Developing Conceptual Glossaries for the Latin Vulgate Bible

Andrew Wilson  
*Linguistics Department*  
*Lancaster University*  
*Lancaster LA1 4YT*  
E-mail: *eiaaw@exchange.lancs.ac.uk*

**ABSTRACT**

A conceptual glossary is a textual reference work that combines the features of a thesaurus and an *index verborum*. In it, the word occurrences within a given text are classified, disambiguated, and indexed according to their membership of a set of conceptual (i.e. semantic) fields. Since 1994, we have been working towards building a set of conceptual glossaries for the Latin Vulgate Bible. So far, we have published a conceptual glossary to the Gospel according to John and are presently completing the analysis of the Gospel according to Mark and the minor epistles. This paper describes the background to our project and outlines the steps by which the glossaries are developed within a relational database framework.
1. Introduction

A conceptual glossary, simply defined, is a reference work in which the words within a given text are classified according to their membership of a set of conceptual (i.e. semantic) fields. It also contains exhaustive references to the locations of words in the text, rather than giving just one or two illustrative citations for each meaning, thus it might crudely be thought of as a hybrid of an *index verborum* and a thesaurus, rather than as a hybrid of a *dictionary* and a thesaurus.¹

The development of conceptual glossaries was pioneered in humanities computing by Klaus Schmidt, who recognized the importance of being able to retrieve word occurrences from literary texts on the basis of related conceptual content rather than just on the basis of surface form (cf. Schmidt 1978). Since the 1970s, Schmidt and his collaborators have been building conceptual glossaries to a substantial number of mediaeval German epics (e.g. Schmidt 1980, 1993, 1994). In recent years, their work has also focussed on making this material available as an online full-text database – the *Mittelhochdeutsche Begriffsdatenbank* (Middle High German Conceptual Database) – which enables users to search for particular concepts and concept co-occurrences within the texts and to retrieve key-word-in-context concordances for these concepts.²

Since 1994, we have been working towards developing a set of conceptual glossaries for the Latin Vulgate Bible, along similar lines to those which Schmidt has built for his German texts.³ Although forms of conceptual analysis have also been applied to several other languages and texts⁴, actual conceptual glossaries are rather rare. Furthermore, very little conceptual analysis has been carried out on Latin texts and virtually none on later Latin texts.⁵ However, texts composed in Latin — for centuries the primary intellectual language of Europe — hold a central position within western intellectual and cultural history. To understand fully the language in which
these texts were written, and how this language evolved over time and across genres, detailed lexical resources and studies are required. Thus Latin, and perhaps more especially later Latin, should be a priority for treatment within the paradigm of conceptual analysis. Furthermore, the massive influence of the Vulgate Bible on religious and secular writing (both in Latin and other languages) and its consequent importance to scholars both in biblical studies and in other fields, suggest that this text in particular should be among the first which are subjected to a conceptual analysis.

The primary motivation for the present work was to take a first small step along the road to a quantitative empirical semantics of later, and especially biblical, Latin. More than twenty years ago, as an adjunct to the work of CETEDOC at Louvain-la-Neuve in producing frequency lists and concordances for later Latin texts, Tombeur (1978: 17) considered recording also "the semantic formula to which it [sc. a given lemma] belongs, and its fundamental meaning, synonyms and antonyms". This is very similar to the kind of analysis to be found in a conceptual glossary, and is an important development for Latin linguistics which needs to be pursued: quantitative, systematic treatments of word meaning in texts will allow us to answer questions such as:

— What are the most frequent and infrequent senses of words?
— What choices does the language provide for expressing concept $x$?
— How do words change their balance of meanings according to genre?
— When do they acquire or lose meanings?
— How and when do totally new concepts or expressions enter the language?

The importance of quantitative data in lexicography and lexicology is increasingly being recognized in work on modern languages (cf. Ooi 1998), but corresponding data do not yet exist for languages such as Latin. To arrive at the kind of databank required to answer such questions about the language and its genres as a whole — that is, a
larger scale conceptual frequency dictionary — it is first necessary to build up empirical treatments of individual texts (i.e. conceptual glossaries).

