

# A Critical Examination of the Intertextual Phrase Matching Module in the *Thesaurus Linguae Graecae* and Its Relevance for Biblical and Patristic Studies

Ernst Boogert

Protestant Theological University, Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam  
eboogert@pthu.nl

**Abstract:** The rich availability of ancient Greek texts in the *Thesaurus Linguae Graecae* (TLG) has opened up new types of research for Biblical and Patristic scholars. A very helpful feature on the TLG's website is the option to trace quotations with the help of an n-grams module (Intertextual Phrase Matching). However, it is virtually unknown how well this module performs and what scholars might expect from the results it produces. The core of this article, therefore, is devoted to a critical examination of the algorithm and of its results. The gospel according to John has been compared with the *Paedagogus* of Clement of Alexandria, with the gospel according to Matthew, and with the complete works of Plutarch. As it turns out, the algorithm performs well in cases of longer quotations with no or very few interpolations. Short quotations, however, are missed while interpolated or adapted quotations are poorly handled. It is suggested that the algorithm might perform better if the team of the TLG were to revisit its decision to ignore stopwords and if the algorithm were to allow for foreign words in its n-grams. Finally, it is advised that more transparency in the algorithm's mechanisms and a possibility for manually tuning its parameters might improve its applicability.

**Keywords:** Thesaurus Linguae Graecae, TLG, N-grams, Intertextual Phrase Matching, Patristic Quotations, Gospel According to John, Gospel According to Matthew, Plutarch

## 1. Introduction

At the beginning of 2015, the team of the *Thesaurus Linguae Graecae* (TLG) at the University of California, Irvine, released a new website, featuring a number of new tools. Next to their brand-new analytics the site now has an n-grams module, the so-called Intertextual Phrase Matching (IPM) module.<sup>1</sup> Both projects are to be developed further in the years to come.<sup>2</sup> Since the IPM module has already been available for four years now, it is an appropriate point in time to evaluate whether it provides useful materials for biblical and patristic studies and whether it deserves some finetuning in the near future.<sup>3</sup>

---

\* I like to thank Arian Verheij ([www.arianverheij.nl/](http://www.arianverheij.nl/)) for his numerous constructive comments.

<sup>1</sup> In this article, the term 'n-grams' refers to the mechanics of intertextual phrase matching, while the term 'IPM module' refers to the application of n-grams as a separate analytical tool. To compare texts in the TLG, one has to click the 'N-GRAMS' tab: <http://stephanus.tlg.uci.edu/Iris/inst/ngram.jsp> (personal login required).

<sup>2</sup> Cf. <http://stephanus.tlg.uci.edu/help.php> (accessed November 16, 2018). While the present article was already in the phase of correction, the TLG's website was updated (November 14, 2018), including its IPM module. Since most of IPM's basic mechanisms and the accompanying documentation did not change, the article has basically remained the same. A quick analysis of the new results has shown that most of the evaluative remarks and recommendations still stand.

<sup>3</sup> The only review of the module that I could find is in a blog post written by an unknown author (~J.): <https://libraryofantiquity.wordpress.com/2016/07/22/help-with-greek-texts-the-new-tlg-part-4-n-grams/> (published on July 22, 2016 and last accessed on August 21, 2018). Cf. also the remarks in <https://classicalstudies.org/scs->

The desirability and huge potential of intertextual phrase matching (i.e. discovering linguistic parallels between texts) is obvious, particularly in biblical and patristic studies. Patristic scholars may remember the time spent in skimming indexes of text editions to find exegetical notes on biblical passages. Biblical scholars – especially in textual criticism – may be familiar with the effort needed to track down patristic quotations referred to in critical apparatuses of the New Testament but lacking precise reference. Now the IPM module of the TLG promises to speed up such processes considerably. Currently, a check as to whether a text contains biblical quotations is only a matter of seconds and the automatic construction of quotation indexes might be within the reach of what is technically possible. This is really impressive, and the team of the TLG deserves much credit for making this technology available.

However, despite the seemingly clear results as they are displayed to the user, it is virtually unknown how accurate these results are. The sparse documentation on the website does not give much indication of the outcome to be expected. Hence, a judicious approach might lead to two types of use: 1) The module will be used, and its results checked by hand,<sup>4</sup> or 2) the module will be used in addition to traditional ways of tracing intertextuality. In both approaches, much of the time-saving will be lost.

This article has two purposes. First, I intend to introduce the IPM module to biblical and patristic scholars, because it provides them with a powerful tool to find and study parallel phrases in ancient Greek texts from Homer (8<sup>th</sup> c. BCE) up to the fall of Byzantium in 1453 CE and beyond; including the Greek Bible and a substantial amount of Greek Patristic texts.<sup>5</sup> Second, I will evaluate some of the results produced by this module, with a view to understanding and improving the way it works.

To initiate the reader into the world of intertextual phrase matching, I will first discuss the use of n-grams in general and its application in the TLG (§2), since understanding this mechanism illuminates its evaluation considerably. Next, I will present and evaluate the results produced by the IPM module in relating the gospel according to John to the *Paedagogus* of Clement of Alexandria, the gospel of Matthew, and all the works of Plutarch (§3). Then, I will check the results of the comparison with the *Paedagogus* against a comprehensive and accurate manual index of gospel passages in that text (§4). Finally, I will present some recommendations for the improvement of the IPM module (§5).<sup>6</sup>

## 2. N-grams in the *Thesaurus Linguae Graecae* (TLG)

The n-gram is one of the most important elements in present-day machine language learning. Although its name has been derived from the Greek word γράμμα ('letter/character'), it usually refers to collocated words, the 'n' indicating the number of words taken at a time. For instance, the sentence "This is a tetragram" can be cut into three pieces of two-word sequences, starting at the first, second, and third words respectively: "This is", "is a", and "a tetragram" (cf. Figure 1). Such pieces are called "bigrams" since they consist of two collocated words. Similarly, the sentence can be cut into two

---

[blog/scott-farrington/review-thesaurus-linguae-graecae](http://blog/scott-farrington/review-thesaurus-linguae-graecae) blogpost written by: Scott Farrington "Review: Thesaurus Linguae Graecae", Januari 2, 2017 (last accessed on August 21, 2018).

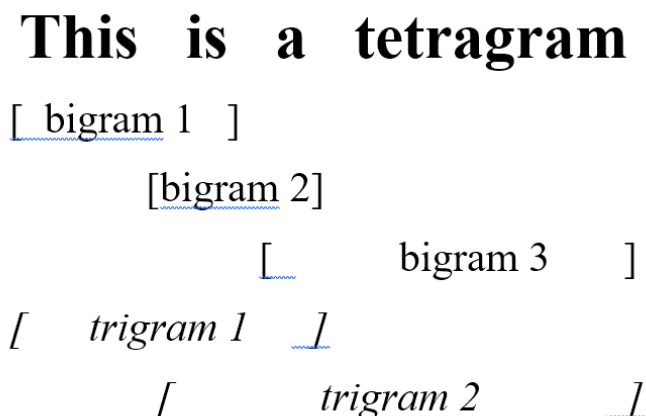
<sup>4</sup> Apart from the time needed to check the results, the most problematic side of this approach is that false negatives (real quotations *not* found by the IPM module) are overlooked.

<sup>5</sup> The process of digitizing Greek texts still continues. At the moment of writing, the TLG contains almost all preserved Greek texts from Homer till the fall of Byzantium and a large number of texts beyond that period (10,000+ works). <http://stephanus.tlg.uci.edu/history.php> (accessed November 16, 2018).

<sup>6</sup> *The Greek materials used in this article are copyrighted by the team of the TLG and the Regents of the University of California and are printed with the explicit permission of the director of the institute: Maria Pantelia. Cf. <http://stephanus.tlg.uci.edu/site.php> (last accessed on October 22, 2018).*

trigrams (“This is a”, “is a tetragram”) and even into four monograms (“This”, “is”, “a”, “tetragram”) or one tetragram (“This is a tetragram”). If another text is compared to this sentence, the computer is instructed to cut its sentences into n-grams of the same size and subsequently to look for identical n-grams and for sequences of identical n-grams. Matching is therefore defined by a number of parameters such as the length of the n-grams and the number of consecutive n-grams required to establish a match.

Figure 1. The concept of n-grams



There are several reasons why n-grams are so useful in comparing texts (cf. Figure 2). First, unlike looking for identical sentences, n-grams provide full flexibility in defining the place where a match may start or end.

Second, the principle of n-grams allows for other types of flexibility too. One can choose for instance to compare n-grams irrespective of word-order or to convert the actual wordforms to their dictionary lemmata. The TLG has implemented both additional techniques.<sup>7</sup> The use of unordered n-grams rather than ordered n-grams is very useful in the case of ancient Greek since this language’s case-system allows for more variation in word order than is permitted in many other languages. The TLG has also implemented a lemmatizer which converts the actual wordforms to their lemmata.<sup>8</sup> This allows for differences in inflection, so as not to tie down the comparison to exact wordforms or specific orthography. Both techniques provide the algorithm with considerable flexibility.

Third, because the text is scanned to build the n-grams, it is also possible to apply word filters. A much-used technique in digital language processing is the use of stopword lists to clean a text from its least meaningful parts. The TLG has employed this technique, which means in practice that only n-grams of ‘non-stopwords’ or ‘content words’ are used to establish matches (cf. Figure 2).<sup>9</sup>

<sup>7</sup> <http://stephanus.tlg.uci.edu/helppdf/ngrams.pdf>, 2 (accessed November 16, 2018). As far as I can judge, the use of unordered n-grams only means that the sequence of words *within* n-grams does not matter. N-grams themselves need to stand in the same sequence to constitute a match.

<sup>8</sup> <http://stephanus.tlg.uci.edu/helppdf/ngrams.pdf>, 2 (accessed February 27, 2018).

<sup>9</sup> In my experience, the following word categories are ignored: definite articles, (all types of) pronouns, particles, conjunctions, numerals, adverbs, and the verb εἶμι. Thus, only sequences of nouns, adjectives, and verbs appear to be used to determine a match. Composite verbs, are reduced to their simple form (eg. εἶσ-έρχομαι = ἔρχομαι). In the documentation, however, it is not clearly specified which words or word classes are meant by “stopwords.” Since the 14<sup>th</sup> of November 2018, the user has the choice between comparing exact wordforms

Figure 2. The concept of n-grams matching in the TLG<sup>10</sup>

John 1:1 (original) Ἐν ἀρχῇ ἦν ὁ λόγος, καὶ ὁ λόγος ἦν πρὸς τὸν θεόν, καὶ θεὸς ἦν ὁ λόγος

Clement Pae. 1.8.62.4 (original) ἐν ἀρχῇ ὁ λόγος ἦν ἐν τῷ θεῷ, καὶ θεὸς ἦν ὁ λόγος

John 1:1 (lemmatized, no stopwords) ἀρχή λόγος λόγος θεός θεός λόγος

Clement Pae. 1.8.62.4 (lemmatized, no stopwords) ἀρχή λόγος θεός θεός λόγος

Trigrams John	Trigrams Clement
ἀρχή λόγος λόγος	ἀρχή λόγος θεός
λόγος λόγος θεός	
λόγος θεός θεός	λόγος θεός θεός
θεός θεός λόγος	θεός θεός λόγος

Fourth, the TLG uses so-called ‘skips’ to allow for interpolations between n-grams in its comparisons. This means that, to some extent, words can be passed over if one of the texts has an addition compared to the other text without losing that place of comparison as a match. This only applies to words that are preceded and followed by at least one matching trigrams.<sup>11</sup>

By using n-grams and additional techniques as just outlined, the TLG has created a flexible comparison algorithm.<sup>12</sup> Users of the TLG can take advantage of its benefits in three different settings:

- **IPM module** (intertextual phrase matching): In this module, the user can select a book or collection (by author) and compare it with another book or collection. The comparison is performed by using trigrams only.<sup>13</sup>

or lemmata, which I consider an important step forward, but it is still not possible to include or exclude stopwords.

<sup>10</sup> Underlining indicates a matching trigram; a dashed line indicates a matching trigram but with different word order. The skip indicates that one unmatched word (red marked word) stands in between the previous and next matching trigram.

<sup>11</sup> I do not know how many content words may stand inbetween, since it is nowhere specified in the documentation. In my experience, however, the TLG deals with interpolations of one content word accurately if it is preceded and followed by a matching trigram.

<sup>12</sup> According to the “Credits and acknowledgements”, it is based on the code of PhiloLine version 0e (developed in 2008-2009); a project of the American and French Research on the Treasury of the French Language (ARTFL). The TLG has customized the code to match the requirements of the Greek language. See <http://stephanus.tlg.uci.edu/credits.php> and <https://code.google.com/archive/p/text-pair/downloads> (both accessed November 5, 2018).

<sup>13</sup> In the version of the TLG that will be analyzed in the following sections, the minimum was set to two trigrams (= four content words), while the maximum was set to four trigrams (= six content words). In the updated version, however, the minimum is set to one trigram (= three content words). As far as I can conclude

- **Browse module** (browse one text): In the browse mode, the n-grams tool is one of the instruments that can be drawn upon to analyze the text (next to a translation, morphological analysis, and lexical tools). After activating the n-grams tool, the text displayed will be analyzed against the whole corpus in the TLG's database. In this mode, both bi- and trigrams are used. Moreover, the user can select the option to perform the comparison on the basis of exact wordforms instead of on lemmata. In the wordform mode, only exact repetitions will be found.
- **Browse module** (parallel browsing): In the parallel browsing mode, the user can analyze the differences and similarities between two texts. In this mode, matches are usually based on a minimum of two trigrams.<sup>14</sup> Moreover, the user can compare the texts in three different settings:
  - Browse two texts: the user is allowed to display and browse in parallel two different texts from the entire TLG database.
  - Highlight similarities: this mode is the same as the previous one, but with matches (based on n-grams) highlighted.
  - Compare editions: This mode is restricted only to those Greek texts that are available in more than one edition. It provides an exact comparison between these editions based on exact wordforms, including features such as accents, capitals, beta escapes, and punctuation. This exact comparison does not use n-grams.

In the following analysis and evaluation, the IPM module has been used to compare various texts with the text of the gospel of John, with the aim of tracing textual matches.<sup>15</sup> Since this gospel has a distinct stylistic and theological profile within the New Testament, and since it is extensive enough to find it quoted or alluded to in a considerable part of later Christian literature, it appears to be an excellent choice for such a comparison.

### 3. Analysis and Evaluation of the Results of the Intertextual Phrase Matching (IPM) Module

What types of matches does the IPM module detect and how can they be characterized? In this section, I compare texts of three different genres with one base text: the gospel according to John. The main point of concern is what types of matches are found and whether these findings can be considered accurate from various perspectives depending on the type of comparison. When appropriate, observations are accompanied by evaluative remarks.

---

from the new results, this decision has primarily increased the number of false positives, at least for the texts that have been analyzed for this article.

Since the last update, the user can manually adjust the maximum number of n-grams from 1 to 6. A higher number, however, does not mean that smaller matches will not be found. It only means that these n-grams are *concatenated* in a match, resulting in a maximum of seven matching content words. The user is therefore advised to set the number of n-grams always to the highest level, because otherwise longer matches will not be concatenated, but presented as separate matches. However, I do also not see why the concatenation is limited to such a low range, since longer matches cannot be expected to fit into it.

<sup>14</sup> An exception will be made if one of the texts is very short. In that case, bigrams consisting of two collocated words will be used. Cf. <http://stephanus.tlg.uci.edu/help/pdf/ngrams.pdf>, 6 (accessed November 16, 2018).

<sup>15</sup> The only edition of the New Testament in the database of the TLG is UBS<sup>2</sup>: K. Aland, M. Black, C.M. Martini, B.M. Metzger, and A. Wikgren, *The Greek New Testament*, 2nd edition., Stuttgart: Württemberg Bible Society, 1968. All quotations from the gospel of John are taken from this edition.

The first comparison concerns the *Paedagogus* of Clement of Alexandria.<sup>16</sup> The primary aim of this comparison is to trace quotations of the gospel of John. Clement is an interesting case since he is known for his free way of quoting the Scriptures.<sup>17</sup> The analysis, therefore, is also an initial check whether the TLG can detect loose quotations. The second comparison concerns the gospel of Matthew, a work of the same genre as John's. Since Matthew does almost certainly not quote John (or vice versa), the goal is to investigate what types of matches are found in cases in which direct literary dependency is absent. The third analysis concerns a comparison with the huge legacy of Plutarch, a contemporary of John, to investigate the matches found that cannot point to any form of textual dependency.

### 3.1. Quotations of John in the *Paedagogus* of Clement of Alexandria

The *Paedagogus* of Clement of Alexandria is a voluminous work on Christian ethics.<sup>18</sup> The intertextual phrase matching tool lists 29 matches between this work and John.<sup>19</sup> In my analysis, I used three context lines to be able to evaluate the matches in the light of their contexts.<sup>20</sup> If still more context was needed, I switched to the parallel browsing mode with “highlighting similarities” enabled.

While studying the matches, several peculiarities attracted my attention:

1) All matches found by the TLG are explicitly marked in the edition of *Paedagogus* by quotation marks. Consider for instance match no. 6 in Table 1.

**Table 1. Match between John 3:36 and *Paedagogus* 1.6.29.2**

6. **N.T.Ev.Jo. {0031.004}** 3.36.2

τὸν υἱόν, καὶ πάντα δέδωκεν ἐν τῇ χειρὶ αὐτοῦ. (36) ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον· ὁ δὲ ἀπειθῶν τῷ υἱῷ οὐκ ὄψεται ζωὴν, ἀλλ' ἡ ὀργὴ τοῦ θεοῦ μένει ἐπ'

6. **Clem.Al.Paed. {0555.002}** 1.6.29.2.1

πιστεύομεν. Πίστις γὰρ μαθήσεως τελειότης· διὰ τοῦτο φησὶν (2) «ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον». Εἰ τοίνυν οἱ πιστεύσαντες ἔχομεν τὴν ζωὴν, τί περαιτέρω τοῦ κεκτησθαι

2) Due to the IPM ignoring stopwords, such words are not highlighted if they stand before or after the matching trigram(s). Consider for instance the non-highlighted article at the start of the quotation in Table 1. Sometimes the loss is more substantial, although the matching parts can easily be retrieved from the context. A case in point is the match of John 6:40 (Table 2), in which the phrase τοῦτο γὰρ ἐστὶν τὸ, in spite of its being a verbatim parallel – except for the final nu in ἐστὶν – is not highlighted.<sup>21</sup>

<sup>16</sup> The *Paedagogus* together with the *Protrepticus* and *Stromata* are the major work (trilogy) of Clement of Alexandria on Christian theology and ethics. The TLG uses the edition of Sources chrétiennes: M. Harl, H.-I. Marrou, C. Matray, and C. Mondésert, *Clément d'Alexandrie. Le pédagogue*, 3 vols. [Sources chrétiennes 70, 108, 158], Paris: Éditions du Cerf, 1:1960; 2:1965; 3:1970. It was originally published around 198 C.E.

<sup>17</sup> According to Cosaert, Clement cites the gospels from memory most of the time, although some quotations are more exact, especially those introduced by a formula such as εὐαγγελίῳ (“according to the gospel”) or φησὶν ὁ κύριος (“as the Lord said”). Cosaert, *The Text of the Gospels in Clement*, 30–31.

<sup>18</sup> According to the statistics module in the TLG, the *Paedagogus* comprises 56,766 words, which is more than three times as many as the gospel of John (15,635 words). <http://stephanus.tlg.uci.edu/Iris/inst/stat.jsp> (Accessed 4 April, 2018).

<sup>19</sup> See the Appendix for a full list of matches found by the TLG.

<sup>20</sup> I was offered the choice between using one context line (standard) or three. Using three context lines means that the results are displayed with one line of additional text before and after the match.

<sup>21</sup> Other examples of substantial omissions are found in matches nos. 1 (ἐν ἀρχῇ ὁ), 2 (οὐδὲ ἐν), 4 (ἵνα πᾶς ὁ), 10 (εἰς τὴν ζωὴν), 11 (Οὐ γὰρ Μωσῆς), 13 (τοῦτο γὰρ ἐστὶ τὸ), 22 (ἐγὼ εἰμι ἡ), 25 (εἰς ἐν). Minor ones occur in nos. 3 (ἡ), 5 (τὸν), 6 (ὁ), 8 (με), 17 (Καὶ ὁ), 24 (ἵνα). Cf. the Appendix.

