

# Discontinuous Coordination in Ancient Greek

Woodley Packard

## 1 Introduction

Ancient Greek features a variety of discontinuous constituent patterns, most commonly adjuncts separated from heads and verbs separated from their arguments, in each case by something that would not typically be thought of as part of the constituent in question. In the present work, I explore what was to me, upon investigation, a more surprising pattern of discontinuous constituency: conjuncts separated from each other by grammatical material that is not subordinate to the coordinated phrase.

I quasi-automatically select a sample of more than a hundred exemplars of this phenomenon from an annotated corpus, and based on observed usage patterns, propose a partial analysis involving extraction. I then implement that analysis in a computational precision grammar and evaluate its effectiveness. The results, while limited in scope, are quite encouraging.

In Section 2, I give a more extended introduction to the phenomenon in question, as well as other supporting topics and resources including the annotated corpus from which the examples are drawn, the precision grammar framework in which my analysis is formulated, and the computational grammar of Ancient Greek in which the analysis is implemented and tested. In Section 3, I describe the procedure used to arrive at the suite of exemplars used for design and evaluation of the analysis. In Section 4, I outline the analysis itself, and in Sections 5 and 6 I discuss the results and offer closing thoughts.

## 2 Background

### 2.1 Ancient Greek Coordination

Coordination in Ancient Greek most commonly takes the form of medial monosyn-ton, and is possible for a broad range of head types. Usually, the coordinands are

both immediately adjacent to the conjunction, as in the following examples:<sup>1</sup>

- (1) πραῦς καὶ ταπεινός  
praus kai tapeinos  
gentle and humble
  
- (2) πορευθέντες καὶ ποιήσαντες ...  
poreuthentes kai poiesantes ...  
go.PERF.PTCP and do.PERF.PTCP ...  
'having gone and having done ...'
  
- (3) χαρά καὶ ἀγαλλίασις  
cara kai hagalliasis  
grace and joy

Much less frequently, however, the left conjunct can be separated from the conjunction by seemingly arbitrary linguistic material, as in these New Testament examples:

- (4) πραῦς εἰμι καὶ ταπεινός  
praus eimi kai tapeinos  
gentle be.1SG and humble  
I am gentle and humble. (Matt. 11:29)
  
- (5) πορευθέντες δὲ οἱ μαθηταὶ καὶ  
poreuthentes de hoi mathetai kai  
go.PERF.PTCP (discourse-particle) the.NOM.PL disciple.NOM.PL and  
ποιήσαντες ...  
poiesantes ...  
do.PERF.PTCP ...  
The disciples, having gone and having done (something), (did something else ...). (Matt. 21:6)

The left conjunct is not always clause-initial:

- (6) ἔσται χαρά σοι καὶ ἀγαλλίασις  
estai cara soi kai hagalliasis  
be.FUT.3SG grace you.DAT and joy  
He will be grace and joy to you. (Luke 1:14)

The phenomenon has a correlate in spoken English, where another conjunct can be expressed as an afterthought, as in Example 7. However, in Greek the use of discontinuous coordination does not appear to be restricted to informal afterthoughts.

- (7) I like eating salmon when it's grilled properly, and trout.

---

<sup>1</sup>These examples were constructed by reordering the actually attested examples below into the more canonical word order.

## 2.2 The PROIEL Treebank

The *Pragmatic Resources of Old Indo-European Languages* (PROIEL) treebank (Haug & Jøhndal, 2008) is a collection of linguistic analyses on ancient texts, covering morphology, syntactic dependencies, and information status. The treebank includes annotations for Herodotus’s *Histories* and the *Greek New Testament*, comprising more than 200,000 words of Ancient Greek text. The morphological annotations in PROIEL give (beside part of speech) gender, case and number for nouns and adjectives, and tense, mood, voice, gender and number for verbs. Participles are annotated with both types of annotation. The syntactic information contained in the treebank consists of a dependency tree, where each word is ascribed a single dependency type and a single head, which is either another word or *ROOT*. Information status is annotated on nouns with tags like *new*, *old* and *non-specific*.