Apart from Fischer’s monumental key-word-in-context concordance of the entire Vulgate (Fischer 1977) and Bergren’s Greek-Latin index to the Vulgate New Testament (Bergren 1991), the amount of broad-coverage attention given to the lexis of the Vulgate (and its predecessor corpus, the Vetus Latina) has been relatively small. In the late 1980s and 1990s, however, a series of doctoral dissertations and other works at the University of Malaga began to develop lexica for some of the books of the Vulgate and the Vetus Latina: for instance, Vulgate Genesis (Sojo Rodríguez 1991), Vulgate Wisdom (Jimenez-Villarejo 1988), Vulgate Isaiah (Alfaro Bech 1989), the Vulgate versions of the Catholic Epistles and of the epistles of Jude, John, and Peter (Rodríguez Martín 1994, 1995a, 1995b, 1995c), and the Vetus Latina of Matthew (Ramirez Olid 1989) have been treated in this way. These lexica follow the traditional approach to philological lexicography, that is to say words are presented alphabetically, with articles discussing their usage, citing parallels in other ancient texts, and listing their locations in the texts. Our project adopts a different approach, which has the potential to supplement the information given in such traditional lexica by providing a conceptually-oriented perspective on the data and allowing the accumulation and comparison of frequency information on the concepts signalled by individual words.

2. Developing the Conceptual Glossaries

2.1 Text

The text used for constructing the conceptual glossaries is that commonly known as the “Stuttgart text” of the Vulgate (Weber 1969). As with the majority of ancient texts, the manuscript history of the Vulgate presents the editor in places with a number of
possible alternative readings, and the readings selected by different editors do, of course, vary. Consequently, the Stuttgart text differs from both the Clementine Vulgate, which functioned as the official text of the Roman Catholic Church for many years, and the Vatican’s *Nova Vulgata*, which has superseded it. Nevertheless, the Stuttgart text is an appropriate base for this project: first, it is also the text used for Fischer’s (1977) key-word-in-context concordance to the Vulgate, and its use in this work thus provides a degree of continuity with this important scholarly tool; second, it is probably the most widely used text of the Vulgate for scholarly work, that is, excluding use in the Roman Catholic liturgy; and third, the existence of such manuscript variants should not detract substantively from the usefulness of the glossaries: the reader will probably want to make use of these works in conjunction with either an edition of the Vulgate text or Fischer’s concordance, both of which indicate clearly where such variant readings occur.

### 2.2 Method

Our first Vulgate conceptual glossary (to the Gospel according to John – Wilson 2000) was produced between 1994 and 1996 on the basis of an initial *index verborum* generated using the Oxford Concordance Program (OCP). However, in more recent work, we have been making use of a full-text relational database, which greatly facilitates the various further processing steps, and it is this work that is described here.

Having separated out the machine-readable text of the relevant book, it is first reformatted so that each word of the running text occurs on a separate line, with its book, chapter and verse reference alongside it. The lines in this file are also given running serial numbers, which enable the text to be reconstructed in its linear order within the database.
Within the database\textsuperscript{6} we thus start off with a vertical text file as shown in Fig. 1:

\textbf{FIGURE 1 HERE}

\begin{quote}
Fig. 1 Initial database text file
\end{quote}

This text is then analysed linguistically, mostly by human analysts but with the help of existing machine-readable resources. The two procedures applied to the text are lemmatization and conceptual classification. As more texts are analysed, we are able to build up lemmatization tables (containing word forms and their corresponding headword forms) and conceptual tables (containing headword forms and their corresponding conceptual categories) to help in this process. However, in each new text, a number of words will remain unclassified and their headword forms and/or conceptual categories must therefore be entered manually; a further set of words will be contained in the existing tables but they may be ambiguous and must therefore still be reviewed by a human analyst. In any case, it is possible that a word is known to the system but is appearing for the first time as part of a new lemma or in a new sense. Fig. 2 shows an example of a lemmatization table and Fig. 3 an example of a conceptual table.

