

The AGNT Project Report—Q3 2019

As a licensee or friend of AGNT or ANLEX, we would like to update you once a quarter about our continuing work to enhance and perfect these databases and about our plans for the future.



The Project. *The AGNT Project Report—Q3 2008* introduced the team, outlined ongoing tasks, and discussed potential tasks.



The Thesaurus Linguae Graecae® Project: Part 2

John J. Hughes

The Thesaurus Linguae Graecae® Digital Library contains virtually all surviving Greek texts from Homer (8 c. B.C.) to the fall of Byzantium in A.D. 1453, as well as a large number of texts up to the twentieth century. The TLG contains “more than 110 million words from over 10,000 works associated with 4,000 authors and is constantly updated and improved with new features and texts. The full corpus is available to more than 2,000 subscribing institutions and thousands of individuals in 73 countries worldwide” ([link](#)).

No other resource in the world can touch the TLG when it comes to finding *all* occurrences of a Greek word or phrase (inflected or lemmatized) in *any* Greek literature in the period between Homer (8 c. B.C.) and the fall of Byzantium in A.D. 1453¹ or any authorial, chronological, or corporeal subset thereof. In short, if you wish to do serious, world-class diachronic or synchronic word studies, you need access to the full TLG corpus.

In the last issue of this newsletter (Q2 2019), I introduced the Thesaurus Linguae Graecae® Project, outlined its history, and discussed these two features in depth:

- Search the canon. [Link](#) and [link](#). Video tutorial [link](#).
- Search the TLG corpus. [Link](#) and [link](#). Video tutorial [link](#).

In this version of the newsletter, I will cover these five features:

- Browse TLG texts. [Link](#) and [link](#). Video tutorial links: [link](#) and [link](#).
- Consult online lexica. [Link](#).
- Create N-gram comparisons (intertextual phrase matching). [Link](#).
- Create word-usage statistics. [Link](#).
- Use the vocabulary tools. [Link](#).

¹ In an e-mail to me dated 06.11.2019, Maria Pantelia, TLG Director, noted: “We are working on A.D. 16–19 and the corpus now contains at least 180 works from that period.” I would like to thank Dr. Pantelia for kindly reading this review and making many helpful comments.

Note: The information and much of the wording in the following review is taken from the TLG site ([link](#)) and the TLG instructional videos ([link](#)).

Browse TLG Texts

The TLG allows you to browse one text or two texts in parallel. To browse a single text, select “Browse” and then “Browse One Text” from the TLG’s main menu.



In the search field, enter the author’s name or the text’s TLG number. The autocomplete feature is enforced. For example, to find works by Justin Martyr, entering “Justin” in the search box, displays this:



Selecting the second entry—“JUSTINUS MARTYR Apol. (0645)”—displays all the works by Justin Martyr:

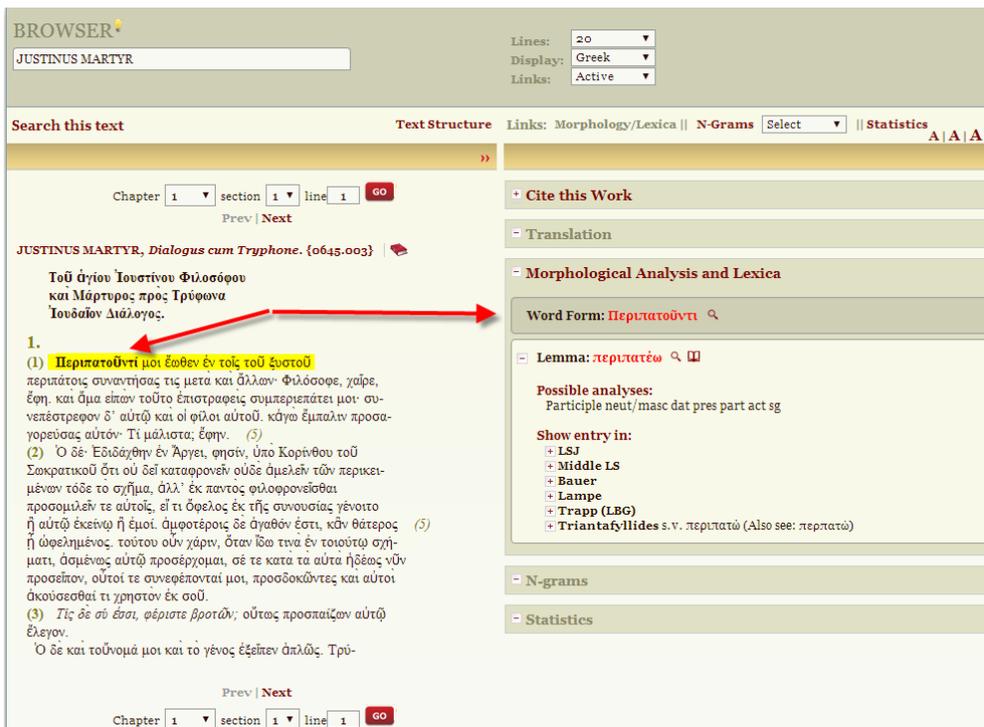


Selecting the first entry—“Dialogus cum Tryphone (0645.003)”—displays the text of the *Dialog* in the left pane, with navigation tools (chapter, section, line, previous, next) positioned immediately above and below that pane, as well as display tools (lines, language, links) positioned above and to the right of the pane.

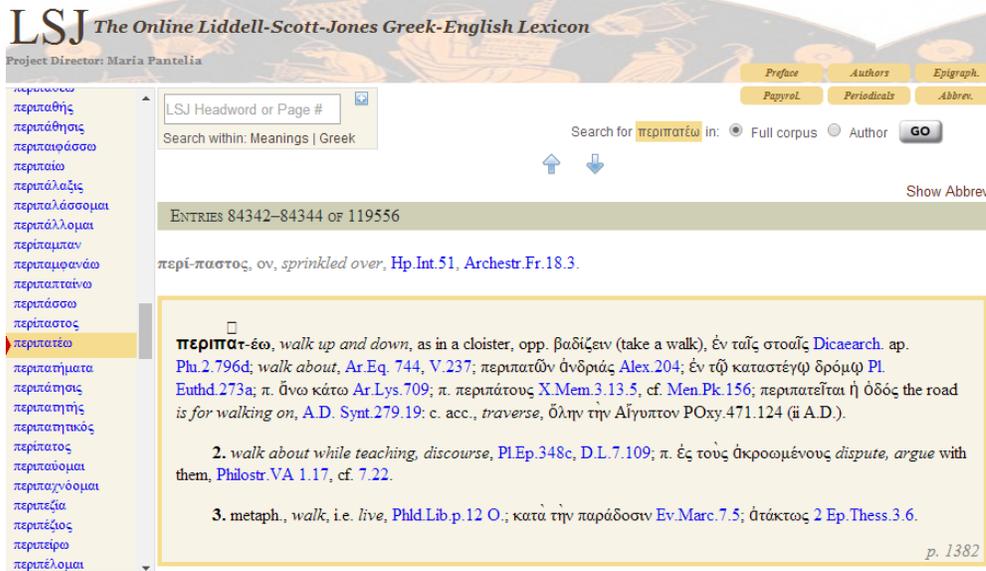


Additional browser features include:

- “Search this text.” This takes you to the Text Search screen, which I covered in last issue of this newsletter (Q2 2019).
- “Text Structure.” This provides an outline of the text for easy navigation. If the citation is “line,” no text structure will be provided. The *Dialog* is structured by chapter.
- “Links.” This turns on the translation, morphological analysis and lexica, n-grams, and statistics links. I will discuss lexica, n-grams, and statistics in separate sections of this review, but note that they are available from the browser screen.
- Selecting a word in the text pane on the left displays the word’s lemma, morphological analysis, and available lexica in the right-hand pane.



Selecting a lexicon opens it to the proper location in a new tab/window.



- “Book icon.” This displays a list of the author’s works. See the second graphic above.
- “Translation.” When the TLG knows of a link to an online translation, this link will be activated. The translation will open in a new window.
- “Cite this Work.” This displays the source of the selected work.



To browse parallel texts, select “Browse” and then “Parallel Browsing” from the TLG’s main menu.



In the search fields, enter the author’s name or the text’s TLG number. The autocomplete feature is enforced. For example, to browse Ephesians and Colossians, enter “NOVUM TESTAMENTUM” in both search fields, and select “Epistula ad Ephesios” from the drop-down list below the left-hand search field and “Epistula ad Colossenses” from the drop-down list below the right-hand search field.