## 2.3 HPSG and the Grammar Matrix

Head-driven Phrase Structure Grammar (Pollard & Sag, 1994) is a constraint-based theory of syntax, morphosyntax and the syntax-semantics interface using feature structures and unification. It is sufficiently descriptive to allow the concise encoding of generalizations, but sufficiently formal to admit computational algorithms for parsing and generation. These properties make HPSG an attractive framework for designing a computational account of the Ancient Greek phenomenon in question. The LinGO Grammar Matrix and customization system (Bender, Flickinger, & Oepen, 2002; Bender, Drellishak, Fokkens, Poulson, & Saleem, 2010) comprise a foundational type hierarchy<sup>2</sup> describing broadly applicable cross-linguistic generalizations together with some finer tuned constraints that can be selected on a per-language basis.

I have been developing a Matrix-based HPSG grammar of Ancient Greek for several years. I chose to use this grammar as a starting point for exploring a possible formal account of discontinuous coordination.

## 3 Data Collection

In order to be able to study discontinuous coordination effectively, a catalogue of examples of the phenomenon was required. To furnish this, I turned to the PROIEL Treebank (Haug & Jøhndal, 2008).

---

<sup>2</sup>In HPSG, generalizations about linguistic structures—specifically the feature structures that are used to describe them—take the form of a type hierarchy.

### 3.1 Automatic search and manual verification

The syntactic dependency annotations in PROIEL made it possible to automatically search for candidate cases of interest. Specifically, for each instance of the conjunction *καί* in the corpus, the sentence containing it was analyzed in the following way. The conjunction and its syntactic dependents were traversed recursively, such that anything that depended even indirectly on the conjunction was given a mark indicating that it was part of the conjoined constituent. Subsequently, the sentence was scanned for any unmarked words in positions with marked words both to the left and to the right. If such words were found, they were taken to be interrupters of the coordination, and as such the sentence was recorded as an apparent example of discontinuous coordination. The hypothesized extent of the coordinated phrase was recorded, along with citation information linking back to the source location within the original treebank.

This procedure resulted in 321 examples from the New Testament portion of the treebank (2.6% of *καί* tokens) and 486 examples from the Herodotus portion (15.2% of *καί* tokens). Manual inspection of all of the New Testament instances showed 142 of them were interesting enough to keep (some were discontinuous in less interesting ways, such as interruptions by light-weight discourse particles or by explicitly parenthetical thoughts). I only completed part of the manual inspection of the Herodotus corpus, since the inventory of exemplars had already reached a size I felt was sufficient to support further inquiry, and the project was lasting longer than anticipated. The acceptance rate on Herodotus was approximately the same (9 accepted of the first 22 instances). This means that roughly 1.6% of all uses of the conjunction *καί* in the New Testament and an estimated 6.2% of all uses of *καί* in Herodotus involve discontinuities. These numbers are summarized in the table below (numbers marked with an asterisk are projected on the basis of the acceptance rate on the manually inspected portion of the Herodotus corpus).

Corpus	<i>καί</i>	Verified discontinuous	Total discontinuous	Fraction
New Testament	8797	142	142	1.6%
Herodotus	3192	9	199*	6.2%*

The following table shows the annotation for Example 4, and how the search algorithm identifies it.

Gloss	gentle	be.1SG	and	humble
Head	and	ROOT	be.1SG	and
Conjunction?			X	
Coordinated?	X		X	X
Discontinuity?		YES		

The automatic search method can overlook certain types of discontinuous coordination. Specifically, it generally misses discontinuously coordinated finite verbs, because the interrupting material is usually annotated as dependent on the conjunction. A study of 50 randomly selected instances of conjunction from PROIEL specifically to probe this problem did not reveal other categories of missed examples beyond the aforementioned finite verbs, but they may exist. An example of finite verbs that are arguably discontinuously coordinated is given below:

- (8) Ἀπεκρίθησαν οἱ Ἰουδαῖοι καὶ εἶπαν  
Apekrithesan hoi Ioudaioi kai eipan  
answer.PAST.3PL the.NOM.PL Jews.NOM.PL and say.PAST.3PL  
αὐτῷ ...  
auto ...  
he.DAT.SG  
The Jews answered and said to him ... (John 8:48)

However, since subjects can be left unexpressed in Ancient Greek, a plausible analysis of this example (and indeed the one taken by the PROIEL annotators) is that *οἱ Ἰουδαῖοι* is directly the subject of *Ἀπεκρίθησαν*, and not of a coordinated verb phrase, and the subject of *εἶπαν* is unexpressed. Because this explanation is broadly applicable (no examples were noted in which the material interrupting a coordinated verb phrase was something other than the subject), I ignore finite verbal coordination in the remainder of the study.