\textbf{FIGURE 2 HERE}

\begin{quote}
Fig. 2 Lemmatization table
\end{quote}

\textbf{FIGURE 3 HERE}

\begin{quote}
Fig. 3 Conceptual table\textsuperscript{7}
\end{quote}

\textbf{2.2.1 Lemmatization}
All main entries in the printed conceptual glossaries are lemmatized headwords (= traditional dictionary citation forms), with the inflexional variants presented in alphabetic order underneath each headword. In general, the lemmatization follows the normal practices for Latin, i.e. verbs are lemmatized to the 1st person singular present indicative active (except where deponent), nouns to the nominative singular (or plural, where the noun occurs only as a plural), adjectives to the nominative singular masculine positive, and so on. However, because our focus is on concepts rather than grammar, the different grammatical functions of homographs within lemmata — e.g. *bona* neut.pl.acc. vs. *bona* fem.sg.nom. — have not been identified, although the major parts of speech have been distinguished and given separate headword entries, e.g. *bonum* (n.) and *bonus* (adj.). Where total frequencies for a form are given, they thus refer to the frequency of that form in that part of speech (e.g. *bonum* as noun), not to the frequency of that form in any part of speech (e.g. *bonum* as both noun and adjective).

### 2.2.2 Conceptual Classification

Following lemmatization, the words in the text are disambiguated and classified according to the conceptual system.

#### 2.2.2.1 Conceptual System

The conceptual system represents the notional set of concepts which may be signalled by the words in the text. The system used in this project is a revised version of that used by Schmidt for his series of conceptual glossaries to mediaeval German epic. Schmidt’s system is, in turn, a revision of the conceptual system devised by Hallig and von Wartburg (1952).
Conceptual systems tend to differ in their arrangement and granularity rather than in their basic contents. The choice of system for a conceptual glossary is therefore more a pragmatic than a theoretical issue. The present system was adopted since (a) it was flexible enough to handle any amendments necessary for treating the world of the Bible, (b) it was not so excessively fine-grained as to result in an almost one-to-one mapping of words and categories, and (c) although some way from being able to be considered a standard, it had already been quite widely used and might thus form the basis of comparative and collaborative work in the future.

Different languages, periods, and texts may require some changes to existing conceptual systems in order for them to be treated adequately in conceptual glossaries. The original Hallig-Wartburg conceptual system was applied initially to the analysis of mid- to late-20th century French vocabulary, and Schmidt’s modifications of it are aimed at a better treatment of the world of the mediaeval German epic. For the Vulgate Bible a number of further modifications were clearly required. These have mainly consisted in amending those parts of the conceptual system that were culture-specific to the context of mediaeval epic. Some of these changes were relatively small: for example, under Foreign Languages, Schmidt’s subcategories of French and Arabic were replaced by, respectively, Greek and Hebrew/Aramaic. Similarly, the categories in Schmidt’s system which relate to chivalry were clearly irrelevant and were deleted. On a somewhat larger scale, the range of industries represented in the system was revised to correspond to those represented in the biblical texts. The classes delineated by Corswant (1960) were used as a basis for these categories. The mediaeval divisions of science and technology were also revised to represent the basic divisions of science in the Hebrew and Hellenistic worlds (cf. the works of Dupont-Sommer 1963 and Lloyd 1973).
The largest revision concerned, not surprisingly, the set of sub-categories under Mythology/Religion/Church. Schmidt’s system, and the original Hallig-Wartburg system, are oriented towards the Christian religion, as they deal with mediaeval and modern western European texts. The biblical world, however, is mainly Jewish, with a clearly separate Christian religion emerging only later. It seemed inappropriate in this context to consign all references to Judaism to an undifferentiated Non-Christian Religion category, whilst leaving Christianity well differentiated but empty. The divisions of Christianity used by Schmidt — for example, hierarchy, festivals, cult places — were therefore re-assigned to a category of Jewish and Judaeo-Christian Religion, which is intended to treat both purely Jewish concepts and those which are held in common with Christianity. The original Hallig-Wartburg category of Irreligion was also reinstated to replace Schmidt’s Theology.

Following Schmidt’s approach, the resulting conceptual system exists in two forms: a format numbered in a traditional way using decimal points, which structures the printed glossaries in a hierarchical manner, and a more concise format based on 8-digit Arabic numbers, which is used in the database and for cross-references in the printed glossaries. The latter format is also needed in order to ensure that the database table may be sorted correctly. Fig. 4 shows an extract from the conceptual system: the decimal-point-based structure is on the left and the Arabic numbers on the right.