Conversely, stopwords occurring within or between matching n-grams are highlighted even if they do not themselves match.<sup>22</sup>

**Table 2. Match between John 6:40 and *Paedagogus* 1.6.28.5**

13. **N.T.Ev.Jo. {0031.004}** 6.40.2

αὐτὸ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ. (40) τοῦτο γὰρ ἐστὶν τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων εἰς αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσῃ αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.

13. **Clem.Al.Paed. {0555.002}** 1.6.28.5.5

κύριος σαφέστατα τῆς σωτηρίας τὴν ἰσότητά ἀπεκάλυψεν εἰπών· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ (5) θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν αἰώνιον, 1.6.29. (1) καὶ ἀναστήσῃ αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.» Καθ' ὅσον μὲν οὖν

3) Since the module uses a maximum sequence of four trigrams with content words, larger matches are broken up.<sup>23</sup> In Table 2, the highlighted part of the comparison ends in the middle of a sentence (at the word ἔχη) although the match still continues verbatim (cf. no. 15 in the Appendix). In *Paedagogus* 1.8.71, Clement has a lengthy quotation from the High Priestly Prayer, containing John 17:24–26a. In the IPM module, however, this quotation appears as four unique matches (nos. 26–29; cf. Appendix). A lookup of the passages in the parallel browsing module demonstrates the unity of these matches (Table 3).<sup>24</sup>

**Table 3. John 17:24–26 and *Paedagogus* 1.8.71.2–3 in the parallel browsing mode (matches nos. 26–29)**

**NOVUM TESTAMENTUM, *Evangeliū secundum Joannem*. {0031.004}** 17.24-26

αὐτοὺς καθὼς ἐμέ ἠγάπησας. (24) Πάτερ, ὃ δέδωκάς μοι, θέλω ἵνα ὅπου εἰμι ἐγὼ κἀκεῖνοι ὧσιν μετ' ἐμοῦ, ἵνα θεωρῶσιν τὴν δόξαν τὴν ἐμὴν ἣν δέδωκάς μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. (25) Πάτερ δίκαιε, καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω, καὶ οὗτοι ἔγνωσαν ὅτι σύ με ἀπέστειλας, (26) καὶ ἐγνώρισα αὐτοῖς τὸ ὄνομά σου καὶ γνωρίσω, ἵνα ἡ ἀγάπη ἣν ἠγάπησάς με ἐν αὐτοῖς ἦ καὶ ἐν αὐτοῖς.

**CLEMENS ALEXANDRINUS, *Paedagogus*. {0555.002}** 1.8.71.2-3

λέγων «πάτερ, οὗς ἔδωκάς μοι, θέλω ἵνα ὅπου εἰμι ἐγὼ, κἀκεῖνοι ὧσιν μετ' ἐμοῦ, ἵνα θεωρῶσιν τὴν δόξαν τὴν ἐμὴν, ἣν ἔδωκάς μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. Πάτερ δίκαιε, καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω, καὶ οὗτοι ἔγνωσαν, ὅτι σύ με ἀπέστειλας· καὶ ἐγνώρισα αὐτοῖς τὸ ὄνομά (10)(3) σου καὶ γνωρίσω». Οὗτός ἐστιν «ὁ ἀποδιδοὺς ἁμαρτίας πατέρων ἐπὶ τέκνα τοῖς μισοῦσι καὶ ποιῶν ἔλεος τοῖς ἀγαπῶσιν».

4) Another observation is the way the TLG handles differences in wordforms. In Table 3, the text of Clement departs a number of times from the text of John (highlighted green).<sup>25</sup> Nonetheless, the system of using lemmata rather than exact wordforms has prevented the TLG from omitting these words.

5) Sometimes, however, a vital part of the match is not detected because of omissions or interpolations by Clement; especially if at the start or at the end of a match. In Table 4 (no. 1), the phrase Ἐν ἀρχῇ (of which only ἀρχῇ counts as a content word) has not been highlighted, because Clement omitted

<sup>22</sup> Cf. Table 6 (no. 19) in which the stopwords καὶ ἡμεῖς and Εἰ τοίνυν οἱ have wrongly been highlighted. Other such examples are: 2 (γὰρ), 14 (καὶ), 16 (ἐπ' αὐτὸν), 17 (δὲ), 18 (καὶ), 22 (εἶτα and πάλιν), 23 (αὐτὸ), 24 (ἐν and καὶ). Meaningful matching words that stand between two matching trigrams are not highlighted, cf. φησὶν in 9 and ἐπήγαγεν in 22.

<sup>23</sup> In some cases, the parallel seems to be longer than four consecutive trigrams, but remember that stop-words do not count. In the new version of the IPM module, the user can manually tune this parameter to a maximum of six consecutive trigrams (=7 content words).

<sup>24</sup> Other parallels to be clustered are: nos. 9–10, 11–12, 13 + 15, 22–23, 24–25.

<sup>25</sup> In the sequence of the sentence: ὁ = οὗς, δέδωκάς = ἔδωκάς (2 times), καὶ οὗτοι = κἀκεῖνοι. In addition, the final nu in UBS2 is two times absent in the text of Clement. The green highlights have been added manually.



the phrase ἦν ὁ λόγος, καὶ in this fairly free quotation of John 1:1. Consequently, the first trigram of Clement's quotation contains only one λόγος instead of two like in John 1:1 and can therefore not be matched. Nevertheless, a match has been established on the basis of two consecutive trigrams: λόγος θεός θεός and θεός θεός λόγος (cf. Figure 2).<sup>26</sup> A similar case is no. 20 (Table 4), in which an interpolation from Matthew 21:9 (τῷ υἱῷ Δαβὶδ) has caused the loss of ὠσαννά at the start of the match.

**Table 4. Result of omission and interpolation in *Paedagogus***

<p>1. <b>N.T.Ev.Jo. {0031.004}</b> 1.1.1</p> <p>(1) Ἐν ἀρχῇ ἦν ὁ λόγος, καὶ ὁ λόγος ἦν πρὸς τὸν θεόν, καὶ θεός ἦν ὁ λόγος. (2) οὗτος ἦν ἐν ἀρχῇ πρὸς τὸν θεόν. (3) πάντα δι' αὐτοῦ ἐγένετο, καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ</p> <p>20. <b>N.T.Ev.Jo. {0031.004}</b> 12.13.4</p> <p><b>Ἄσαννά·</b> εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου, καὶ ὁ βασιλεὺς τοῦ Ἰσραήλ. (5)</p>	<p>1. <b>Clem.Al.Paed. {0555.002}</b> 1.8.62.4.2</p> <p>(4) ἀλλ' οὐδὲ ἐπὶ τοῦ λόγου· ἐν γὰρ ἄμφω, ὁ θεός, ὅτι εἶπεν «ἐν ἀρχῇ ὁ λόγος ἦν ἐν τῷ θεῷ, καὶ θεός ἦν ὁ λόγος». Εἰ δὲ οὐ μισεῖ τῶν ὑπ' αὐτοῦ γενομένων οὐδέν, λείπεται φιλεῖν αὐτό.</p> <p>20. <b>Clem.Al.Paed. {0555.002}</b> 1.5.12.5</p> <p>ἐξῆλθον εἰς ὑπάντησιν κυρίῳ καὶ ἐκέκραγον λέγοντες, ὠσαννά τῷ υἱῷ Δαβὶδ, εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου», φῶς καὶ δόξα καὶ αἶνος μεθ' ἱκετηρίας τῷ κυρίῳ· τουτὶ γὰρ (5)</p>
--	--

6) Despite some of my criticisms, the examples till now have shown accurate quotations by Clement. In other cases, however, the IPM module lists less accurate matches. An interesting one is the match of *Paedagogus* 1.6.28.5 with John 3:15, 3:36, and 6:47 (cf. nos. 4, 5, and 16 in Table 5). While it is completely clear from the initial τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν... that Clement quotes John 6:40, it is matched with the other verses too, because part of John 6:40 is parallel to these other verses in John. The same applies to 1.6.29.2, which contains a literal quotation of John 3:36 (no. 6), but has also been matched with 6:40 because of the same reason (no. 14). From the perspective of quotation, these matches count as false positives.

**Table 5. Matches and false positives related to *Paedagogus* 1.6.28.5 and 1.6.29.2<sup>27</sup>**

<p>4. <b>N.T.Ev.Jo. {0031.004}</b> 3.15.1</p> <p>οὕτως ὑψωθῆναι δεῖ τὸν υἱὸν τοῦ ἀνθρώπου, ἵνα πᾶς ὁ πιστεύων ἐν αὐτῷ ἔχη ζωὴν αἰώνιον. Οὕτως γὰρ ἠγάπησεν ὁ θεὸς τὸν κόσμον,</p> <p>5. <b>N.T.Ev.Jo. {0031.004}</b> 3.36.2</p> <p>καὶ πάντα δέδωκεν ἐν τῇ χειρὶ αὐτοῦ. ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον· ὁ δὲ ἀπειθῶν τῷ υἱῷ οὐκ ὄψεται ζωὴν,</p> <p>6. <b>N.T.Ev.Jo. {0031.004}</b> 3.36.2</p> <p>καὶ πάντα δέδωκεν ἐν τῇ χειρὶ αὐτοῦ. ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον· ὁ δὲ ἀπειθῶν τῷ υἱῷ οὐκ ὄψεται ζωὴν</p> <p>13. <b>N.T.Ev.Jo. {0031.004}</b> 6.40.2</p>	<p>4. <b>Clem.Al.Paed. {0555.002}</b> 1.6.28.5.6</p> <p>εἰπὼν· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.»</p> <p>5. <b>Clem.Al.Paed. {0555.002}</b> 1.6.28.5.6</p> <p>εἰπὼν· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.»</p> <p>6. <b>Clem.Al.Paed. {0555.002}</b> 1.6.29.2.1</p> <p>Πίστις γὰρ μαθήσεως τελειότης· διὰ τοῦτο φησιν «ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον». Εἰ τοίνυν οἱ πιστεύσαντες ἔχομεν τὴν ζωὴν, τί περαιτέρω τοῦ κεκτηθῆσθαι</p> <p>13. <b>Clem.Al.Paed. {0555.002}</b> 1.6.28.5.5</p>
---	---

<sup>26</sup> From other matches, I conclude that forms of the verb εἰμί do not count as content word (cf. Appendix nos. 13, 22).

<sup>27</sup> *Paedagogus* 1.6.28.5 has also been mismatched with John 6:54, which compares closely to 6:40, and 6:68 (see match no. 18 in the Appendix). In Table 5, unnecessary text elements have been suppressed.



τοῦτο γάρ ἐστιν τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων εἰς αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω αὐτὸν ἐγὼ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ.

14. **N.T.Ev.Jo. {0031.004}** 6.40.2

τοῦτο γάρ ἐστιν τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων εἰς αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω αὐτὸν ἐγὼ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ.

15. **N.T.Ev.Jo. {0031.004}** 6.40.3

θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων εἰς αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω αὐτὸν ἐγὼ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ. Ἐγόγγυζον οὖν οἱ Ἰουδαῖοι περὶ αὐτοῦ ὅτι εἶπεν,

16. **N.T.Ev.Jo. {0031.004}** 6.47.2

ἀμὴν ἀμὴν λέγω ὑμῖν, ὁ πιστεύων ἔχει ζωὴν αἰώνιον. ἐγὼ εἰμι ὁ ἄρτος τῆς ζωῆς. οἱ πατέρες ὑμῶν ἔφαγον ἐν τῇ ἐρήμῳ

κύριος σαφέστατα τῆς σωτηρίας τὴν ἰσότητα ἀπεκάλυψεν εἰπών· «τοῦτο γάρ ἐστι τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.»

14. **Clem.Al.Paed. {0555.002}** 1.6.29.2.1

Πίστις γὰρ μαθήσεως τελειότης· διὰ τοῦτο φησιν «ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον». Εἰ τοίνυν οἱ πιστεύσαντες ἔχομεν τὴν ζωὴν, τί περαιτέρω τοῦ κεκτησθαι

15. **Clem.Al.Paed. {0555.002}** 1.6.28.5.6

εἰπών· «τοῦτο γάρ ἐστι τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.» Καθ' ὅσον μὲν οὖν δυνατὸν ἐν τῷδε τῷ κόσμῳ,

16. **Clem.Al.Paed. {0555.002}** 1.6.28.5.6

εἰπών· «τοῦτο γάρ ἐστι τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.»

7) Even more interesting is no. 19 (Table 6), which shows a false positive between John 6:68 and *Paedagogus* 1.6.29.2. Contrary to the false positives in Table 5, this one results from the decision to ignore stopwords, not from internal parallels in John itself. If the stopwords Εἰ τοίνυν οἱ (which as such, by the way, should not have been highlighted) had not been ignored, this case would not have been considered a match. This case therefore, clearly calls for a critical examination of the list of stopwords.

**Table 6. Mismatch with *Paedagogus* 1.6.29.2 due to ignored stopwords**

19. **N.T.Ev.Jo. {0031.004}** 6.68.2

ὁμοῖς θέλετε ὑπάγειν; (68) ἀπεκρίθη αὐτῷ Σίμων Πέτρος, Κύριε, πρὸς τίνα ἀπελευσόμεθα; ῥήματα ζωῆς αἰωνίου ἔχεις, (69) καὶ ἡμεῖς πεπιστεύκαμεν καὶ ἐγνώκαμεν ὅτι σὺ εἶ ὁ ἅγιος τοῦ θεοῦ. (70) ἀπεκρίθη αὐτοῖς ὁ Ἰησοῦς,

19. **Clem.Al.Paed. {0555.002}** 1.6.29.2.1

πιστεύομεν. Πίστις γὰρ μαθήσεως τελειότης· διὰ τοῦτο φησιν (2) «ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον». Εἰ τοίνυν οἱ πιστεύσαντες ἔχομεν τὴν ζωὴν, τί περαιτέρω τοῦ κεκτησθαι ζωὴν αἰδίων ὑπολείπεται; οὐδὲν δὲ ἐνδεῖ τῇ πίστει τελεία

8) A final case is match no. number 21 (Table 7). Although the match is perfect, the preceding context in *Paedagogus* shows that this is certainly a quotation, not from John but rather from Matthew 23:37–39. Filtering out such matches is technically possible but it would require an additional closest-match-filter and a comparison not just against John but against the entire New Testament.

**Table 7. Mismatch with *Paedagogus* 1.9.79.3 due to an improper parallel with John 12:13**

21. **N.T.Ev.Jo. {0031.004}** 12.13.4

Ὠσαννά·  
εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου,  
καὶ ὁ βασιλεὺς τοῦ Ἰσραήλ. (5)

21. **Clem.Al.Paed. {0555.002}** 1.9.79.3.3

οἶκος ὑμῶν ἔρημος, λέγω γὰρ ὑμῖν· ἀπάρτι οὐ μὴ ἴδῃτέ με, ἕως ἄν εἴπητε· εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου.»  
Εἰ γὰρ οὐ δέχεσθε τὴν φιланθρωπῖαν, ἐπιγνώσεσθε τὴν ἐξουσίαν.

We are now in a position to draw up some conclusions with regard to the matches found by the IPM module. The primary aim of this undertaking was to trace quotations of John in the *Paedagogus* of

Clement. Of the 29 matches found, 22 matches are (part of) quotations of John.<sup>28</sup> However, 14 of these actually constitute six long quotations.<sup>29</sup> Of the remaining 7 matches, one is incorrect as a result of stopwords being ignored (no. 19), while another one contains in fact a quotation from Matthew which happens to be partly equal to the passage in John (no. 21). The remaining 5 matches have wrongly been connected to multiple passages in John because of internal matches within John itself.<sup>30</sup> In conclusion, the number of false positives in this test is approximately 25% (7 out of 29), which is not too bad. The flexibility provided by the application of n-grams has prevented the loss of several quotations that do not verbally match the UBS<sup>2</sup> text, mostly due to Clement's free way of quoting. False positives appear to be mostly due to internal parallels in John and the ignorance of stopwords. The possibility of quotations not being detected will be discussed in Section 4.

### 3.2. Matches with the Gospel according to Matthew

In the previous section, I analyzed a case in which literary dependency could be expected. In this section, however, I examine a situation in which a relationship in genre and theme is evident, but in which literary dependency is highly unlikely: the gospel according to Matthew compared to John. My aim is not to provide an evaluation of the IPM module, but to analyze and categorize the matches found to get an idea about the performance of the IPM module in such comparisons relevant to Biblical Studies. As it happens, the TLG finds 16 matches between the two.<sup>31</sup> In my view, these matches can be broken down into several categories, that I labeled as 'strong lexical parallels', 'matching narrative constructions', and 'loose lexical matches'.

#### 3.2.1. Strong lexical parallels between John and Matthew

'Strong lexical parallels' are matches that share a set of specialized vocabulary suggesting that the match represents a true textual parallel or even a shared quotation. The TLG traces eight matches that fall into this category. Three turn out to be shared quotations of the Old Testament (Table 8). As such, these matches are very similar in nature to those found in section 3.1, although they do not contain direct quotations, but quotations from a common source.

**Table 8. Shared Old Testament quotations in John and Matthew**

9. **N.T.Ev.Jo. {0031.004}** 12.13.4

Ωσαννά·

εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου,  
καὶ ὁ βασιλεὺς τοῦ Ἰσραήλ. (5)

9. **N.T.Ev.Matt. {0031.001}** 21.9.4

Ωσαννά τῷ υἱῷ Δαβὶδ·

Εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου·  
Ωσαννά ἐν τοῖς ὑψίστοις. (5)

<sup>28</sup> Here is the full list of matches that contain actual quotations from John by Clement: nos. 1, 2, 3, 6, 7, 8, 9–10, 11–12, 13 + 15, 17, 20, 22–23, 24–25, 26–29. They can easily be reviewed in the Appendix.

<sup>29</sup> Nos. 9–10, 11–12, 13 + 15, 22–23, 24–25, 26–29.

<sup>30</sup> *Paedagogus*, 1.6.28.5: nos. 4, 5, 16, 18; *Paedagogus*, 1.6.29.2: no. 14

<sup>31</sup> Last checked on March 2, 2018. For each of the gospels of Mark and Luke, the TLG finds nine parallels. Due to the reduction of two required trigrams in the old version to one trigram in the new version, the number of matches between John and Matthew has increased enormously. The TLG currently finds 461 matches, with lemma comparison enabled and the maximum number of concatenated n-grams set to 6 (last checked March 8, 2019). However, most of them fall into the category of 'matching narrative constructions' in the form *someone speaks to someone*. This construction has numerous internal parallels both in John and Matthew, thus leading to this exceptional increase of matches. Nevertheless, one may find a number of interesting matches that have not been found in the old version. This, however, requires a fresh treatment of the current comparison. The big difference in matches also shows that a seemingly slight change in the algorithm's parameters (one required trigram instead of two required trigrams) has a huge bearing on its output.

10. **N.T.Ev.Jo. {0031.004}** 12.13.4

Ἰσσαννά·

εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου,  
καὶ ὁ βασιλεὺς τοῦ Ἰσραήλ. (5)11. **N.T.Ev.Jo. {0031.004}** 12.15.1ἐστὶν γεγραμμένον, (15) Μὴ φοβοῦ, **θυγάτηρ Σιών·**  
**ἰδοὺ ὁ βασιλεὺς σου ἔρχεται**, καθήμενος ἐπὶ πῶλον  
ὄνου.10. **N.T.Ev.Matt. {0031.001}** 23.39.2ἔρημος. (39) λέγω γὰρ ὑμῖν, οὐ μὴ με ἴδωτε ἀπ' ἄρτι ἕως  
ἂν εἴπητε, **εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου.**  
(1) Καὶ ἐξελθὼν ὁ Ἰησοῦς ἀπὸ τοῦ ἱεροῦ ἐπορεύετο, καὶ11. **N.T.Ev.Matt. {0031.001}** 21.5.1γένονεν ἵνα πληρωθῇ τὸ ῥηθὲν διὰ τοῦ προφήτου λέγοντος,  
(5) Εἶπατε τῇ **θυγατρὶ Σιών, ἰδοὺ ὁ βασιλεὺς σου**  
**ἔρχεται** σοι, πραῦς καὶ ἐπιβεβηκὼς ἐπὶ ὄνον,

The phrase εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι Κυρίου (nos. 9, 10) originates from Ps. 117:26 [118:26] in the Septuagint and is quoted verbally in all four canonical gospels, probably pointing to a shared interpretation of this verse in early Christianity.<sup>32</sup> The third match (no. 11), however, is most interesting, since both Matthew 21:5 and John 12:15 quote Zechariah 9:9 differently. Although the TLG finds the parallel on the basis of the first part, which is equally quoted while leaving out κήρυσσε, θύγατερ Ἱερουσαλήμ, it does not recognize the equality in the second part in which ὄνος and πῶλος occur (cf. Table 9).<sup>33</sup> This second part of the sentence is quoted quite literally in Matthew, but is only alluded to in John 12:15.