### 3.2 Establishing a baseline

My existing computational grammar of Ancient Greek had low coverage on open running text. Using the sentence segmentation from the PROIEL treebank, none of the 151 examples identified in the previous phase received an analysis. While this was encouraging in the sense that it indicated a lack of overgeneration, it was in part due to lacking lexical coverage. To correct this, I augmented the grammar with a full-form lexicon automatically derived from the morphological data contained in the PROIEL annotations, so that at least every noun, verb, adjective and adverb in the test suite was within the grammar's lexical scope. Additionally, I trimmed irrelevant context

away from the left and right periphery of the sentences in the testsuite in order, to the extent possible, to measure coverage changes only related to the phenomenon under study (i.e. discontinuous coordination). This typically involved removing additional independent clauses before or after the clause in question, and removing adjuncts from the front of the clause. These changes led to a baseline result of parsing on 2 out of the 151 examples. The following statistics describe the final test suite:

Corpus	Sentences	Words	Vocabulary Size
New Testament	142	1406	656
Herodotus	9	124	91

The frequency of discontinuous coordination of different phrase types is described below:<sup>3</sup>

Corpus	Noun	Preposition	Adjective	Non-finite Verb
New Testament	89	8	14	30
Herodotus	6	1	0	2

## 4 Analysis

In considering the patterns that appear in connection with this phenomenon, the first observation to be made is that although the coordination is discontinuous, the relative order of the constituents of the coordination are the same as in the continuous case. That is, of the following conceivable discontinuous schemata, most do not appear in the data:

- (9) X interruption and X
- (10) X and interruption X
- (11) X interruption X and
- (12) X X interruption and
- (13) and X interruption X
- (14) and interruption X X

---

<sup>3</sup>One instance from the New Testament, in Luke 8:47, involves discontinuous coordination of two indirect questions, which didn't seem to fit into this table.

In fact, of these six, the only schema that is attested is schema 9. I take this as evidence that “and X” is a constituent in the discontinuous variants of coordination, just as it is in the ordinary, continuous variant. This would be compatible with schema 13 as well; the fact that this is unattested leads to the following principle: *the right-hand conjunct always moves further to the right, relative to the continuous variant.*

I brought a subset of the test suite data before the participants at the 2017 DELPH-IN<sup>4</sup> summit, and am indebted to the keen observations made by several people present at that meeting. In particular, Emily Bender, Dag Haug, and Guy Emerson together observed that the moving “and X” constituent generally moves all the way to the right of the clause in which it would canonically appear, but not further, and suggested that an analysis based on extraposition from the left conjunct, analogous to other more conventional gapped constituents, might work.

This analysis involves two new syntactic rules: one to allow the introduction of a gap indicating an extraposed conjunct, and one to allow the filling of that gap higher up in the syntax tree by just such a conjunct, provided it appears on the right.

$$\begin{array}{c}
 \left[ \begin{array}{l}
 \text{extracted-conjunct-xp} \\
 \text{LOCAL} \quad \left[ \begin{array}{l} \text{COORD} \quad - \\ \text{CAT} \quad \boxed{5} \end{array} \right] \\
 \text{SLASH} \quad \left\langle \left[ \begin{array}{l} \text{COORD} \quad + \\ \text{CAT.HEAD} \quad \boxed{6} \end{array} \right] \right\rangle \\
 \text{ARGS} \quad \left\langle \left[ \begin{array}{l} \text{LOCAL} \quad \left[ \begin{array}{l} \text{COORD} \quad - \\ \text{CAT} \quad \boxed{5} \end{array} \right] \left[ \begin{array}{l} \text{HEAD} \quad \boxed{6} +nvjp \\ \text{VAL} \quad \text{saturated} \end{array} \right] \right] \right\rangle
 \end{array} \right] \\
 \\
 \left[ \begin{array}{l}
 \text{head-filler} \\
 \text{LOCAL} \quad \boxed{1} \\
 \text{SLASH} \quad \langle \rangle \\
 \text{ARGS} \quad \left\langle \left[ \begin{array}{l} \text{LOCAL} \quad \boxed{1} \\ \text{SLASH} \quad \langle \boxed{2} \rangle \end{array} \right], \boxed{2} \right\rangle
 \end{array} \right]
 \end{array}$$