**FIGURE 4 HERE**

Fig. 4 Extract from conceptual system

2.2.2.2 Classification
Each word in the text is examined in context by the analyst, disambiguated, and classified under one or more headings of the conceptual system. In general, we treat vocabulary at the level of the individual word, even where a strong collocational bond exists between two words (e.g. between testimonium and perhibeo, or between gratias and ago). However, a small number of words are listed in the form of phrases underneath a headword entry. These cases consist of: (a) certain proper names, where there is a particularly strong bond between the two elements (e.g. Mons Oliveti), and (b) terms which only make sense when presented in phrasal form (e.g. the special sense of amicus in the phrase amicus sponsi ≈ ‘best man’). It is recognized that the cooccurrence patterns of words are important items of information, but these should strictly be treated in collocational or valency dictionaries, rather than in conceptual glossaries.

2.2.2.3 Cross-referencing

Conceptual relationships between words are complex. Not only can the same word signal a variety of different independent concepts (e.g. terra may refer to the planet earth; land [as opposed to sea]; a geographical/political region; or the substance “earth”): the same word can also simultaneously represent more than one concept (e.g. caecus signifies both sight and illness/deformity). These instances of simultaneous conceptual content are a consequence of the fact that concepts do not appear to form discrete, Aristotelian categories in the mind. As the research of the prototype theorists has shown (e.g. Rosch 1977), mental categories seem rather to correspond to the theory of “fuzzy sets” (cf. Zadeh 1965, 1982), where some entities may belong clearly within a single category, but others may belong, to varying degrees, to more than one category.

The issue of fuzzy sets is handled through the cross-referencing into more than one category of the same occurrence of a word. Thus, for example, the same
occurrences of *caecus* are classified under both “Sight” and “Illness/Deformity”. This practice has two consequences. First, conceptual categories are represented in their entirety, so that all words which belong within them are actually categorized within them, i.e. no false exclusions have been made in order to achieve a correlation of one category per occurrence. Second, as a result of this, conceptual frequencies for each word do not always add up to the total number of occurrences of that word in the text: sometimes they will exceed that total. Thus, one cannot just take all the frequency information for a given word produce a straightforward “pie chart” from it, but one can take the individual word-concept frequencies and see what proportion of the total number of occurrences of the word signals that particular concept. It is considered that this exhaustive representation of meaning is more important than mathematical convenience.

### 2.2.2.4 Figurative Language

A further problem that is posed in classifying words for conceptual glossaries is that of how far one should treat figurative language such as metaphor and metonymy. For example, when Jesus refers to his disciples’ “bringing forth fruit”, should this expression be categorized in relation to the source domain (plants), the target domain (result), or both? In our project, the following approach has been adopted:

(a) if a word clearly and conventionally signifies a different meaning to the “core” sense of the word, the instance is treated in the same way as any other word that can have more than one meaning: thus, for example, *video* is assigned the possible senses of “see” [visually], “experience”, “understand”, and “visit”.

(b) if a figurative expression appears to be *consciously active*, and clearly possesses both figurative and non-figurative senses, the figurative sense is again treated in
the same way as (a) above, but it is additionally cross-referenced to both conceptual categories involved.\textsuperscript{10} Thus \textit{pater}, referring to God, is assigned both to Judaeo-Christian Religion and to Relations.

(c) if an expression makes sense figuratively only as a part of an entire “mental model” of a passage — as is the case, for example, with the word \textit{pastor} in the parable of the good shepherd — then it is treated literally.

\textbf{2.2.3 Frequency Analysis and Glossary Production}

At the end of the analysis stage, we have a text completely annotated with lemma headwords and conceptual categories, as shown in Fig. 5. For the printed glossaries, we then also need to extract the necessary frequencies from this. We extract:

- the frequency of each headword in the text
- the frequency of each inflexional variant in the text
- the frequency of headwords within conceptual categories
- the frequency of inflexional variant within conceptual categories

In the printed glossaries, the results of the conceptual analysis are then arranged in two ways: firstly according to conceptual categories, so that, for example, all words related to the concepts of ‘joy’ and ‘sadness’ are presented together (cf. Appendix A), and secondly in alphabetical order, so that individual words can be looked up and their semantic range examined (cf. Appendix B). As previously discussed, words are indexed primarily as lemmatized headwords, with the frequencies of the inflexional forms also shown beneath them. For every content word in the text, its conceptual frequencies are also shown, i.e. the fraction of the total occurrences of headword (or variant) \(X\) falling into conceptual category \(Y\). In the left-hand margin of the
conceptually arranged glossary, we indicate the other categories into which the word has been classified (the cross-references).

