

Spelling variation in Codex Bezae

1. The Latin text

1.1. Identifying types of variation

Classification of spelling variants:

- Substitution of one letter for another
- Insertion of letters
- Doubling of letters
- Deletion of letters
- Undoubling of letters

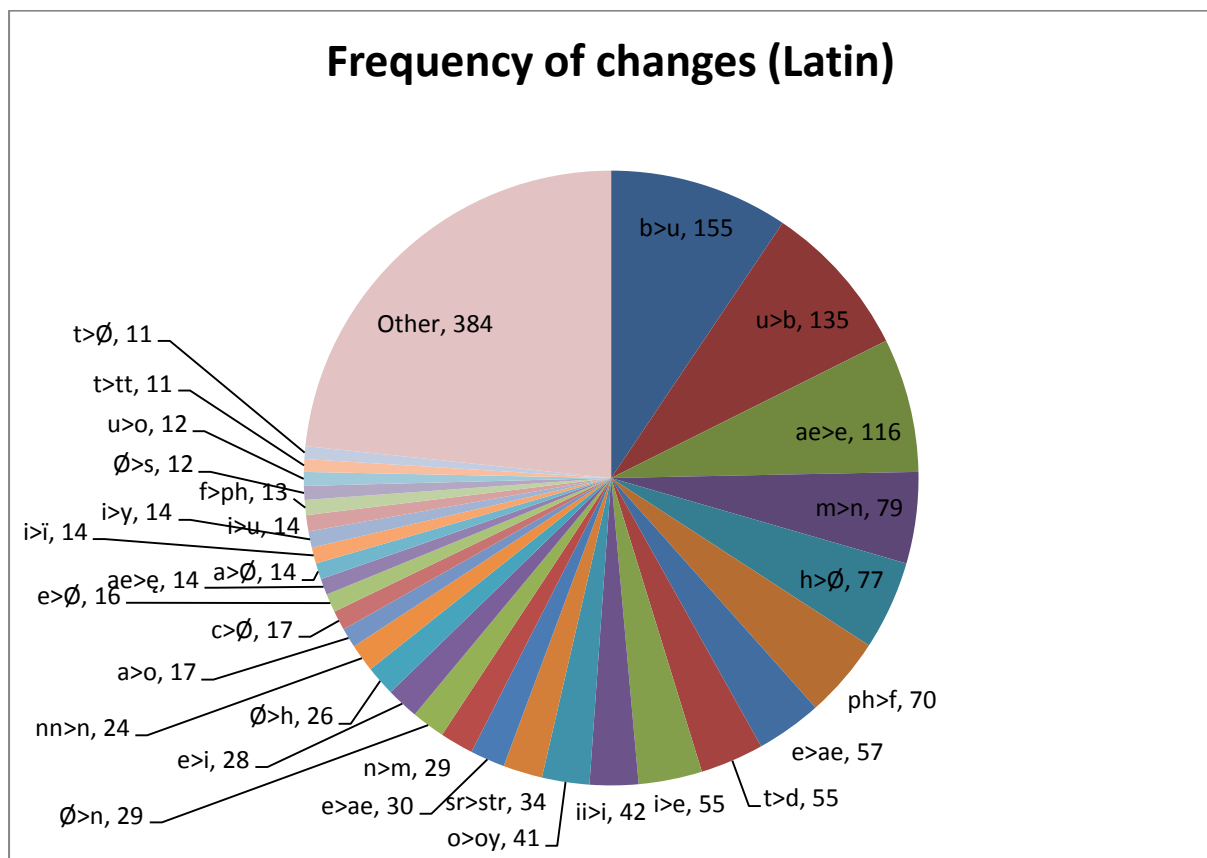


Chart 1

About 224 types of variant spelling in the Latin text altogether, affecting ~1624 words

- 3 types account for ~25%
- 10 types account for 50%
- 29 types account for 75%
- 384 (24%) of changes occur 10 times or fewer
- 208 (18%) of changes occur 4 times or fewer
- 130 types of variation occur only once
- 35 types of variation occur only twice

1.2. Identifying causes of variation

Reasons for variation:

- Phonological
- Graphological
- Morphological
- Lexical

- Borrowing (esp. names)
- Diachronic

Factors affecting variation with phonological/graphological causes:

- Neighbouring sounds/letters
- Other sounds/letters in the word
- Sounds/letters in neighbouring words?
- Vowel length, stress?

