

## Argumentative meanings and their stylistic configurations in clinical research publications

Olga L. Gladkova<sup>a\*</sup>, Chrysanne DiMarco<sup>b</sup> and Randy Allen Harris<sup>c</sup>

<sup>a</sup>*Department of English Language and Literature, Department of Drama and Speech Communication, University of Waterloo, Waterloo, ON, Canada;* <sup>b</sup>*Cheriton School of Computer Science, Department of English Language and Literature, University of Waterloo, Waterloo, ON, Canada;* <sup>c</sup>*Department of English Language and Literature, University of Waterloo, Waterloo, ON, Canada*

(Received 3 May 2015; accepted 23 March 2016)

The paper reports on the results of an exploratory study into the topical organisation and stylistic features of argumentation in a corpus of ophthalmic clinical research papers. The study responds to the need for systematised and generalisable argumentation models in knowledge-intensive fields. We present here a schematised superstructure of the arguments from the corpus, charting the configurations of stylistic features, which signal the elements of this superstructure, epistemic topoi. We pay special attention to the role of lexical categories (or semantic fields) in the configurations, to the relations between the fields, and to their interactions with other elements of the configurations, including semantic, grammatical, syntagmatic, deictic, and coreferential features. Epistemic topoi are a promising discourse constituent in argumentation because, as we found, they are distinct from syntagmatic units, such as phrases, clauses, or argumentative zones, and because they are signalled with substantially distinctive stylistic features despite having no fixed order in the superstructure. They hold considerable promise for computational argumentation analysis and processing, perhaps especially in scientific and technical discourses, where the need for reliable detection and summarisation is particularly high. Our investigation shows that despite the complex and interpenetrating semantic and stylistic attributes of argumentation, there are significant, computationally tractable regularities.

**Keywords:** argument domain; argumentative superstructure; clinical research; epistemic topoi; lexical (semantic) fields; linguistic features; metadiscourse; stylistic configurations

Argumentation analysis is gaining popularity (e.g. Association for Computational Linguistics, 2014; Palau & Moens, 2009; Saint-Dizier, 2012; Wyner, Mochales-Palau, Moens, & Milward, 2010)<sup>1</sup> and there is a growing consensus about the need for this work, as recent lively debates at the London Argumentation Forum have shown (<http://www.dcs.kcl.ac.uk/pg/hadjinik/LAF/>). It is generally agreed in this community that argumentative functions and meanings are signalled with the linguistic features of the statements communicating them, but it is increasingly clear that no type of linguistic ‘markers’, taken in isolation, can reliably communicate argumentative meanings and functions (Association for Computational Linguistics, 2014; Taboada, 2009; Teufel, 1999). Indeed, as we demonstrate in this paper, at the level of individual statements, the links between linguistic and argumentative organisation are quite complex. Nonetheless, we are convinced they are computationally tractable. We use the concept of epistemic topoi as shorthand for statement types that function as the semantic elements of problem-solving, decision-making, and interpersonal argumentation in research publications. Such elements include statement categories like STUDY DESIGN, MOTIVATION, and KEY FINDINGS, with conventional functions in the

---

\*Corresponding author. Email: [ogladkov@uwaterloo.ca](mailto:ogladkov@uwaterloo.ca)

argumentative *superstructure* (Van Dijk, 1980) of a particular genre of papers in a particular research domain.

Topoi (singular, *topos*), deriving from classical rhetoric, are most fully articulated in the early literature by Aristotle, who distinguishes between common topoi, lines of argument present in all genres and discourses (such as CONTRAST, COMPARE, and FROM THE IMPOSSIBLE), and specific topoi, lines of argument present in particular genres and/or argument fields.<sup>2</sup> More recently topoi have drawn some attention in rhetoric and linguistics, where they are mostly known as *speech acts* (Myers, 1992) or *topics* (Van Dijk, 1980, esp. pp. 94–98), as well as in NLP, where they may be called *content elements* (Trawiński, 1989), *components of information* (Liddy, 1991), *sentence classes* (Paice, 1990), *generalized propositions* (Teufel, 2014), *argument categories* or *moves* (Association for Computational Linguistics, 2014).

In this paper we report on the results of the first, exploratory part of our project inquiring into the topical organization of arguments in research papers and its surface manifestations. We offer an analytic framework that can be deployed profitably in computational argumentation research. In specific, we chart the argumentative functions of epistemic topoi and their stylistic features in a corpus of clinical ophthalmic publications. During the first stage of the project we restricted our analysis to manual annotation of the corpus by one researcher with subsequent linguistic analysis of the annotation results. Despite the obvious limitations of the exploratory format of the study, we believe the findings it produced are robust enough to be deployed computationally. Another necessary caveat is that the superstructures and the specific features of particular topoi will likely show significant variation between argument domains and genres. For example, in our corpus we did not find a tense shift similar to the one that Malcolm (1987) and Myers (1992) established in their materials. Despite such variation, however, we are confident other practitioners will derive helpful insights from our description of the features of argumentative meanings and the functions of the meanings in biomedical research papers, both in our approach and in our findings.

The particular type of surface features that we present here is the recognizable configurations of stylistic elements identifying the epistemic topoi. Such configurations are not random combinations of linguistic cues but sets or interrelated semantic, lexico-grammatic, deictic, and coreferential features. We start by situating our approach in an overview of earlier research on the metadiscursive signals of statement types in research papers. The next two methodological sections outline our study design and offer insights into the stylistic properties of epistemic topoi. The fourth body section describes the system of topoi in our corpus, and the last two summarize and interpret these findings in the context of computational argumentation.

## 1. Theoretical background: metadiscourse analysis

Introduced in Teun A. van Dijk's important early text-linguistics study, *Macrostructures* (1980), the term *superstructure* refers to “the schematic form that organizes the global meaning of a text” (pp. 108–9). This form “indicates which textual ‘functions’ are relevant for this kind of discourse” (p. 69).

The superstructure (which van Dijk also calls [*conventional*] *schema*) is a semantic, not syntagmatic<sup>3</sup> construct, so its elements have limited connections with the sequential organisation of the text, such as the IMRD<sup>4</sup> sections (that is, the ‘Introduction’, ‘Methods’, ‘Results’, and ‘Discussion’). For example, we can confidently expect statements of theme or purpose to occur at the end of research paper introductions (Myers, 1992) and at the beginning of discussions. Yet there are no predictable ‘slots’ for such statements in the linear structure of either part of the argument (as reflected in our corpus). In the absence of reliable positional markers, analysis must

turn to the linguistic properties of the topoi. The traditional terms for such surface manifestations of argumentative meanings and functions is *metadiscourse*.