FIGURE 5 HERE

Fig. 5 Fully analysed text

3. Progress and Prospects

3.1 Project Publications

Our first conceptual glossary was to the Gospel according to John and was published in 2000 by Olms-Weidmann (Wilson 2000). With the help of a British Academy Small Research Grant\(^1\), we have recently completed the analysis of the Gospel according to Mark. We are presently undertaking the final checks and typesetting of this work, which we hope to have with the publishers by the end of 2002 (Wilson and Worth, i.p.). We are also in the final stages of analysing the Petrine Epistles (Wilson, i.p.) and have begun work on the Johannine Epistles and Jude.

3.2 Prospects for Computational Developments

The conceptual analysis of texts is a slow and labour-intensive process. Any procedure which can accelerate this task and reduce the labour requirement, without an appreciable loss of accuracy in the analysis, is therefore to be welcomed. The results of our work so far already go some way towards this goal by providing ready-made classifications for a number of words within the domain of the Latin Bible. However, further work is still required to facilitate the conceptual treatment of Latin texts.

One labour-saving procedure which has a good chance of success is automatic part-of-speech identification and lemmatization. Although morphological analysis and lemmatization software already exists for Latin (e.g. Bozzi 1990, Courtois 1996), it
typically, at present, does not work fully automatically: where only one possible part of
speech or lemma can be identified, it is able to resolve the case deterministically, but,
where more than one part of speech or lemma is possible, this is left for human analysts
to resolve from a set of possible options suggested by the program. Experience with the
analysis of English and other languages suggests, however, that software may be
developed to resolve automatically all parts of speech and lemmata in texts at an
accuracy rate in excess of some 97%.

The second labour-saving technique which should be developed for Latin is
automated, or semi-automated, word-sense discrimination. If means for the automated
disambiguation of a sufficiently large number of common words can be developed, it
will be much easier and quicker to process texts from their raw state into conceptual
glossaries. A number of methods for sense discrimination are possible. In the ACASD
project at Lancaster (Rayson & Wilson 1996), a high rate of success (ca. 91%) was
obtained, based primarily on empirical or intuitive sense-frequency rankings, including
the promotion of senses according to the domain of the text (for example, in a religious
text, mass would receive its religious sense, although in English overall it is perhaps
more likely to mean weight or quantity). Such global methods are, however, bound to
fail in a given number of cases, since they are only roughly constrained for context.
Alternative methods, which have greater contextual constraints, are those based on
templates of the surrounding context (both syntactic and semantic) and those which
attempt to model the surrounding context statistically. Statistical modelling for
semantic disambiguation is still an active research area and has not as yet demonstrated
high accuracy results on unconstrained text. However, contextual templates already
have a proven history of success (cf. Kelly and Stone 1975, Schmidt 1991) and it should
be a relatively straightforward task to define such templates for a number of high frequency ambiguous Latin words.

One possible application of conceptual glossaries such as these is in translation studies and multilingual lexicology. For example, one might wish to study the Latin Vulgate translation of the Gospel according to John in comparison with the original Greek version, Luther’s German translation, or the various Old English gospel translations. However, in order for all the necessary information to be made readily available to the scholar, this application of conceptual glossaries strictly requires glossaries to be available for all the versions of the text being studied, rather than for just one version. It might be thought that the production of such parallel glossaries would be as substantial a task as the production of, say, the original Latin glossary. In practice, however, the application of text alignment software — such as that developed by McEnery and Oakes (1996) — can greatly reduce the labour involved in generating such parallel glossaries, once a glossary to one version of the text has been produced.

The alignment software works by initially aligning — that is establishing a formal equivalence between — segments of text, such as sentences or verses, by means of simple statistical calculations based on their lengths and the a-priori likelihood of particular alignment types (e.g., one to one, two to one, one to two, etc.). Results on various language pairs have ranged from 54% to 100%, with most being in excess of 75% (McEnery and Oakes 1996). With the use of hard-coded regions, such as Bible chapters and verses, straightforward segment alignment should be able to approximate towards 100% success.