In several cases what may superficially appear to be a spelling variation likely has a purely morphological cause, with a different part of a word's paradigm being used with only a small change in meaning:

- e.g. 3ps. present indicative **-at** vs. subjunctive **-et** in the first conjugation
- other variants placed in this category include confusion of present and future tenses and of accusative and ablative cases
- these variants have not been considered here

Sometimes, however, although the form used is strictly part of the same paradigm, the different in meaning is considered great enough that a phonological explanation seems more likely:

- use of future tense ending **-bit** in place of perfect tense **-uit** and vice versa
- **-es** (nom./acc./voc. pl.) in place of gen. sg. **-is**
- certain uses of passive infinitive **-ri** for active infinitive **-re**
- + a small number of others
- these spellings *have* been included in the survey

1.3. Identifying patterns of variation

Some interesting patterns include:

*Variation between **b** and **u***

- **b > u** is particularly likely i_e, o_e, #_i, a_i, e_i, a_o
- **b > u** is 1.64 times more likely than chance in words containing another **b**, and only 0.55 times as likely in words containing **m**
- 73 instances (48%) of **b > u** affect inflectional endings
 - 26 of these affect the future ending **-bit** (cf. perfect **-uit**) (15% of instances of -bit affected)
 - Also examples of **-bimus > -uimus** and **-beris > -ueris**

- **-bo** and **-bitis** affected 24% and 27% of the time respectively
- All but two examples affect first conjugation verbs or forms of **VIDEO**
- 41% of **b > u** substitutions affect one of **BIBO, DO, LIBERO, PARABOLA** or **VIDEO**
- Words of the form **IVbVr-** account for 9% of the total
- **u > b** is particularly likely a_a, i_a, u_a, l_a, o_e, u_e, u_i and in words containing **b**
- It is less likely than might be expected in words containing **p** and **m**
- 25 instances (19%) affect inflectional endings, all but two affecting **-bit**
 - 92% affect first conjugation verbs, **INTROEO** or **EXEO**
- 40 instances (30%) affect words of the form **sv{r,l}u-** (**SERVO, SALVVS** etc.)

*Variation between **ae**, **e** and **ę***

- 52% of instances of **e > ae** affect the sequence **pre** or the sequence **que**
- 46% of instances of **ae > e** affect the sequence **prae** or the sequence **quae**
- 22% of instances of **QVAERO** are affected by **ae > e**, and make up 26% of instances of this substitution
- **ae > e** seems particularly likely in words containing another instance of **ae**
- The spelling **ę** for **ae** occurs 14 times in the manuscript's supplementary leaves

Variation between double and single letters

- about 83 cases where a letter that is written double in the Vulgate occurs singly in Codex Bezae
- 42 instances of **ii > i**, including 32 affecting **fili** (affected in 91% of possible instances)
- 24 cases of **nn > n**, all affecting **IOHANNES** (affected in 24% of possible instances)
- 25 instances of a letter being doubled
- Primarily lexically conditioned: **COTIDIE > cottidi-** in 11 instances, **camelum > camellum** in 4 instances, **osanna > ossan(n)a** in 4 instances

*Variation between **m** and **n***

- Two main types: variation affecting a prefix, and variation affecting final consonant (almost always on proper nouns)
- Wide variation between **im-/com-** and **in-/con-** spellings (chart 2) (also variation in spelling of these prefixes before **r-** and **l-**)
 - NB also **emmanuhel > inmanuel** (Mt 1:23)