There are three options under Parallel Browsing: (1) browse any two texts, (2) browse two texts with the option to highlight the similarities between them (an expanded version of the N-grams feature; see discussion below), and (3) compare two editions (in cases when the TLG® has more than one edition of the same text, this option allows you to compare the two editions

and quickly identify the differences). In the following example, I have selected “Highlight similarities” from the “Select” menu. Clicking the “Show Texts” button displays this:

The screenshot shows the 'PARALLEL BROWSING' interface. At the top, there are search filters for 'NOVUM TESTAMENTUM {0031}' and 'Epistula Pauli ad Ephesios {010}' on the left, and 'NOVUM TESTAMENTUM {0031}' and 'Epistula Pauli ad Colossenses {012}' on the right. A 'Select' menu is set to 'Highlight similarities', and a 'Show Texts' button is visible. Below the filters, the 'Text Structure' section shows navigation options for Chapter, section, and line. The main content area displays two columns of text. The left column is the source text (Epistula Pauli ad Ephesios) and the right column is the target text (Epistula Pauli ad Colossenses). Similarities between the two texts are highlighted in yellow. For example, in the left column, '(3) Εὐλογητός ὁ θεὸς καὶ πατὴρ τοῦ κυρίου ἡμῶν Ἰησοῦ Χριστοῦ, ὁ εὐλογησθεὶς ἡμῶς ἐν πάσῃ εὐλογίᾳ πνευματικῇ ἐν τοῖς ἑπουρανίοις ἐν Χριστῷ.' is highlighted. In the right column, '(3) Εὐχαριστοῦμεν τῷ θεῷ πατρὶ τοῦ κυρίου ἡμῶν Ἰησοῦ Χριστοῦ πάντοτε περὶ ὑμῶν προσευχόμενοι,' is highlighted. The interface also includes navigation buttons like 'Prev', 'Next', and 'Jump'.

The “Jump” button jumps you to the next similarity between the two texts. The left hand text is considered the source text. Use normal navigation to move to different sections of the text.

The TLG browser help page is located [here](#), and a YouTube video is available [here](#). The parallel browsing help page is located [here](#), and a YouTube video is available [here](#).

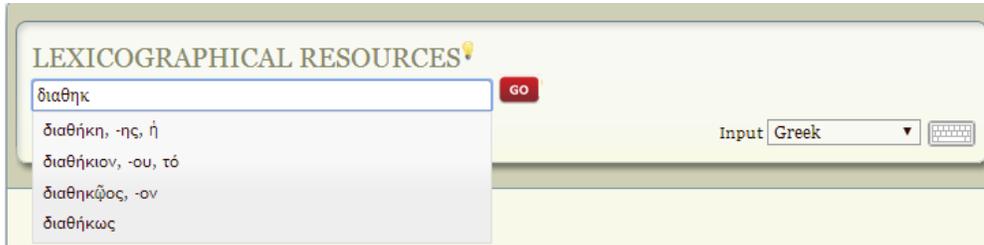
Consult Online Lexica²

Selecting “Lexica” from the TLG’s main menu allows users to access online lexica by first entering a word. Input can be proper Greek, Beta code, or transliteration. (The TLG help page for lexica is located [here](#).)

² Because of the highly inflected nature of Greek, and the linguistic and orthographic heterogeneity of the corpus, searching the TLG for individual word forms can be tedious. In early 2003, the TLG undertook the project to lemmatize its corpus. The first effort to employ computers in the morphological analysis of Greek was undertaken by David W. Packard in the early 1970s. The TLG lemmatizer is an extensive system that at present can recognize automatically 98.3% of all wordforms in the TLG corpus. To reach this goal, the TLG digitized and extracted a large number of head words from dictionaries (such as LSJ, Lampe, Bauer, Trapp, Kriaras). The lemmatized search engine was made available in beta form in 2006 and is now a regular feature of the TLG search engine. As part of the lemmatization project, the TLG became active in developing online lexica. The first major effort was the digitization of the Liddell-Scott-Jones (LSJ) dictionary, the premier lexicon for ancient Greek. Cunliffe’s *Lexicon of Homeric Greek* and Powell’s *Lexicon to Herodotus* followed. In 2013 the TLG and the Austrian Academy of Sciences joined forces to create a digital version of the *Lexikon zur byzantinischen Gräzität* (LBG). Source: <http://stephanus.tlg.uci.edu/history.php>.