Once segment equivalences have been identified, the alignment software would then go on to attempt to indicate which words are equivalent to each other. It might do this by means of a number of methods: truncation (a rough procedure akin to
stemming); measures of the proportion of shared character pairs between two words; and cooccurrence measures, which show how significant the pairing of two words is between the two parallel texts. These techniques have so far shown a high degree of success: in experiments with French, Spanish, and English, pairings with shared character pair proportions of 0.7 and above (measured using a Dice [a.k.a. Sørensen] coefficient) have been correct translations in more than 80% of cases (McEnery and Oakes 1996).

By using methods such as these, a large proportion of the word-level translation equivalents in a pair of texts may rapidly be identified and, once this has been done, the words extracted from the new translation text may be linked directly to the conceptual classifications for the corresponding words in the original text. Of course, the validity of these classifications for the second language will need to be checked as the conceptual glossary is completed, but this is still likely to be a quicker process than attempting to classify words from scratch in a translation text. Preliminary experiments are currently being planned to investigate the feasibility of this approach.

4. Conclusion

Clearly, there is still a lot of work to be done before the Vulgate Bible, or even just the New Testament, is fully covered by a conceptual analysis, and we gladly invite other scholars to become involved in this effort. We will be very pleased to share our resources and expertise with those interested in undertaking the analysis of further books of the Vulgate. Ultimately, we would like to aim towards putting the conceptual databases online, in a similar way to the Mittelhochdeutsche Begriffsdatenbank.

The Mittelhochdeutsche Begriffsdatenbank can be found at: http://mhdbdb.sbg.ac.at/

I am grateful to the two research assistants who have worked with me on this project – Celia Worth and Aleksandra Tarkowska.

E.g. Shakespearean English (Spevack 1993); Mycenaean Greek (Kazanskiene and Kazanskij 1986); and New Testament Greek (Louw and Nida 1989).

Those conceptually-oriented works which do exist fall mainly into one of two categories: (1) traditional philological monographs, which use conceptual fields simply as a means of delimiting their scope (e.g. Crampon 1985), and (2) pedagogically-oriented works, which are not concerned with the analysis of texts (e.g. Caron 1992). A notable exception is Najock’s (forthcoming) conceptual index to Virgil’s Eclogues.

We are currently working with MS Access 2000 for Windows.

The field containing the conceptual categories has been called “Klauscats” for short, reflecting the source of the original 8-digit category format.

Trailing zeroes on the 8-digit numbers contained in the database are superfluous for the printed glossaries and are removed at the typesetting stage.

The full conceptual system can be examined in Wilson (2000) and also on the internet at: http://www.ling.lancs.ac.uk/staff/andrew/vulgate/glosscats.htm.

I use the terms “conventionally” (in point a) and “consciously active” (in point b) to avoid the mistake of considering conventionalized metaphors to be “dead” metaphors. Lakoff and Turner (1989), for example, show that conventionalized metaphors may still be active, even though they may be unconscious. However, if every conventionalized yet active metaphor were to be cross-referenced between its source and target domains, then much of the meaning discrimination work, which forms the basis of the conceptual glossaries, would be veiled. Metaphor is strictly the domain of a different kind of reference work.

SG-32004 (Conceptual glossary and index to the Vulgate translation of the Gospel according to Mark) to Dr. A. Wilson.
References


Wilson, A. (i.p.) *Conceptual Glossary and Index to the Vulgate Translation of the Petrine Epistles*.

Wilson, A. and Worth, C.A. (i.p.) *Conceptual Glossary and Index to the Vulgate Translation of the Gospel according to Mark*.