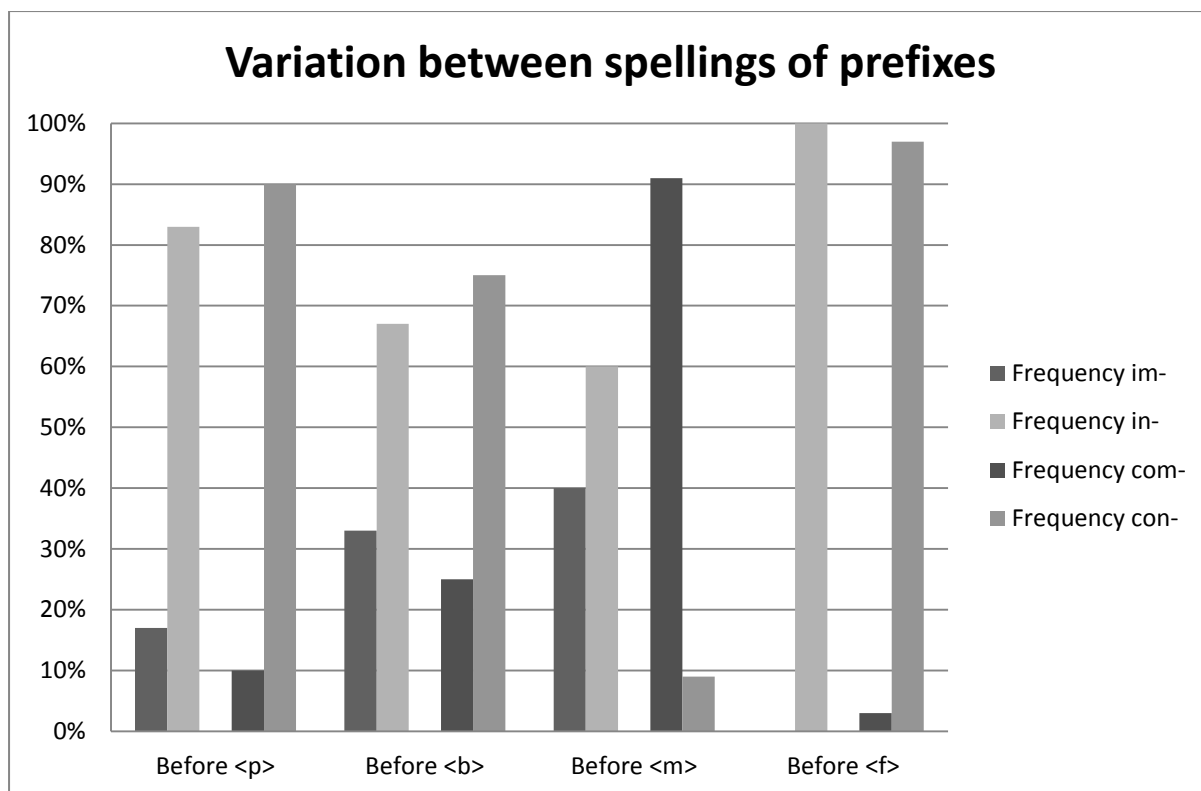


Chart 2

Deletion and insertion of h

- 52 instances (68%) of **h**-deletion affect proper names
- 28 involve the loss of initial **h**
- 22 instances of **th** > **t** (12 affecting the sequence **eth**); 6 of **ch** > **c**
- 12 instances of **h**-deletion before word-final **el** – excluding **is(t)rahel**, 92% of instances of final **-hel** are affected
- **h**-insertion occurs 26 times: 10 times initially and 11 times after **c**
- 17 instances (65%) of **h**-insertion affect names

Other spelling variation

- 71% of instances of the 70 instances of **f** > **ph** affect **capharnaum**, **PHARISAEVS**, **PROPHETA** or **PROPHETO**
- 85% of instances of **t** > **d** occur word-finally
- 73% of instances of **t** > **d** affect one of **at**, **caput**, **quot**, **quotquot** or **reliquit**
- About 38% of instances of **i** > **e** occur where there is some form of morphological alternation between these letters in a word's paradigm, mostly examples of **-es** being used in place of **-is**
- The form **-id-** in verbs is sometimes spelled **-ed-**, especially after **d** or before **i**
- 14 instances of **n**-insertion occur before **s**, in all but one of these the preceding letter is **e**