### 1.1. *Isolated linguistic features and their distributions*

The linguistic ‘signposts’ of argumentative organisation are typically analysed in terms of their distributions in texts. Their various types have enjoyed much attention in the literature, from passages of explicit authorial commentary (Crismore & Farnsworth, 1990) to typographic features (Kumpf, 2000) and intonation (Thompson, 2003). Hyland (2005) provides an extensive inventory of metadiscursive signals, which includes phrases, content and form words. Of all these types, content words have since long been analysts’ favourites. Such lexicalised, explicit, references to argumentative meanings are often referred to as *metalanguage* (Berry, 2005).

Despite its great significance, however, metalanguage has limited direct correlations with the argumentative meanings of sentences or clauses. It is not present in every statement, and, even when present, it does not necessarily label epistemic topoi. For example, most academic readers will probably recognize the semantic types of the following two statement types from our corpus. The first type talks about consistency between the reported results and earlier findings:

1a. Crichton et al. confirmed the same issue. (E10)

(We label this topos RESULTS CONSISTENCY; for a full list of our topoi, their classes, and their argumentation categories, see Appendix 3.) The other familiar statement type talks of an issue that is open and in need of attention or solution (the topos, OPEN ISSUE):

1b. To the best of our knowledge, the effect of brimonidine on POBF has not yet been reported. (G24)

Note that there is no direct link between the metalanguage used in these statements (*issue, knowledge, effect, and reported*) and their topical designations. By simple lexical search, one might expect that the metalanguage of 1.a signals the OPEN ISSUE topos, since it is the only one here that uses the term *issue*. On the other hand, the term *reported* of 1.b might suggest the presence of *results* and would thus be expected to correlate with the RESULTS CONSISTENCY topos. Yet instead the obverse is true: the RESULTS CONSISTENCY statement makes no mention of either *results* or *consistency* but invokes *open issues*. The OPEN ISSUES statement explicitly mentions *reporting* but not *open issues*.

The question is then how to reconcile common intuitions about the metadiscursive functions of lexis with the negative results of empirical findings based on metalanguage. The answer to this question lies, we think, in the notion of stylistic configurations, which include lexis along with other features.

### 1.2. *Configurations of stylistic features*

Statistical correlations between certain linguistic features and the IMRD structure are frequently reported in the literature to be (e.g. Channell, 1990; Malcolm, 1987; Salager-Meyer, 1992). Yet distribution studies, such as Teufel’s (1999), have also demonstrated that at the level of statements no linguistic feature taken in isolation is sufficient for analysis of argumentative meanings. Even the distributions of multiple types of metadiscourse are not helpful for this purpose. The smallest units that can be identified statistically by the distributions of their linguistic features appear to be about a hundred words long (Biber, Csomay, Jones, & Keck, 2004, p. 57). This is an order of magnitude larger than an average sentence in most texts, including argumentative texts (and certainly including our corpus).

Comprehensive descriptions of the linguistic signatures of particular statement types are laborious and rare. But from such work we draw confidence that combinations of linguistic features can provide access to text semantics at the statement level. Swales's influential 1990 CARS model of rhetorical moves in research papers suggests that the linguistic features of statements can index their meanings even in abstraction from their particular contents and textual environments. For instance, statements that evoke recent developments in the relevant field are signalled with recognizable sets of features:

- The possibility ... has generated interest in ...
- Recently, there has been wide interest in ...
- The study of ... has become an important aspect of ...
- The theory that ... has led to hope that ...
- The effect of ... has been studied extensively in recent years.
- Many investigators have recently turned to ... (p. 144)

There is no infallible, one-to-one term-function mapping. Yet in some of these statements the idea is made explicit with the lexeme *recent(ly)*, while in others it is signalled through a combination of the present perfect tense and positively valenced diction: *generated interest, important, hope*.

Similarly valenced diction (*classic, great importance, central*), however, when combined with the present indefinite tense, creates a different profile:

- The time development ... is a classic problem in fluid mechanics.
- The explication of the relationship between ... is a classic problem of ...
- Knowledge of ... has a great importance for ...
- A central issue in ... is the validity of ... (p. 144)

Such statements evoke timeless or long-standing, paradigmatic issues that are endowed with the status of disciplinary touchstones.

Swales' examples demonstrate that argumentative meanings are communicated by combinations of linguistic features, not by specific isolated markers. Constellations of features, like tense, aspect, and diction, create the stylistic profiles of statements. Following Swales' analysis, several other authors have confirmed his insights that configurations of features may indeed operate as metadiscourse (Hersh, 2009, p. 406; Liakata, Thompson, Waard, Nawaz, Maat, & Ananiadou, 2012; Litman, 1996; Stirling, Fletcher, Mushin, & Wales, 2001; Taboada, 2009).

Our findings join these confirmations of Swales' thesis. And we go further. Work in these areas has up till now been carried out as two parallel lines of analysis, one focused on identifying the meanings, and the other on describing (or computing) their surface features. We do both in concert. We catalogue argumentative meanings *based on* their surface manifestations. Our contribution, that is, consists in bringing these two lines of analysis together. Our analysis covers the entire argumentative superstructure along with the stylistic profiles of its elements as they are represented in our corpus.

## 2. Study design and corpus

Our study design had significant differences from the methods currently popular in computational argumentation. Rather than evaluate an imported taxonomy based on inter-annotator agreement, we developed our typology from the bottom up using the methods of analytic induction (Strong, 1988) and triangulation (Lazaraton, 2002). The corpus annotation results produced

by one annotator were triangulated with the results of linguistic analysis and insights gained by studying the field's metatheory. Thus, we conceived our study along the lines of Willard's *human as scientist* approach (1989, p. 18), on the assumption that analysts can learn about significant textual and discursive patterns of a domain in the same way its novices do (cf. Wilbur, Rzhetsky, & Shatkay, 2006). As is usually the case with exploratory research, our study was iterative and cyclical (Brown, 2004). We view this approach as necessary for the development of a typology that is based both on human readers' perceptions of argumentative meanings *and* their surface features, which may allow for more elaborate annotation guidelines and more effective machine training materials in the future.

