various points in time.

The database is also useful for scholars working on usage from a perspective of (critical) discourse analysis. The fact that the huge database provides full-text entries allows for investigations of the language and rhetoric of usage advice. The database allows for this data to be re-used in these kinds of studies. The information on secondary sources also allows such investigations to be related to scientific discussions in the literature by linguists at various points in time.

For educators, writers and teachers of English to native speakers as well as to those who learn English as a foreign language, the database provides an overview of and insights into the problems of usage in general, as well as into those concerning specific usage problems. The advantage of this information is that it has a sound basis in up-to-date linguistic research.
The general public can benefit from this database to find answers to questions on specific points of usage by either consulting individual usage guides or by comparing advice from different guides.
2. CONSTRUCTION AND DATA

As noted above, “[t]he usage guide addresses itself to a tiny fragment of the language, or rather to a number of tiny fragments, for the subjects it treats are not inherently linked together” (Weiner 1988: 173). In addition, the usage guide also addresses itself to the language in a fragment of time, which is that of the period in which it is published. The HUGE database links all these fragments together; not just those from one guide, from one level of the language, or from one period of time, but across all of these variables. Fragmentation, but fragmentation of an ordered and very specific nature, is precisely the strength of a relational database such as the Hyper Usage Guide of English. The main parts of the database are represented by its principal tables containing information on usage guides, usage problems, entries, secondary references, and persons, as shown below in Tables 1–5.

<table>
<thead>
<tr>
<th>Usage Guide</th>
<th>Usage Problem</th>
<th>Entry</th>
<th>Secondary Reference</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>tags</td>
<td>usage guide</td>
<td>title</td>
<td>last name</td>
</tr>
<tr>
<td>year</td>
<td>problem term</td>
<td>usage problem</td>
<td>year</td>
<td>first name</td>
</tr>
<tr>
<td>edition</td>
<td>description</td>
<td>page</td>
<td>edition</td>
<td>nationality</td>
</tr>
<tr>
<td>original year</td>
<td>example</td>
<td>text</td>
<td>in title</td>
<td>occupation</td>
</tr>
<tr>
<td>impression</td>
<td></td>
<td></td>
<td>volume</td>
<td></td>
</tr>
<tr>
<td>author</td>
<td></td>
<td></td>
<td>issue</td>
<td></td>
</tr>
<tr>
<td>editor</td>
<td></td>
<td></td>
<td>author</td>
<td></td>
</tr>
<tr>
<td>place</td>
<td></td>
<td></td>
<td>editor</td>
<td></td>
</tr>
<tr>
<td>publisher</td>
<td></td>
<td></td>
<td>place</td>
<td></td>
</tr>
<tr>
<td>language variety</td>
<td></td>
<td></td>
<td>publisher</td>
<td></td>
</tr>
<tr>
<td>organisation</td>
<td></td>
<td></td>
<td>download link</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. HUGE database principal tables

These tables will be discussed in the remainder of this chapter, after a few words on the varieties of English included in the database.

Varieties of English

English is the most widely used language in the world and consequently there are many varieties of the language. The largest varieties in terms of number of native speakers are British and American English; while Canadian, Australian, and New Zealand English are other main varieties in terms of native speakers. In addition, there are other varieties commonly known as ‘World Englishes’, such as Jamaican or Singaporean English. An interesting variety is Indian English, which is probably the fastest growing variant at this time, and may become a main variant.
The varieties of English included in the HUGE database, that is, the language varieties described by the usage guides in the database, are only the two largest varieties of English: British and American. This was done mainly for practical reasons. Planned extensions of the database are to incorporate usage guides discussing the other main varieties: Canadian, Australian, and New Zealand English. The HUGE database is constructed in such a way that the number of language varieties can be expanded virtually infinitely.

Most usage guides concern themselves with only one variety of English. Some usage guides explicitly mention which variety that is, such as Bryan Garner’s *A Dictionary of American Usage*. Some usage guides address more than one variety of English – in the current situation that means both British and American English. In cases in which the language variety was not explicitly mentioned, the variety was usually assigned based on the country of publication and the nationality of the author. Whether the guide described more than one variant was determined by scanning its contents: if both variants were repeatedly referred to, the guide was categorised as pertaining to both British and American English. If only sporadic references were made to a variety that was not the main variety, it was categorised as pertaining to only either British or American English.