The *extracted-conjunct-xp* rule is the more complex and interesting of the pair. It is a unary rule that launches a gap from an XP, i.e. any valence-saturated phrase, with two conditions: first, it cannot already be part of a coordination structure, and second,

<sup>4</sup>Deep Linguistic Processing with HPSG Initiative: <http://www.delph-in.net/>.

its head type must be one of a specifically permitted set: noun, verb, adjective or preposition.<sup>5</sup> When this condition is met, a new constituent is licensed which matches the daughter in every way, with the exception that it additionally possesses a nonempty SLASH list, indicating that another constituent is expected higher up the tree. That constituent is relatively underspecified, but must match the daughter in head type, and must be marked for coordination (this ensures that the conjunction is present).

The *head-filler* rule simply discharges the gap from the SLASH list by stipulating a matching constituent immediately following its daughter, the discontinuity already having been absorbed by other pre-existing rules (e.g. *head-complement*), which preserve the SLASH value.

The semantic analysis of (continuous or discontinuous) coordination in Matrix-derived grammars asserts that two indices are coordinated by virtue of a coordination predicate such as *and\_rel*, which takes the coordinated indices as arguments, as well as a third index (the C-ARG) which signifies the coordinated whole. The latter index is what further linguistic content uses to refer to the coordinated content.

To preserve legibility, the above schemata do not indicate how the rules manage the semantic portion of the analysis, i.e. how different argument roles of the conjuncts, conjunctions, and interrupting material are linked and filled. The *head-filler* rule requires no further changes. In the actual implementation, the *extracted-conjunct-xp* rule is overlaid with the following additional constraints:

$$\left[ \begin{array}{l} \text{extracted-conjunct-xp} \\ \text{LOCAL.CONT.HOOK} \left[ \begin{array}{ll} \text{LTOP} & \#lbl \\ \text{INDEX} & \#carg \\ \text{XARG} & \#xarg \end{array} \right] \\ \text{SLASH} \left\langle \left[ \begin{array}{ll} \text{LOCAL.COORD-REL} & \left[ \begin{array}{ll} \text{LBL} & \#lbl \\ \text{C-ARG} & \#carg \\ \text{L-INDEX} & \#lind \\ \text{R-INDEX} & \#rind \end{array} \right] \\ \text{CONT.HOOK} & \left[ \begin{array}{ll} \text{INDEX} & \#rind \\ \text{XARG} & \#xarg \end{array} \right] \end{array} \right] \right\rangle \\ \text{ARGS} \left\langle \left[ \begin{array}{ll} \text{LOCAL.CONT.HOOK} & \left[ \begin{array}{ll} \text{INDEX} & \#lind \\ \text{XARG} & \#xarg \end{array} \right] \end{array} \right] \right\rangle \end{array} \right]$$

<sup>5</sup>It seems likely that adverbs are also admissible heads, but this is unattested in the test suite.



In words, the mother constituent behaves semantically as though it were itself the complete conjoined phrase. Its INDEX is the index of the conjoined phrase, i.e. the C-ARG of the coordination predicate (COORD-REL), and its label (LTOP) is the label of that predicate as well. The daughter constituent fills the role of the left conjunct, so its own INDEX is identified with the coordination predicate’s L-INDEX. The displaced constituent fills the role of the right conjunct, so its INDEX is identified with the coordination predicate’s R-INDEX.