The choice of the publication topic for the corpus was motivated by a combination of the lead researcher's professional background and personal interest in glaucoma. Our work with the corpus consisted of a survey of a set of NTG articles followed with close reading of the literature reviews from this set and manual annotation of a smaller subset of clinical research papers. The survey helped us to gain understanding of the nature of NTG and its treatment and to achieve insight into the conceptual and methodological tools of the field's research and the argumentative features of the publications. The literature reviews introduced us to the insiders' perspectives on the state of the art. Finally, the annotation and linguistic analysis allowed us to identify, describe, and classify the most salient and significant argumentative patterns in the clinical studies and identify their stylistic manifestations.

The larger set consisted of fifty-seven NTG papers: all MEDLINE (PubMed) full-text English-language articles published by 1994 and all MEDLINE (PubMed) free full-text English-language articles with abstracts published after 1994. It included several research genres, most notably case studies, literature reviews, methodological inquiries, clinical, experimental, and laboratory investigations. This set was narrowed down to a manageable size and format for manual annotation, based on the genre and technical parameters of the publications. Specifically, the annotation corpus was restricted to clinical research papers (the largest subset in the larger set), from which two papers were excluded because of their no-copy format. It consisted of seventeen articles (45,599 words), listed in Appendix 1.

To concentrate on the textual mechanics of the argument, the papers were stripped of figures, tables, end-of-text citations, and front- and end-matter. Parenthetical citations were replaced with ellipses. Using the technique of visual annotation (Gladkova, 2010), the lead author identified the recurrent statement types comprising the argumentative superstructure of the papers.<sup>5</sup> This annotation was based on the distinctive meanings and configurations of stylistic features of the statements. At the first stage, the statement types were marked in the corpus and classified based on their rhetorical functions. We then analysed the obtained materials for two kinds of correspondences: (1) stylistic patterns correlated with argumentative meanings and (2) their typical clusterings and distributions in the texts. The former analysis served as a test of whether semantic designations could be paired with stylistic patterns. The latter analysis generated a functional categorisation of the patterns. During the second stage our work was motivated by the questions of (1) whether or not the statement types could be recognised in isolation from the texts, based on their internal stylistic configurations, and (2) whether such configurations were pronounced enough to claim the status of distinct categories for the identified meanings. The annotation results were tabulated and subjected to comparative linguistic analysis. Ambiguities were resolved, the categories lacking formal distinctiveness were merged, and the categories showing irreconcilable features were divided. Once the classification of the topoi was complete, we went on to classify the stylistic features associated with them. The purpose of this work was to formalise 'the internal structure' of the identified categories (Ide & Romary, 2004, p. 223) and to verify our list of topoi against a set of methodologically sound criteria.

### 3. Methodological insights: stylistic properties of epistemic topoi

#### 3.1. Syntagmatic indeterminacy of topoi

Syntagmatic realisations of argumentative functions and meanings may lie anywhere between a lexical unit and a text. This is because meanings occupy a different plane of organisation from syntax. Halliday and Hasan (1976) usefully observe that they are communicated by ‘texture’, a network of semantic links in the text (p. 8), and are subject to a different ‘kind of STRUCTURAL integration’ than clauses and sentences (Halliday and Hasan, 1976, p. 2; their emphasis).

Consider the STUDY DESIGN topos from our corpus. This topos describes the authors’ purposes and/or methods, but it is seldom if ever referenced explicitly. We found only two instances of its metalinguistic tag in our corpus, both in the same paper:

- 2a. We cannot directly compare our results with the previous data, because our work has a different study design<sup>6</sup> based on the subgroup comparison within patients with NTG. (G5)
- 2b. Our study design was not based on the comparison between the case and control groups, but based on the comparison of subgroups in patients with NTG. (G5)

We see here that an argumentative meaning can in principle be expressed with a syntagmatic unit as short as a phrase. More strikingly, as we demonstrated above, metalanguage does not necessarily index the argumentative meanings of the whole statement where it occurs. In these examples, *study design* is not an indication of the statement meaning – not even in 2b where it is the sentence topic (Van Dijk, 1980, esp. pp. 94–98). Instead, in both examples *study design* functions as part of another topos: a statement type explaining the distinction of the present study from other similar investigations. For comparison, here is a straightforward STUDY DESIGN statement from the same paper:

- 2c. The purpose of the present study was to classify patients with untreated NTG by the degree of nocturnal BP reduction; to study BP, IOP, and MOPP parameters in each classification; and to investigate predictor variables of circadian MOPP fluctuation (CMF). (G5)

Unlike 2a and 2b, this sentence contains no lexicalised reference to its own topos. Instead, it names the cohort type (*patients with untreated NTG*), the objectives and methods of the investigation (*to classify patients . . . by the degree of nocturnal BP reduction; to study BP, IOP, and MOPP parameters in each classification; and to investigate predictor variables of circadian MOPP fluctuation*). These meanings are the major semantic elements that constitute the concept of study design in our corpus. Here are two other statements of this type, each containing a combination of the same meanings:

- 2d. This is a retrospective review of a large number of NPG patients referred to a single hospital based glaucoma service. (G33)
- 2e. In this study, we tried to evaluate the difference in nerve fibre layer (NFL) defects between NTG and POAG through analysis of RNFL photographs. (G21)

The idea of study design can have an even longer expression, straddling several sentences. In our corpus, one paper (E1) has an entire section entitled ‘Analysis Design’. Like the STUDY DESIGN statements above, this section consists of semantic blocks related to data selection, analysis objectives and methods.

The flexible unit sizes that can perform argumentative functions highlight both the difference and the interaction between semantic and syntagmatic organization and show that argumentative meanings, or epistemic topoi, are primarily associated with the former, rather than the latter.



The syntagmatic indeterminacy of topoi creates challenges for analysis and calls for methodological decisions. One obvious decision concerns the size of analytic units. The current practice for information-retrieval and knowledge-management systems favours lifting whole sentences from texts, rather than generating synthetic content from words and phrases. In keeping with this practice, we decided to focus on sentence-size units and dismiss argumentative meanings manifest below or above the sentence level. In statements with ambiguous sentence boundaries, our default analysis unit was the clause. A different decision was required when we encountered loose lexically cohesive links between sentences (connectives, anaphoric pronouns, adverbs, and adjectives). When extracted from the text, statements with such links seem incomplete or even incomprehensible:

- 3a. The latter probably because the treatment with pilocarpine was ceased after surgery. (E15)
- 3b. However, no statistically significant change in IOP range was found. (G16)

Unless loose links were a recurrent feature of a topos (such as RESULTS CONSISTENCY and EXTRAPOLATIONS), we considered linked series of sentences as composite statements, similar to compound sentences.