- There are also 9 instances of **THESAVRVS** > **thensaur-**; this word does not occur in Codex Bezae without
- **m** is lost before **pt** on 4 occasions (8% of possible instances)
- **MOSES** is spelled **moys-** in 98% of instances, encompassing all instances of **o** > **oy**
- The 34 instances of **sr** > **str** affect only **israhel** and **israhelitae**
- Of the 17 instances of **a** > **o**, 8 affect **NAZARENVS** and 6 affect **SALOMON**
 - Note that **NAZARETH** is always spelled with **a**
- Of the 14 instances of **a**-deletion, 7 affect **CAIAPHAS** and 4 affect **CLAVSVM**.
- There are about 14 examples of **i** being written **ī**
 - All but 3 affect names and all but 2 occur initially
 - There is also 1 instance of **y** > **ÿ**
- 11 out of the 14 examples of **u** > **i** affect **MONVMENTVM** (> **moniment-**)
- Of the 17 instances of **c**-deletion, 9 affect **ZACCHARIAS** and 5 affect **SANCTVS**
- **th** > **d** occurs 9 times, all affecting names containing **-eth**
- 8 of the 9 examples of **u** > **o** occur immediately adjacent to either **r** or **l**
- etc.

1.4. Defining search parameters

This information can be used to help in the design of a search facility that will pick up variant spellings.

At the simplest level, a search for a letter *X* which is frequently replaced by a letter *Y* in Codex Bezae can be constructed by allowing all searches for words spelled *aXb* (where *a* and *b* are strings of zero or more letters) to also return words spelled *aYb*. This works in some instances, e.g. a search for words containing **i** that also provides words spelled the same except that they contain **y** in place of **i** will not return any false results and will account for variant spellings such as **sytheon** for **simeon**.

However, in many instances such a simplistic substitution will return many false results, e.g. allowing searches for **e** to return all results with **a** will produce **cantum** for **centum**, **ago** for **ego** etc.

Ways around this – one or more of:

- Define longer substitution strings
 - E.g. instead of **e** returning all results with **a**, **ne** should return results with **na**, **disperg** should return results with **disparg** and **iaiu** should return results with **ieiu** (these specifications account for 15 out of 31 instances of **a** > **e** with very few unwanted results)
- Restrict substitutions to the beginnings or ends of words
 - E.g. searches for **d** should only return results with **t** word-finally
- Exclude certain results, e.g. **eth** should return results with **ed** but **sed** should be excluded (or else searches for **SETH** will return many unwanted instances of **SED**)

These points can be informed by the survey summarised in 1.3, by taking into account:

- Environments in which variant spellings are particularly common or uncommon

- Words which are affected particularly often

The crucial factor in determining whether substituting one string for another is the presence or absence of *minimal pairs*, words which differ from each other by one letter (or string of letters).

- If the number of minimal pairs for two strings is zero or very small (and the affected words are infrequent), then to allow one string to return results for another can be permitted with the knowledge that the number of false results will be negligible.
- Searches for minimal pairs were undertaken using the program Minimal Pair Finder. Example output: minimal pairs for **c** and **g**:

```

c          g
d i c i m u s      d i g i m u s
c 1          g 1
c 2          g 2
c 3          g 3
c 4          g 4
c l o r i a      g l o r i a
c l a d i u m    g l a d i u m
m a c i s      m a g i s
c r a e c i    g r a e c i
c r e c i s    g r e c i s
c r e c i s    g r e g i s

```

Resulting specification: words spelled with **c** should also return words spelled with **g** (to account for instances of **c > g** in the manuscript), though results with **gl** should be excluded

In some instances there will still be a small number of false results, e.g. searches for **VIVO** will give some false results with **BIBO** due to the rule that means searches for **u** also return results with **b**.

- However, no given rule will return more than about 3% false results if all words it will affect are searched for in turn (and most will give significantly less than 3%)
- Some individual words will give higher rates of false results, although care has been taken to ensure that the rate should never be so great as to overly inconvenience the user (i.e. not normally greater than 50%)
- Caveat: assumption that only words that actually occur in Codex Bezae will be searched for, though they may be searched for in the standard spelling where they only occur spelled otherwise
 - Potentially words which do not occur may be searched for and return 100% false results (e.g. a search for the nonce word **aego** will return over 200 false results with **ego**).