**Usage guides**

In the project it was necessary to address the issue of the usage guide as a genre, for primarily two practical reasons. First, a definition of the genre was necessary in order to delimit the scope of the project’s investigation. And second, while creating the HUGE-database, this necessity became more practical as it was necessary to determine which titles would be put in the database, and which would not. This meant deciding which works were usage guides and which were not. It should be noted that because not all usage guides published are in the database; when a title is not in the database this doesn’t mean that it is not a usage guide according to our definition. In addition, it turned out that the genre was actually not very clearly defined, so part of our thinking about usage guides has been about defining the genre. Our main criterion was that since the HUGE-database should primarily deal with usage problems of a grammatical nature, a work should deal with these kinds of problems in order to be considered a usage guide.

The database contains selected usage guides from 1770 to the present. The year 1770 was chosen because that year saw the publication of Robert Baker’s *Reflections on the English Language*. A usage guide is defined as a work that merely offers advice on usage; therefore normative grammars which include usage advice – of which many were published before 1770 – were not included. Baker’s book is also the only usage guide to be included from the eighteenth century. The HUGE database contains eight usage guides that were originally published in the nineteenth century, which covers much of that century. The twentieth century saw a great increase in the number of titles published, and consequently it was not possible to be exhaustive. We have tried to represent the twentieth-century as best as possible by including usage guides from every decade, and as many from each year as were available. The problem of comprehensiveness is also there for the twenty-first century. However, because we are still in the beginning of this century, there are still far fewer usage guides to include in the HUGE database, and consequently the problem is proportionally smaller.
As a starting principle, only one edition of each of the usage guides was to be selected for inclusion in the HUGE database. Ideally that edition would be the first edition, but whether this happened obviously depended on whether the first edition was available. Most of the guides in the database are first editions, but in some cases later editions were used. Wherever the information was available, the edition and impression numbers were included in the bibliographical part of the database. For the field edition, the values ‘revised’ and ‘unknown’ were available besides cardinal numbers.

Apart from edition and impression, other fields with information were included in the description of the usage guides. These are: the year in which it was published and the original year of when the work was first published (usually the same in case of a first edition), the name of the author or editor of the work, given in separate fields, as some usage guides have both, and the publisher and place of publication. The field language variety indicated whether the guide deals with British or American English. The field organisation refers to whether the usage guide is organised alphabetically or by topic.

Usage problems

And as also noted above, the usage guide traditionally covers usage problems related to “spelling, punctuation, phonology, morphology, syntax, and lexis, and involving sociolinguistic considerations” (Weiner 1988: 173). The HUGE database, however, deals primarily with usage problems of the – very broadly speaking – grammatical kind. This mostly means that usage problems regarding only pronunciation, etymology and punctuation were generally not selected for inclusion in the HUGE database. Usage problems related to lexis and spelling were sometimes included, based on how pervasive or salient we judged them to be.

The first usage problems with which the database was started were those 55 investigated in Attitudes to English Usage by Mittins et al. (1970), to which were added those from David Crystal’s Grammatical Top Ten (Crystal 1995: 194). This resulted in a total of 60 usage problems, with which the data-entry process was started. A tally was kept of usage problems not appearing in the original list, but which appeared repeatedly in the usage guides entered into the database. In addition notice was taken of other usage problems that seemed to be particularly socially salient. This led to a second list of 63 usage problems, for a total of 123.

The usage problems were given a problem term, the name by which the problem is commonly known (for instance, split infinitive, who/whom). The usage problems were also provided with a description, an example sentence – usually the one given in Mittins et al. or Crystal – and a number of tags indicating the nature of the usage problem on both the linguistic and metalinguistic level. Tags used include morphology, concord, logic, number, semantics, comparative, adjective, pronoun and various others. As an example, the data for the usage problem of their being used as a common-sex singular pronoun would have the following information.

<table>
<thead>
<tr>
<th>Problem term</th>
<th>singular they</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>use of ‘they or ‘their’ as a common-sex singular pronoun</td>
</tr>
<tr>
<td>Example</td>
<td>Everyone has their / his or her off-days.</td>
</tr>
<tr>
<td>Tags</td>
<td>agreement, concord, gender, personal, pronoun, syntax</td>
</tr>
</tbody>
</table>
This usage problem is about personal pronouns, but as can be seen, there is no single tag for that. We use two separate tags, personal and pronoun, to minimise redundancy and maximise flexibility in the tag system. Lists of both the usage problems and the tags used can be found in the search module of the database – see also the following chapters.