The treatment of XARG is of particular interest. The XARG feature is used in semantic composition to denote an external (i.e. not yet realized) argument. The two main uses are to link a verb to the index of its subject and to link an adjunct to the index that it modifies. Both of these cases are fair game for discontinuous coordination. In the case of verbs, the coordinated verbs are implied to have the same subject, and in the case of adjuncts, the coordinated adjuncts are implied to modify the same head. By the time the dislocated conjunct is realized, the relevant index may well be sealed away inside of the preceding constituent and therefore inaccessible. This conundrum is avoided by linking the XARG of the dislocated conjunct at the time the gap is stipulated.

## 5 Results

The table below gives parse coverage on the discontinuous coordination test suite for the baseline Greek grammar, by head type, as a proportion of all examples with that head type.

Corpus	Noun	Preposition	Adjective	Non-finite Verb	Total
New Testament	1 / 89	1 / 8	0 / 14	0 / 30	2 / 142
Herodotus	0 / 6	0 / 1	0 / 0	0 / 2	0 / 9

Note that coverage of the baseline grammar on the complete New Testament corpus is 2268 / 11504 sentences, using the sentence segmentation given in the PROIEL treebank. As a percentage, that is 19.7%. For Herodotus, it is only 18 / 5491 sentences, or 0.3%. However, the vast majority of the Herodotus inputs lack lexical coverage, since a full form lexicon was extracted from the PROIEL New Testament treebank but not from the Herodotus portion. On the portions of each corpus that enjoy complete lexical coverage, the numbers are much more similar: 2268 / 8940 = 25.4% for the New Testament, and 18 / 103 = 17.5% for Herodotus. The remaining disparity is perhaps due to Herodotus’s employment of the Ionic dialect, whereas the grammar is tuned more toward the Koine dialect used in the New Testament.

Two sentences received analyses before implementation of the new rules. These were both legitimately structurally ambiguous—there was an (incorrect) analysis available using continuous coordination. The table below gives parse coverage for the augmented grammar, again as a proportion of all examples with that head type.

Corpus	Noun	Preposition	Adjective	Non-finite Verb	Total
New Testament	19 / 89	2 / 8	4 / 14	2 / 30	27 / 142
Herodotus	0 / 6	0 / 1	0 / 0	0 / 2	0 / 9

The new analysis achieves coverage for a nontrivial portion of the examples of discontinuous coordination, including some from each head type. The most pertinent question is how coverage of the revised grammar on sentences containing discontinuous coordination compares with coverage of the original grammar on sentences that do not. Since the phenomenon in question is syntactic rather than morphological or lexical, some noise can be excluded by measuring coverage just on the subset for which complete lexical coverage is available. For the New Testament, that corresponds to 94% of the testsuite inputs and 78% of all of the New Testament;<sup>6</sup> for Herodotus, it is 55% of the testsuite inputs and 1.8% of all of Herodotus. The following table shows parse coverage as a proportion of inputs with full lexical coverage.

Corpus	Noun	Preposition	Adjective	Non-finite Verb	Total
New Testament	19 / 77	2 / 7	4 / 12	2 / 28	27 / 134
Herodotus	0 / 4	0 / 1	0 / 0	0 / 0	0 / 5

Syntactic coverage of the baseline grammar on the lexically covered subset of the full New Testament is 25.4%; this is comparable to the  $27 / 134 = 20.1\%$  observed with the augmented grammar on the discontinuous coordination testsuite, although somewhat lower. For Herodotus, the baseline grammar parses 17.5% of lexically covered inputs. The observed value of zero out of five for the augmented grammar, while disappointing, does not differ from 17.5% by enough to be troubling (even  $1 / 5 = 20\%$  would surpass the baseline).

These data alone say quite a lot about the effectiveness of the proposed analysis: without it, a stark difference is observed in coverage between the discontinuous

<sup>6</sup>Lexical coverage is not 100% because the automatically extracted full-form lexicon can only cover open class words; there are still some rarer function words that need to be analyzed by hand.

coordination testsuite and the rest of the PROIEL data. When the proposed analysis is employed, however, the coverage difference essentially disappears. This indicates that the gapping strategy is able to accommodate nearly as many sentences from the testsuite as could be hoped for: it is sufficiently general to explain the great majority of the observations. There are at least two important things this does not tell us, however.<sup>7</sup> The first is whether the new analysis might be *too* general, i.e. whether it might overgenerate and apply in situations that it was not intended for. The second is whether *correct* analyses are produced for the testsuite examples, i.e. whether the trees found use the new rules in the expected way. I take the liberty of investigating both of these questions on the New Testament data alone, since (1) coverage is much higher on open text, so small amounts (percentage-wise) of overgeneration should be easier to detect, and (2) the lack of analyses found for discontinuous coordination in Herodotus renders verifying correctness moot.