**Table 9. Comparison of Zechariah 9:9 in John 12:15 and Matthew 21:5<sup>34</sup>**

<b>Zechariah 9:9</b>	<b>John 12:14b–15</b>	<b>Matthew 21:4b–5</b>
Χαῖρε σφόδρα,	καθὼς ἐστὶν γεγραμμένον,	ἵνα πληρωθῇ τὸ ῥηθὲν διὰ τοῦ
	(15)	προφήτου λέγοντος, (5)
<b>θυγάτηρ Σειών</b> κήρυσσε,	Μὴ φοβοῦ, <b>θυγάτηρ Σιών·</b>	Εἶπατε τῇ <b>θυγατρὶ Σιών,</b>
<b>θύγατερ Ἱερουσαλήμ ἰδοὺ ὁ βασιλεὺς</b>	<b>ἰδοὺ ὁ βασιλεὺς σου ἔρχεται,</b>	<b>ἰδοὺ ὁ βασιλεὺς σου ἔρχεται σοι,</b>
<b>σου ἔρχεται σοι</b> δίκαιος καὶ σώζων	καθήμενος <b>ἐπὶ πῶλον ὄνου.</b>	<b>πραῦς καὶ ἐπιβεβηκὼς ἐπὶ ὄνον,</b>
αὐτός, <b>πραῦς καὶ ἐπιβεβηκὼς ἐπὶ</b>		<b>καὶ ἐπὶ πῶλον</b> υἱὸν <b>ὑποζυγίου.</b>
<b>ὑποζυγίου καὶ πῶλον</b> νέον.		

By way of excursus, I was surprised to find that a comparison (both in the IPM module and in the parallel browsing mode) between John, Matthew, and Zechariah fails to detect any match between the gospels and the prophet.<sup>35</sup> Three characteristics of the use of n-grams in the TLG can explain this result. First, in Zechariah 9:9, θύγατερ Σειών is separated from ἰδοὺ ὁ βασιλεὺς σου ἔρχεται σοι by too much interpolation (κήρυσσε, θύγατερ Ἱερουσαλήμ). Second, the phrase ἰδοὺ ὁ βασιλεὺς σου ἔρχεται contains only three content words (ἰδοὺ, βασιλεὺς, and ἔρχεται), so the required minimum of two trigrams was not reached. Finally, in Zechariah 9:9 the phrases ἰδοὺ ὁ βασιλεὺς σου ἔρχεται σοι and **πραῦς καὶ ἐπιβεβηκὼς** are separated by the phrase δίκαιος καὶ σώζων αὐτός which is lacking in Matthew 21:5. Again, in the TLG this interpolation containing two content words is too extensive for

<sup>32</sup> Matthew 21:9; 23:39; Mark 11:9; Luke 13:35; John 12:13. According to Brown, John has probably recorded a similar tradition, independently from the Synoptics. Raymond E. Brown, *The Gospel According to John* (i-xii), The Anchor Bible 29 (Garden City, New York: Doubleday, 1966) 460–1. According to Bultmann, “[d]ie Quelle kann nicht einer der Synoptiker sein...”. D. Rudolf Bultmann, *Das Evangelium des Johannes*, 17. Auflage, Kritisch-exegetischer Kommentar über das Neue Testament (Göttingen: Vandenhoeck & Ruprecht, 1962) 319.

<sup>33</sup> πῶλος in Matthew is only visible in the parallel browsing mode, since it appears in the next sentence.

<sup>34</sup> The yellow highlighted words show shared words among all three versions, the green highlighted words show shared words by Zechariah and Mathew, while the red highlighted word shows a parallel word in the versions of John and Matthew.

<sup>35</sup> The edition of the Septuagint used in the TLG is: A. Rahlfs, *Septuaginta*, vol. 1, 9th edition (Stuttgart: Württembergische Bibelanstalt, 1935 [repr. 1971]).

a match to be established.<sup>36</sup> In sum, the match with Zechariah 9:9 was skipped because of too much interpolation, either in Zechariah 9:9, or in the gospel texts. A better assessment of stopwords and a better way to deal with interpolations could have prevented the TLG from skipping over this parallel. Nevertheless, the lexically closer match between Matthew and John was found.<sup>37</sup>

The TLG finds five additional matches that are very close in wording (Table 10), showing strong lexical parallelisms. All five may possibly go back to a shared traditions or sources. In John 2:19 (no. 1) it is Jesus who says: “break this temple down and in three days I will raise it up”, while in Matthew 27:40 it is the people who recall those same words albeit in a slightly different phrasing. In no. 8, the same logion of Jesus is told with a small difference in sequence. The last two matches (nos. 14 and 15) are both parts of a parallel narrative about Jesus before Pontius Pilate.

The most interesting match, however, is no. 12. Although the context is completely different, there is a strong lexical resemblance between the two verses. In John 12:34, it is the agitated crowd that asks Jesus to reveal who he is: τίς ἐστὶν οὗτος ὁ υἱὸς τοῦ ἀνθρώπου; (“Who is this Son of man?”) In Matthew 16:13 it is Jesus himself who asks his disciples to tell how the people think about him: τίνα λέγουσιν οἱ ἄνθρωποι εἶναι τὸν υἱὸν τοῦ ἀνθρώπου; (“Who do the people say that the Son of man is?”) Though these questions are asked by different persons in two completely different settings, it is clear that they can both be understood as parallel in requesting the identity of the Son of man.

**Table 10. Strong lexical parallels between John and Matthew  
(minus Old Testament quotations)**

1. **N.T.Ev.Jo. {0031.004}** 2.19.2

δεικνύεις ἡμῖν, ὅτι ταῦτα ποιεῖς; (19) ἀπεκρίθη Ἰησοῦς καὶ εἶπεν αὐτοῖς, Λύσατε τὸν ναὸν τοῦτον καὶ ἐν τρισὶν ἡμέραις ἐγερῶ αὐτόν. (20) εἶπαν οὖν οἱ Ἰουδαῖοι, Τεσσαράκοντα καὶ ἕξ ἔτεσιν οἰκοδομήθη ὁ ναὸς οὗτος, καὶ σὺ

8. **N.T.Ev.Jo. {0031.004}** 12.8.1

αὐτήν, ἵνα εἰς τὴν ἡμέραν τοῦ ἐνταφιασμοῦ μου τηρήσῃ αὐτό· (8) τοὺς πτωχοὺς γὰρ πάντοτε ἔχετε μεθ’ ἑαυτῶν, ἐμὲ δὲ οὐ πάντοτε ἔχετε. (9) Ἔγνω οὖν [ὁ] ὄχλος πολὺς ἐκ τῶν Ἰουδαίων ὅτι

12. **N.T.Ev.Jo. {0031.004}** 12.34.4

νόμου ὅτι ὁ Χριστὸς μένει εἰς τὸν αἰῶνα, καὶ πῶς σὺ λέγεις ὅτι δεῖ ὑψωθῆναι τὸν υἱὸν τοῦ ἀνθρώπου; τίς ἐστὶν οὗτος ὁ υἱὸς τοῦ ἀνθρώπου; (35) εἶπεν οὖν αὐτοῖς ὁ Ἰησοῦς, Ἔτι μικρὸν χρόνον τὸ φῶς ἐν ὑμῖν ἐστίν.

14. **N.T.Ev.Jo. {0031.004}** 19.2.1

(1) Τότε οὖν ἔλαβεν ὁ Πιλάτος τὸν Ἰησοῦν καὶ ἐμαστίγωσεν. (2) καὶ οἱ στρατιῶται πλέξαντες στέφανον ἐξ ἀκανθῶν ἐπέθηκον αὐτοῦ τῇ κεφαλῇ, καὶ ἰμάτιον πορφυροῦν περιέβαλον αὐτόν, (3) καὶ ἤρχοντο πρὸς αὐτόν καὶ

1. **N.T.Ev.Matt. {0031.001}** 27.40.1

παραπορευόμενοι ἐβλασφήμουν αὐτὸν κινεῦντες τὰς κεφαλὰς αὐτῶν (40) καὶ λέγοντες, Ὁ καταλύων τὸν ναὸν καὶ ἐν τρισὶν ἡμέραις οἰκοδομῶν, σώσον σεαυτόν, εἰ υἱὸς εἶ τοῦ θεοῦ, [καὶ] κατάρβηθι ἀπὸ τοῦ σταυροῦ. (41) ὁμοίως

8. **N.T.Ev.Matt. {0031.001}** 26.11.1

εἶπεν αὐτοῖς, Τί κόπους παρέχετε τῇ γυναικί; ἔργον γὰρ καλὸν ἠργάσατο εἰς ἐμέ· (11) πάντοτε γὰρ τοὺς πτωχοὺς ἔχετε μεθ’ ἑαυτῶν, ἐμὲ δὲ οὐ πάντοτε ἔχετε· (12) βαλοῦσα γὰρ αὕτη τὸ μύρον τοῦτο ἐπὶ τοῦ σώματός μου πρὸς τὸ

12. **N.T.Ev.Matt. {0031.001}** 16.13.3

Φιλίππου ἠρώτα τοὺς μαθητὰς αὐτοῦ λέγων, Τίνα λέγουσιν οἱ ἄνθρωποι εἶναι τὸν υἱὸν τοῦ ἀνθρώπου; (14) οἱ δὲ εἶπαν, Οἱ μὲν Ἰωάννην τὸν βαπτιστὴν, ἄλλοι δὲ Ἡλίαν, ἕτεροι δὲ Ἰερεμίαν ἢ ἕνα τῶν προφητῶν. (15) λέγει αὐτοῖς,

14. **N.T.Ev.Matt. {0031.001}** 27.29.1

σπεῖραν. (28) καὶ ἐκδύσαντες αὐτὸν χλαμύδα κοκκίνην περιέθηκον αὐτῷ, (29) καὶ πλέξαντες στέφανον ἐξ ἀκανθῶν ἐπέθηκον ἐπὶ τῆς κεφαλῆς αὐτοῦ καὶ κάλαμον ἐν τῇ δεξιᾷ αὐτοῦ, καὶ γονυπετήσαντες ἔμπροσθεν αὐτοῦ ἐνέπαιξαν

<sup>36</sup> As far as I could observe, the algorithm used by the TLG can only deal with interpolations of one content word, plus an unlimited number of stopwords.

<sup>37</sup> In the new version, the TLG fails to find these quotations too (last checked March 8, 2019).

15. **N.T.Ev.Jo. {0031.004}** 19.3.2

περιέβαλον αὐτόν, (3) καὶ ἤρχοντο πρὸς αὐτόν καὶ ἔλεγον, Χαῖρε, ὁ βασιλεὺς τῶν Ἰουδαίων· καὶ ἐδίδοσαν αὐτῷ ῥαπίσματα. (4) Καὶ ἐξήλθεν πάλιν ἔξω ὁ Πιλάτος

15. **N.T.Ev.Matt. {0031.001}** 27.29.4

αὐτοῦ, καὶ γονυπετήσαντες ἔμπροσθεν αὐτοῦ ἐνέπαιξαν αὐτῷ λέγοντες, Χαῖρε, βασιλεὺ τῶν Ἰουδαίων, (30) καὶ ἐμπτύσαντες εἰς αὐτόν ἔλαβον τὸν κάλαμον καὶ ἔτυπτον

3.2.2. *Shared narrative constructions and loose lexical matches between John and Matthew*

Of another nature are the matches that I have labelled ‘shared narrative constructions’ and ‘loose lexical matches’. Both share very common vocabulary, but are distinguished by the presence or absence of a shared syntactic construction. Moreover, shared narrative constructions usually match very closely in wording representing a common narrative phrase, while loose lexical matches are mainly found because of the algorithm’s feature to use lemma forms rather than exact wordforms.

In Table 11, I have listed all matches which comprise a form of ἀποκρίνομαι + (particle) + subject/object + (καί) + a form of λέγω + object/subject + saying. They all represent an introduction of direct speech in the narrative.<sup>38</sup>

**Table 11. Shared narrative phrases in John and Matthew**

2. **N.T.Ev.Jo. {0031.004}** 3.3.1

τὰ σημεῖα ποιεῖν ἃ σὺ ποιεῖς, ἐάν μὴ ἦ ὁ θεὸς μετ’ αὐτοῦ. (3) ἀπεκρίθη Ἰησοῦς καὶ εἶπεν αὐτῷ, Ἀμὴν ἀμὴν λέγω σοι, ἐάν μὴ τις γεννηθῆ ἄνωθεν, οὐ δύναται ἰδεῖν τὴν

3. **N.T.Ev.Jo. {0031.004}** 4.10.1

οὔσης; (οὐ γὰρ συγχρῶνται Ἰουδαῖοι Σαμαρίταις.) (10) ἀπεκρίθη Ἰησοῦς καὶ εἶπεν αὐτῇ, Εἰ ἦδεις τὴν δωρεὰν τοῦ θεοῦ καὶ τίς ἐστὶν ὁ λέγων σοι, Δός μοι πεῖν, σὺ

4. **N.T.Ev.Jo. {0031.004}** 4.10.1

οὔσης; (οὐ γὰρ συγχρῶνται Ἰουδαῖοι Σαμαρίταις.) (10) ἀπεκρίθη Ἰησοῦς καὶ εἶπεν αὐτῇ, Εἰ ἦδεις τὴν δωρεὰν τοῦ θεοῦ καὶ τίς ἐστὶν ὁ λέγων σοι, Δός μοι πεῖν, σὺ

5. **N.T.Ev.Jo. {0031.004}** 5.19.1

ἔλεγεν τὸν θεόν, ἴσον ἑαυτὸν ποιῶν τῷ θεῷ. (19) Ἀπεκρίνατο οὖν ὁ Ἰησοῦς καὶ ἔλεγεν αὐτοῖς, Ἀμὴν ἀμὴν λέγω ὑμῖν, οὐ δύναται ὁ υἱὸς ποιεῖν ἀφ’ ἑαυτοῦ οὐδὲν ἐάν μὴ τι βλέπῃ τὸν πατέρα ποιοῦντα·

6. **N.T.Ev.Jo. {0031.004}** 6.26.1

εὐρόντες αὐτὸν πέραν τῆς θαλάσσης εἶπον αὐτῷ, Ῥαββί, πότε ὤδε γέγονας; (26) ἀπεκρίθη αὐτοῖς ὁ Ἰησοῦς καὶ εἶπεν, Ἀμὴν ἀμὴν λέγω ὑμῖν, ζητεῖτέ με οὐχ ὅτι εἶδετε σημεῖα ἀλλ’ ὅτι ἐφάγετε ἐκ τῶν ἄρτων καὶ ἐχορτάσθητε.

2. **N.T.Ev.Matt. {0031.001}** 21.21.1

χρῆμα ἢ συκῆ. (20) καὶ ἰδόντες οἱ μαθηταὶ ἐθαύμασαν λέγοντες, Πῶς παραχρῆμα ἐξηράνθη ἡ συκῆ; (21) ἀποκριθεὶς δὲ ὁ Ἰησοῦς εἶπεν αὐτοῖς, Ἀμὴν λέγω ὑμῖν, ἐάν ἔχητε πίστιν καὶ μὴ διακριθῆτε, οὐ μόνον τὸ τῆς συκῆς

3. **N.T.Ev.Matt. {0031.001}** 20.22.1

ἵνα καθίσωσιν οὗτοι οἱ δύο υἱοὶ μου εἷς ἐκ δεξιῶν σου καὶ εἷς ἐξ εὐωνύμων σου ἐν τῇ βασιλείᾳ σου. (22) ἀποκριθεὶς δὲ ὁ Ἰησοῦς εἶπεν, Οὐκ οἴδατε τί αἰτεῖσθε· δύνασθε πιεῖν τὸ ποτήριον ὃ ἐγὼ μέλλω πίνειν; λέγουσιν αὐτῷ,

4. **N.T.Ev.Matt. {0031.001}** 21.27.1

μεθὰ τὸν ὄχλον, πάντες γὰρ ὡς προφήτην ἔχουσιν τὸν Ἰωάννην. (27) καὶ ἀποκριθέντες τῷ Ἰησοῦ εἶπαν, Οὐκ οἴδαμεν. ἔφη αὐτοῖς καὶ αὐτός, Οὐδὲ ἐγὼ λέγω ὑμῖν ἐν ποίᾳ ἐξουσίᾳ ταῦτα ποιῶ.

5. **N.T.Ev.Matt. {0031.001}** 21.21.1

χρῆμα ἢ συκῆ. (20) καὶ ἰδόντες οἱ μαθηταὶ ἐθαύμασαν λέγοντες, Πῶς παραχρῆμα ἐξηράνθη ἡ συκῆ; (21) ἀποκριθεὶς δὲ ὁ Ἰησοῦς εἶπεν αὐτοῖς, Ἀμὴν λέγω ὑμῖν, ἐάν ἔχητε πίστιν καὶ μὴ διακριθῆτε, οὐ μόνον τὸ τῆς συκῆς

6. **N.T.Ev.Matt. {0031.001}** 21.21.1

χρῆμα ἢ συκῆ. (20) καὶ ἰδόντες οἱ μαθηταὶ ἐθαύμασαν λέγοντες, Πῶς παραχρῆμα ἐξηράνθη ἡ συκῆ; (21) ἀποκριθεὶς δὲ ὁ Ἰησοῦς εἶπεν αὐτοῖς, Ἀμὴν λέγω ὑμῖν, ἐάν ἔχητε πίστιν καὶ μὴ διακριθῆτε, οὐ μόνον τὸ τῆς συκῆς

<sup>38</sup> It should be noted that in the new version of the TLG, matches that fall into this category are increased enormously, especially because (short) narrative phrases have many internal parallels both in John and Matthew. Cf. note 31.



16. **N.T.Ev.Jo. {0031.004}** 21.21.1

δείπνω ἐπὶ τὸ στήθος αὐτοῦ καὶ εἶπεν, Κύριε, τίς ἐστὶν ὁ παραδιδούς σε; (21) τοῦτον οὖν ἰδὼν ὁ Πέτρος λέγει τῷ Ἰησοῦ, Κύριε, οὗτος δὲ τίς; (22) λέγει αὐτῷ ὁ Ἰησοῦς, Ἐὰν αὐτὸν θέλω μένειν ἕως ἔρχομαι, τί πρὸς σέ; σύ

16. **N.T.Ev.Matt. {0031.001}** 17.4.1

τὸ φῶς. (3) καὶ ἰδοὺ ὠφθη αὐτοῖς Μωϋσῆς καὶ Ἥλιος συλλαλοῦντες μετ' αὐτοῦ. (4) ἀποκριθεὶς δὲ ὁ Πέτρος εἶπεν τῷ Ἰησοῦ, Κύριε, καλὸν ἐστὶν ἡμᾶς ὥδε εἶναι· εἰ θέλεις, ποιήσω ὥδε τρεῖς σκηνάς, σοὶ μίαν καὶ Μωϋσεὶ μίαν καὶ

The remaining two matches fall into the category of loose lexical matches. These matches comprise of similar vocabulary, but do not show a shared syntactic construction or theme, nor do they show up in comparable contexts (Table 12). The use of lemmata rather than exact wordforms, together with the omission of stopwords has led to these two nonrelated matches. In most types of research, these matches would be considered false positives.