Entries

An entry in the database refers to a bit of text in a particular usage guide which discusses a particular usage problem. Before any type of data entry could be done, the usage guide first had to be converted from a paper book into an electronic format. We chose PDF as this is probably the most widely used format.

In order to be able to feed the usage guides into the automatic document feed of the scanner, the books needed to be disassembled into separate sheets, which was accomplished by removing their spines with an industrial paper cutter. The usage guides were then scanned with a HP Scanjet Enterprise 7500 scanner as black-and-white TIFFs with a resolution of 600dpi. On these image files, optical character recognition (OCR) was performed on these scans using Abbyy Finereader 11 and the resulting files were saved as searchable PDFs. In the case of many of the older usage guides, typically those from before the late-nineteenth century, the OCR was too inaccurate to render a viable searchable PDF. These texts were retyped into an MS-Word file and from there the relevant entries were copied into the HUGE database.

These PDFs were searched manually and the text of the relevant entries were copied into the database, with as much preservation of the formatting as possible. The entries were then linked to the specific usage problem which they are about (possibly more than one), and to the usage guide in which they appear, including the page number.

Sometimes an entry refers to more than one usage problem. These were dealt with in two ways, depending on the structure of the lemma in the usage guide. One structure is where the lemma was undivided, presented as a continuous piece of text, uninterrupted by sub-headings indicating different aspects of the lemma relating to different usage problems. In these cases, the complete text of the entry was put into the database, and the relevant usage problems were connected to this single entry. The other structure is where the entry contained sub-headings, such as letters or numbers, related to different usage problems. In those cases, the relevant sub-parts of the lemma were put into the database as separate entries, each representing one different usage problem. An exception to this is a case in which such a sub-part of a lemma refers to more than one usage problem. In those cases, the protocol for undivided lemmas was applied to these sub-parts.

It should be noted here that despite checking of the formatting, and because of limitations of the HUGE database’s text editor (such as not supporting the use of small capitals, which are represented by regular capitals), errors may still be present in individual entries so when citing from the database the text should be checked against an original of the work.

Secondary references
Inclusion of the references also started from Mittins et al. (1970) and the sources they mention in their list of references. To these were added known publications from that year onward. This was done by searching academic databases such as the MLA and LLBA, for books, book chapters and articles related to specific usage guides or specific usage problems. In addition, entire publishing histories of selected journals, *American Speech, English Studies, English Today*, and *The English Journal*, were searched by our research assistants for relevant articles. The bibliographic data for the secondary reference was typed into a field of the database similar to that used for the usage guides. And similar to how the texts of entries were linked to a specific guide and a specific usage problem, the secondary reference was linked to the usage guides and usage problems discussed in it.

So, the secondary reference table contains the book or article’s **title** and **year** of publication, as well as the **edition** in case of a book, and the title in which the article appears, **in title**, in case of a chapter or article in a collection. For journal articles, the **volume** and **issue** of the journal is given. Furthermore, the **place** of publication and the **publisher** (for books) are included. Most importantly, the **usage guides** and/or **usage problems** discussed are listed, including the page numbers where the actual citation occurs. Finally, even though the secondary references themselves are primarily available as bibliographic data – as it was not part of the design of the HUGE database to include full-text versions of scholarly publications – **links** to the full-texts at external locations were included where these were available. Note that these links to not all refer to open access publications, and consequently some locations may not be available without a subscription.

**Persons**

There are two types of persons in the HUGE database: authors and editors. From a technical, database point of view they are the same, which is why there is one menu to search for authors and editors in the search module of the HUGE database, called **persons**. These are the authors and editors of usage guides as well as those of the secondary references in the database. The information on authors and editors is fairly simple. The person’s **last name** and **first name** are given, plus, as far as this is known, their **nationality** and **occupation**. There is an extendable list of occupations of which one or more can be assigned to a person. These occupations are: **academic**, **editor**, **educator**, **civil servant**, **cleric**, **grammarian**, **journalist**, **lawyer**, **lexicographer**, **linguist**, **other language professional**, **teacher**, **writer**, **other** or **[unknown]**. Some of these occupations describe more general categories than others, but this seems to be unavoidable.