## 5.1 Correctness

I used the *full forest treebanker* (Packard, 2015) to inspect the analyses produced for the 27 exemplars that were parsable. Of these, 24 received analyses that I judged either fully correct or correct apart from issues unrelated to discontinuous coordination. The 3 inputs which were parsable but did not receive any correct analyses are displayed below.

- (15) καὶ            κύριον            αὐτὸν καὶ            Χριστὸν  
kurion            auton            kai    Criston            epoiesen  
lord.ACC.SG him.ACC.SG and    Christ.ACC.SG make.PAST.3SG  
ἐποίησεν    ὁ            θεός  
ho            theos            ...  
the.NOM.SG God.NOM.SG  
God has made him both Lord and Christ. (Acts 2:36)

Here, ἐποίησεν is ditransitive, with one of the complements (αὐτὸν) interrupting the other (καὶ κύριον καὶ Χριστὸν). In this case, it appears the dislocated conjunct does not move all the way to the right of its clause, at least according to traditional HPSG clause boundaries. It is worth noting that some theories of syntax consider the two complements of these types of ditransitive to form a clause-like constituent together. However, in the present experiment, this sentence did not receive a suitable analysis because my grammar does not recognize a gapful constituent immediately to the left

<sup>7</sup>A third consideration which is mostly ancillary to the present discussion is efficiency. The augmented grammar takes about 4 times longer to parse the New Testament than the baseline grammar, which is a considerable cost for a somewhat rare phenomenon.

of the dislocated καὶ Χριστόν. It would be worthwhile to try to find more instances of discontinuous coordination involving ditransitives to see if this pattern is common.

- (16) ζητήματα δέ τινα περὶ τῆς ἰδίας  
zetemata de tina peri tes idias  
question.ACC.PL - certain.ACC.PL about the.GEN.SG own.GEN.SG  
δεισιδαιμονίας εἶχον πρὸς αὐτόν καὶ περὶ  
deisidaimonias eicon pros auton kai peri  
religion.GEN.SG have.PAST.3PL toward him.ACC.SG and about  
τινος Ἰησοῦ  
tinou iesou  
certain.GEN.SG Jesus.GEN.SG  
They had certain questions about their own religion and about a certain Jesus.  
(Acts 25:19)

- (17) ὁ γὰρ κύριος ὁ θεός  
ho gar kurios ho theos  
the.NOM.SG for lord.NOM.SG the.NOM.SG God.NOM.SG  
ὁ παντοκράτωρ ναὸς αὐτῆς  
ho pantokrator naos autes  
the.NOM.SG almighty.NOM.SG temple.NOM.SG it.GEN.SG  
ἐστὶν καὶ τὸ ἀρνίον  
estin kai to arnion  
be.PRES.3SG and theNOM.SG lamb.NOM.SG  
For the Lord God Almighty is its temple, and the lamb.

These two examples could not be analyzed correctly for much more mundane reasons: in the former, the adjective τινα was not in the lexicon (its inclusion in the full-form lexicon was masked by a homophonous interrogative pronoun). In the final example, the grammar has no analysis for tripartite apposition. I would expect both of these to receive suitable analyses if those unrelated shortcomings were corrected.

The remaining 24 exemplars all received analyses featuring reasonable tree structures and correct semantic connections regarding the position of the conjoined entity in the sentence's argument structure and the relationship between the conjuncts.

## 5.2 Overgeneration

In regards to overgeneration, the augmented grammar offers analyses for 2421 sentences in the New Testament. This is an increase of 153 inputs relative to the

baseline. While up to 27 are expected,<sup>8</sup> at least 126 new sentences (i.e. 1.1% of the corpus) that were not previously identified as involving discontinuous coordination receive parses courtesy of the new rules.