**Table 12. Loose lexical matches between John and Matthew**

7. **N.T.Ev.Jo. {0031.004}** 10.34.3

κρίθη αὐτοῖς [ὁ] Ἰησοῦς, Οὐκ ἔστιν γεγραμμένον ἐν τῷ νόμῳ ὑμῶν ὅτι Ἐγὼ εἶπα, Θεοὶ ἐστε; (35) εἰ ἐκείνους εἶπεν θεοὺς πρὸς οὓς ὁ λόγος τοῦ θεοῦ ἐγένετο, καίου δύναται λυθῆναι ἡ γραφή, (36) ὃν ὁ πατὴρ ἡγάσεν

7. **N.T.Ev.Matt. {0031.001}** 22.31.2

εἰσιν. (31) περὶ δὲ τῆς ἀναστάσεως τῶν νεκρῶν οὐκ ἀνέγνωτε τὸ ῥηθὲν ὑμῖν ὑπὸ τοῦ θεοῦ λέγοντος, (32) Ἐγὼ εἶμι ὁ θεὸς Ἀβραάμ καὶ ὁ θεὸς Ἰσαὰκ καὶ ὁ θεὸς Ἰακώβ; οὐκ ἔστιν [ὁ] θεὸς νεκρῶν ἀλλὰ ζώντων. (33) καὶ

13. **N.T.Ev.Jo. {0031.004}** 13.18.2

οἴδατε, μακάριοι ἐστε ἐὰν ποιῆτε αὐτά. (18) οὐ περὶ πάντων ὑμῶν λέγω· ἐγὼ οἶδα τίνας ἐξελεξάμην· ἀλλ' ἵνα ἡ γραφή πληρωθῇ, Ὁ τρώγων μου τὸν ἄρτον ἐπῆρεν

13. **N.T.Ev.Matt. {0031.001}** 26.70.1

αὐτῷ μία παιδίσκη λέγουσα, Καὶ σὺ ἤσθα μετὰ Ἰησοῦ τοῦ Γαλιλαίου. (70) ὁ δὲ ἠρνήσατο ἔμπροσθεν πάντων λέγων, Οὐκ οἶδα τί λέγεις. (71) ἐξελθόντα δὲ εἰς τὸν πυλῶνα εἶδεν αὐτὸν ἄλλη καὶ λέγει τοῖς ἐκεῖ, Οὗτος ἦν μετὰ

### 3.2.3. Conclusions

In conclusion, despite its built-in flexibility, the TLG returns mainly longer matches that are closely resembling each other's vocabulary with the maximum of one interpolating meaningful word. For a comparison of John with Matthew, this appears not to be very helpful, even if the low yield seems to confirm the idea that John is substantially different from Matthew.<sup>39</sup> In a comparison of for instance Matthew with Mark, in which literary dependency is almost certain, the TLG might offer much more help in discovering the longer parallel that resemble each other closely in wording.<sup>40</sup> However, verbally looser parallels that might be of importance for synoptic analysis are unlikely to be detected. The same applies probably to comparisons of the New Testament letters.<sup>41</sup>

<sup>39</sup> Kurt Aland et al. identified around a hundred smaller and larger parallel passages between John and Matthew. Kurt Aland (ed.), *Synopsis Quattuor Evangeliorum*, 15<sup>th</sup> revised edition, Stuttgart: Deutsche Bibelgesellschaft, 1997.

<sup>40</sup> In the old version, a comparison between Matthew and Mark yields 259 matches. Many of these are due to internal parallels, or are parallels consisting of very common vocabulary, or are actually part of a longer parallel passage. Nevertheless, a substantially greater number of parallel passages can be traced by using the IPM module. The resulting parallels have a value by themselves since they probably concern the most literal parallel phrases. In the new version, a comparison between Matthew and Mark yields a stunning number of 709 matches (lemma forms) even if the concatenation of matches is set to 6 n-grams. In the exact wordforms mode, the IPM module yields 218 matches.

<sup>41</sup> A comparison between the closely related letters to the Ephesians and the Colossians yields 4 matches with very common vocabulary, but also 6 matches with more specialized vocabulary (the new version has 12 matches in total). A comparison between Ephesians and Galatians yields 3 matches consisting of very general vocabulary (the new version has 7 matches in total). A comparison between Romans and Galatians yields only 4 matches, of which 2 consist of common vocabulary, 1 contains a shared Old Testament quotation, and 1

### 3.3. Intertextual matches between John and the works of Plutarch

During the last decades, much attention has been paid to the question how Plutarch's ideas relate to the New Testament.<sup>42</sup> However, according to Helge Almquist, who has carefully investigated some three hundred New Testament parallels in the works of Plutarch, no *direct* dependence can be traced.<sup>43</sup> These matches contain mainly phrases that show similarity in thought and highlight the general hellenic character of the New Testament.

Since Plutarch's work is so extensive, it would be worthwhile to see whether the TLG is able to locate some examples that underscore these similarities in thought and character. Interestingly, the TLG lists only five matches with John (Table 13) among the more than one million words of Plutarch in the database.<sup>44</sup> None of them point to similarity in thought.

Two conclusions can be drawn:

- 1) All matches consist of very common vocabulary.<sup>45</sup>

---

consists of a shared phrase with one of Paul's key thoughts: ἐξ ἔργων νόμου οὐ δικαιωθήσεται πᾶσα σὰρξ (Romans 3:20; Galatians 2:16). The new version, however, yields 37 matches. Most of them comprise very common vocabulary like the names of Jesus. All comparisons in the new version have been conducted with lemma search enabled and maximum match concatenation set to six n-grams.

<sup>42</sup> The university of Bern has dedicated a special project to the study of Plutarch in relation to the New Testament, called: Plutarch und das Neue Testament. Cf. for a bibliography on research devoted to the study of Plutarch from the perspective of the New Testament: [http://www.plutarch.unibe.ch/assets/bibliographie\\_plutarch-nt.pdf](http://www.plutarch.unibe.ch/assets/bibliographie_plutarch-nt.pdf) (last accessed 8 March, 2019).

<sup>43</sup> Helge Sigvard Almquist, *Plutarch und das Neue Testament: Ein Beitrag zum Corpus Hellenisticum Novi Testamenti*, Uppsala: Appelbergs Boktryckeri, 1946. About the Johanneic parallels in Plutarch he writes: "Die johanneischen Schriften zeigen, wie zu erwarten, nur wenige Berührungen mit der Welt des Plutarch. Es handelt sich hauptsächlich um kulturgeschichtliche Details und wenig bedeutende Sprach- und Stilelemente, die in den erzählenden Teilen des Evangeliums auftreten. Gedankenwelt und Sprache des Johannes sind durchaus der echt griechischen Kultursphäre fremd, (...). Mit dem Hellenismus des Plutarch hat der vierte Evangelist nichts Wesentliches gemeinsam." Almquist, *Plutarch und das Neue Testament*, 144. However, Almquist indicates 16 verses in John that are somehow parallel to 25 passages in Plutarch. Almquist, *ibidem*. 71-77.

<sup>44</sup> At the moment of writing (last checked on 1 March 2018), Plutarch has 1,036,815 words distributed over 147 works in the TLG. (Cf. <http://stephanus.tlg.uci.edu/Iris/inst/stat.jsp> Author → Plutarch). It should be noted that comparisons with other well-known Greek authors also show remarkably low numbers of matches. For the whole of the New Testament, a comparison with the work of Plutarch lists 17 parallels. (last checked October 31, 2018). The new version finds 51 matches if compared with John and 726 if compared with the whole of the New Testament. Most of them are false positives, due to causes explained in this section. When exact wordforms are used, however, the results drop to 22 for the whole of the New Testament and 5 for John (last checked March 18, 2019).

<sup>45</sup> The percentages refer to the ranking of the word among all the lemmata used in the gospel of John and the corpus of Plutarch (excluding stop-words and in the sequence John|Plutarch). The TLG lists 1,107 lemmata for John and 28,511 lemmata for Plutarch. Percentages have been rounded up (e.g. 2.115...% = 3%).

For instance, the lemma λόγος appears 40 times in the gospel of John. It ranks 34<sup>th</sup> in the frequency ranking of lemmata in the gospel of John. Since there are 1,107 lemmata in John, there are 1,107 ranks and therefore λόγος is among the 4% most frequent words in John ( $34/1,107 * 100\% = 3.071... \approx 4\%$ ). *A lower percentage thus indicates a higher relative frequency in the corpus.* Numbers are based on the statistics module in the TLG: <http://stephanus.tlg.uci.edu/Iris/inst/stat.jsp> (accessed March 9, 2018).



**verbs:** ἀκούω [2%|0.3%], (ἀνα)γι(γ)νώσκω [3%|3%], γί(γ)νωμαι [3%|0.02%], ἐθέλω [6%|2%], ἐρωτάω [5%|1%], (ἐπ)έχω [1%|4%], λέγω (6) [0.1%|0.004%]

**nouns:** θεός (3) [2%|0.1%], λόγος [4%|0.04%], υἱός [3%|1%], χρόνος [34%|0.2%]

**adjectives:** πολὺς [3%|0.008%].

In all instances – except for χρόνος in John – these words are among the 6% most common words in John and among 4% most common words in Plutarch. The fact that χρόνος – that only appears 4 times in John – could have been part of a match is due to the fact that it is among the most favorite words of Plutarch (rank 2).

2) If stopwords and verbal prefixes had not been neglected, all these matches would not have been found, since exactly these surrounding words vary a lot.

The investigation of these matches, therefore, underscores the conclusion of the prior sections that the IPM module performs best if used for close phrase matching. In other cases, the IPM module finds mainly false positives due to the exclusion of stopwords. However, the result of only 5 matches in this comparison with a one million words corpus is a very low and shows that the IPM module is quite accurate in eliminating false positives.<sup>46</sup>

**Table 13. Matching phrases between John and the work of Plutarch**

1. **N.T.Ev.Jo. {0031.004}** 9.27.1

Τί ἐποίησέν σοι; πῶς ἥνοιξέν σου τοὺς ὀφθαλμούς; (27) ἀπεκρίθη αὐτοῖς, **εἶπον ὑμῖν ἤδη καὶ οὐκ ἤκούσατε· τί πάλιν θέλετε ἀκοῦναι; μὴ καὶ ὑμεῖς θέλετε** αὐτοῦ μαθηταὶ γενέσθαι; (28) καὶ ἐλοιδόρησαν αὐτὸν καὶ εἶπον,

2. **N.T.Ev.Jo. {0031.004}** 5.6.2

τῇ ἀσθενεῖα αὐτοῦ· (6) τοῦτον ἰδὼν ὁ Ἰησοῦς κατακείμενον, καὶ **γνοὺς ὅτι πολὺν ἤδη χρόνον ἔχει**, λέγει αὐτῷ, Θέλεις ὑγιῆς γενέσθαι; (7) ἀπεκρίθη αὐτῷ ὁ

3. **N.T.Ev.Jo. {0031.004}** 9.19.1

ἦν τυφλὸς καὶ ἀνέβλεψεν, ἕως ὅτου ἐφώνησαν τοὺς γονεῖς αὐτοῦ τοῦ ἀναβλέψαντος (19) καὶ **ἠρώτησαν αὐτοὺς λέγοντες, Οὗτός ἐστιν ὁ υἱὸς ὑμῶν, ὃν ὑμεῖς λέγετε** ὅτι τυφλὸς ἐγεννήθη; πῶς οὖν βλέπει ἄρτι; (20) ἀπεκρίθησαν

4. **N.T.Ev.Jo. {0031.004}** 10.34.3

κρίθη αὐτοῖς [ὁ] Ἰησοῦς, Οὐκ ἔστιν γεγραμμένον ἐν τῷ νόμῳ ὑμῶν ὅτι Ἐγὼ **εἶπα, Θεοὶ ἐστε;** (35) **εἰ ἐκείνους εἶπεν θεοὺς** πρὸς οὓς ὁ λόγος τοῦ θεοῦ ἐγένετο, καὶ οὐ δύναται λυθῆναι ἡ γραφή, (36) ὃν ὁ

5. **N.T.Ev.Jo. {0031.004}** 10.35.2

τῷ νόμῳ ὑμῶν ὅτι Ἐγὼ εἶπα, Θεοὶ ἐστε; (35) εἰ ἐκείνους **εἶπεν θεοὺς πρὸς οὓς ὁ λόγος τοῦ θεοῦ ἐγένετο**, καὶ οὐ δύναται λυθῆναι ἡ γραφή, (36) ὃν ὁ πατήρ ἠγάσεν

1. **Plu. De capienda ex inimicis utilitate {0007.072}** 89.A.9

ἔοικε προστάττειν ὁ θεὸς ὡς τῷ μέλλοντι ψέγειν ἕτερον τὸ “γνώθι σαυτὸν,” ἵνα μὴ **λέγοντες ἂ θέλουσιν ἀκούωσιν ἂ μὴ θέλουσι**. “φιλεῖ” γὰρ ὁ (10) τοιοῦτος κατὰ τὸν Σοφοκλέα

2. **Plu. Septem sapientium convivium {0007.079}** 151.C.5

γενέσθαι παρ’ ἡμῶν οὐ τὰμὰ κωλύσει.” Τούτων **ἀναγνωσθέντων οὐ πολὺν χρόνον ἐπι-(5)σχῶν** ὁ Βίας, ἀλλὰ μικρὰ μὲν αὐτὸς πρὸς αὐτῷ γενόμενος μικρὰ δὲ τῷ Κλεοβούλῳ

3. **Plu. Regum et imperatorum apophthegmata {0007.081}** 189.A.11

διππε, μετὰ Φωκίῳ ἀποθανοῦμενος;’ (10) “Ἦδη δὲ τῆς κύλικος αὐτῷ προσφερομένης **ἐρωτηθεὶς εἶ τι λέγει πρὸς τὸν υἱὸν ‘ἔγωγε’ εἶπεν** ‘ἐντέλλομαι (B) καὶ παρακαλῶ μηδὲν Ἀθηναίους μνησικακεῖν.’

4. **Plu. Apophthegmata Laconica {0007.082}** 236.D.4

ἔφη· ἐπικειμένου δὲ μᾶλλον καὶ λέγοντος ‘πάντως σε δεῖ εἰπεῖν’, ὁ Λάκων ἀντηρώτησε ‘τίμι με δεῖ **εἰπεῖν, σοὶ ἢ τῷ θεῷ;** τοῦ δ’ εἰπόντος ‘τῷ θεῷ’, ‘σὺ τοίνυν’ ἔφη (5) ‘ἀποχώρησον.’

5. **Plu. Quaestiones convivales {0007.112}** 718.B.10

Ἐκ δὲ τούτου γενομένης σιωπῆς, πάλιν ὁ Διογενιανὸς ἀρξάμενος ‘βούλεσθ’ **εἶπεν, ‘ἐπεὶ λόγοι περὶ (10) (C) θεῶν γεγόνασιν**, ἐν τοῖς Πλάτωνος γενεθλίοις

<sup>46</sup> It should be noted, that in the new version, the number of false positives has been increased (cf. Note 44).

αὐτὸν Πλάτωνα κοινωνὸν παραλάβωμεν, ἐπισκεψάμενοι τίνα λαβῶν

## 4. The Accuracy of the IPM Module

In section 3, we discussed the matches found by the IPM module between John on the one hand and Clement's *Paedagogus*, Matthew, and Plutarch, respectively, on the other. In this section, returning to *Paedagogus*, I compare the results of the IPM module with a manually compiled index of gospel parallels (Cosaert 2008). The choice for a text by Clement is motivated by his free way of quoting the New Testament, which provides a perfect heuristic tool to explore the accuracy and the limitations of the IPM module. I would expect that suggestions for improvements that apply to Clement's work will probably perform even better on texts that have more accurate quotations.

### 4.1. Comparison with quotations in the *Paedagogus* of Clement of Alexandria according to Cosaert

In 2008, a thoroughly revised edition of Carl P. Cosaert's doctoral dissertation on Clement's gospel quotations was published.<sup>47</sup> It provides a thorough analysis of gospel quotations in the extant works of Clement of Alexandria.<sup>48</sup> Since the IPM module has a built-in flexibility in finding matches, I expected it to find at least the exact and slightly adapted citations, and probably part of the more extensive allusions. Unfortunately, this turns out not to be the case (cf. Table 14).<sup>49</sup>

First of all, it should be noted that Cosaert lists three times as many quotations (42) as does the TLG (14). All citations and adaptations found by the TLG are present in the list of Cosaert. The opposite, however, is not true: the TLG lists only 63% (12 out of 19) of the citations that are labeled by Cosaert as "a verbally exact quotation of a particular Gospel passage."<sup>50</sup> The percentages of the adapted quotations and the allusions are even lower: 17% (2 out of 12) and 0% (0 out of 11) respectively.<sup>51</sup>

<sup>47</sup> Carl P. Cosaert, *The Text of the Gospels in Clement of Alexandria* (Atlanta: Society of Biblical Literature, 2008). According to the *Series Editor's Foreword* by Michael W. Holmes, this study renders all previous work on Clement's citations obsolete. Cosaert, *The Text of the Gospels in Clement*, x.

<sup>48</sup> In my analysis, I follow the classification of Cosaert (citations, adaptations, allusions, [lemmata]), without getting into debates about the classification of intertextuality: cf. Carl P. Cosaert, *The Text of the Gospels in Clement of Alexandria*, Atlanta: Society of Biblical Literature, 2008, 45-46, and using the term 'quotations' as a general reference to all three categories.

<sup>49</sup> It should be noted that I have taken from Cosaert the established gospel references only, unless the TLG came up with a parallel that is classified as undeterminable by Cosaert. This is only the case with no. 20 (Matthew 21:8-9/John 12:13 [Paedagogus 1.12.5] cf. the Appendix and my discussion in section 3.2.1). In his list of indeterminable gospel references (Cosaert, *The Text of the Gospels in Clement*, 311-335), Cosaert has 14 additional references to John, totaling 15 possible references to John (1:23 [2.112.1]; 14:13-14 [3.40.2]; 6:11 [2.13.2]; 2:14-16 [3.79.2]; 19:2, 5 [2.73.3; 2.74.1; 2.75.2]; 1:27 [2.117.4]; 1:29, 36 [1.24.4]; 6:32-33, 41; 6:49, 58 [3.40.1]; 10:2, 11, 14 [1.84.1]; 10:11, 14 [1.37.3; 1.53.2; 1.85.2]). Almost all these references are due to Clement's use of rich biblical language. Some of them, e.g. 1.24.4 (quoting John 1:29 or 36), could have been counted as a quotation of John. Quotations covering more than one verse are taken as one, both in Cosaert and the TLG.

<sup>50</sup> Cosaert, *The Text of the Gospels in Clement*, 46.

<sup>51</sup> The new version of the IPM module (cf. Note 2) performs slightly better (lemma mode): 14 citations, 1 adaptation, and 1 allusion are now found. However, the price to be paid is that the TLG now lists 84 matches (compared to 29 in the previous version) that need to be scanned to find the real quotations (last accessed 24 January, 2019). This means that the number of false positives has been increased substantially. The other

**Table 14. A comparison of quotations found by Cosaert and the TLG**

	Number of quotations found by Cosaert	Number of quotations found by TLG	Coverage TLG by Cosaert	Coverage Cosaert by TLG
Citations [C]	19	12	100%	63%
Adaptations [Ad]	12	2	100%	17%
Allusions [All]	11	0	-	0%
<b>Total/Average</b>	<b>42</b>	<b>14</b>	<b>100%</b>	<b>33%</b>

These discouraging figures prompt the question whether the underlying algorithm of the TLG could have performed better if different settings had been used. The remainder of this section is therefore devoted to analyzing the results and suggesting improvements to the IPM module. In section 4.2, I try to figure out why some of the *exact citations* have been detected while others have not, and how the method used to trace these parallels can be improved. In section 4.3, I discuss the conditions under which the *adaptations* could have been found. In section 4.4, I briefly explore the possibility of finding *allusions* digitally with the n-grams technique.

#### 4.2. Intertextual phrase matching and the search for citations

Since the TLG uses a flexible method to address matches among texts, it is striking that in the case of Clement, 7 out of 19 of his most literal citations have not been found. How can this be, given the use of lemmata rather than exact wordforms and despite the tolerance of interpolations and omissions? Before addressing that question, I will first consider the citations actually detected by TLG.