Where available, biographical information was taken from the *Oxford Dictionary of National Biography* and the *American National Biography*. Additional biographical information was gathered from notes in journals in which the authors publish or the institutions for which they work, as well as by simply searching the internet. Unfortunately, only incomplete biographical information was available for many persons in the HUGE database.

**Types of access**

There are two levels of public access to the database: **limited** and **full**. Except for the full texts of the entries of the usage guides in the HUGE database, which are not accessible for the limited accounts, both account types include all primary and metadata described in this chapter. It is still possible to search for entries
using a limited account. However, the texts themselves will not be shown; a message "restricted access account – text withheld" is shown instead.
3. SEARCHING THE DATABASE

This chapter discusses the basics of using the database, that is, performing searches or queries. Additional features of the search module will be discussed in more detail in a practical, step-by-step manner in the example searches in Chapter 4. The website is located on http://huge.ullet.net.

Home screen, login and search module

The website of the HUGE database opens on the basic Home screen. You can log in by entering a username and password in the grey box to the right of the bar at the top of the page, followed by clicking the log in button. A session remains active for two hours, after which you will have to log in again, and the box also shows how much time is left on your session. You can avoid having to log in again by clicking the link refresh next to the time before the session expires. You can end your session by clicking the link logout. After logging in you are in the extended Home screen, which shows some additional information regarding the sources used to build the database.

Main search menu

On the left side of the Home screen is the main Menu. This is opened by letting the cursor hover over it. The search module can be accessed by opening the Menu and selecting Query. This will open the search screen, which looks as in Figure 1.

![Figure 1. Search module main menu and default search window](image)

The bar at the top called Search for is the main search menu of the search module and shows the primary search functions, represented by the five buttons on the top row of the window. These buttons are from left to right: Usage Guides, Entries, References, Usage Problems and Persons. These allow users to search for usage guides, entries in usage guides for specific usage problems, references in studies such as Mittins et
al. 1970, usage problems and persons (authors and editors). The text to the left of the bar shows which part of the main search menu is currently active. The search module of the HUGE database allows for searches of virtually all of the database's fields and combinations of these fields using the logical operators AND, OR and NOT (for more on this, see the section ‘using operators’).

**Constructing queries**

A new query starts by clicking one of the buttons in the main search menu. The default search window is called **Query 1**, which starts with a single search row. The query is built from left to right, and starts by opening the leftmost drop-down list. This menu contains all the items out of which the HUGE database is made up, described in Chapter 2, such as year, usage guide title or edition, language variety, usage problem term or tag, author name or occupation, entry text, or reference title. Some of these use drop-down lists from which items need to be selected, others have fields in which text can be typed. In text fields, the wildcards (?) and (*) can be used to replace a single character or several characters respectively.

Within this window, simple queries can be combined into more complex ones using the appropriate logical operator AND or OR (the default operator is AND). This is done by clicking the button **Add condition**. This adds a search row underneath the existing search rows, as shown in Figure 2 for an example in which we want to search for usage guides that deal with both British and American English.

![Figure 2. Query box with two search rows combined by operator AND](image)

Clicking on the button with the **green plus sign** of a row adds a search row **above** that row. This makes it easier to change the order of conditions, which is useful because, depending on the operators used, the order in which conditions are processed affects the outcome of the search. A condition in the query can be removed by clicking the button with the **red minus sign** button in front of the relevant search row.

**Using operators**

There are two logical operators with which query lines can be combined: AND and OR. The most important thing to remember about them is that the operators work exactly opposite to the 'normal' way we naturally interpret and use the words **and** and **or**. In normal language, the expression \(x \text{ or } y \text{ and } z\) is usually interpreted as \((x \text{ or } y) \text{ and } z\). For example, if I say 'I want to see usage guides written by American
or British authors, and after the year 2000", then I expect the answer to be usage guides written after the year 2000 by British authors as well as usage guides written after the year 2000 by American authors.