Appendix A

Annotated Sample Entry : Conceptual Index

<table>
<thead>
<tr>
<th>4.</th>
<th>War/Conflict</th>
<th>2324</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA)</td>
<td>General</td>
<td>23241</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>21072</th>
<th>alapa</th>
<th>**** 2 / 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>alapam</td>
<td>* 1 / 1</td>
</tr>
<tr>
<td>18.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>alapas</td>
<td>* 1 / 1</td>
</tr>
<tr>
<td>19.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

cross-reference(s)

text references (chapter.verse)

freq. of variant in this category / freq. of variant overall

**** cumul. freq. of lemma in this category / cumul. freq. of lemma overall
Appendix B

Annotated Sample Entry : Alphabetic Index

adorator
Respect/Admiration/Worship  2270407
adoratores  * 1 / 1  
****  1 / 1
Jew./Jud.-Chr. Religion: Rites And Cults  24455
adoratores  * 1 / 1  
****  1 / 1

* freq. of variant in this category / freq. of variant overall

**** cumul. freq. of lemma in this category / cumul. freq. of lemma overall
<table>
<thead>
<tr>
<th>Runner</th>
<th>Textrefs</th>
<th>Textwords</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1Pe-1:1</td>
<td>Petrus</td>
</tr>
<tr>
<td>2</td>
<td>1Pe-1:1</td>
<td>apostolus</td>
</tr>
<tr>
<td>3</td>
<td>1Pe-1:1</td>
<td>Iesu</td>
</tr>
<tr>
<td>4</td>
<td>1Pe-1:1</td>
<td>Christi</td>
</tr>
<tr>
<td>5</td>
<td>1Pe-1:1</td>
<td>electis</td>
</tr>
<tr>
<td>6</td>
<td>1Pe-1:1</td>
<td>advenis</td>
</tr>
<tr>
<td>7</td>
<td>1Pe-1:1</td>
<td>dispersionis</td>
</tr>
<tr>
<td>8</td>
<td>1Pe-1:1</td>
<td>Ponti</td>
</tr>
<tr>
<td>9</td>
<td>1Pe-1:1</td>
<td>Galatiae</td>
</tr>
<tr>
<td>10</td>
<td>1Pe-1:1</td>
<td>Cappadociae</td>
</tr>
<tr>
<td>11</td>
<td>1Pe-1:1</td>
<td>Asiae</td>
</tr>
<tr>
<td>12</td>
<td>1Pe-1:1</td>
<td>et</td>
</tr>
<tr>
<td>13</td>
<td>1Pe-1:1</td>
<td>Bithyniae</td>
</tr>
<tr>
<td>14</td>
<td>1Pe-1:2</td>
<td>secundum</td>
</tr>
<tr>
<td>15</td>
<td>1Pe-1:2</td>
<td>praescientiam</td>
</tr>
<tr>
<td>16</td>
<td>1Pe-1:2</td>
<td>Dei</td>
</tr>
<tr>
<td>17</td>
<td>1Pe-1:2</td>
<td>Patris</td>
</tr>
<tr>
<td>Lemma</td>
<td>Klauscats</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Abraham</td>
<td>24411100</td>
<td></td>
</tr>
<tr>
<td>Aenon</td>
<td>24222000</td>
<td></td>
</tr>
<tr>
<td>Andreas</td>
<td>23112500</td>
<td></td>
</tr>
<tr>
<td>Annas</td>
<td>23112500</td>
<td></td>
</tr>
<tr>
<td>Arimathia</td>
<td>24212000</td>
<td></td>
</tr>
<tr>
<td>Barabbas</td>
<td>23112500</td>
<td></td>
</tr>
<tr>
<td>Bethania</td>
<td>24222000</td>
<td></td>
</tr>
<tr>
<td>Bethlehem</td>
<td>24222000</td>
<td></td>
</tr>
<tr>
<td>Bethsaida</td>
<td>12021100</td>
<td></td>
</tr>
<tr>
<td>Bethsaida</td>
<td>24222000</td>
<td></td>
</tr>
<tr>
<td>Caesar</td>
<td>23112500</td>
<td></td>
</tr>
<tr>
<td>Caesar</td>
<td>24321000</td>
<td></td>
</tr>
<tr>
<td>Caiaphas</td>
<td>23112500</td>
<td></td>
</tr>
<tr>
<td>Cana</td>
<td>24212000</td>
<td></td>
</tr>
<tr>
<td>Capharnaum</td>
<td>24212000</td>
<td></td>
</tr>
<tr>
<td>Lemma</td>
<td>Textwords</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>ab</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>abs</td>
<td></td>
</tr>
<tr>
<td>abeo</td>
<td>abeuntes</td>
<td></td>
</tr>
<tr>
<td>abeo</td>
<td>abissent</td>
<td></td>
</tr>
<tr>
<td>abeo</td>
<td>abibant</td>
<td></td>
</tr>
<tr>
<td>abeo</td>
<td>abierant</td>
<td></td>
</tr>
<tr>
<td>abeo</td>
<td>abierio</td>
<td></td>
</tr>
<tr>
<td>abeo</td>
<td>abierunt</td>
<td></td>
</tr>
<tr>
<td>abeo</td>
<td>abii</td>
<td></td>
</tr>
<tr>
<td>abeo</td>
<td>abiit</td>
<td></td>
</tr>
<tr>
<td>abeo</td>
<td>abire</td>
<td></td>
</tr>
<tr>
<td>Abraham</td>
<td>Abrahae</td>
<td></td>
</tr>
<tr>
<td>Abraham</td>
<td>Abraham</td>
<td></td>
</tr>
</tbody>
</table>
II. Mind And Soul