Overlapping and nesting topoi also called for methodological decisions. We had to decide which configurations of features to count as basic types and which as composite. Our method consisted of treating as a basic topos any statement type with an identifiable and recurrent semantic makeup. Combinations of such basic topoi were classified as composite categories. For example, the major semantic elements of the topos STUDY DESIGN, discussed above, are the cohort type, the objectives and methods of the investigation. All these elemental meanings are present in our corpus as basic topoi dealing with the specific aspects of study designs (COHORT SCREENING, THEME/PURPOSE, INTERVENTIONS, etc.), each with its own typical configuration of stylistic features. Since any element of STUDY DESIGN can be elaborated into an independent statement, we categorised the statements dealing with its isolated aspects as basic topoi, and statements dealing with more than one of these aspects, as the composite STUDY DESIGN topos.<sup>7</sup>

### 3.2. *The lexical elements of stylistic configurations*

The elements of the stylistic configurations of topoi do not function as unmalleable nuggets of semantic information but interact with one another and with their semantic and syntagmatic environments in rich but regular ways. To illustrate this property of the topoi from our corpus, we will first consider some of their lexical attributes.

Most of the lexemes that we found to be significant for our purposes provide expressions for people's activities and for the theory that they construct about health and disease. The expressions allowing authors to talk about their activities concern clinical practice, data acquisition, and argument and discourse organisation. Theory expressions include language related to observations and knowledge.

Clinical practice lexis containing professional terminology is a prominent category in our corpus:

- Clinical procedures (e.g. follow-up, management, operate, surgery, therapy, wash-out)
- Diseases, syndromes, symptoms, clinical instruments, and medications (e.g. latanoprost, NTG, POAG, [visual] field loss).

The next category, which also accounts for much volume in the papers and is associated with a substantial body of expressions, is data acquisition:

- Analysis (e.g. assess, calculate, determine, divide)

- Examination (e.g. [de]note, examine, measure, monitor, record, test)
- Designation (e.g. calculate/define/record/take as, classify as/in[to], consider as/to be, criteria, exclude, include, judge).

Epistemic lexis is related to the organisation of arguments and discourse:

- Reasoning (e.g. conclude, consider, data, find, know, propose, reveal, suggest, support)
- Research (e.g. address, investigate, literature, paper, publish, question, report).

Observations of natural phenomena inform the next large body of expressions:

- Phenomena and attributes (e.g. *age, deteriorate, duration, high, range, time, value*)
- Circumscription (e.g.  $\geq$ , *at least, maximum, only, or better*)
- Generalisation (e.g. *all, average, both, each, few, majority, mean, most, none, range*)
- Numerals (e.g. *two,  $14.6 \pm 1.7$* )
- Participants (e.g. *controls, eye, patients, subjects*).

Knowledge lexis indexes available theory and relations between what the community considers as relevant aspects and parameters of the observed phenomena:

- Association (e.g. associated, correlated, involve, linked, more likely, predictor)
- Cause and effect (e.g. cause, contribute, influence, lead to, mechanism, pathogenesis, risk [factor], role, susceptible; affect verbs: e.g. affect, improve, increase, reduce)
- Identity (e.g. characteristic, distinct, form, normal, reliable, reproducible, subset, typical).
- Presence and appearance (e.g. have, indicator, present with/as, reflect, show, sign).

Several categories of expressions communicate more abstract meanings than those listed above and are used in the stylistic configurations of topoi across the board:

- Congruity and consistency (e.g. agreement, comparable, confirm, consistent, similar, surprising)
- Enablement and possibility (e.g. able, can, easy, likely, may, possible)
- Deontic modality (e.g. have to, need, must, should)
- Diminution or negation (e.g. few, hardly, not, small)
- Evaluation (e.g. advantage, ideal, successful)
- Time and aspect (e.g. further, future, recent, new, no longer, remain, today's).

In our corpus, we found that most lexemes function within their topical configurations as members of their semantic categories (or fields) rather than as unique individuals, which suggests powerful computational applications. But, as always, there are complications. We found exceptions to this principle that require special attention.

First, as we learned, lexical synonymy does not always imply identical semantic functions. For example, *significant* and *important* are close synonyms in most contexts, and this is true to an extent in our corpus as well. That is why both lexemes are frequent in MOTIVATION statements, which typically occur in introductions and conclusions. Yet *significant* – unlike *important* – is also part of the STATISTICAL SIGNIFICANCE topos. In fact in each of its topoi, *significant* conveys the meaning of either statistical or functional significance. As a result, the use of *significant* overlaps with *important* only in one topos: MOTIVATION.

Second, we found a few expressions working as unique identifiers of their argumentative meanings. Such are the phrases *in summary* and *in conclusion*, which in our corpus occur only in the KEY FINDINGS topos. As members of the reasoning field, neither *summary* nor *conclusion* matter for the configuration of this topos. Either of them used in a statement without *in* does not suggest talk of KEY FINDINGS. They signal this topos only as part of an *in*-phrase. Somewhat



similarly to such unique signals, several configurations include topos-specific lexis, which may be drawn from one or more fields above. Such are words like *clinical*, *retrospective*, *review*, or *trial* (signalling the RESEARCH TYPE topos) and words like *figure* and *table* (signalling the DATA PRESENTATION topos).

Overall, however, stand-alone or unique lexical markers were anomalies to the general stylistic complexity of epistemic topoi. First, most configurations signalling the topoi in our corpus include not single but multiple lexical field members, typically two or more per topos. Second, lexical signals are usually combined with other stylistic features. Third, in a significant number of topoi some lexical fields were found to be interchangeable within their stylistic configurations with other types of features.