However, in the language of logical operators this way of asking a question does not work. The operation AND always takes priority over the operation OR. The query $x \text{ OR } y \text{ AND } z$ can only be interpreted as $x \text{ OR } (y \text{ AND } z)$. This priority of AND over OR is indicated in the query box by the appearance of brackets around the conditions combined by AND. There is, however, a different way create a query, one that will be interpreted as $(x \text{ OR } y) \text{ AND } z$. Rather than cramming as many conditions as possible in a single query, this is done by stacking two queries. What do we mean by stacking queries? Stacking queries means that the results of one query will be used as the input data for a subsequent query, which will have exactly the same functionalities as the original one. Creating stacked queries can be done by clicking the button Add query below the search window. An additional search window called Query 2 (and so on) will open up, which can be used in the way shown above. This can also be done after the results of the search have been displayed, by clicking the button Search within results in the bottom of the window, below the search results adds an additional query. A concrete example of stacking searches is given in Example search 6 in Chapter 4.

Any query can be removed by clicking on the cross in the top right-hand corner of that query box. To start with a complete new set of queries, click the button Clear all below the query window(s). This will clear all active queries. After the search is run and the results are displayed, the query can be cleared or altered by clicking the button Clear query or Alter query, respectively. Both of these buttons will return an empty query window. Note that the results remain on the screen; these will only be replaced once a new query has been run.

In order to change the original query without having to construct it completely anew, there is the option to alter the original query by clicking the button Alter query in the bar underneath the query window(s). To keep the bottom queries of query stacks visible on the screen, the ones on top can be minimised by clicking the dash button in the top right corner of the window of each separate query. This option toggles to a full window button which can be used to restore the full query box, just as in Microsoft Windows™.

**Running queries**

Clicking the button Search at the bottom right-hand corner of the query window runs the search. A box above the box Query 1 appears called Active Query. This box shows the last query that was run, and can act as a check for whether the right combination of operators has been used. As mentioned above, the box with search results continues to be displayed even after the query has been cleared, until a new query is run.

Clicking the Search button on the right-hand side below the search window runs the query. The default columns with search results displayed include those categories that were part of the actual query. For a query on usage guides, for instance, this includes the title, year of publication, original year of publication, and author, plus all the variables searched for. In addition, the number of matches is displayed at the bottom of the search results box.
We can take our earlier example a little further. When we search for usage guides that deal with both British and American English, and that were originally published before the year 2000, we get the following default search results, shown in Figure 3.

![Figure 3. Example search results](image)

Search results can be sorted, alphabetically or chronologically for instance, by each column displayed by clicking the up/down sorting arrows next to their labels, circled for the column Author(s) in this example. Just like the box Active Query, the box Search Results also continues to be displayed even after the active query has been cleared, until a new search is run.

**Displaying additional search results**

To display additional columns after a query has already been run, start by clicking the button Alter query, and then clicking the button Columns to display. Clicking this button opens a window where the desired categories can be selected by a series of check boxes, as shown in Figure 4 below.

![Figure 4. Checkboxes to display additional search results](image)

Select the categories to be displayed, and click on the Search button again to re-run the query with the new settings. This will display the search results with the desired items of additional information. Which options are available depends on which part of the main search menu is active. Table 1 on the following page shows which items of additional information are available for which parts of the main search menu.
Some of the columns in the search result show no values, but instead a button called show on each row. Clicking on this button gives you additional information for that column, usually a list or table that cannot be displayed as a nested table within the search results. One such column is the one with a list of usage guides when searching for usage problems, which is called Guide list. Clicking on the show button of a line of a particular usage problem opens a little pop-up window with the usage guides in which that usage problem is discussed. For example, if we search for usage problems that are tagged with the tag lexis, and we select Guide list as an additional column to display, we get a short list of usage problems with a column Guide list containing show buttons. Figure 5 shows such a list of results, including what is shown when the show button has been clicked.
Other columns that display only show buttons are entry list and reference list, as well as author details and editor details.

Displaying entry text

One of the most important ways to display additional information in the search results is with searches for entries for specific usage problems or in specific usage guides. The default display for the results for entry searches shows a column with lines of part of the text of the entry. For example, if we search for the usage problem meet (up) with in usage guides published after the year 2000, we get four hits from the database, as shown in Figure 6.