A search for διαθήκη, for example, would display a list of words beginning with the letters you have typed. The autocomplete feature is enforced.



Selecting the first entry shown above, displays διαθήκη, ἥ in LSJ.



Sadly, the full text of BAGD and of Lampe’s *Patristic Lexicon* are not available, though basic definitions are.

LEMMA

Lemma: διαθήκη, -ης, ἡ

Show entry in:

- + LSJ
- + Middle LS
- + DGE
- Bauer
- Lampe
- + Kriaras
- + Triantafyllides

1 last will and testament so exclusively in Hellenistic times 3 The mng. compact, contract seems firmly established for Gr-Rom. times

1. testament, will, 2. compact, covenant, 3. written covenant, scriptures of Old or New Testament, 4. record

Selecting “Geographical Distribution” from the left-hand sidebar displays a map of countries whose texts in the TLG database include the selected word.

LEMMA

Lemma: διαθήκη, -ης, ἡ

Show entry in:

- + LSJ
- + Middle LS
- + DGE
- + Bauer
- + Lampe
- + Kriaras
- + Triantafyllides

See also: διάθεσις, -εως, ἡ

Show entry in:

- + LSJ
- + Middle LS
- + DGE
- + WIP
- + Lampe
- + Kriaras
- + Triantafyllides s.v. διάθεσις

GEOGRAPHIC DISTRIBUTION

Location	Count
Abdera	2
Achrida	169
Aegyptus	18
Aenus	3
Agrigentum [vel Acragas]	3
Alexandria	1,170
Alexandria (Troadis)	2
Amasia	7
Amida	2
Ancyra	30
Antiochia	658
Antiochia (Pisidiae)	1
Antiochia vel Damascus	1