Table 15 lists all citations that are both mentioned in Cosaert's list and detected by the TLG. Consecutive matches are manually concatenated.<sup>52</sup>

**Table 15. Citations from the list of Cosaert and found by the TLG**

##### 2. N.T.Ev.Jo. {0031.004} 1.3.1

καὶ θεὸς ἦν ὁ λόγος. (2) οὗτος ἦν ἐν ἀρχῇ πρὸς τὸν θεόν. (3) πάντα δι' αὐτοῦ ἐγένετο, καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ ἓν. ὃ γέγονεν (4) ἐν αὐτῷ ζωὴ ἦν, καὶ ἡ ζωὴ ἦν τὸ

##### 3. N.T.Ev.Jo. {0031.004} 1.17.1

αὐτοῦ ἡμεῖς πάντες ἐλάβομεν, καὶ χάριν ἀντὶ χάριτος· (17) ὅτι ὁ νόμος διὰ Μωϋσέως ἐδόθη, ἡ χάρις καὶ ἡ ἀλήθεια διὰ Ἰησοῦ Χριστοῦ ἐγένετο. (18) θεὸν οὐδεὶς ἑώρακεν πώποτε· μονογενὴς θεὸς ὁ ὢν εἰς τὸν κόλπον τοῦ πατρὸς

##### 2. Clem.Al.Paed. {0555.002} 1.11.97.3.5

«πᾶσα σοφία παρὰ κυρίου καὶ μετ' αὐτοῦ ἐστὶν εἰς τὸν αἰῶνα»· παρρησίᾳ δὲ ὅτι θεὸς καὶ δημιουργός, «πάντα γὰρ δι' (5) αὐτοῦ ἐγένετο καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ ἓν»· εὐνοία δὲ ὅτι μόνος ὑπὲρ ἡμῶν ἱερεῖον ἑαυτὸν ἐπιδέδωκεν, «ὁ γὰρ

##### 3. Clem.Al.Paed. {0555.002} 1.7.60.2.1

θεράποντος αὐτοῦ· διὸ καὶ πρόσκαιρος ἐγένετο, «ἡ δὲ αἰδῖος (2) χάρις καὶ ἡ ἀλήθεια διὰ Ἰησοῦ Χριστοῦ ἐγένετο.» Ὅρατε τὰς λέξεις τῆς γραφῆς· ἐπὶ μὲν τοῦ νόμου «ἐδόθη» φησὶ μόνον,

option is to choose the wordform comparison mode. This results in 32 matches by which 11 citations are accurately traced, while no adaptations or allusions are found.

<sup>52</sup> The first sign (e.g. |2|) shows the start of the highlighting of the second parallel. The next lower sign (e.g. |1|) shows the end of the highlighting of the first parallel.

## 6. N.T.Ev.Jo. {0031.004} 3.36.2

τὸν υἱόν, καὶ πάντα δέδωκεν ἐν τῇ χειρὶ αὐτοῦ. (36) ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον· ὁ δὲ ἀπειθῶν τῷ υἱῷ οὐκ ὄσεται ζωὴν, ἀλλ' ἡ ὀργὴ τοῦ θεοῦ μένει ἐπ'

## 7. N.T.Ev.Jo. {0031.004} 4.32.1

(31) Ἐν τῷ μεταξύ ἠρώτων αὐτὸν οἱ μαθηταὶ λέγοντες, Ῥαββί, φάγε. (32) ὁ δὲ εἶπεν αὐτοῖς, Ἐγὼ βρωῶσιν ἔχω φαγεῖν ἢν ὑμεῖς οὐκ οἴδατε. (33) ἔλεγον οὖν οἱ μαθηταὶ πρὸς ἀλλήλους, Μὴ τις ἤνεγκεν αὐτῷ φαγεῖν; (34) λέγει

## 8. N.T.Ev.Jo. {0031.004} 4.34.2

πρὸς ἀλλήλους, Μὴ τις ἤνεγκεν αὐτῷ φαγεῖν; (34) λέγει αὐτοῖς ὁ Ἰησοῦς, Ἐμὸν βρωμά ἐστιν ἵνα ποιήσω τὸ θέλημα τοῦ πέμψαντός με καὶ τελειώσω αὐτοῦ τὸ ἔργον. (35) οὐχ ὑμεῖς λέγετε ὅτι Ἔτι τετράμηνός ἐστιν καὶ ὁ

## 9–10. N.T.Ev.Jo. {0031.004} 5.24.1–2

τὸν υἱὸν οὐ τιμᾷ τὸν πατέρα τὸν πέμψαντα αὐτόν. (24) Ἀμὴν ἀμὴν λέγω ὑμῖν ὅτι ὁ τὸν λόγον μου ἀκούων καὶ πιστεύων τῷ πέμψαντί με |10| ἔχει ζωὴν |9| αἰώνιον, καὶ εἰς κρίσιν οὐκ ἔρχεται ἀλλὰ μεταβέβηκεν ἐκ τοῦ θανάτου εἰς τὴν ζωὴν. (25) ἀμὴν ἀμὴν λέγω ὑμῖν ὅτι ἔρχεται

## 11–12. N.T.Ev.Jo. {0031.004} 6.32.2–4

ἔδωκεν αὐτοῖς φαγεῖν. (32) εἶπεν οὖν αὐτοῖς ὁ Ἰησοῦς, Ἀμὴν ἀμὴν λέγω ὑμῖν, οὐ Μωϋσῆς δέδωκεν ὑμῖν τὸν ἄρτον ἐκ τοῦ οὐρανοῦ, ἀλλ' ὁ πατήρ μου δίδωσιν ὑμῖν τὸν ἄρτον ἐκ τοῦ οὐρανοῦ τὸν |12| ἀληθινόν· (33) ὁ γὰρ ἄρτος |11| τοῦ θεοῦ ἐστὶν ὁ καταβαίνων ἐκ τοῦ οὐρανοῦ καὶ ζωὴν διδοὺς τῷ κόσμῳ. (34) Εἶπον οὖν πρὸς αὐτόν, Κύριε, πάντοτε δὸς ἡμῖν

## 13 + 15. N.T.Ev.Jo. {0031.004} 6.40.2–3

αὐτὸ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ. (40) τοῦτο γὰρ ἐστὶν τὸ θέλημα τοῦ πατρός μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ |15| πιστεύων εἰς αὐτὸν ἔχη |13| ζωὴν αἰώνιον, καὶ ἀναστήσω αὐτὸν ἐγὼ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ. (41) Εὐγγυζον οὖν οἱ Ἰουδαῖοι περὶ αὐτοῦ ὅτι εἶπεν,

## 17. N.T.Ev.Jo. {0031.004} 6.51.3

καταβάς· ἐάν τις φάγη ἐκ τούτου τοῦ ἄρτου ζήσει εἰς τὸν αἰῶνα· καὶ ὁ ἄρτος δὲ ὃν ἐγὼ δώσω ἡ σὰρξ μου ἐστὶν ὑπὲρ τῆς τοῦ κόσμου ζωῆς. (52) Ἐμάχοντο οὖν πρὸς ἀλλήλους οἱ Ἰουδαῖοι λέγοντες,

## 6. Clem.Al.Paed. {0555.002} 1.6.29.2.1

πιστεύομεν. Πίστις γὰρ μαθήσεως τελειότης· διὰ τοῦτο φησιν (2) «ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον». Εἰ τοίνυν οἱ πιστεύσαντες ἔχομεν τὴν ζωὴν, τί περαιτέρω τοῦ κεκτῆσθαι

## 7. Clem.Al.Paed. {0555.002} 1.6.45.4.1

Ἀλλὰ καὶ τοῖς ὑποτιτθίοις παιδίοις ἀρκεῖ μόνον τὸ γάλα καὶ (4) ποτὸν εἶναι καὶ τροφήν. «Ἐγὼ», φησὶν ὁ κύριος, «βρωῶσιν ἔχω φαγεῖν, ἢν ὑμεῖς οὐκ οἴδατε· ἐμὸν βρωμά ἐστὶν, ἵνα ποιήσω τὸ θέλημα τοῦ πέμψαντός με». Ὅρατε ἄλλο βρωμά

## 8. Clem.Al.Paed. {0555.002} 1.6.45.4.2

(4) ποτὸν εἶναι καὶ τροφήν. «Ἐγὼ», φησὶν ὁ κύριος, «βρωῶσιν ἔχω φαγεῖν, ἢν ὑμεῖς οὐκ οἴδατε· ἐμὸν βρωμά ἐστὶν, ἵνα ποιήσω τὸ θέλημα τοῦ πέμψαντός με». Ὅρατε ἄλλο βρωμά ἀλληγορούμενον παραπλησίως γάλακτι τὸ θέλημα τοῦ θεοῦ.

## 9–10. Clem.Al.Paed. {0555.002} 1.6.27.1.5–6

ἔπεσθαι Χριστῷ. «Ὁ γὰρ γέγονεν ἐν αὐτῷ, ζωὴ ἐστὶν.» «Ἀμὴν ἀμὴν λέγω ὑμῖν», φησὶν, «ὁ τὸν λόγον μου ἀκούων (5) καὶ πιστεύων τῷ πέμψαντί με |10| ἔχει ζωὴν |9| αἰώνιον καὶ εἰς κρίσιν οὐκ ἔρχεται, ἀλλὰ μεταβέβηκεν ἐκ τοῦ θανάτου εἰς τὴν (2) ζωὴν.» Οὕτω τὸ πιστεῦσαι μόνον καὶ ἀναγεννηθῆναι τελειώσις

## 11–12. Clem.Al.Paed. {0555.002} 1.6.46.2.2–4

(2) γάλα. Ἔτι δὲ καὶ ἄρτον αὐτὸν οὐρανῶν ὁμολογεῖ ὁ λόγος. «Οὐ γὰρ Μωσῆς», φησὶν, «ἔδωκεν ὑμῖν τὸν ἄρτον ἐκ τοῦ οὐρανοῦ νοῦ [sic], ἀλλ' ὁ πατήρ μου δίδωσιν ὑμῖν τὸν ἄρτον ἐκ τοῦ οὐρανοῦ τὸν |12| ἀληθινόν· ὁ γὰρ ἄρτος |11| τοῦ θεοῦ ἐστὶν ὁ ἐκ τοῦ οὐρανοῦ καταβαίνων καὶ ζωὴν διδοὺς τῷ κόσμῳ. Καὶ ὁ ἄρτος, ὃν ἐγὼ (5) (3) δώσω, ἡ σὰρξ μου ἐστὶν ὑπὲρ τῆς τοῦ κόσμου ζωῆς.» Ἐνταῦθα

## 13 + 15. Clem.Al.Paed. {0555.002} 1.6.28.5.5–6

κύριος σαφέστατα τῆς σωτηρίας τὴν ἰσότητά ἀπεκάλυψεν εἰπών· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρός μου, ἵνα πᾶς ὁ (5) θεωρῶν τὸν υἱὸν καὶ |15| πιστεύων ἐπ' αὐτόν ἔχη |13| ζωὴν αἰώνιον, (1) καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.» Καθ' ὅσον μὲν οὖν δυνατὸν ἐν τῷδε τῷ κόσμῳ, ὃν ἐσχάτην ἡμέραν ἠνίξαστο εἰς

## 17. Clem.Al.Paed. {0555.002} 1.6.46.2.5

τὸν ἀληθινόν· ὁ γὰρ ἄρτος τοῦ θεοῦ ἐστὶν ὁ ἐκ τοῦ οὐρανοῦ καταβαίνων καὶ ζωὴν διδοὺς τῷ κόσμῳ. Καὶ ὁ ἄρτος, ὃν ἐγὼ (5) (3) δώσω, ἡ σὰρξ μου ἐστὶν ὑπὲρ τῆς τοῦ κόσμου ζωῆς.» Ἐνταῦθα τὸ μυστικὸν τοῦ ἄρτου παρασημειωτέον, ὅτι σάρκα αὐτὸν λέγει

20. **N.T.Ev.Jo. {0031.004}** 12.13.4

Ἵωσαννά·  
εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου,  
καὶ ὁ βασιλεὺς τοῦ Ἰσραήλ. (5)

22–23. **N.T.Ev.Jo. {0031.004}** 15.1–2

ποιῶ. Ἐγείρεσθε, ἄγωμεν ἐντεῦθεν. **15. (1)** Ἐγὼ εἰμι ἡ ἄμπελος ἡ ἀληθινή, καὶ ὁ πατήρ μου ὁ γεωργὸς ἐστίν. **(2)** πᾶν κλήμα ἐν ἐμοὶ μὴ φέρον καρπὸν, **|23|** αἶρει **|22|** αὐτό, καὶ πᾶν τὸ καρπὸν φέρον καθαίρει αὐτό ἵνα καρπὸν πλείονα φέρῃ. **(3)** ἤδη ὑμεῖς καθαροὶ ἐστέ διὰ τὸν λόγον ὃν λελάληκα ὑμῖν· **(4)** μείνατε ἐν ἐμοί,

24–25. **N.T.Ev.Jo. {0031.004}** 17.21–22

τῶν πιστευόντων διὰ τοῦ λόγου αὐτῶν εἰς ἐμέ, **(21)** ἵνα πάντες ἐν ὧσιν, καθὼς σύ, πάτερ, ἐν ἐμοὶ κἀγὼ ἐν σοί, ἵνα καὶ αὐτοὶ ἐν ἡμῖν ὧσιν, ἵνα ὁ κόσμος πιστεύῃ ὅτι σύ με ἀπέστειλας. **(22)** κἀγὼ τὴν **|25|** δόξαν ἣν δέδωκάς **|24|** μοι δέδωκα αὐτοῖς, ἵνα ὧσιν ἐν καθὼς ἡμεῖς ἐν, **(23)** ἐγὼ ἐν αὐτοῖς καὶ σύ ἐν ἐμοί, ἵνα ὧσιν τετελειωμένοι εἰς ἐν, ἵνα γινώσκῃ ὁ κόσμος ὅτι σύ με ἀπέστειλας καὶ

26–29. **N.T.Ev.Jo. {0031.004}** 17.24–25

ἵνα γινώσκῃ ὁ κόσμος ὅτι σύ με ἀπέστειλας καὶ ἠγάπησας αὐτούς καθὼς ἐμέ ἠγάπησας. **(24)** Πάτερ, ὃ δέδωκάς μοι, θέλω ἵνα ὅπου εἰμι ἐγὼ κἀκεῖνοι ὧσιν μετ' ἐμοῦ, ἵνα θεωρῶσιν τὴν δόξαν τὴν **|27|** ἐμήν ἣν δέδωκάς **|26|** μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. **(25)** **|28|** πάτερ δίκαιε, **|27|** καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω, καὶ οὗτοι **|29|** ἔγνωσαν ὅτι σύ με ἀπέστειλας, **|28|** **(26)** καὶ ἐγνώρισα αὐτοῖς τὸ ὄνομά σου καὶ γνωρίσω, ἵνα ἡ ἀγάπη ἣν ἠγάπησάς με ἐν αὐτοῖς ἦ κἀγὼ ἐν αὐτοῖς.

20. **Clem.Al.Paed. {0555.002}** 1.5.12.5

ἐξήλθον εἰς ὑπάντησιν κυρίῳ καὶ ἐκέκραγον λέγοντες, ὡσαννά τῷ υἱῷ Δαβίδ, εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου», φῶς καὶ δόξα καὶ αἴνος μεθ' ἱκετηρίας τῷ κυρίῳ· τουτί γάρ (5)

22–23. **Clem.Al.Paed. {0555.002}** 1.8.66.4.3–5

ἐκφαίνει, τὴν πολύτροπον καὶ πολυφελῆ θεραπείαν ἀλληγορῶν, ὀπηνίκα εἰπὼν «ἐγὼ εἰμι ἡ ἄμπελος ἡ ἀληθινή, καὶ ὁ πατήρ μου ὁ γεωργὸς ἐστίν», εἶτα ἐπήγαγεν πάλιν «πᾶν κλήμα ἐν ἐμοὶ μὴ φέρον **|23|** καρπὸν αἶρει **|22|** αὐτό, καὶ πᾶν τὸ καρ- **(5)**ποφοροῦν καθαίρει, ἵνα καρπὸν πλείω φέρῃ». Καθυλομανεῖ γὰρ μὴ κλαδευομένη ἡ ἄμπελος, οὕτως δὲ

24–25. **Clem.Al.Paed. {0555.002}** 1.8.71.1.4–6

ἔτι λόγων τὴν εὐαγγέλιον τοῦ κυρίου παραθεμένῳ φωνῆν· ἐν μὲν αὐτὸν λέγει· «ἵνα πάντες ἐν ὧσι, καθὼς σύ, πάτερ, ἐν ἐμοὶ κἀγὼ ἐν σοί, ἵνα καὶ αὐτοὶ ἐν ἡμῖν ἐν ὧσιν, ἵνα καὶ ὁ **(5)** κόσμος πιστεύῃ, ὅτι σύ με ἀπέστειλας. Κἀγὼ τὴν **|25|** δόξαν, ἣν ἔδωκάς **|24|** μοι, δέδωκα αὐτοῖς, ἵνα ὧσιν ἐν καθὼς ἡμεῖς ἐν· ἐγὼ ἐν αὐτοῖς καὶ σύ ἐν ἐμοί, ἵνα ὧσι τετελειωμένοι εἰς ἐν»· ἐν **(2)** δὲ ὁ θεὸς καὶ ἐπέκεινα τοῦ ἐνὸς καὶ ὑπὲρ αὐτὴν μονάδα.

26–29. **Clem.Al.Paed. {0555.002}** 1.8.71.2.6–10

καὶ δίκαιός ἐστιν ἐν τῷ αὐτῷ μαρτυρήσει κύριος εὐαγγελίῳ **(5)** λέγων «πάτερ, οὓς ἔδωκάς μοι, θέλω ἵνα ὅπου εἰμι ἐγὼ, κἀκεῖνοι ὧσι μετ' ἐμοῦ, ἵνα θεωρῶσι τὴν δόξαν τὴν **|27|** ἐμήν, ἣν ἔδωκάς **|26|** μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. **|28|** Πάτερ δίκαιε, **|27|** καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω κἀκεῖνοι **|29|** ἔγνωσαν, ὅτι σύ με ἀπέστειλας. **|28|** καὶ ἐγνώρισα αὐτοῖς τὸ ὄνομά **(10)** **(3)** σου καὶ γνωρίσω». Οὗτός ἐστιν «ὁ ἀποδιδούς ἀμαρτίας πατέρων ἐπὶ τέκνα τοῖς μισοῦσι καὶ ποιῶν ἔλεος τοῖς ἀγαπῶσιν».

One of the first things that is clear from this list is that the citations found by the TLG are primarily the extensive ones. This is because the TLG starts finding matches with the minimum of two trigrams of content words, resulting in a minimum of four content words for any match. Even the shortest citation meets this requirement (no. 20).<sup>53</sup> This requirement, however, has also led to the loss of a number of exact citations. Table 16 has the citations found by Cosaert, but not by the TLG. All of these citations contain fewer than four content words. A clear example is no. 3. This match, though it consists of seven words, is missed by the TLG, since it has only two content words: γέγονεν and ζωή.<sup>54</sup> In the current system, this citation would have been found by using a minimum of one matching bigram. Even more extreme is no. 1, which only counts as one content word (ἐγένετο), even if it

<sup>53</sup> See John 12:13 and *Paedagogus* 1.5.12.5: εὐλογημένος **|1a|** ὁ **|ignored|** ἐρχόμενος **|1b/2a|** ἐν **|ignored|** ὀνόματι **|1c/2b|** κυρίου **|2c|**.

<sup>54</sup> ὁ **|ignored|** γέγονεν **|1a|** ἐν **|ignored|** αὐτῷ **|ignored|** ζωή **|1b|** ἣν **|ignored|**. As far as I could observe, the verb εἰμί does not count as a content word.

has four words verbatim parallel to John 1:3. The choice to negate stopwords has reduced this citation to one content word only.