To illustrate some of these points, consider the MOTIVATION topos, which provides motivation for the readers to attend to the studies by explaining their relevance. Positive motivation is typically provided by either stressing the significance of the issue or the benefits of the findings:

- 4a. Therefore, POAG patients with uniocular field loss represent an ideal population in which to investigate factors influencing the onset of field loss over a period of time. (G36)
- 4b. In view of these, the importance of assessing the impact on ocular haemodynamics of a glaucoma medication becomes evident. (G24)
- 4c. As the eye with the most serious progression was operated on, the fact that in the operated eye progression was stopped and in the non-operated eye progression went on has double significance. (E15)
- 4d. The importance of being able to evaluate as accurately as possible intraocular blood circulation and any changes in it resulting from different forms of treatment is clear. (E3)
- 4e. This makes measurement of POBF relevant in evaluating ocular haemodynamic effects of glaucoma medications. (G24)
- 4f. Our analysis may be more easily extrapolated to clinical practice in that the risk of future visual field progression can be estimated from 'current' IOP, taken as the median of readings done in the past 6 months. (G31)

As we explained above, many MOTIVATION statements have the lexeme *importan\** or *significan\** as part of their stylistic configurations. This is the case with statements 4b, 4c, and 4e. Yet there are other lexical means as well that communicate the relevance of the study. For instance, one can explicitly state this idea, as 4f does. Another option is to emphasise the significance with evaluative lexis, such as *ideal* in 4a. Yet another way is to use enablement and possibility expressions, such as *being able* and *possible* in 4d or *may* and *easily* in 4f.

Of course a relevance expression is not sufficient on its own to make a MOTIVATION statement. Here, for instance, are two statements whose argumentative meanings are other than MOTIVATION despite the presence of relevance lexis in them:

- 5a. On the other hand, there are also a number of studies by different authors that describe the importance of local or general vascular factors as the primary cause of the ocular damage. (E3)
- 5b. It is unknown whether this effect is of relevance *in vivo* and in humans. (G10)

A careful look at the MOTIVATION set and its comparative analysis with other topoi reveals a rather complex configuration of features composing its stylistic profile. The relevance expressions are complemented with lexis from the fields of clinical practice (*POAG*, *field loss*, *glaucoma medication*, *progression*, *operated [on]*, *treatment*, *POBF*, *clinical practice*, *IOP*), data acquisition (*population*, *evaluate*, *measurement*, *analysis*, *estimated*, *readings*), argument and discourse organisation (*fact*, *extrapolated*, *investigate*), observation (*onset*, *period of time*, *haemodynamics*,

*changes, intraocular blood circulation, median*), or knowledge (*factors, impact, resulting, effects, risk*). Many of these lexemes take the form of abstract nouns. Another important feature of MOTIVATION, one that is particularly tractable computationally, is co-referential links with the paper title. The topos contains the same words as the title, or their synonyms, hyponyms, or hypernyms. Consider how statement 4a echoes the title of its paper ('Clinical Factors Influencing the Visual Prognosis of the Fellow Eyes of Normal Tension Glaucoma Patients with Unilateral Field Loss'). We repeat 4a here for convenience:

Therefore, POAG patients with unilateral field loss represent an ideal population in which to investigate factors influencing the onset of field loss over a period of time.

Thus, the identification of MOTIVATION statements hinges on a stylistic configuration that includes several lexical fields (emphasis, evaluation, enablement and possibility, clinical practice, data acquisition, argument and discourse organisation, observation, and knowledge). At least two of these fields are interchangeable. The configuration also includes morphological features (abstract nouns) and co-referential links with titles.

### 3.3. Interaction between the elements of stylistic configurations

Lexical analysis is indispensable for the computational processing and modelling of arguments. Yet not only is lexis seldom the only significant attribute of a topos, it is also hardly the most reliable one. We found that lexicalised expressions may be interchangeable with morphological and grammatical meanings. Also frequent are interactions of the meanings and functions of lexis with grammar, morphology, and syntax.

One striking feature of the topoi in our corpus is the morphological fluidity of their stylistic configurations. Consider these two THEME/PURPOSE statements:

- 6a. This study is aimed at assessing the effects of therapy on POBF and functional parameters in patients with NTG. (E3)
- 6b. This study was conducted to determine the longer term effect of latanoprost on the IOP of patients with newly diagnosed NTG. (G16)

In terms of their semantic makeups, both statements have expressions of purpose. Yet, while in 6a the idea is communicated with an explicit metalinguistic tag, *aimed*, in 6b the same effect is achieved syntactically, with the presence of an adverbial modifier (or adjunct) of purpose, *to determine*. The particular lexical meaning of the adjunct plays no role here. A different word with the same grammatical meaning would work just as well: *This study was conducted to analyse/ to test/ to study/ to inquire*, etc.

In the same way, the idea of possibility, ability, or enablement may be expressed either explicitly, with the metalinguistic tag *possible*, or with a modal verb such as *can*. It may even be expressed morphologically, with the suffix *-ible/-able*. The idea of identity can be conveyed with identity lexis (e.g. *characteristic, subtype*) or with the help of a copula establishing a relationship of identity between the subject and predicative (e.g. *Phospholipids are constituents of all membranes*). Comparison can be expressed lexically (e.g. *different, similar*) or metalinguistically (*compar-*), as well as with comparative grammatical forms or comparative syntactic constructions. Meanings can easily traverse the boundaries between types and levels of linguistic organisation.

Another linguistic phenomenon that analysis of topical organisation must take into account is lexical polysemy. In our corpus, we found that some rather frequent or significant lexemes shift their meanings depending on the configurations in which they are integrated to form a particular topos. One such lexeme is *group*, a definitive feature of a number of topoi, which is typically

associated with participant lexis but also communicates the idea of organisation. In particular, *group* is a lexical attribute of the INTERVENTION DATA topos, where it is combined with clinical practice or examination lexis and with numerals or generalisation lexis. In such contexts, *group* performs the same role as participant lexis, such as *participants*, *patients*, or *subjects*:

- 7a. In 10 eyes an argon laser trabeculoplasty had been performed. (E15)
- 7b. Of the remaining 83 patients, 28 had fixation threatening field defects and were started on treatment. (G16)
- 7c. Six eyes, four from the nil adjunct group, one from the 5-FU group, and one from the MMC group, had further glaucoma drainage surgery (see Table 1). (G35)

On the other hand, where the organisation motifs take the upper hand, we find *group* in the company of such abstract analytic notions as *category*, *sample*, *parameter*, *factor*, and *event*. Consider this representative example of the DATA HANDLING topos:

- 7d. These ‘visual field failure’ events were modelled on baseline values for factors that did not change with time, such as sex and adjunct group. (G31)

Here the authors categorise *group* as a *factor* in parallel with another abstraction, *sex*. In such statements, *group* does not mean a collective of study participants but an analytic entity with certain characteristics deemed significant for the investigation.