Clicking on each of the lines in the column Text will display the text of that entry. This will be illustrated further in the next chapter. Clicking on a line in the column Text will open a pop-up window with the full
text of the entry. The text can then easily be copied by clicking the button Select entire text and selecting Copy from the pop-up window's context menu (right click).

Note that two of the four lines do not actually show part of the text for the entries retrieved. This is because the contents of the guides from which they come have not yet been cleared for publication with the publisher of those titles.

Exporting search results

All the search results chosen in the search windows, as well as those chosen in the window Columns to display can be exported for use in other applications. This allows the information in the HUGE database to be used as raw data for further analysis. A search for entries in the main search menu Entry provides the extra option of exporting both plain and full-text search results. Exporting the full-text of the query’s search results is done by checking the box full text at the bottom left of the window. The HUGE database supports exporting results in two file formats: a comma separated value format (.csv), which can be read by most spreadsheet applications, and eXtended Markup Language (.xml), which can be used in more specialised software such as concordancing programs. In the case of entry texts, there is the additional option of exporting them as either plain entries with only text extracts, or as full-text. Finally, the results from a query can also be printed. Clicking on the button Print version opens an on-screen preview of your print in a new window. Use your browser’s print function to print the results or save them as a PDF.
4. EXAMPLE SEARCHES

Now that we have covered the basics of querying the database in Chapter 3, let us look at a few examples of searches to get a better idea of how the database works in practice. We will start with some relatively simple searches and work our way to more complex ones. Each of these example searches will start with a different part of the main search menu, so as to illustrate them all.

Example search 1: Entries

For the first example search, we will start with a very simple query that has just one search condition. Say that you will want to see all the full-text entries in the database about the adverbial use of the word likely. Since we want to find entries, we start by clicking the button Entries in the main search menu. Then go to the window Query 1 and choose problem term as the type of search condition from the leftmost drop-down list, as shown in Figure 7.

![Figure 7. Selecting type of search condition](image)

The default operator is appears; leave this as it is. Click on the empty box next to it, in which a list appears with all the usage problems in the database, from which you can select the term likely. This is shown in Figure 8. Note that below the box Query 1, there is a small box called Active Query, indicated by the red arrow, which shows you in a shorthand form what the current query looks like.
Click the button **Search** to run this search. The default display of the results of this search will look like Figure 9, which shows a line from each entry retrieved, as well as the page numbers in the usage guide which that entry comes from.

![Figure 8. Selecting problem term](image1)

![Figure 9. Default results for entries search](image2)
In the bottom of the window it also shows how many hits the search yielded: in this case 21 entries on *likely* were retrieved out of the total 5579 entries in the database. Click on a line in the column *Text* to display a pop-up window with the full text of that specific entry, as shown in Figure 10. Note that some of the entries will not be visible as they have not yet been cleared with the publisher of the usage guide from which they were taken.

![Figure 10. Full text pop-up window in search results](image)

If you want to select the text in the pop-up window, click the button **Select entire text**. Click the button **Close** or the *X* in the top right corner to close the window.

**Example search 2: Persons and the operator AND**

For the second example search, we will start at the option *Persons* in the main search menu, for a simple query that has just two search conditions, combined with the operator AND. Imagine that you wish to find out which British authors have written or edited usage guides that deal with American English. Since we
want to find persons, we start by clicking the button **Persons** in the main search menu. Then go to the window **Query 1** and choose *nationality* as the type of condition from the leftmost drop-down list, as illustrated in Figure 11.

![Figure 11. Selecting type of search condition](image)

The default operator *is* appears; leave this as it is. Then choose *United Kingdom* from the rightmost drop-down list of countries, shown in Figure 12.

![Figure 12. Selecting nationality](image)

Click the button **Add condition**, keep the default operator *AND* that appears, and choose *guide language variety* from the left drop-down list, as shown in Figure 13.
As in the previous step, keep the default operator *is* that appears, and click on the empty box next to it, from which you then select the language variety *American* from the drop-down list that appears, as shown in Figure 14.

After following these steps, we now have constructed a simple search for persons with two conditions. Click the button **Search** to run this search. The default display of the results of this search will look like Figure 15.
The table of results shows the default information for persons, name, nationality and occupation, as well as how many usage guides each has written or edited. The search results can be sorted in various ways by clicking on the sorting arrows next to the columns’ headers. The last column Guide list has no data but only a show button. Behind this button is another table with information on the usage guides associated with this person, and the nature of that association. Clicking this show button opens this table in a pop-up window. This is shown in Figure 16.