**Table 16. Citations in the list of Cosaert that are not found by the TLG**

Text John (UBS <sup>2</sup> )	Citations found in <i>Paedagogus</i> (Cosaert)	Matching words (excluding stopwords)	Matching words (including stopwords)
1. John 1:3 (3) πάντα δι' αὐτοῦ ἐγένετο, καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ ἓν. ὃ γέγονεν	1. ( <i>Paed.</i> 1.60.2) ου αὐτου εγενετο ουδε εν	1	4
2. John 1:3–4 (3) πάντα δι' αὐτοῦ ἐγένετο, καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ ἓν. ὃ γέγονεν (4) ἐν αὐτῷ ζωὴ ἦν, καὶ ἡ ζωὴ ἦν τὸ φῶς τῶν ἀνθρώπων.	2. ( <i>Paed.</i> 1.27.1) ο γαρ γεγονεν εν αυτω ζωη εστιν	2	6
3. John 1:3–4 (3) πάντα δι' αὐτοῦ ἐγένετο, καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ ἓν. ὃ γέγονεν (4) ἐν αὐτῷ ζωὴ ἦν, καὶ ἡ ζωὴ ἦν τὸ φῶς τῶν ἀνθρώπων.	3. ( <i>Paed.</i> 2.79.3) ο γαρ γεγονεν εν αυτω ζωη ην	2	6
4. John 6:51 ἐγὼ εἰμι ὁ ἄρτος ὁ ζῶν ὁ ἐκ τοῦ οὐρανοῦ καταβάς· ἐάν τις φάγη ἐκ τούτου τοῦ ἄρτου ζήσκει εἰς τὸν αἰῶνα· καὶ ὁ ἄρτος δὲ ὃν ἐγὼ δώσω ἡ σὰρξ μου ἐστίν ὑπὲρ τῆς τοῦ κόσμου ζωῆς.	4. ( <i>Paed.</i> 1.47.1) (ειπεν) και ο αρτος ον εγω δωσω η σαρξ μου εστιν	3	9
5. John 6:55 ἡ γὰρ σὰρξ μου ἀληθὴς ἐστίν βρῶσις, καὶ τὸ αἷμά μου ἀληθὴς ἐστίν πόσις.	5. ( <i>Paed.</i> 1.36.5) το αιμα μου γαρ (φησιν ο κυριος) αληθης εστιν ποσις	3	6
6. John 16:27 αὐτὸς γὰρ ὁ πατὴρ φιλεῖ ὑμᾶς, ὅτι ὑμεῖς ἐμὲ πεφιλήκατε καὶ πεπιστεύκατε ὅτι ἐγὼ παρὰ [τοῦ] θεοῦ ἐξῆλθον.	6. ( <i>Paed.</i> 1.8.2) αυτος γαρ ο πατηρ φιλει υμας, οτι υμεις εμε πεφιληκατε	3	10
7. John 17:23 ἐγὼ ἐν αὐτοῖς καὶ σὺ ἐν ἐμοί, ἵνα ὧσιν τετελειωμένοι εἰς ἓν, ἵνα γινώσκῃ ὁ κόσμος ὅτι σὺ με ἀπέστειλας καὶ ἠγάπησας αὐτούς καθὼς ἐμὲ ἠγάπησας.	7. ( <i>Paed.</i> 1.8.2) και ηγαπησας αυτους καθως, εμε ηγαπησας	3	6

In the light of these findings, the team of the TLG can do two things to improve the accuracy of their IPM module. The first possibility would be to reduce the minimum requirement of a match to one trigram (= 3 consecutive content words) –which will result in 4 (out of 7) additional citations found– or to one bigram which will result in only one citation (no. 1) not being found, but probably also in a massive number of false positives.<sup>55</sup> The other possibility would be to reconsider the decision to

<sup>55</sup> This can be seen in the browse mode with one text selected and the n-grams tool switched on. Tried on the *Paedagogus*, the TLG highlights on average more than half of each page to indicating parallels with its full corpus. As mentioned in section 2, both bi- and trigrams are used in this module.



ignore stopwords. If the algorithm takes stopwords into account, while leaving the other conditions as they are, only no. 5 would not be found because of a rather extensive interpolation: γαρ (φησιν ο κυριος). The suggestion to include stopwords will be discussed more extensively in the next section, but for now it suffices to say that the inclusion of stopwords shows a promising increase in results with regard to the citations of Clement.

### 4.3. Intertextual phrase matching and the search for adapted quotations

The first adaptation that is both listed by Cosaert and found by the TLG concerns the interesting quotation of John 1:1 in *Paedagogus* 1.8.62.4: ἐν ἀρχῇ ὁ λόγος ἦν ἐν τῷ θεῷ, καὶ θεὸς ἦν ὁ λόγος *in the beginning the logos was in God and the logos was God*. In this adaptation, Clement contracts the first part of the sentence (ἐν ἀρχῇ ἦν ὁ λόγος, καὶ ὁ λόγος ἦν ... *in the beginning the logos was and the logos was ...*) to ἐν ἀρχῇ ὁ λόγος ἦν *in the beginning the logos was ...*. The second part of the sentence is rendered differently: ἐν τῷ θεῷ instead of πρὸς τὸν θεόν, while the third part of the sentence is left as it is: καὶ θεὸς ἦν ὁ λόγος *and the logos was God*. Although this match was not highlighted in full by the TLG (cf. Table 4), it was found because of five consecutive matching lemmata λόγος εἰμί θεός θεός λόγος (=3 trigrams).

A second adaptation found by the TLG and listed by Cosaert as indeterminable (Matthew 21:8–9/John 12:13) has already been discussed in Section 3.1. Now the remaining ten adaptations listed by Cosaert are not found by the TLG (Table 17). This raises the question as to whether these adaptations could have been found, had the algorithm been tuned differently. To answer that question, we have to explore the nature of Clement's adaptations first.

As can be observed from the yellow highlighted parts of these parallels in Table 17, even these adaptations have much in common with the quoted text. Nevertheless, the varying extent of the adaptations and especially its interpolations appear to have posed difficulties to the algorithm in its current state. Seven adaptations (nos. 1–5, 9–10) have been skipped because they do not meet the minimum requirement of 2 trigrams. However, it is exactly these adaptations that agree closest with the text of John. Except for no. 9, these adaptations would have been found if stopwords had been taken into account.

The remaining adaptations are more complex because they are interrupted by more extensive modifications (green highlighted text) that substitute parts of the quotation, as well as by interpolations or deletions (blue highlighted text) that interrupt the flow of the parallel. Because of these (extensive) adaptations, it is conceivable that some of these quotations might have been missed even with the stopwords included.<sup>56</sup> However, the inclusion of stopwords would have increased the amount of adaptations found with more than 50 percent (6 out of 11).

**Table 17. Adaptations in the list of Cosaert not found by the TLG**

Text John UBS <sup>2</sup>	Adaptation by Clement	Number of matching content words
1. John 1:3 πάντα δι' αὐτοῦ ἐγένετο, καὶ <b>χωρὶς αὐτοῦ ἐγένετο οὐδὲ</b> <b>ἐν</b> . ὃ γέγονεν	1. <i>Paed.</i> 3.33.3 <b>ἐγενετο</b> <b>ἀνευ</b> <b>αυτου ουδε εν</b>	1
2. John 1:5	2. <i>Paed.</i> 1.28.3	

<sup>56</sup> Adaptations that would have been missed if stopwords had been taken into account (and other settings remain equal) are nos. 1, 6, 7, 9, 11.



καὶ τὸ φῶς ἐν τῇ σκοτίᾳ φαίνει, καὶ ἡ σκοτία αὐτὸ οὐ κατέλαβεν.	καὶ το σκοτος αυτον ου καταλαμβανει	2
3. John 1:5 καὶ τὸ φῶς ἐν τῇ σκοτίᾳ φαίνει, καὶ ἡ σκοτία αὐτὸ οὐ κατέλαβεν.	3. Paed. 2.79.3 καὶ ἡ σκοτια αυτον ου καταλαμβανει	2
4. John 1:5 καὶ τὸ φῶς ἐν τῇ σκοτίᾳ φαίνει, καὶ ἡ σκοτία αὐτὸ οὐ κατέλαβεν.	4. Paed. 2.99.6 καὶ ἡ σκοτια (φησιν) αυτο ου καταλαμβανει	2
5. John 1:14 καὶ ὁ λόγος σὰρξ ἐγένετο καὶ ἐσκήνωσεν ἐν ἡμῖν, καὶ ἐθεασάμεθα τὴν δόξαν αὐτοῦ, δόξαν ὡς μονογενοῦς παρὰ πατρός, πλήρης χάριτος καὶ ἀληθείας.	5. Paed. 2.20.1 δι ἡν ο λογος γεγονεν σαρξ	3
6. John 6:53 εἶπεν οὖν αὐτοῖς ὁ Ἰησοῦς, Ἄμην ἀμὴν λέγω ὑμῖν, ἐὰν μὴ φάγητε τὴν σάρκα τοῦ υἱοῦ τοῦ ἄνθρώπου καὶ πίητε αὐτοῦ τὸ αἷμα, οὐκ ἔχετε ζωὴν ἐν ἑαυτοῖς.	6. Paed. 1.38.2 (ο κυριος εν τω κατα Ιωαννην ευαγγελιω...) φαγεσθε μου τας σαρκας (ειπων) και πιεσθε μου το αιμα	4
7. John 6:53 εἶπεν οὖν αὐτοῖς ὁ Ἰησοῦς, Ἄμην ἀμὴν λέγω ὑμῖν, ἐὰν μὴ φάγητε τὴν σάρκα τοῦ υἱοῦ τοῦ ἄνθρώπου καὶ πίητε αὐτοῦ τὸ αἷμα, οὐκ ἔχετε ζωὴν ἐν ἑαυτοῖς.	7. Paed. 1.42.3 φαγεσθε μου (φησι) την σαρκα και πιεσθε μου το αιμα	4
8. John 10:11 Ἐγὼ εἰμι ὁ ποιμὴν ὁ καλός· ὁ ποιμὴν ὁ καλός τὴν ψυχὴν αὐτοῦ τίθησιν ὑπὲρ τῶν προβάτων.	8. Paed. 1.97.3 ο γαρ αγαθος ποιμην την ψυχην εαυτου τιθησιν υπερ των προβατων	4
9. John 10:16 καὶ ἄλλα πρόβατα ἔχω ἃ οὐκ ἔστιν ἐκ τῆς αὐλῆς ταύτης· κάκεῖνα δεῖ με ἀγαγεῖν, καὶ τῆς φωνῆς μου ἀκούσουσιν, καὶ γενήσονται μία ποίμνη, εἷς ποιμὴν.	9. Paed. 1.53.3 καὶ γενησονται (φησιν) οι παντες μια ποιμνη και εις ποιμην	3
10. John 13:33 τεκνία, ἔτι μικρόν μεθ' ὑμῶν εἰμι· ζητήσετέ με, καὶ καθὼς εἶπον τοῖς Ἰουδαίοις ὅτι Ὅπου ἐγὼ ὑπάγω ὑμεῖς οὐ δύνασθε ἐλθεῖν, καὶ ὑμῖν λέγω ἄρτι.	10. Paed. 1.13.3 (φησι γαρ) παιδια ετι μικρον μεθ υμων ειμι	1
11. John 21:4-5 πρωῖας δὲ ἤδη γενομένης ἔστη Ἰησοῦς εἰς τὸν αἰγιαλόν· οὐ μέντοι ᾔδεισαν οἱ μαθηταὶ ὅτι Ἰησοῦς ἐστίν. (5) λέγει οὖν αὐτοῖς [ὁ] Ἰησοῦς, Παιδιά, μὴ τι προσφάγιον ἔχετε; ἀπεκρίθησαν αὐτῷ, Οὐ.	11. Paed. 1.12.2 σταθεις, (φησιν), ο κυριος επι τω αιγιαλω προς τους μαθητας (αλιευοντες δε ευτυχον) ενεφωνησεν τε, παιδια, μη τι οψον εχετε;	5

As far as I can judge, the IPM module is able to deal with one interpolated content word, which needs to be preceded as well as followed by at least one trigram (cf. nos. 9, 22 and 23 in the Appendix), making up the required minimum of 2 trigrams. But consider no. 8 in Table 17, which is interpolated differently in both parallels.

John 10:11	ὁ ποιμὴν ὁ καλός	τὴν ψυχὴν αὐτοῦ	τίθησιν ὑπὲρ τῶν προβάτων
Paed. 1.97.3	ο γαρ αγαθος ποιμην	την ψυχην εαυτου	τιθησιν υπερ των προβατων

According to the n-grams system of the TLG the comparison is rendered internally with lemma forms and without stopwords, resulting in the following comparison:

John 10:11 ποιμήν καλός ψυχή τίθημι πρόβατον

Paed. 1.97.3 αγαθός ποιμήν ψυχή τίθημι πρόβατον

The quotation in *Paedagogus* meets the requirements of a match perfectly, since – ignoring αγαθός – it contains two trigrams (ποιμήν ψυχή τίθημι and ψυχή τίθημι πρόβατον). In John 10:11, however, due to the word καλός the algorithm notices just one of the two trigrams of *Paed.* (ψυχή τίθημι πρόβατον) and therefore sees no match. The addition of stopwords in the process makes the comparison even more complicated since in that case, the match is interpolated at the end (αὐτοῦ|εαυτου) too. One promising solution that might do the job is the use of trigrams that allow for one foreign lexeme (=1-skip-bigram).<sup>57</sup> In that case, both ποιμήν καλός ψυχή (John) and αγαθός ποιμήν ψυχή (Clement) would be matched as a valid 1-skip-bigram and a match would be established on the basis of three 1-skip-bigrams.

However, in preceding paragraphs, I have suggested several times that the inclusion of stopwords might be necessary to find small parallels that have only one or two content words (cf. Table 16). Moreover, stopwords have been shown to be necessary to exclude matches that, although they meet the prerequisites of two trigrams (because of their ‘content’ lexemes), appear to be false positives when stopwords are taken into account (cf. Table 13). Both scenarios seem to call for a reconsideration of the exclusion of stopwords.<sup>58</sup>

If we add stopwords, the case of Paed. 1.97.3 will look like this:

John 10:11 ὁ ποιμήν ὁ καλός ὁ ψυχή αὐτός τίθημι ὑπέρ ὁ πρόβατον

Paed. 1.97.3 ὁ γὰρ αγαθός ποιμήν ὁ ψυχή ἑαυτοῦ τίθημι ὑπέρ ὁ πρόβατον

By using skipgrams, this quotation would be found of five 1-skip-bigrams, since the foreign word between ψυχή and τίθημι will be ignored. Only the start of the match will not give matching 1-skip-bigrams, because of two interpolating words. However, by using 1-skip-bigrams, a large number of false positives can be expected. It might therefore be even better to use 1-skip-trigrams: n-grams consisting of four continuous words with one foreign element. By using 1-skip-trigrams, this match

<sup>57</sup> Cf. David Guthrie, Ben Allison et al. “A Closer Look at Skip-gram Modelling” in: *Proceedings LREC’2006* (Genoa, 2006): 1222–1225. <http://www.lrec-conf.org/proceedings/lrec2006/> (accessed April 6, 2018). They define skip-grams as “a technique where by n-grams are still stored to model language, but they allow for tokens to be skipped” (1222). A ‘token’ can mean any size of text of which the n-gram is built, like a word, or a (set of) character(s). In my case, a trigram with one skippable foreign token will be called a 1-skip-bigram (‘bigram’ since the *resulting* ngram consists of two words). The sentence ποιμήν καλός ψυχή τίθημι πρόβατον will result in the following 1-skip-bigrams: ποιμήν καλός; ποιμήν ψυχή; καλός ψυχή; καλός τίθημι; ψυχή τίθημι; ψυχή πρόβατον; τίθημι πρόβατον.

<sup>58</sup> “We ignore stop-words that do not contribute much meaning, and which can distract from the underlying similarity of two texts.” <http://stephanus.tlg.uci.edu/helppdf/ngrams.pdf>, 1 (accessed March 13, 2018) As already noted (note 9), in practice this means that all words but nouns, verbs, and adjectives are ignored.

will be found on the basis of four 1-skip-trigrams: [1] ὁ ψυχὴ αὐτός|ἐαυτοῦ τίθημι [2] ψυχὴ αὐτός|ἐαυτοῦ τίθημι ὑπέρ [3] αὐτός|ἐαυτοῦ τίθημι ὑπέρ ὁ [4] τίθημι ὑπέρ ὁ πρόβατον.<sup>59</sup>

But what if quotations are even more interpolated, like no. 7? Reduced to its lexemes with stopwords included, the match looks like this:

John 6:53 ἐσθίω ὁ σὰρξ ὁ υἱὸς ὁ ἄνθρωπος καὶ πίνω αὐτός ὁ αἷμα

Paed. 1.42.3 ἐσθίω ἐγὼ (φημί) ὁ σὰρξ καὶ πίνω ἐγὼ ὁ αἷμα

Despite interpolations at the start of both sentences, a match can be established on the basis of two consecutive matching 1-skip-trigrams at the end. But to complicate matters once more, let us have a final look at no. 11, which is extensively interpolated on both sides of the parallel (Table 18).

**Table 18. John 21:4–5 and Paedagogus 1.12.2**

John 21:4–5	Paed. 1.12.2
ἴστημι Ἰησοῦς εἰς ὁ αἰγιαλός οὐ μέντοι οἶδα ὁ μαθητῆς ὅτι Ἰησοῦς εἰμί λέγω οὖν αὐτός [ὁ] Ἰησοῦς παιδίον μή τις προσφάγιον ἔχω	ἴστημι (φημί) ὁ κύριος ἐπὶ ὁ αἰγιαλός πρὸς ὁ μαθητῆς (ἀλιεύω δε τυγχάνω) ἐνφωνέω τε παιδίον μή τις ὄψον ἔχω

Although a match can be established on the basis of two 1-skip-trigrams at the end, it is clear that words in the remaining part of the sentence are parallel too (ἴστημι, ὁ αἰγιαλός, and ὁ μαθητῆς). Due to long interpolations (blue highlighted text) and the use of different words (green highlighted text), it is very difficult to match this part of the sentences with the help of n-grams, which means that other techniques need to be employed to add these parallel words to the match.<sup>60</sup>

#### 4.4. Intertextual phrase matching and the search for allusions

Since the results achieved in the previous section approach already the limits of what is technically possible in finding parallels with n-grams, let us take finally a short look at the allusions listed by Cosaert. As can be seen in Table 19, allusions pose an even bigger challenge than adaptations, since allusions typically contain few verbal parallels and many interpolations (cf. nos. 3, 6, 7, 10, 11).

**Table 19. Allusions in the list of Cosaert and not found by the TLG<sup>61</sup>**

##### Text John UBS2

###### 1. John 1:1

Ἐν ἀρχῇ ἦν ὁ λόγος, καὶ ὁ λόγος ἦν πρὸς τὸν θεόν, καὶ θεὸς ἦν ὁ λόγος.

###### 2. John 1:14

καὶ ὁ λόγος σὰρξ ἐγένετο καὶ ἐσκήνωσεν ἐν ἡμῖν, καὶ ἐθεασάμεθα τὴν δόξαν αὐτοῦ, δόξαν ὡς μονογενοῦς παρὰ πατρός, πλήρης χάριτος καὶ ἀληθείας.

##### Allusion by Clement

###### 1. Paed. 1.4.1

λογος θεος ο εν τω πατρι

###### 2. Paed. 1.9.4

και γαρ ο λογος αυτος εναργως σαρξ γενομενος

<sup>59</sup> The underlined words show the foreign element. According to my experience, the use of 1-skip-trigrams leads to much less false positives than the use of 1-skip-bigrams. However, the general influence of skipgrams on the results needs to be investigated. This requires a separate study.

<sup>60</sup> A possible solution is word for word matching of surrounding context words if a match is already established.

<sup>61</sup> Highlights are on the basis of lemma forms.

3. John 1:18

θεὸν οὐδεὶς ἑώρακεν πώποτε· **μονογενῆς** θεὸς ὁ ὢν εἰς τὸν **κόλπον** τοῦ **πατρὸς**· ἐκεῖνος ἐξηγήσατο.

4. John 1:27

ὁ ὀπίσω μου ἐρχόμενος, οὗ **οὐκ εἰμί [ἐγὼ] ἄξιος** ἵνα **λύσω** αὐτοῦ **τὸν ἱμάντα τοῦ ὑποδήματος**.