In some cases, the meaning that a lexeme brings to a topos depends on its morphological form. For example, the adjective *normal* communicates the meaning of identity, in the same way as *characteristic* or *typical*. Witness the similarity of function between the underscored words in these COHORT SCREENING statements:

- 8a. To make the diagnosis of NTG, patients must have had ... a reproducible visual field defect typical of glaucoma ... Patients with a normal visual field in one eye and a field defect in the contralateral eye, at the time of diagnosis, were selected for this study. (G36)
- 8b. A diagnosis of NTG was made if ... glaucomatous optic disc changes and visual field defects characteristic of glaucoma ... were present in one or both eyes of the patient ... (G16)

The adverb *normally*, on the other hand, plays the same semantic role as *may* or *usually* in our corpus, all of them expressing enablement and possibility, not identity. Witness the functioning of this lexical field as part of the KNOWN CAUSES/EFFECTS configuration:

- 8c. Levels of IOP, which are normally well tolerated, may produce damage to the optic nerve. (E15)
- 8d. Compressive optic neuropathy is usually caused by intracranial lesions and not by normal blood vessels ... (G12)

Within lexical fields, the semantic distinctions between abstract nouns and other morphological forms turned out to play an important part in the organisation of arguments. Consider the role of the word *women* in this OBSERVATION DATA statement:

- 9a. Thirty-four patients (63%) were women. (G36)

It talks about the sex of the patients using a concrete noun, *women*. Most of such statements are found in the results section. On the other hand, when the authors analyse and interpret observations – as distinct from reporting them – more abstract nouns come into play. In the following ASSOCIATIONS/CORRELATIONS statement from the discussion section of the same paper, *sex* is a condensed reiteration of the idea that the majority of patients were women:

- 9b. Analysis using Cox univariate and multivariate regression techniques revealed strong evidence of independent associations between time to onset of field loss and both the *sex* of the patient and the severity of field loss of the fellow eye (AGIS score) at presentation (Table 1). (G36)

Similarly, the presence of *results* in ‘It should be noted that the results presented here come from a retrospective analysis of data’ (G35) does not mean that the sentence presents the study results but rather that it comments on them. Not only does the statement contain no results, it occurs outside of the results section. This mismatch represents a pattern: The abstract term *results* is almost seven times more frequent in discussions compared to other sections (including the ‘Results’ section!).

Syntagmatic environments influence meanings in a number of ways. First, some lexemes take on different meanings in different syntactic structures. For example, the word *classify* may refer to an analytic activity, along with *calculate*, *define*, or *divide*. However, in combination with *as* or *into*, the same word takes on the meaning of designation and thus falls into the same field as to *define as*, *exclude*, or *include*. Second, the syntactic roles that certain lexical elements play or the positions they occupy in their sentences is a factor in our corpus. One group of such elements is the analytic abstractions *difference*, *significance*, and *correlation*, some of the most frequent abstract nouns in our corpus. They are particularly prominent in the STATISTICAL SIGNIFICANCE topos, which we already touched on above. The central role of the abstractions in this topos is signalled by their role as the sentence subject. This is in fact the only syntactic role in which these three nouns function as a definitive feature of the topos in our corpus.

Another syntagmatic feature that may interact with the meanings of some lexical elements is the word order. For example, autoreferential phrases like *our study* or *this paper* are one of the elements indicating the THEME/PURPOSE and KEY FINDINGS topoi, but only when used in the first segment of the sentence.

Last but not least, the interaction between the lexical features of topoi and their temporal and modal profiles creates rather interesting semantic patterns. Some of the most important distinctions between various kinds of primary and secondary information in our corpus are signalled through the modal, temporal, and aspectual features of the statements: verbal tenses, infinitive forms, modal expressions, and adverbial modifiers. For examples of these distinctions, we will turn to two groups of statements: one talking about well-known relationships among the phenomena at issue, and the other about relationships that have been found or are proposed by the authors. These relationships include causal links, correlations, associations, differences, similarities, groupings, and divisions. In our corpus, we have divided the statements addressing such relationships into the following topoi:

KNOWN CAUSES/EFFECTS  
 KNOWN ASSOCIATIONS/CORRELATIONS  
 FOUND CAUSES/EFFECTS  
 FOUND ASSOCIATIONS/CORRELATIONS  
 EXTRAPOLATIONS

The first two topoi talk about the state of knowledge. The last three present the authors’ own findings and thoughts. To communicate these argumentative meanings, the authors deploy a system of lexical signals combined with certain syntactic structures.

Some statements conveying received knowledge are presented as more or less unproblematic information:

- 10a. Included as positive factors are a higher incidence of disc haemorrhage ... , more pronounced peripapillary atrophy ... , higher incidence of retinal occlusive vascular diseases ... , coexistence of immunocompromised conditions ... , increased resistance index in orbital vessels ... , and alterations in the diurnal variation of systemic blood pressure ... (G12)
- 10b. In normal subjects, a higher IOP is associated with a higher degree of myopic refraction, and myopia is more prevalent in patients with primary open-angle glaucoma or NTG than in normal subjects ... (E1)

Oftentimes, however, the authors will use hedging to tone down their sense of confidence in the propositions:

- 10c. It has been shown that apoptosis can be induced by antiphosphatidylserine antibodies, which results in occlusion of small vessels by thromboemboli and finally leads to disturbance of the microcirculation in the inner ear and eye. (G14)
- 10d. Both of these studies found that the patients with higher initial IOPs showed greater IOP reductions. (G6)

In each of these pairs, the first statement (10a and 10c) talks about CAUSES/EFFECTS, and the second (10b and 10d) of ASSOCIATIONS/CORRELATIONS. The lexical signals of the CAUSES/EFFECTS meaning used in this set are the words *factors*, *induced*, *results in*, and *leads to*. The ASSOCIATIONS/CORRELATIONS meaning in 10b is even more obvious due to the metalinguistic tag *associated*. Statement 10d is devoid of such clear lexical signals. Instead, its stylistic configuration includes the verb *found*, two comparative adjectives (*higher* and *greater*), and the word *patients*. The frequent references to the sources of the cited information (represented as ellipses in 10a and 10b) is of course yet another indicator of secondary evidence.