In this example we can see that the linguist Sidney Greenbaum is connected to the 1988 *Longman Guide to English Usage* as its author, but not as its editor. Click the ✗ in the top right corner to close the pop-up window (or click the show button again).

**Example search 3: References and the operator OR**

In this example search, we will start with the option References in the main search menu. Again, we will perform a simple search, but this time we will use the operator OR to combine our search conditions. In this example we want to find secondary references that deal with the usage problems *pretty* (as an adverb of degree) and *very unique* (the use of an intensifier with absolute adjectives). Since we want to find references, we start by clicking the button References in the main search menu. Then go to the window Query 1 and choose problem term as the type of search condition from the leftmost drop-down list, as shown in Figure 17.
The default operator \textit{is} appears; leave this as it is. Click on the empty box next to it, and the complete list of usage problems will be shown, from which you then select the term \textit{pretty}. This is shown in Figure 18.

Click the button \textit{Add condition}, and change the operator that combines the two query lines from \textit{AND} to \textit{OR} in the way illustrated in Figure 19.

The second part of the query is constructed the same as the first. Again, choose \textit{problem term} from the left drop-down list, leave the default operator \textit{is} as it appears, and select the term \textit{very unique} from the list of...
usage problems that appears by clicking on the box on the right-hand side. With that, we have constructed the full query, which looks as in Figure 20.

![Figure 20. Full search for references](image)

As before, click **Search** to run the search. The results of the search are shown in Figure 21. Again, the search results can be sorted in various ways by clicking on the **sorting arrows** in the headers of the columns.

![Figure 21. Default results for reference search](image)

The last column **Problem list** has no data but only a **show** button. Behind this button is another table with information on all the usage problems discussed in that secondary reference work. Clicking this **show** button opens this table in a pop-up window, as shown in Figure 22.
In this example we see that Pam Peters’ chapter “English Usage: Prescription and Description” in the *Handbook of Linguistics*, published in 2006, mentions the usage problem *very unique*, among others. These others, which are usage problems we have not searched for, can be hidden by clicking the link *matches only* in the title bar of the *Usage Problem list* pop-up window. This will show only those usage problems that match the query. To reverse this, click the link *show all* that appears.

**Example search 4: usage problems**

In this example search we will start with the option *Usage Problems* in the main search menu. We will again perform a relatively simple search with two conditions. We will also again use the operator OR to combine our search conditions, and we will search with tags. As you have read earlier in this manual, the majority of the usage problems in the HUGE database are of a grammatical nature. Say that you want to find those that specifically deal with *comparatives* or *superlatives*. Since we want to find usage problems, start by clicking the button *Usage Problems* in the main search menu. Then go to the window *Query 1* and choose *problem tag* from the leftmost drop-down list, as illustrated in Figure 23.
Keep the operator `is`, and then click on the empty box next to it. Clicking on the box will open a drop-down list with all of the problem tags in the database from which you then select `comparative`, as shown in Figure 24.

![Figure 23. Selecting type of search condition](image)

Click the button `Add condition`, and change the operator that combines the two query lines from `AND` to `OR`. Again, choose `tag` from the left drop-down list. Complete this second query line in the same way as the first one, except choosing the tag `superlative` in the rightmost box. The complete query looks like Figure 25.

![Figure 24. Selecting problem tag](image)
Again, the query can be run by clicking the **Search** button; the result of this query is the list of usage problems shown in Figure 26.

The results show the seven usage problems in the HUGE database that deal with either comparatives or superlatives.

**Example search 5: usage guides**

This example search will be a little more complicated. We will combine both operators *AND* and *OR* to form more complex searches within a single query. Say that you will want to see all usage guides that deal with American English and whose authors were writers by profession, as well as all usage guides that deal with British English, and that were written by linguists.

Because we want to find usage guides, start by clicking the button **Usage Guides** in the main search menu. Then, from the leftmost drop-down list in the window **Query 1** choose *guide language variety*, keep the default constraint *is* from the centre drop-down list, and click on the empty box next to it, from which you then select *American* from the drop-down list that appears. Then click the button **Add condition**, keep the
default operator AND that appears, choose author occupation from the left drop-down list and keep the default constraint is. Clicking on the text box on the right-hand side of the query window opens a list of occupations, from which you choose writer. So far, the first part of our query looks as shown in Figure 27.