I have produced a set of 75 rules accounting for about 72% of variant spellings. The erratic nature of the variation in the text is such that it does not seem likely that a particularly more comprehensive set of rules could be produced. The 500 or so variants which are not accounted for must be dealt with in some other way (e.g. having the standard spelling marked directly in the XML).

Potential problem: words which require more than one substitution to be accounted for, e.g. **iohannem > iohannen (i > i̇, m > n)**

- Allowing multiple substitutions could potentially lead to false results: if “for X give Y” and “for Y give Z” are both allowed, searches of the form aXb might give unwanted results for aZb , or searches of the form $aXbYc$ might give unwanted results of the form $aYbZc$.
- A test run on a sample of 100 words suggests that this shouldn’t be a major problem, however
- Should it turn out to be, the number of words which vary from the standard spelling in more than one way that the rules I have drawn up will return is fairly small (around 50), so if the number of substitutions a word can undergo is limited to one this will not greatly restrict the power of the search engine
- It may be sensible in any case to restrict the number to two, as no more than two of the rules I have drawn up are needed in combination to return any of the non-standard spellings for which they can account

2. Extending the survey: the Greek text

2.1. Distribution of variants in the Latin text

Problem: time limitations prevented the production of a full survey of the Greek text. However:

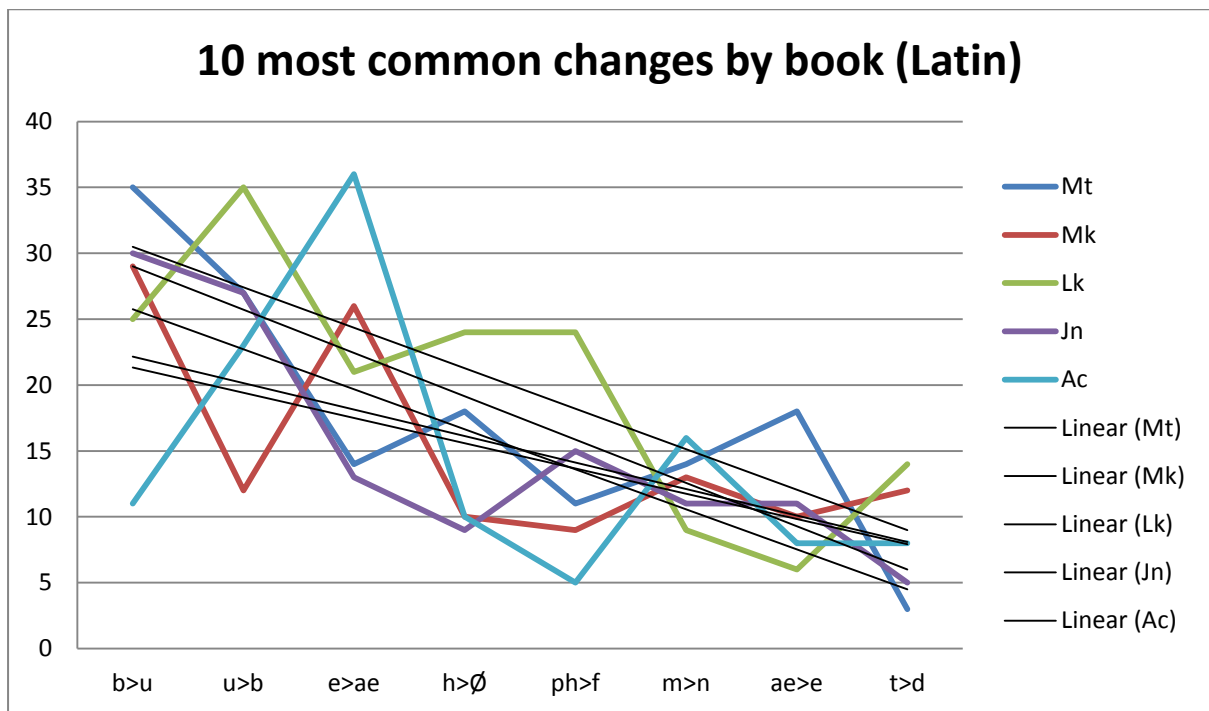


Chart 3(a)