The stylistic configurations of the set of statements below are somewhat similar to the ones above in that the first one talks about CAUSES/EFFECTS and the second about ASSOCIATIONS/CORRELATIONS. Yet these statements convey not secondary information but primary findings:

- 11a. Dorzolamide lead [*sic.*] to a significant acceleration of systolic blood flow in the short posterior ciliary artery (Table 4). (G10)
- 11b. Eyes from older patients were more likely to lose visual acuity over the follow-up period. (G33)

What are the features telling the readers that statements 10a through 10d refer to KNOWN information and statements 11a and 11b to what was FOUND during the study? This difference is mostly signalled by the verb tenses in the statements. In the simple sentences from the KNOWN set (10a and 10b), all finite verbs are used in present indefinite (also referred to as simple present). This represents the relationships discussed in the sentences as enduring, objective facts of reality. In the compound sentences from the same group (10c and 10d), the main clauses have either present perfect or past indefinite verbs. The subordinate clauses of these compound sentences have either verbs coordinated with the past tense of the main clause (as in 10d) or present indefinite verbs (as in 10c). In contrast, in the FOUND set (11a and 11b) all finite verbs are used in past indefinite. This highlights the authors' reluctance to extrapolate their findings beyond their study until they get corroborated by other researchers. When, however, the authors consider that EXTRAPOLATIONS are warranted, here is how they may represent them:

- 12a. Probably the deficient perfusion of the optic nerve head is due to an imbalance between intraocular pressure and the blood-pressure in the small branches of the short, ciliary arteries. (E15)
- 12b. This suggests that an increased POBF may be associated with favourable prognosis of glaucoma . . . (G24)

The former of these statements hypothesises about a causal link, and the latter points to a possible correlation. In contrast with the FOUND CAUSES/EFFECTS and FOUND ASSOCIATIONS/CORRELATIONS statements above, where the finite verbs are past indefinite throughout, here all verbs are present indefinite. Statement 12a is a simple sentence. The present tense of its predicate indicates the authors' invitation for their readers to consider generalising their EXTRAPOLATION to other cases. Yet the authors are careful not to misrepresent this EXTRAPOLATION as a causal link that is taken for granted or that has been suggested by other authors. So they use the modal adverb *probably* to assume ownership of both the EXTRAPOLATION and the caution. Statement 12b is a compound sentence. In its subordinate clause, the compound modal predicate used in present indefinite (*may be associated*) communicates a cautious generalisation, much like the verb and adverb do in 12a (*Probably . . . is due*). In the main clause, the present indefinite tense verb (*suggests*) underscores grammatically the fact that the EXTRAPOLATION is happening before the readers' eyes, as it were, rather than borrowed from the literature.

In summary, it is safe to say that topoi consist of meanings, rather than words. Many of these meanings can be expressed not only by lexical means but also morphologically and grammatically. Lexical meanings also frequently depend on their morphological and grammatical forms and syntagmatic environments.

#### 4. Corpus annotation and analysis results: topical organisation of ophthalmic research papers

An important motive behind the studies reported on in the papers from our corpus, as one would expect, is understanding the nature of the disease, its treatment, and management options. Such understanding requires a complex conceptual system involving phenomenal, methodological, and technical knowledge. The phenomenal knowledge includes the description of NTG in terms of its signs and symptoms, unique cases and general patterns, causes, effects, and risk factors, types and distinctions from other similar diseases, diagnosis protocols, treatment methods and their effects. The knowledge also accounts for interactions among these aspects and the ways that the disease affects lives and society.

Of all modes of reasoning used in biomedical research, medical metatheory especially favours problem-solving and decision-making (Connelly & Johnson, 1980; Levene, 1980; see also Kneale, 1949, where they are called, respectively, primary and secondary induction). In addition to inductive reasoning, we identified a significant number of topoi associated with the projected interpersonal relations between the authors and readers (cf. Hyland, 2005).

Each of these three argumentation modes (problem-solving, decision-making, and interpersonal) performs a number of functions. Problem-solving (primary) induction is aimed at the solutions of the specific biomedical problems addressed in the papers. Decision-making (secondary) induction allows the authors to make research decisions, interpret findings, and propose recommendations for future research and clinical practice. Interpersonal argumentation is used to impress the significance of the studies on the readers, to engage them in the integration of the findings into the field's theory, and to make the arguments reader-friendly. The distinctions between the reasoning modes are loosely linked to the conventional IMRD structure of the papers. The methods and results sections mostly deal with problem-solving induction, while introductions



and discussions are dominated by decision-making induction. Interpersonal argumentation also tends to gravitate towards introductions and discussions. The layers of argument created by the three reasoning modes have various degrees of cohesion between them. Interpersonal argumentation is the most autonomous. Problem-solving and decision-making are also fairly independent in their objectives, materials, and results, but one would make little sense without the other.

#### 4.1. *Problem-solving topoi*

Problem-solving, based on primary information, is comparatively straightforward. It follows a set of highly standardised procedures, including statistical analysis and comparison. The relative simplicity of its operations means that, within their methodological frameworks, the authors have little influence on their results (though, of course, they have great influence over the input to such procedures). Such straightforwardness is underscored by rather uniform temporal features of this mode of reasoning. Problem-solving reports on what was done and found during the investigation, and it is overwhelmingly written in the past tense. The problem-solving topoi are divided between the method and results narratives depending whether they deal with methodological or observational content.

##### 4.1.1. *Method narratives*

The method narratives, mostly found in the methods sections, communicate information about the researchers' actions, stipulations, and decisions. The following basic topoi are used to convey these meanings:

COHORT SCREENING  
 INTERVENTIONS  
 INFORMATION  
 DATA HANDLING  
 INSTRUMENTS  
 DATA PROCESSING/ANALYSIS TOOLS  
 STIPULATED CONCEPTS/CLASSIFICATIONS

This part of the argument is constructed as a matter-of-fact account of procedures and techniques most of which are expected to be familiar to the readers. It has highly standardised terminology and few citations. Despite their transparent coding, such narratives play an important part in the community's discourse. Detailed accounts of study designs are present in each paper, which suggests that verifiability and replicability (at least in principle) are highly valued appeals. An average methods section in our corpus is almost as long as an average discussion, twice as long as a results section and thrice that of an introduction.