![Figure 27. First part of query](image)

Click the button Add condition once more, and change the operator to OR. The query is automatically grouped into two parts, shown by the brackets that appear. They indicate that conditions joined by the operator AND are executed before those joined by the operator OR. Choose guide language variety from the left drop-down list, keep the default constraint and choose British from the rightmost drop-down list.

Again, click the button Add condition, keeping the default operator AND. The brackets are immediately extended to include this condition. Choose author occupation from the left drop-down list and keep the constraint is. Then choose linguist from the box on the right-hand side of the query window. After following these steps, the complete query now looks like Figure 28.

![Figure 28. Full query](image)
Click the button **Search** to run this search. The default display of the results of this search will look like in Figure 29. Note that in this case, since the default results display for this search already includes the name and occupation of the authors, the only extra information that the **show** buttons in the column **Author details** provide is their nationality.

Figure 29 also shows that the results of this search include usage guides written by authors who have additional occupations besides those given in the conditions of the search.

It is possible to add additional information to this search without fundamentally changing the query. We have not set any conditions with regard to usage problems, but it is possible to show which ones are discussed in the guides resulting from the search by displaying additional columns. Click on **Alter query**, which reveals the original query. Then click on **Columns to display**, check the box called **problem list**, and rerun the query by clicking the **Search** button. Figure 30 shows part of the results with the **Problem list show** button revealing a list of usage problems associated with Edward Gould's 1867 *Good English*.
Now let us take this query a little further. Say that we want to limit our results from the previous query to those guides originally published after the year 2000. This actually cannot be done in the same query by using **Add condition**. If we do this, by adding the condition line **AND guide year, original is greater than 2000**, as shown in Figure 31, we actually don’t get the right query.

![Figure 30. Results with additional information](image1)

![Figure 31. Advanced search – not the intended query](image2)

Try running this query and you will see that the results are not the ones you may expect. So how do we get the right query structure, one that gets us the results we want? The solution is to stack a second query onto the first one. This is started by clicking the button **Add query** in the bar underneath **Query 1**, or by clicking the button **Search within results** in the bottom bar of the **Search Results** window. The way this is done is shown in the next example, as part of a search for the full texts of entries of usage guides.
Example search 6: entries

For the last example, we will do an even more complicated search. Say that you will want to see all entries about the use of the word *literally* as an intensifier, as well as the word *pretty* as a degree adverb, appearing in guides whose writers are not American, and for which there are secondary sources in the HUGE database. Since this time we want to find entries, start by clicking the button Entries in the main search menu. Rather than creating a complicated, unnecessarily long and repetitive query, the easiest way to do this is to stack queries. So let us look at our first conditions: the usage problems we want to search for: *literally* and *pretty*, and our simple first query using the operator OR. If we were to run this query, it would give us all entries about the usage problems *literally* and *pretty*.

Rather than repeating search conditions as in the example in Figure 31, we can easily add the other conditions by adding a second query. Click the button Add query, and a new query box called Query 2 opens up. Choose author nationality from the left drop-down list, choose is NOT, and choose United States from the list of countries. Click the button Add condition, keep the default operator AND that appears, choose [special] from the left drop-down list, then choose is from the centre drop-down list, and choose connected to any reference in the rightmost drop-down list. These steps result in the query structure shown in Figure 32, composed of two simple queries rather than one complicated and repetitive one.

![Two stacked queries](image)

As usual, clicking the Search button runs the query, which gives rise to the search results shown in Figure 33. Note that although the query included the condition that the usage problem *literally* should appear in references, there is no column that shows information about these references. This is because the query does not relate to individual entries but only to the usage problems *literally* and *pretty*, whichever usage guide it appears in.
To display the secondary references, click the button **Alter query** and then **Columns to display**, check the box **reference list** and re-run the query by clicking the **Search** button. The results look like in Figure 34, which shows the references discussing Henry Fowler’s 1926 *A Dictionary of Modern English Usage*.

The full texts of any specific entry can be displayed as explained in example search 1.
REFERENCES