A) General

B) Perception
   1. General
   2. Extra-Sensory Perception

C) Consciousness

D) Memory

E) Imagination

F) Thought
   1. General
   2. Concept/Idea/View
   3. Thought Process
   4. Opinion/Decision
      AA) Reasoning/Agreement/Disagreement
      BB) Truth/Falsehood/Lies
<table>
<thead>
<tr>
<th>Runner</th>
<th>Textrefs</th>
<th>Textwords</th>
<th>Lemma</th>
<th>Klauscats</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1Pe-1:1</td>
<td>Petrus</td>
<td>Petrus</td>
<td>23112400</td>
</tr>
<tr>
<td>2</td>
<td>1Pe-1:1</td>
<td>apostolus</td>
<td>apostolus</td>
<td>23307400</td>
</tr>
<tr>
<td>2</td>
<td>1Pe-1:1</td>
<td>apostolus</td>
<td>apostolus</td>
<td>23123100</td>
</tr>
<tr>
<td>2</td>
<td>1Pe-1:1</td>
<td>apostolus</td>
<td>apostolus</td>
<td>24452000</td>
</tr>
<tr>
<td>3</td>
<td>1Pe-1:1</td>
<td>lesu</td>
<td>lesus</td>
<td>23112400</td>
</tr>
<tr>
<td>4</td>
<td>1Pe-1:1</td>
<td>Christi</td>
<td>Christus</td>
<td>24410000</td>
</tr>
<tr>
<td>5</td>
<td>1Pe-1:1</td>
<td>electis</td>
<td>elego</td>
<td>22825000</td>
</tr>
<tr>
<td>6</td>
<td>1Pe-1:1</td>
<td>advenis</td>
<td>advena</td>
<td>23111100</td>
</tr>
<tr>
<td>6</td>
<td>1Pe-1:1</td>
<td>advenis</td>
<td>advena</td>
<td>24100000</td>
</tr>
<tr>
<td>6</td>
<td>1Pe-1:1</td>
<td>advenis</td>
<td>advena</td>
<td>31500000</td>
</tr>
<tr>
<td>7</td>
<td>1Pe-1:1</td>
<td>dispersionis</td>
<td>dispersio</td>
<td>21072000</td>
</tr>
<tr>
<td>7</td>
<td>1Pe-1:1</td>
<td>dispersionis</td>
<td>dispersio</td>
<td>31500000</td>
</tr>
<tr>
<td>8</td>
<td>1Pe-1:1</td>
<td>Ponti</td>
<td>Pontus</td>
<td>24120000</td>
</tr>
<tr>
<td>9</td>
<td>1Pe-1:1</td>
<td>Galatiae</td>
<td>Galatia</td>
<td>24120000</td>
</tr>
<tr>
<td>10</td>
<td>1Pe-1:1</td>
<td>Cappadociae</td>
<td>Cappadocia</td>
<td>24120000</td>
</tr>
<tr>
<td>11</td>
<td>1Pe-1:1</td>
<td>Asiae</td>
<td>Asia</td>
<td>24120000</td>
</tr>
<tr>
<td>12</td>
<td>1Pe-1:1</td>
<td>et</td>
<td>et</td>
<td>99999999</td>
</tr>
<tr>
<td>13</td>
<td>1Pe-1:1</td>
<td>Bithyniae</td>
<td>Bithynia</td>
<td>24120000</td>
</tr>
</tbody>
</table>