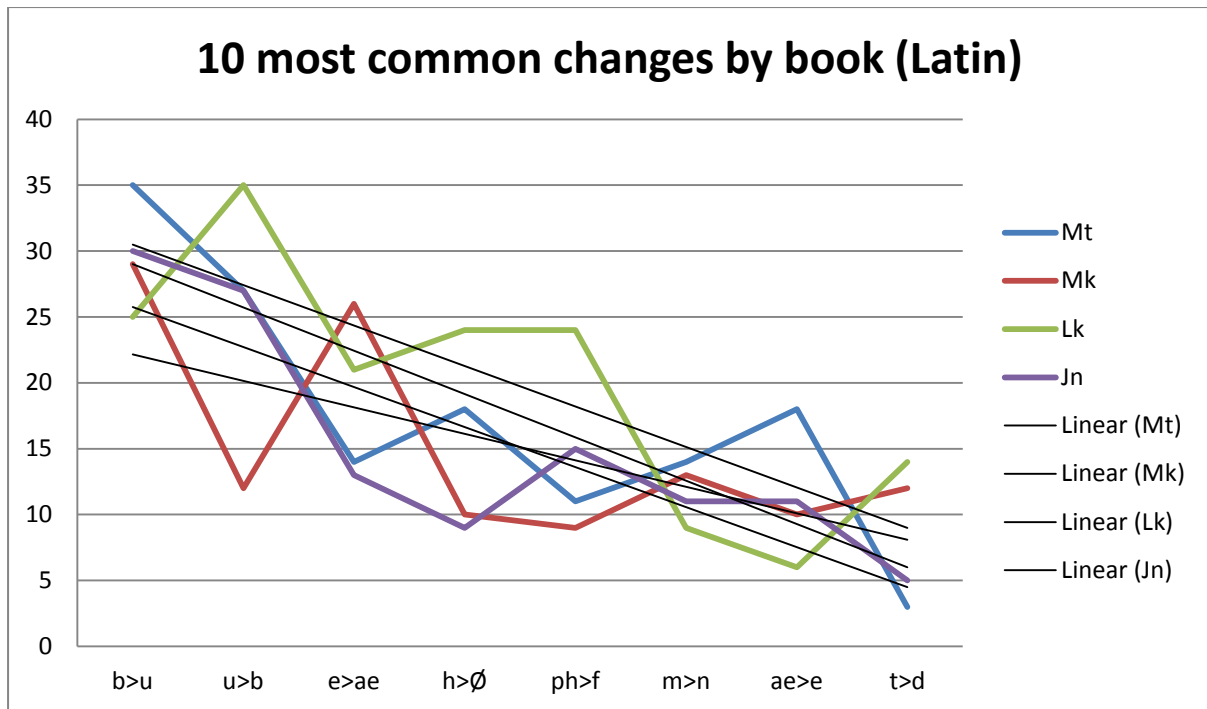


Chart 3(b)

The patterns of variation in the four gospels seem to be reasonably indicative of that in the MS as a whole, though Acts is less typical.

- Correlation coefficients for the Gospels: 0.81, 0.71, 0.82, 0.87
- Correlation coefficient for Acts: 0.58

This suggests one gospel in the Greek may provide a good sample of the entire Greek text.