5. John 2:1–11 / John 4:46<sup>62</sup>

ἦλθεν οὖν πάλιν εἰς τὴν Κανὰ τῆς Γαλιλαίας, ὅπου **ἐποίησεν τὸ ὕδωρ οἶνον**.

6. John 4:7

Ἔρχεται γυνὴ ἐκ **τῆς Σαμαρείας** ἀντλήσαι ὕδωρ. λέγει αὐτῇ ὁ Ἰησοῦς, Δός μοι **πεῖν**.

7. John 11:43

καὶ ταῦτα **εἰπὼν** φωνῇ μεγάλῃ ἐκραύγασεν, **Λάζαρε**, δεῦρο ἕξω. 11:44 **ἐξῆλθεν ὁ τεθνηκώς** δεδεμένος τοὺς πόδας καὶ τὰς χεῖρας κειρίαις, καὶ ἡ ὄψις αὐτοῦ σουδαρίω περιεδέδετο. λέγει αὐτοῖς ὁ Ἰησοῦς, Λύσατε αὐτὸν καὶ ἄφετε αὐτὸν ὑπάγειν.

8. John 13:5

εἶτα βάλλει ὕδωρ εἰς τὸν νιπτῆρα **καὶ ἤρξατο νίπτειν τοὺς πόδας τῶν μαθητῶν** καὶ ἐκμάσσειν τῷ λεντίῳ ᾧ ἦν **διεζωσμένος**.

9. John 13:5

εἶτα βάλλει ὕδωρ εἰς τὸν νιπτῆρα **καὶ ἤρξατο νίπτειν τοὺς πόδας τῶν μαθητῶν** καὶ ἐκμάσσειν τῷ λεντίῳ ᾧ ἦν **διεζωσμένος**.

10. John 17:21

ἵνα πάντες ἔν ὧσιν, καθὼς σύ, **πάτερ**, **ἐν ἐμοὶ κάγω ἐν σοὶ**, ἵνα καὶ **αὐτοὶ ἐν** ἡμῖν ὧσιν, ἵνα ὁ κόσμος πιστεύῃ ὅτι σύ με ἀπέστειλας.

11. John 19:17

καὶ **βαστάζων** αὐτῷ τὸν σταυρὸν ἐξῆλθεν εἰς τὸν λεγόμενον Κρανίου Τόπον, ὃ λέγεται Ἑβραϊστὶ Γολγοθᾶ,

3. *Paed.* 1.8.2

διον ο **μονογενῆς** εκ **κολπων πατρος** κατατεμπεται

4. *Paed.* 2.117.4

**ουκ αξιος ειναι** ομολογων **τον ιμαντα των υποδηματων λυειν** του κυριου

5. *Paed.* 2.29.1

ει γαρ και **το υδωρ οινον** εν τοις γαμοις **πεποιηκεν**

6. *Paed.* 2.38.2

και **την Σαμαριτιν** ητει **πιειν** σκευει κεραμεω του φρεατος ανιμωσαν

7. *Paed.* 1.6.3

και **τω τεθνεωτι Λαζαρε**, **ειπεν**, εξιθι: + ο δε **εξηλθεν**

8. *Paed.* 2.38.1

**και τους ποδας ενιπτειν** αυτων σαβανω περιζωσαμενος

9. *Paed.* 2.63.2

**και αυτος ο σωτηρ απονιπτων τους ποδας των μαθητων**

10. *Paed.* 1.53.1

**εν αυτω** γαρ ο υιος **και εν** τω υιω ο **πατηρ**

11. *Paed.* 1.23.1

**εβαστασε** τα ξυλα της ιερουργιας ο Ισαακ, ως ο κυριος το ξυλον

Some allusions can be found if stopwords are added to the comparison and if skipgrams are used (cf. nos. 4, 8, 9). Nevertheless, in his allusions, Clement adapts the text of John to such an extent, that the proposed modifications of the TLG's algorithm in section 4.3 do not suffice to find most of the remaining allusions. For these cases techniques other than n-grams are to be explored.

## 5. Recommendations and Conclusions

One of the hardest aspects of writing this article has been the sparse documentation of the IPM module as provided by the TLG. Nevertheless, the analysis has revealed much of the mechanics involved behind the scenes. In this section, I will wrap up the suggestions presented and conclude with a number of remarks.

My first recommendation would be for TLG to provide more transparency about the methods and parameters used by the IPM algorithm, its possibilities, and its limitations. In my view, transparency is of major concern in digital humanities, since a lack of transparency turns the methods used into a

<sup>62</sup> John 4:46 is not mentioned by Cosaert.

black box for those academics that are not familiar with programming. This might result in an unfounded reliance on the outcomes of a digital tool.

Second, I would recommend the team of the TLG to reconsider its choice to simply exclude stopwords. As I have shown, the exclusion of stopwords results in false positives (non-matches in which stopwords are the most distinguishing factor) as well as false negatives, because of parallel passages containing too few content words. A ‘switch’ to turn stopwords exclusion on or off would be most welcome, as would be the option of editing the list of stopwords.

Third, I would like to see a mechanism that concatenates extensive matches. In its current state, the algorithm cuts them down into smaller pieces, which is undesirable and unnecessary, while disturbing statistics if one would count them.<sup>63</sup>

Fourth, the TLG might improve its way of displaying matches. Currently, stopwords at the start and the end of the match are not highlighted properly, while stopwords in the middle of the parallel are highlighted, even if they are dissimilar. One way of doing this is to run a separate comparison after the n-grams comparison, in which the n-grams comparison determines the existence of the match, while the second comparison determines the text to be highlighted.

Fifth, to increase the quality of the matches found, the team of the TLG might review the idea to supply a(n optional) probability filter, to connect results with the best parallel possible.

Sixth, I would like to see the user more involved in defining the parameters of the algorithm. As the evaluation has shown, different types of matches require an algorithm tuned with different parameters. I can imagine two scenarios, that might prove to be helpful to both the standard user and the more demanding one:

- A **standard mode** in which the user can choose between one or more predefined settings of the algorithm, for which the results to be expected are carefully defined in accompanying documentation.
- An **advanced mode**, in which more informed users can select their own parameters, knowing that they might get less expected results. After investigating the results produced by the current algorithm, I propose the following parameters as good candidates to be included in an advanced mode:
  - choice between lemma or exact word comparison<sup>64</sup>
  - size of n-grams
  - size of skips
  - required minimum number of continuous n-grams to constitute a match
  - allowed maximum number of interpolating words
  - optional use of a probability filter
  - manually adjustable number of content lines displayed
  - various modes of highlighting the result:
    - lemma comparison
    - exact word comparison
    - both with a different color

---

<sup>63</sup> In the new version of the IPM module, this has been improved. Now, the user can choose between the concatenation of one to six n-grams (=7 content words). However, longer quotations are still cut into pieces.

<sup>64</sup> It should be noted that in the new version of the IPM module, the user can already select these two options.

With such an advanced mode, the use of the algorithm would become much more flexible and transparent. Moreover, users would be allowed to experiment with different settings to obtain, use, and describe the results they like. In the current state of the IPM module, the user can only use the settings of the module as it is, without even knowing exactly what is happening in the black box of the algorithm.

To conclude, I would like to stress that the IPM module of the TLG does already a good job in tracing and displaying parallels. My suggestions should by no means be taken to imply that the IPM module is useless. What they do show, however, is that a careful analysis of the results reveals a number of possible improvements and that the current results require a proper procedure of post-processing if the acquisition of an accurate list of quotations is intended.<sup>65</sup> “As with many other features of the TLG, the ultimate value of N-gram searching will be revealed by the ends it achieves.”<sup>66</sup>

---

<sup>65</sup> Unfortunately, the recent update of the module has not improved the overall accuracy of the module.

<sup>66</sup> <https://classicalstudies.org/scs-blog/scott-farrington/review-thesaurus-linguae-graecae> blogpost written by: Scott Farrington “Review: Thesaurus Linguae Graecae”, Januari 2, 2017 (last accessed on August 21, 2018).

## Appendix: Matches between the *Paedagogus* of Clement of Alexandria and John (TLG full list)<sup>67</sup>

### 1. N.T.Ev.Jo. {0031.004} 1.1.1

#### ΚΑΤΑ ΙΩΑΝΝΗΝ

(1) Ἐν ἀρχῇ ἦν ὁ λόγος, καὶ ὁ λόγος ἦν πρὸς τὸν θεόν, καὶ θεὸς ἦν ὁ λόγος. (2) οὗτος ἦν ἐν ἀρχῇ πρὸς τὸν θεόν. (3) πάντα δι' αὐτοῦ ἐγένετο, καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ

### 2. N.T.Ev.Jo. {0031.004} 1.3.1

καὶ θεὸς ἦν ὁ λόγος. (2) οὗτος ἦν ἐν ἀρχῇ πρὸς τὸν θεόν. (3) πάντα δι' αὐτοῦ ἐγένετο, καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ ἐν. ὁ γέγονεν (4) ἐν αὐτῷ ζωὴ ἦν, καὶ ἡ ζωὴ ἦν τὸ

### 3. N.T.Ev.Jo. {0031.004} 1.17.1

αὐτοῦ ἡμεῖς πάντες ἐλάβομεν, καὶ χάριν ἀντὶ χάριτος· (17) ὅτι ὁ νόμος διὰ Μωϋσέως ἐδόθη, ἡ χάρις καὶ ἡ ἀλήθεια διὰ Ἰησοῦ Χριστοῦ ἐγένετο. (18) θεὸν οὐδεὶς ἑώρακεν πώποτε· μονογενὴς θεὸς ὁ ὢν εἰς τὸν κόλπον τοῦ πατρὸς

### 4. N.T.Ev.Jo. {0031.004} 3.15.1

τῆ ἐρήμῳ, οὕτως ὑψωθῆναι δεῖ τὸν υἱὸν τοῦ ἀνθρώπου, (15) ἵνα πᾶς ὁ πιστεύων ἐν αὐτῷ ἔχη ζωὴν αἰώνιον. (16) Οὕτως γὰρ ἠγάπησεν ὁ θεὸς τὸν κόσμον, ὥστε

### 5. N.T.Ev.Jo. {0031.004} 3.36.2

τὸν υἱόν, καὶ πάντα δέδωκεν ἐν τῇ χειρὶ αὐτοῦ. (36) ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον· ὁ δὲ ἀπειθῶν τῷ υἱῷ οὐκ ὄψεται ζωὴν, ἀλλ' ἡ ὀργὴ τοῦ θεοῦ μένει ἐπ'

### 6. N.T.Ev.Jo. {0031.004} 3.36.2

τὸν υἱόν, καὶ πάντα δέδωκεν ἐν τῇ χειρὶ αὐτοῦ. (36) ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον· ὁ δὲ ἀπειθῶν τῷ υἱῷ οὐκ ὄψεται ζωὴν, ἀλλ' ἡ ὀργὴ τοῦ θεοῦ μένει ἐπ'

### 7. N.T.Ev.Jo. {0031.004} 4.32.1

(31) Ἐν τῷ μεταξὺ ἡρώτων αὐτὸν οἱ μαθηταὶ λέγοντες, Ῥαββί, φάγε. (32) ὁ δὲ εἶπεν αὐτοῖς, Ἐγὼ βρώσιν ἔχω φαγεῖν ἢν ὑμεῖς οὐκ οἴδατε. (33) ἔλεγον οὖν οἱ μαθηταὶ πρὸς ἀλλήλους, Μὴ τις ἤνεγκεν αὐτῷ φαγεῖν; (34) λέγει

### 8. N.T.Ev.Jo. {0031.004} 4.34.2

πρὸς ἀλλήλους, Μὴ τις ἤνεγκεν αὐτῷ φαγεῖν; (34) λέγει αὐτοῖς ὁ Ἰησοῦς, Ἐμὸν βρώμᾳ ἐστὶν ἵνα ποιήσω τὸ θέλημα τοῦ πέμψαντός με καὶ τελειώσω αὐτοῦ τὸ ἔργον. (35) οὐχ ὑμεῖς λέγετε ὅτι Ἔτι τετράμηνός ἐστιν καὶ ὁ

### 1. Clem.Al.Paed. {0555.002} 1.8.62.4.2

(4) ἀλλ' οὐδὲ ὑπὸ τοῦ λόγου· ἐν γὰρ ἄμφω, ὁ θεός, ὅτι εἶπεν «ἐν ἀρχῇ ὁ λόγος ἦν ἐν τῷ θεῷ, καὶ θεὸς ἦν ὁ λόγος». Εἰ δὲ οὐ μισεῖ τῶν ὑπ' αὐτοῦ γενομένων οὐδέν, λείπεται φιλεῖν αὐτό.

### 2. Clem.Al.Paed. {0555.002} 1.11.97.3.5

«πᾶσα σοφία παρὰ κυρίου καὶ μετ' αὐτοῦ ἐστὶν εἰς τὸν αἰῶνα»· παρρησία δὲ ὅτι θεὸς καὶ δημιουργός, «πάντα γὰρ δι' (5) αὐτοῦ ἐγένετο καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ ἐν»· εὐνοία δὲ ὅτι μόνος ὑπὲρ ἡμῶν ἱερεῖον ἑαυτὸν ἐπιδέδωκεν,

### 3. Clem.Al.Paed. {0555.002} 1.7.60.2.1

θεράποντος αὐτοῦ· διὸ καὶ πρόσκαιρος ἐγένετο, «ἡ δὲ αἰδῖος (2) χάρις καὶ ἡ ἀλήθεια διὰ Ἰησοῦ Χριστοῦ ἐγένετο.» Ὅρατε τὰς λέξεις τῆς γραφῆς· ἐπὶ μὲν τοῦ νόμου «ἐδόθη» φησὶ μόνον,

### 4. Clem.Al.Paed. {0555.002} 1.6.28.5.6

εἰπών· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ (5) θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν αἰώνιον, (1) καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.» Καθ' ὅσον μὲν οὖν

### 5. Clem.Al.Paed. {0555.002} 1.6.28.5.6

εἰπών· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ (5) θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν αἰώνιον, (1) καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.» Καθ' ὅσον μὲν οὖν

### 6. Clem.Al.Paed. {0555.002} 1.6.29.2.1

πιστεύομεν. Πίστις γὰρ μαθήσεως τελειότης· διὰ τοῦτο φησιν (2) «ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον». Εἰ τοίνυν οἱ πιστεύσαντες ἔχομεν τὴν ζωὴν, τί περαιτέρω τοῦ κεκτηθῆναι

### 7. Clem.Al.Paed. {0555.002} 1.6.45.4.1

Ἀλλὰ καὶ τοῖς ὑποτιθησὶν παιδίοις ἀρκεῖ μόνον τὸ γάλα καὶ (4) ποτὸν εἶναι καὶ τροφήν. «Ἐγὼ», φησὶν ὁ κύριος, «βρώσιν ἔχω φαγεῖν, ἢν ὑμεῖς οὐκ οἴδατε· ἐμὸν βρώμᾳ ἐστὶν, ἵνα ποιήσω τὸ θέλημα τοῦ πέμψαντός με». Ὅρατε ἄλλο βρώμα

### 8. Clem.Al.Paed. {0555.002} 1.6.45.4.2

(4) ποτὸν εἶναι καὶ τροφήν. «Ἐγὼ», φησὶν ὁ κύριος, «βρώσιν ἔχω φαγεῖν, ἢν ὑμεῖς οὐκ οἴδατε· ἐμὸν βρώμᾳ ἐστὶν, ἵνα ποιήσω τὸ θέλημα τοῦ πέμψαντός με». Ὅρατε ἄλλο βρώμα

<sup>67</sup> The discussion of the matches in this article is based on this list. These results were produced by the TLG till November 14, 2018. Currently the TLG displays a new list of matches because of a different set of parameters used. The most important change is that only one trigram is used to establish a match instead of two trigrams as it was in the old version. For readability reasons, the matches displayed have been cleaned of superfluous elements.



## 9. N.T.Ev.Jo. {0031.004} 5.24.1

τὸν υἱὸν οὐ τιμᾷ τὸν πατέρα τὸν πέμψαντα αὐτόν.  
(24) Ἀμὴν ἀμὴν λέγω ὑμῖν ὅτι ὁ τὸν λόγον μου ἀκούων  
καὶ πιστεύων τῷ πέμψαντί με ἔχει ζωὴν αἰώνιον, καὶ  
εἰς κρίσιν οὐκ ἔρχεται ἀλλὰ μεταβέβηκεν ἐκ τοῦ θανάτου

## 10. N.T.Ev.Jo. {0031.004} 5.24.2

(24) Ἀμὴν ἀμὴν λέγω ὑμῖν ὅτι ὁ τὸν λόγον μου ἀκούων  
καὶ πιστεύων τῷ πέμψαντί με ἔχει ζωὴν αἰώνιον, καὶ  
εἰς κρίσιν οὐκ ἔρχεται ἀλλὰ μεταβέβηκεν ἐκ τοῦ θανάτου  
εἰς τὴν ζωὴν. (25) ἀμὴν ἀμὴν λέγω ὑμῖν ὅτι ἔρχεται

## 11. N.T.Ev.Jo. {0031.004} 6.32.2

ἔδωκεν αὐτοῖς φαγεῖν. (32) εἶπεν οὖν αὐτοῖς ὁ Ἰησοῦς,  
Ἀμὴν ἀμὴν λέγω ὑμῖν, οὐ Μωϋσῆς δέδωκεν ὑμῖν τὸν  
ἄρτον ἐκ τοῦ οὐρανοῦ, ἀλλ' ὁ πατήρ μου δίδωσιν ὑμῖν  
τὸν ἄρτον ἐκ τοῦ οὐρανοῦ τὸν ἀληθινόν· (33) ὁ γὰρ ἄρτος  
τοῦ θεοῦ ἐστὶν ὁ καταβαίνων ἐκ τοῦ οὐρανοῦ καὶ ζωὴν

## 12. N.T.Ev.Jo. {0031.004} 6.32.4

ἄρτον ἐκ τοῦ οὐρανοῦ, ἀλλ' ὁ πατήρ μου δίδωσιν ὑμῖν  
τὸν ἄρτον ἐκ τοῦ οὐρανοῦ τὸν ἀληθινόν· (33) ὁ γὰρ ἄρτος  
τοῦ θεοῦ ἐστὶν ὁ καταβαίνων ἐκ τοῦ οὐρανοῦ καὶ ζωὴν  
διδούς τῷ κόσμῳ. (34) Εἶπον οὖν πρὸς αὐτόν, Κύριε,  
πάντοτε δὸς ἡμῖν

## 13. N.T.Ev.Jo. {0031.004} 6.40.2

αὐτὸ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ. (40) τοῦτο γὰρ ἐστὶν τὸ  
θέλημα τοῦ πατρός μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν  
καὶ πιστεύων εἰς αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω  
αὐτὸν ἐγὼ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ.

## 14. N.T.Ev.Jo. {0031.004} 6.40.2

αὐτὸ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ. (40) τοῦτο γὰρ ἐστὶν τὸ  
θέλημα τοῦ πατρός μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν  
καὶ πιστεύων εἰς αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω  
αὐτὸν ἐγὼ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ.

## 15. N.T.Ev.Jo. {0031.004} 6.40.3

θέλημα τοῦ πατρός μου, ἵνα πᾶς ὁ θεωρῶν τὸν υἱὸν  
καὶ πιστεύων εἰς αὐτὸν ἔχη ζωὴν αἰώνιον, καὶ ἀναστήσω  
αὐτὸν ἐγὼ [ἐν] τῇ ἐσχάτῃ ἡμέρᾳ. (41) Ἐγόγγυζον οὖν οἱ  
Ἰουδαῖοι περὶ αὐτοῦ ὅτι εἶπεν,

## 16. N.T.Ev.Jo. {0031.004} 6.47.2

οὗτος ἐώρακεν τὸν πατέρα. (47) ἀμὴν ἀμὴν λέγω ὑμῖν,  
ὁ πιστεύων ἔχει ζωὴν αἰώνιον. (48) ἐγὼ εἰμι ὁ ἄρτος  
τῆς ζωῆς. (49) οἱ πατέρες ὑμῶν ἔφαγον ἐν τῇ ἐρήμῳ τὸ

## 17. N.T.Ev.Jo. {0031.004} 6.51.3

καταβάς· ἐάν τις φάγη ἐκ τούτου τοῦ ἄρτου ζήσει εἰς  
τὸν αἰῶνα· καὶ ὁ ἄρτος δὲ ὃν ἐγὼ δώσω ἢ σὰρξ μου

ἀλληγορούμενον παραπλησίως γάλακτι τὸ θέλημα τοῦ  
θεοῦ.