STATISTICS

RESULTS

Map Satellite

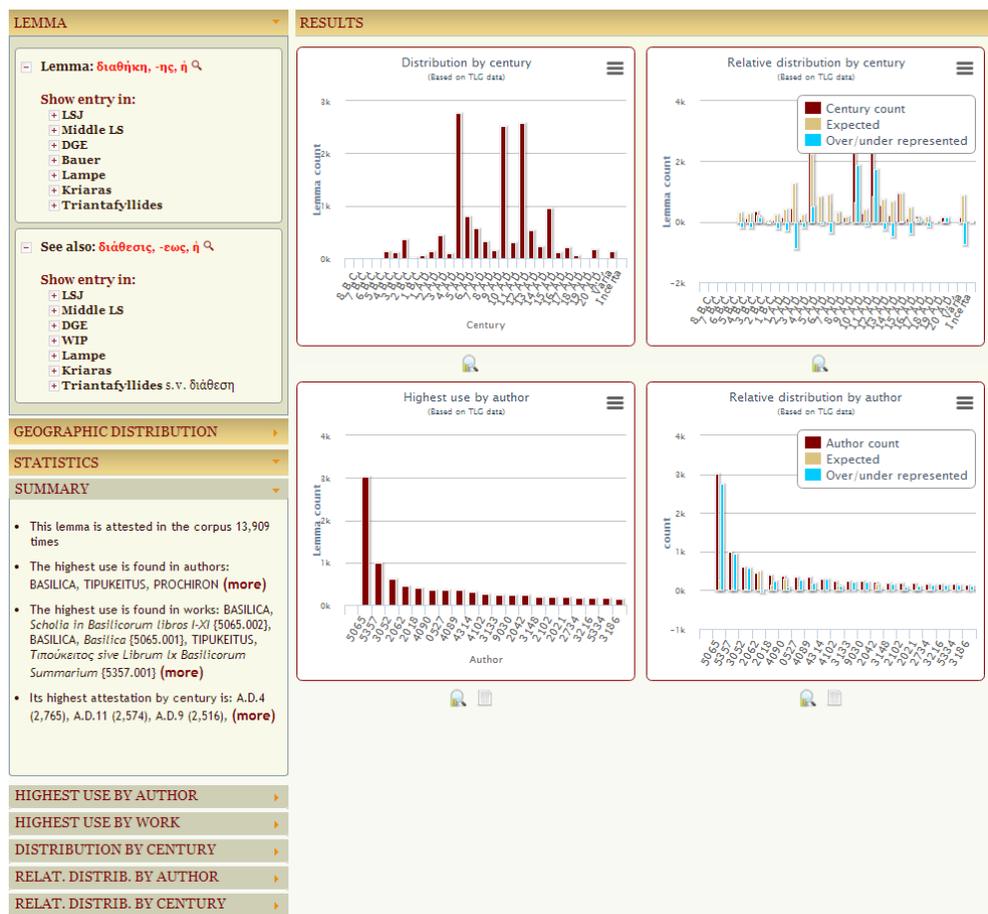
Google

Map data ©2019 Imagery ©2019 NASA, TerraMetrics Terms of Use

Selecting “Statistics” from the left-hand sidebar displays statistical information about each lemma included in the corpus and lemma usage by geographic location. Statistical information can be expanded by clicking the appropriate section in the left-hand sidebar or by clicking on the icon underneath each diagram. Statistical information is displayed both as prose and in bar-graph form. For more information, see the statistics help [file](#).

The statistical information includes the following (examples come from the lemma διαθήκη, ης, ἡ; selecting “more” displays additional information):

- Number of occurrences of the lemma in the TLG corpus. For example, this lemma is attested in the corpus 13,909 times.
- Highest use by author. For example, the highest use is found in authors: BASILICA, TIPUKEITUS, PROCHIRON (more).
- Highest use by work. For example, the highest use is found in works: BASILICA, Scholia in Basilicorum libros I-XI {5065.002}, BASILICA, Basilica {5065.001}, TIPUKEITUS, Τιπούκειτος sive Librum Ix Basilicorum Summarium {5357.001} (more).
- Distribution by century. For example, its highest attestation by century is: A.D.4 (2,765), A.D.11 (2,574), A.D.9 (2,516), (more).
- Relative distribution by author. This shows how many times the lemma occurs in each TLG author, how many times the lemma would be expected to occur in each author’s works, and how over or under represented the lemma is in each author’s works.
- Relative distribution by century. Shows information similar to that for “Relative distribution by author” but expressed by century, rather than by author.



Create N-Gram Comparisons

Selecting “N-Grams” from the TLG’s main menu allows users to create N-grams—intertextual phrase matching—by first entering an author name. I’ll use “NOVUM TESTAMENTUM {0031},” select the default of 1 N-gram, and compare Ephesians and Colossians, using the “Wordform” (= inflected) option.



Doing so displays 23 instances where identical phrases occur in the UBS2 (2nd ed., 1968) GNT of Ephesians and Colossians. Switching from “Wordform” to “Lemma” displays 35 instances where the same lemmas are used in Ephesians and Colossians. Increasing the number of N-grams (to 2–6) decreases the number of matches, whether using “Wordform” or “Lemma.” As you can see from the last screen shot, n-grams are based on trigrams (combinations of three words). If there are none, the program uses bigrams (combination of two words).

N-grams are overlapping sequences of content words in text.³ They provide an efficient mechanism for identifying common passages between texts: by identifying sequences of two or three content words shared between two texts, we can quickly identify text passages in common.

N-gram comparison uses trigrams common between two selected works or authors, to identify common passages. The words in common between the two texts are highlighted. The comparison allows for small interpolations in the common passages; the words actually in common are highlighted, so un-highlighted words between two highlighted sections indicate interpolations (or deletions, if the apparent interpolation is in the original text). The comparison is based on lemmata rather than wordforms, so it allows for differences of inflection between the two texts, as well as slight differences in word order (through unordered n-grams). There is a limit to the size of common passages identified between words, so that an extended passage common between two texts may be shown chopped up into several consecutive matches. If more lines of context are selected for the match, the maximum size of the match is increased, as is the maximum size of allowed interpolations.

The TLG website uses n-grams in three places/sections. (1) N-Grams: Compare two texts or two authors side by side. (2) Browse (Browse one text): Enable N-grams and click on its phrase to see its attestation in the corpus. (3) Browse (Parallel Browsing): Select two passages and view their similarities.

Browsing a text in n-gram mode (Browse, Links: N-Grams, Select: lemma or wordforms) highlights all the n-grams in a passage that also occur elsewhere in the TLG corpus. The n-grams can be based on wordforms, or on lemmata: if you select wordforms, only exact repetitions of the wordforms are reported, whereas lemmata allow for changes in inflection. The n-grams highlighted include both trigrams and bigrams; the trigrams can span across sentences, but the bigrams cannot. (This helps contain the number of matches reported.)