2.2. Variation in the Greek text of Matthew and preliminary search specifications

Matthew: about 264 types of spelling changes

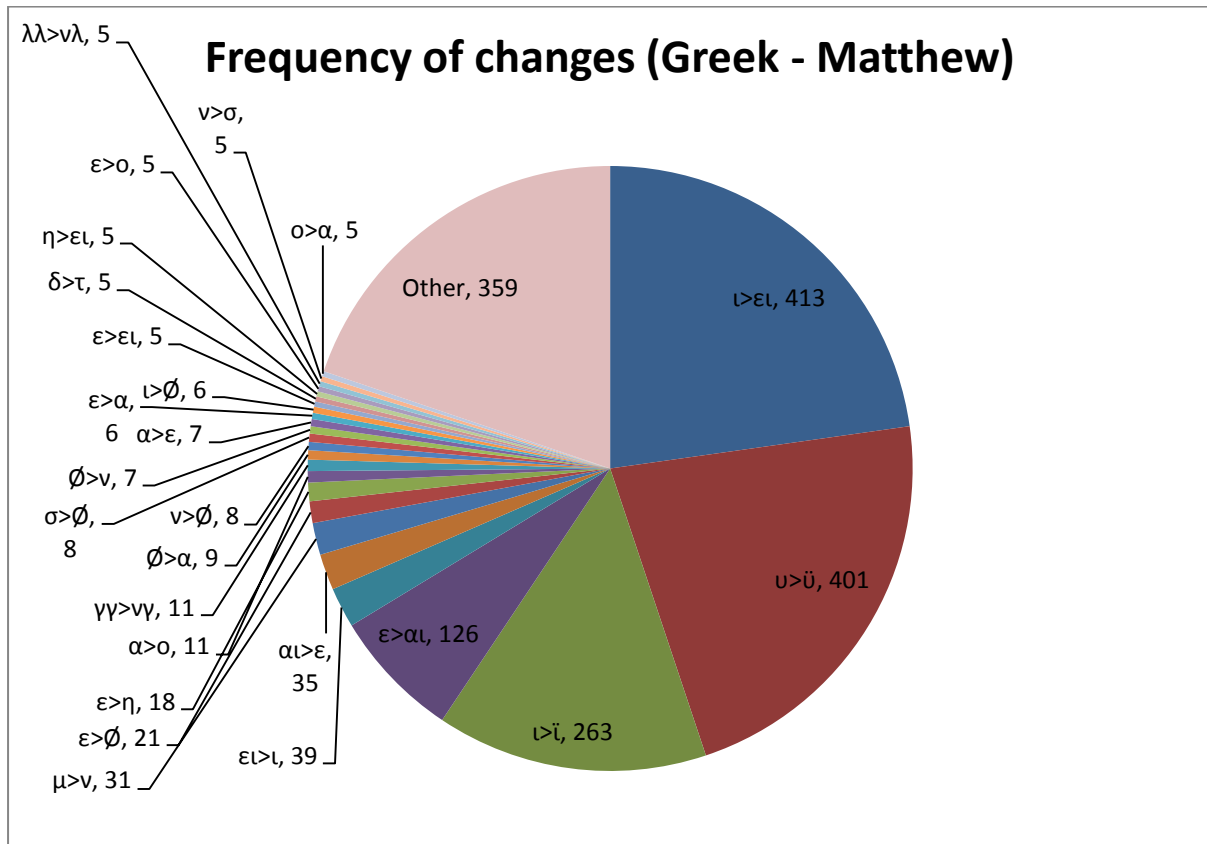


Chart 4

Note differences between this chart and the corresponding one for Latin (chart 1).

No in-depth survey of the patterning to the variation in the Greek text of Matthew has been undertaken.

- However, it has still been possible to produce search specifications using the Minimal Pair Finder and with a closer look at conditioning factors for certain changes where relevant.

I have produced a list of 35 rules which should account for 86% of variant spellings in Matthew (64% if the common substitutions ι > ι̂, υ > υ̂ and ι > ει are excluded).

- Number of false results these rules will return in Matthew is minimal.
- Based on the evidence of section 2.1, these rules should also account for most of the variant spellings in the remainder of the Greek text, although complete accuracy cannot be guaranteed.
- A more detailed survey of the rest of the Greek text would allow:
 - Identification of some patterns that are not apparent in Matthew alone, and the creation of new rules based upon these
 - Testing of the applicability of the rules drawn up for Matthew to the rest of the text (particularly with regards to the number of false results produced)