## 9. Clem.Al.Paed. {0555.002} 1.6.27.1.5

ἔπεσθαι Χριστῷ. «Ὁ γὰρ γέγονεν ἐν αὐτῷ, ζωὴ ἐστίν.»  
«Ἀμὴν ἀμὴν λέγω ὑμῖν», φησὶν, «ὁ τὸν λόγον μου  
ἀκούων (5) καὶ πιστεύων τῷ πέμψαντί με ἔχει  
ζωὴν αἰώνιον καὶ εἰς  
κρίσιν οὐκ ἔρχεται, ἀλλὰ μεταβέβηκεν ἐκ τοῦ θανάτου εἰς

## 10. Clem.Al.Paed. {0555.002} 1.6.27.1.6

«Ἀμὴν ἀμὴν λέγω ὑμῖν», φησὶν, «ὁ τὸν λόγον μου  
ἀκούων (5) καὶ πιστεύων τῷ πέμψαντί με ἔχει ζωὴν  
αἰώνιον καὶ εἰς  
κρίσιν οὐκ ἔρχεται, ἀλλὰ μεταβέβηκεν ἐκ τοῦ θανάτου εἰς  
τὴν (2) ζωὴν.» Οὕτω τὸ πιστεῦσαι μόνον καὶ  
ἀναγεννηθῆναι

## 11. Clem.Al.Paed. {0555.002} 1.6.46.2.2

(2) γάλα. Ἔτι δὲ καὶ ἄρτον αὐτὸν οὐρανῶν ὁμολογεῖ ὁ  
λόγος. «Οὐ γὰρ Μωσῆς», φησὶν, «ἔδωκεν ὑμῖν τὸν ἄρτον ἐκ  
τοῦ οὐρανοῦ νοῦ, ἀλλ' ὁ πατήρ μου δίδωσιν ὑμῖν τὸν ἄρτον  
ἐκ τοῦ οὐρανοῦ τὸν ἀληθινόν· ὁ γὰρ ἄρτος τοῦ θεοῦ ἐστὶν  
ὁ ἐκ τοῦ οὐρανοῦ καταβαίνων καὶ ζωὴν διδούς τῷ κόσμῳ.  
Καὶ ὁ ἄρτος, ὃν ἐγὼ (5)

## 12. Clem.Al.Paed. {0555.002} 1.6.46.2.4

ἀλλ' ὁ πατήρ μου δίδωσιν ὑμῖν τὸν ἄρτον ἐκ τοῦ  
οὐρανοῦ τὸν ἀληθινόν· ὁ γὰρ ἄρτος τοῦ θεοῦ ἐστὶν ὁ ἐκ  
τοῦ οὐρανοῦ καταβαίνων καὶ ζωὴν διδούς τῷ κόσμῳ. Καὶ ὁ  
ἄρτος, ὃν ἐγὼ (5) (3) δώσω, ἡ σὰρξ μου ἐστὶν ὑπὲρ τῆς τοῦ  
κόσμου ζωῆς.» Ἐνταῦθα

## 13. Clem.Al.Paed. {0555.002} 1.6.28.5.5

κύριος σαφέστατα τῆς σωτηρίας τὴν ἰσότητα ἀπεκάλυψεν  
εἰπών· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρός μου, ἵνα πᾶς  
ὁ (5) θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν  
αἰώνιον, (1) καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.»  
Καθ' ὅσον μὲν οὖν

## 14. Clem.Al.Paed. {0555.002} 1.6.29.2.1

πιστεύομεν. Πίστις γὰρ μαθήσεως τελειότης· διὰ τοῦτο  
φησὶν (2) «ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον». Εἰ  
τοίνυν οἱ πιστεύσαντες ἔχομεν τὴν ζωὴν, τί περαιτέρω τοῦ  
κεκτησθαι

## 15. Clem.Al.Paed. {0555.002} 1.6.28.5.6

εἰπών· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρός μου, ἵνα πᾶς  
ὁ (5) θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν  
αἰώνιον, (1) καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.»  
Καθ' ὅσον μὲν οὖν δυνατὸν ἐν τῷδε τῷ κόσμῳ, ὃν ἐσχάτην  
ἡμέραν ἤνιξατο εἰς

## 16. Clem.Al.Paed. {0555.002} 1.6.28.5.6

εἰπών· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρός μου, ἵνα πᾶς  
ὁ (5) θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν  
αἰώνιον, (1) καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.»  
Καθ' ὅσον μὲν οὖν

## 17. Clem.Al.Paed. {0555.002} 1.6.46.2.5

τὸν ἀληθινόν· ὁ γὰρ ἄρτος τοῦ θεοῦ ἐστὶν ὁ ἐκ τοῦ  
οὐρανοῦ

ἔστιν ὑπὲρ τῆς τοῦ κόσμου ζωῆς.

(52) Ἐμάχοντο οὖν πρὸς ἀλλήλους οἱ Ἰουδαῖοι λέγοντες,

18. **N.T.Ev.Jo. {0031.004}** 6.54.2

ἐν ἑαυτοῖς. (54) ὁ τρώγων μου τὴν σάρκα καὶ πίνων μου τὸ αἶμα ἔχει ζωὴν αἰώνιον, καὶ γὰρ ἀναστήσω αὐτὸν τῇ ἐσχάτῃ ἡμέρᾳ. (55) ἡ γὰρ σὰρξ μου ἀληθὴς ἐστὶν βρωσίς, καὶ τὸ αἶμά μου ἀληθὴς ἐστὶν πόσις. (56) ὁ

19. **N.T.Ev.Jo. {0031.004}** 6.68.2

ὕμεις θέλετε ὑπάγειν; (68) ἀπεκρίθη αὐτῷ Σίμων Πέτρος, Κύριε, πρὸς τίνα ἀπελευσόμεθα; ῥήματα ζωῆς αἰώνιου ἔχεις; (69) καὶ ἡμεῖς πεπιστεύκαμεν καὶ ἐγνώκαμεν ὅτι σὺ εἶ ὁ ἅγιος τοῦ θεοῦ. (70) ἀπεκρίθη αὐτοῖς ὁ Ἰησοῦς,

20. **N.T.Ev.Jo. {0031.004}** 12.13.4

Ἦσαννά·

εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου, καὶ ὁ βασιλεὺς τοῦ Ἰσραήλ. (5)

21. **N.T.Ev.Jo. {0031.004}** 12.13.4

Ἦσαννά·

εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου, καὶ ὁ βασιλεὺς τοῦ Ἰσραήλ. (5)

22. **N.T.Ev.Jo. {0031.004}** 15.1.1

ποιῶ. Ἐγείρεσθε, ἄγωμεν ἐντεῦθεν. 15. (1) Ἐγὼ εἰμι ἡ ἄμπελος ἡ ἀληθινή, καὶ ὁ πατήρ μου ὁ γεωργὸς ἐστίν. (2) πᾶν κλῆμα ἐν ἐμοὶ μὴ φέρον καρπὸν, αἶρει αὐτό, καὶ πᾶν τὸ καρπὸν φέρον καθαίρει αὐτό ἵνα καρπὸν πλείονα φέρῃ. (3) ἦδη ὑμεῖς καθαροὶ ἐστε

23. **N.T.Ev.Jo. {0031.004}** 15.2.2

ὁ γεωργὸς ἐστίν. (2) πᾶν κλῆμα ἐν ἐμοὶ μὴ φέρον καρπὸν, αἶρει αὐτό, καὶ πᾶν τὸ καρπὸν φέρον καθαίρει αὐτό ἵνα καρπὸν πλείονα φέρῃ. (3) ἦδη ὑμεῖς καθαροὶ ἐστε διὰ τὸν λόγον ὃν λελάληκα ὑμῖν. (4) μείνατε ἐν ἐμοί,

24. **N.T.Ev.Jo. {0031.004}** 17.21.2

τῶν πιστευόντων διὰ τοῦ λόγου αὐτῶν εἰς ἐμέ, (21) ἵνα πάντες ἐν ὧσιν, καθὼς σύ, πάτερ, ἐν ἐμοὶ καὶ ἐν σοί, ἵνα καὶ αὐτοὶ ἐν ἡμῖν ὧσιν, ἵνα ὁ κόσμος πιστεῦῃ ὅτι σύ με ἀπέστειλας. (22) καὶ γὰρ τὴν δόξαν ἣν δέδωκάς μοι δέδωκα αὐτοῖς, ἵνα ὧσιν ἐν καθὼς ἡμεῖς ἐν, (23) ἐγὼ ἐν

25. **N.T.Ev.Jo. {0031.004}** 17.22.1

ἵνα καὶ αὐτοὶ ἐν ἡμῖν ὧσιν, ἵνα ὁ κόσμος πιστεῦῃ ὅτι σύ με ἀπέστειλας. (22) καὶ γὰρ τὴν δόξαν ἣν δέδωκάς μοι δέδωκα αὐτοῖς, ἵνα ὧσιν ἐν καθὼς ἡμεῖς ἐν, (23) ἐγὼ ἐν

καταβαίνων καὶ ζωὴν διδοὺς τῷ κόσμῳ. Καὶ ὁ ἄρτος, ὃν ἐγὼ (5) (3) δώσω, ἡ σὰρξ μου ἐστὶν ὑπὲρ τῆς τοῦ κόσμου ζωῆς.» Ἐνταῦθα τὸ μυστικὸν τοῦ ἄρτου παρασημειωτέον, ὅτι σάρκα αὐτὸν λέγει

18. **Clem.Al.Paed. {0555.002}** 1.6.28.5.6

εἰπών· «τοῦτο γὰρ ἐστὶ τὸ θέλημα τοῦ πατρὸς μου, ἵνα πᾶς ὁ (5) θεωρῶν τὸν υἱὸν καὶ πιστεύων ἐπ' αὐτὸν ἔχη ζωὴν αἰώνιον, (1) καὶ ἀναστήσω αὐτὸν ἐν τῇ ἐσχάτῃ ἡμέρᾳ.» Καθ' ὅσον μὲν οὖν δυνατὸν ἐν τῷδε τῷ κόσμῳ, ὃν ἐσχάτην ἡμέραν ἠνίξατο εἰς

19. **Clem.Al.Paed. {0555.002}** 1.6.29.2.1

πιστεύομεν. Πίστις γὰρ μαθήσεως τελειότης· διὰ τοῦτο φησιν (2) «ὁ πιστεύων εἰς τὸν υἱὸν ἔχει ζωὴν αἰώνιον». Εἰ τοίνυν οἱ πιστεύσαντες ἔχομεν τὴν ζωὴν, τί περαιτέρω τοῦ κεκτησθαι ζωὴν αἰδίων ὑπολείπεται; οὐδὲν δὲ ἐνδεῖ τῇ πίστει τελεία

20. **Clem.Al.Paed. {0555.002}** 1.5.12.5.4

ἐξῆλθον εἰς ὑπάντησιν κυρίῳ καὶ ἐκέκραγον λέγοντες, ὡσαννὰ τῷ υἱῷ Δαβίδ, εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου», φῶς καὶ δόξα καὶ αἶνος μεθ' ἱκετηρίας τῷ κυρίῳ· τουτὶ γὰρ (5)

21. **Clem.Al.Paed. {0555.002}** 1.9.79.3.3

οἶκος ὑμῶν ἔρημος, λέγω γὰρ ὑμῖν· ἀπάρτι οὐ μὴ ἴδητέ με, ἕως ἂν εἴπητε· εὐλογημένος ὁ ἐρχόμενος ἐν ὀνόματι κυρίου.»

Εἰ γὰρ οὐ δέχεσθε τὴν φιλανθρωπίαν, ἐπιγνώσεσθε τὴν

22. **Clem.Al.Paed. {0555.002}** 1.8.66.4.3

ἐκφαίνει, τὴν πολύτροπον καὶ πολυωφελεῖ θεραπείαν ἀλληγορῶν, ὁπηνίκα εἰπὼν «ἐγὼ εἰμι ἡ ἄμπελος ἡ ἀληθινή, καὶ ὁ πατήρ μου ὁ γεωργὸς ἐστίν», εἶτα ἐπήγαγεν πάλιν «πᾶν κλῆμα ἐν ἐμοὶ μὴ φέρον καρπὸν αἶρει αὐτό, καὶ πᾶν τὸ καρ- (5)ποφοροῦν καθαίρει, ἵνα καρπὸν πλείονα φέρῃ».

23. **Clem.Al.Paed. {0555.002}** 1.8.66.4.5

πατήρ μου ὁ γεωργὸς ἐστίν», εἶτα ἐπήγαγεν πάλιν «πᾶν κλῆμα ἐν ἐμοὶ μὴ φέρον καρπὸν αἶρει αὐτό, καὶ πᾶν τὸ καρ- (5)ποφοροῦν καθαίρει, ἵνα καρπὸν πλείονα φέρῃ». Καθυλομανεῖ γὰρ μὴ κλαδευομένη ἡ ἄμπελος, οὕτως δὲ καὶ ὁ

24. **Clem.Al.Paed. {0555.002}** 1.8.71.1.4

ἔτι λόγων τὴν εὐαγγέλιον τοῦ κυρίου παραθεμένῳ φωνῆν· ἐν

μὲν αὐτὸν λέγει· «ἵνα πάντες ἐν ὧσιν, καθὼς σύ, πάτερ, ἐν ἐμοὶ καὶ ἐν σοί, ἵνα καὶ αὐτοὶ ἐν ἡμῖν ἐν ὧσιν, ἵνα καὶ ὁ (5)

κόσμος πιστεῦῃ, ὅτι σύ με ἀπέστειλας. Καὶ γὰρ τὴν δόξαν, ἣν ἔδωκάς μοι, δέδωκα αὐτοῖς, ἵνα ὧσιν ἐν καθὼς ἡμεῖς ἐν· ἐγὼ

ἐν αὐτοῖς καὶ σὺ ἐν ἐμοί, ἵνα ὧσι τετελειωμένοι εἰς ἐν»· ἐν

25. **Clem.Al.Paed. {0555.002}** 1.8.71.1.6

ἐμοὶ καὶ ἐν σοί, ἵνα καὶ αὐτοὶ ἐν ἡμῖν ἐν ὧσιν, ἵνα καὶ ὁ (5)

κόσμος πιστεῦῃ, ὅτι σύ με ἀπέστειλας. Καὶ γὰρ τὴν δόξαν, ἣν

αὐτοῖς καὶ σὺ ἐν ἐμοί, ἵνα ὧσιν τετελειωμένοι εἰς ἕν, ἵνα γινώσκῃ ὁ κόσμος ὅτι σὺ με ἀπέστειλας καὶ

26. **N.T.Ev.Jo. {0031.004}** 17.24.1

ἵνα γινώσκῃ ὁ κόσμος ὅτι σὺ με ἀπέστειλας καὶ ἠγάπησας αὐτοὺς καθὼς ἐμέ ἠγάπησας. (24) Πάτερ, ὃ δέδωκάς μοι, θέλω ἵνα ὅπου εἰμί ἐγὼ κάκεῖνοι ὧσιν μετ' ἐμοῦ, ἵνα θεωρῶσιν τὴν δόξαν τὴν ἐμήν ἣν δέδωκάς μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. (25) πάτερ

27. **N.T.Ev.Jo. {0031.004}** 17.24.3

κάς μοι, θέλω ἵνα ὅπου εἰμί ἐγὼ κάκεῖνοι ὧσιν μετ' ἐμοῦ, ἵνα θεωρῶσιν τὴν δόξαν τὴν ἐμήν ἣν δέδωκάς μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. (25) πάτερ δίκαιε, καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω, καὶ οὗτοι ἔγνωσαν ὅτι σὺ με ἀπέστειλας, (26) καὶ

28. **N.T.Ev.Jo. {0031.004}** 17.25.1

ἐμοῦ, ἵνα θεωρῶσιν τὴν δόξαν τὴν ἐμήν ἣν δέδωκάς μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. (25) πάτερ δίκαιε, καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω, καὶ οὗτοι ἔγνωσαν ὅτι σὺ με ἀπέστειλας, (26) καὶ ἐγνώρισα αὐτοῖς τὸ ὄνομά σου καὶ γνωρίσω, ἵνα ἡ ἀγάπη

29. **N.T.Ev.Jo. {0031.004}** 17.25.3

δίκαιε, καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω, καὶ οὗτοι ἔγνωσαν ὅτι σὺ με ἀπέστειλας, (26) καὶ ἐγνώρισα αὐτοῖς τὸ ὄνομά σου καὶ γνωρίσω, ἵνα ἡ ἀγάπη ἣν ἠγάπησάς με ἐν αὐτοῖς ἢ κάγῳ ἐν αὐτοῖς.

ἔδωκάς μοι, δέδωκα αὐτοῖς, ἵνα ὧσιν ἐν καθὼς ἡμεῖς ἐν· ἐγὼ

ἐν αὐτοῖς καὶ σὺ ἐν ἐμοί, ἵνα ὧσι τετελειωμένοι εἰς ἕν· ἐν (2) δὲ ὁ θεὸς καὶ ἐπέκεινα τοῦ ἐνὸς καὶ ὑπὲρ αὐτὴν μονάδα.

26. **Clem.Al.Paed. {0555.002}** 1.8.71.2.6

καὶ δίκαιός ἐστιν ἐν τῷ αὐτῷ μαρτυρήσει κύριος εὐαγγελίῳ (5) λέγων «πάτερ, οὐς ἔδωκάς μοι, θέλω ἵνα ὅπου εἰμί ἐγὼ, κάκεῖνοι ὧσι μετ' ἐμοῦ, ἵνα θεωρῶσι τὴν δόξαν τὴν ἐμήν, ἣν ἔδωκάς μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. Πάτερ δίκαιε, καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω κάκεῖνοι

27. **Clem.Al.Paed. {0555.002}** 1.8.71.2.7

λέγων «πάτερ, οὐς ἔδωκάς μοι, θέλω ἵνα ὅπου εἰμί ἐγὼ, κάκεῖνοι ὧσι μετ' ἐμοῦ, ἵνα θεωρῶσι τὴν δόξαν τὴν ἐμήν, ἣν ἔδωκάς μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. Πάτερ δίκαιε, καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω κάκεῖνοι ἔγνωσαν, ὅτι σὺ με ἀπέστειλας· καὶ ἐγνώρισα αὐτοῖς τὸ ὄνομά (10)

28. **Clem.Al.Paed. {0555.002}** 1.8.71.2.8

κάκεῖνοι ὧσι μετ' ἐμοῦ, ἵνα θεωρῶσι τὴν δόξαν τὴν ἐμήν, ἣν ἔδωκάς μοι, ὅτι ἠγάπησάς με πρὸ καταβολῆς κόσμου. Πάτερ δίκαιε, καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω κάκεῖνοι ἔγνωσαν, ὅτι σὺ με ἀπέστειλας· καὶ ἐγνώρισα αὐτοῖς τὸ ὄνομά (10) (3) σου καὶ γνωρίσω». Οὗτός ἐστιν «ὁ ἀποδιδούς ἁμαρτίας

29. **Clem.Al.Paed. {0555.002}** 1.8.71.2.10

δίκαιε, καὶ ὁ κόσμος σε οὐκ ἔγνω, ἐγὼ δέ σε ἔγνω κάκεῖνοι ἔγνωσαν, ὅτι σὺ με ἀπέστειλας· καὶ ἐγνώρισα αὐτοῖς τὸ ὄνομά (10) (3) σου καὶ γνωρίσω». Οὗτός ἐστιν «ὁ ἀποδιδούς ἁμαρτίας πατέρων ἐπὶ τέκνα τοῖς μισοῦσι καὶ ποιῶν ἔλεος τοῖς ἀγαπῶσιν».