³ The text from this paragraph to the end of this section is taken from this TLG help [page](#).

- Most Frequent Lemmata. “The most frequent lemmata in the corpus are: λέγω, πᾶς, πᾶσα, πᾶν, ἔχω (more).”
- Most Frequent Wordforms. “The most frequent wordforms in the corpus are: Θεοῦ, πάντα, μόνον (more).”
- Lemmata by Century. “Lemmata are distributed by century as follows: A.D.2 (102,954), A.D.12 (94,908), A.D.6 (92,816) (more).”
- Wordforms by Century. “Wordforms are distributed by century as follows: A.D.2 (513,897), A.D.12 (503,158), A.D.4 (499,175) (more).”

Statistics can be sorted by full corpus, author, and corpus, as well as by century. This is a powerful, feature-rich tool. For additional information see the statistics help [file](#).

Use the Vocabulary Tools

Selecting “Vocab. Tools” from the TLG’s main menu displays the Vocabulary Tools screen. Entering “Justin Martyr” in the search field by using the autocomplete function and then selecting “Dialogus cum Tryphone (003)” from the drop-down menu displays this screen, which provides three sets of information: (a) lemmata in this author/work, (b) wordforms in this author/work, and (c) flashcard.

Lemma	Count
1. λέγω	876
2. θεός, -οῦ, ὁ	655
3. πᾶς, πᾶσα, πᾶν	442
4. γίνομαι	369
5. κύριος, -α, -ον	335
6. Χριστός, -οῦ, ὁ	283
7. ἡμέτερος, -α, -ον	273
8. φημί	249
9. λόγος, -ου, ὁ	247
10. ἀνθρώπος, -ου, ὁ	226
11. ποίω	206
12. κύρις, -εως, ἡ	192

“Lemmata List” (identical with “Lemmata in this author/work”) is the default information set. By default, lemmas are arranged by count number, highest to lowest. Selecting “Alphabetical” from the “Sort by” menu displays the lemmas alphabetically.

Selecting “Wordform List” (identical with “wordforms in this author/work”) displays each inflected form in the selected text, arranged by count number, highest to lowest. Selecting “Alphabetical” from the “Sort by” menu displays the inflected forms alphabetically.

Selecting “Flashcard” displays a Greek lemma in the text. Lemmas are displayed in a random (not alphabetical) order



Clicking the lemma displays the discussion of the word in the default lexicon (usually the Middle Liddell-Scott) and provides links to other lexica. After viewing the discussion, clicking the back arrow and then clicking the forward arrow displays the same lemma, but without the discussion. This is a helpful and powerful learning tool.



For more help on using the vocabulary features, see [here](#).

Conclusion

The Thesaurus Linguae Graecae® Digital Library is an indispensable tool for students of Greek New Testament lexicography. As I stated at the beginning of this review: “No other resource in the world can touch the TLG when it comes to finding *all* occurrences of a Greek word or phrase (inflected or lemmatized) in *any* Greek literature in the period between Homer (8 c. B.C.) and the fall of Byzantium in A.D. 1453 or any authorial, chronological, or corporeal subset thereof. In short, if you wish to do serious, world-class diachronic or synchronic word studies, you need access to the full TLG corpus.”

From the perspective of Greek New Testament scholarship, the TLG has two main (but by no means fatal!) limitations: (1) Only the UBS2 (2nd ed., 1968) text of the Greek NT is available—the latest version is UBS5 (5th ed., 2014)—and (2) the full text of BAGD is not available, though basic definitions are. I highly recommend the TLG and commend all who have labored to make this terrific tool available to the scholarly community.



As always, we remain open to developing AGNT and ANLEX in ways that are most useful to the needs of students and readers of God's Word.

Thank you for your continued support of *The AGNT Project*, for faithfully marketing the AGNT and ANLEX databases, and for making these state-of-the-art tools for studying the Greek New Testament available to students, scholars, pastors, translators, and laymen worldwide.

John Hughes
Agent for *The AGNT Project*
johnhughes@centurytel.net
Phone: 406.862.7289
FAX: 406.862.0917