Manual inspection of the unexpected analyses showed that the *head-filler* rule was applying following *xsubj* and *xcomp* rules, i.e. for extracted subjects and complements, not only for extracted conjuncts. I revised the analysis to use a new RSLASH list instead of SLASH. The new list, RSLASH, is treated identically to SLASH by the rest of the grammar. This removed the possibility of crossover between extracted arguments (dislocated to the left) and extracted conjuncts (dislocated to the right), eliminating 45 of the instances of overgeneration. Another large chunk of spurious analyses involved a unary rule which pumps neuter plural  $\bar{N}$  to singular. This rule neglected to maintain the COORD feature on the mother, and also spuriously applied to relative pronouns. Correcting this rule eliminated analyses for 32 more items that used the discontinuous coordination analysis. It also eliminated spurious analyses for 2 inputs that were parsable with the baseline grammar, reducing the baseline to 2266 / 11504 from 2268 / 11504. Small additional improvements were had by further improvements of RSLASH management, for instance sealing off MOD...RSLASH when converting a participle to a substantive, and causing subordinators to require empty COMPS...RSLASH. The majority of the remaining cases seem to fall into three categories: (1) legitimate examples of discontinuous coordination (including one that was missed by the automatic search), (2) cases where discontinuous coordination does not seem to be the intended reading but is arguably available as an (unlikely) ambiguity, and (3) cases in which extracting a conjunct and immediately realizing it results in an analysis that would have been impossible using the ordinary coordination rules, apparently due to leaked constraints.

## 6 Conclusion

My corpus search showed that discontinuous coordination occurs with nontrivial frequency in Ancient Greek: more than 1% of all uses of the conjunction *kai* are discontinuous, leading to a test suite of 151 exemplars extracted from the PROIEL treebank. The phenomenon showed significant regularities: the marked conjunct is always dislocated to the right, typically to the end of its clause. This suggested an analysis parallel to that typically used in HPSG for topicalization, i.e. positing an extraction site, marking intervening constituents as gapped, and filling the gap with the marked conjunct higher up in the tree. I presented just such an analysis, and explored

---

<sup>8</sup>The sentences used in the testsuite are in some cases shortened versions of the full New Testament sentences, so it is likely that less than 27 of the 153 are actually the ones from the discontinuous coordination testsuite.

its performance on a test suite and on sentences that do not exhibit discontinuous coordination. My baseline grammar parsed just two out of the 151 test suite sentences, while the grammar augmented with the proposed analysis parsed 27 of them, which is roughly the same rate at which the grammar parses New Testament sentences not containing discontinuous coordination. Furthermore, the analysis was able to identify another instance of discontinuous coordination which was overlooked when creating the original test suite. I take this as evidence that the proposed analysis is a reasonable account of the main facts about discontinuous coordination in Ancient Greek.

## Acknowledgements

I am grateful to Emily Bender for many useful discussions about the direction and methodology of this study, and to Dag Haug both for his helpful advice and encouragement and for the availability of the excellent PROIEL treebank.

## References

- Bender, E. M., Drellishak, S., Fokkens, A., Poulson, L., & Saleem, S. (2010). Grammar customization. *Research on Language & Computation*, 8(1), 23–72. Retrieved from <http://dx.doi.org/10.1007/s11168-010-9070-1> (10.1007/s11168-010-9070-1)
- Bender, E. M., Flickinger, D., & Oepen, S. (2002). The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In J. Carroll, N. Oostdijk, & R. Sutcliffe (Eds.), *Proceedings of the workshop on grammar engineering and evaluation at the 19th international conference on computational linguistics* (pp. 8–14). Taipei, Taiwan.
- Haug, D., & Jøhndal, M. (2008). Creating a parallel treebank of the old Indo-European Bible translations. In *Proceedings of the second workshop on language technology for cultural heritage data (LaTeCH 2008)* (pp. 27–34).
- Packard, W. (2015). *Full forest treebanking*. Unpublished master’s thesis, University of Washington, Seattle, WA.
- Pollard, C., & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago, IL, USA: The University of Chicago Press.