There is also a conventionalised order to the topoi. Authors first tend to talk about their patients (COHORT SCREENING), then the organisation of the study (INTERVENTIONS and INFORMATION), and, finally, processing and analysis of the data (DATA PROCESSING /ANALYSIS TOOLS and STIPULATED CONCEPTS/CLASSIFICATIONS). The stylistic patterns signalling the topoi are quite well defined.

The COHORT SCREENING<sup>8</sup> topos describes the study participants in general terms:

- 13a. The patients had normal open angles with untreated IOP levels less than or equal to 22 mmHg in both eyes, and showed glaucomatous optic disc changes and reproducible visual field defects with reliable measurements in at least one eye. (E1)
- 13b. Exclusion criteria included history of allergy to fluorescein and a refractive error > -8 dioptres. (G34)

Like all methodological narrative topoi, COHORT SCREENING statements have past tense predicates in the main clause. They are marked up with identity lexis (*normal, reproducible, reliable*) and expressions of circumscription (*less than or equal, at least, >*), generalisation (*both*), or designation (*exclusion, criteria, included*). They also contain numerous expressions and abstractions from the fields of clinical practice, examination, observation, or presence and appearance (*had, open angles, IOP, levels, glaucomatous, optic disc, changes, visual field, defects, measurements, history, allergy, fluorescein, refractive error*). Finally, they have either participant lexis (*patients*) or designation abstractions (*criteria*) in the subject slot.

INTERVENTIONS<sup>9</sup> is the most general of the topoi dealing with the organisation of the study. The writers use it to explain how they treated or managed the disease, how they scheduled and performed examinations, tests, and measurements:

- 14a. Isolated peaks of 26 mmHg were allowed in a diurnal IOP curve without therapy. (E15)
- 14b. The patients underwent CDI measurements of ocular perfusion of the right eye by CDI shortly before and 3–5 weeks after initiation a local therapy with either latanoprost or bimatoprost. Both eye drops were applied once a day between 6 p.m. and 8 p.m. (G10)

This topos is signalled with clinical practice or examination lexis (*therapy, measurements, applied*). Most sentences in this category have past passive verbs; many have numerals and time expressions.

INFORMATION<sup>10</sup> statements are also mostly passive in the main clause. They are used to explain how the authors collected and organised the data:

- 15a. The relationship between the intensity decrease and the intensity variance was examined to determine the difference between the pattern of RNFL loss in the two groups (Figure 3). (E10)
- 15b. For each patient, the relative NRR area was calculated. (G36)

Like the INTERVENTIONS topos, these statements deal with the researchers' actions. Yet here the actions are focused on analytic entities and medical data. So, apart from clinical practice or examination lexis (*examined, relative, calculated*), the recurrent features of INFORMATION statements include analysis lexis (*determine*) combined with participant lexis, the word *group*, or observation abstractions (*relationship, intensity, decrease, variance, difference, pattern, area*). Also typical of this topos are generalisation lexis (*both, each*) or small natural numbers, often with the definite article (*the two*).

INSTRUMENTS<sup>11</sup> is another predominantly passive topos. It usually contains the names of equipment linked to the verb with the word *using* or with the preposition *on* or *with* (*the aid of*). Its lexical features include clinical practice or examination expressions:

- 16a. Stereophotographs of the optic discs were taken with the simultaneous stereo fundus camera (Topcon TRC-SS2), using Kodak Ektachrome 100 HC film. (E1)
- 16b. Visual field examinations were performed with the 24-2 full-threshold program on the Humphrey field analyzer (HFA; Carl Zeiss Meditec, Inc., Dublin, CA). (G5)

The next group of methodological topoi refers to the processing and analysis of data. The DATA PROCESSING/ANALYSIS TOOLS<sup>12</sup> statements continue the list of predominantly passive topoi sharing many features with INSTRUMENTS. Yet where the latter talks about examination and measurement equipment, the former has names of computer software, models, templates, or formulae:

- 17a. Student's *t*-test for paired data was used. (G10)
- 17b. Statistical analyses were performed using SAS/STAT software 8.1. (G6)

The STIPULATED CONCEPTS/CLASSIFICATIONS<sup>13</sup> topos refers to the methodological frameworks, terms, and categories adopted for the studies:

- 18a. The intensity decrease was an index for the diffuse retinal damage, whereas the intensity variance indicated an index for estimating the localized retinal damage ... (E10)
- 18b. Early complications were those occurring in the perioperative or early postoperative period. Late complications were after the initial healing phase had been completed, and were considered to be those seen 3 months or more after surgery. (G35)
- 18c. Altitudinal visual field asymmetry was present in patients who showed different stages of the visual field score (Aulhorn criteria) when comparing the lower and upper hemispheres. (G34)

In differentiating such statements from the rest of past-tense methodological topoi, we found expressions of presence and appearance (*index, indicated, present*) and designation (*criteria, considered to be*) to be the most reliable. Designation may also be expressed grammatically, by means of the compound nominal predicate with the copula *to be*, as in 18c. All main clause predicates in such statements from our corpus have either lexical or grammatical expressions of designation.

#### 4.1.2. *Results narratives*

In the results sections, the authors develop their arguments from ‘raw’ data to problem-solving inferences. Here the authors present their observations in the form of quantified data of various levels of specificity. This content relies on a relatively small number of topoi:

INTERVENTION DATA  
 PARTICIPATION DATA  
 OBSERVATION DATA  
 DEMOGRAPHICS  
 SUMMATED OBSERVATIONS

The outcomes of problem-solving induction drawn from the data presented in the methods and results narratives take the following forms in our corpus:

COMPARISON  
 FOUND ASSOCIATIONS/CORRELATIONS  
 FOUND CAUSES/EFFECTS  
 STATISTICAL SIGNIFICANCE

Continuing from the methods topoi, such statements are written overwhelmingly in the past tense, which identifies their content as strictly local, confined to the study.

The data for the most part refer to the study participants, clinical interventions, and their effects. In our corpus, we identified the following data topoi:

INTERVENTION DATA:

- 19. Twelve patients had less than five visual field examinations after surgery. (G31)

PARTICIPATION DATA:

- 20. None of the patients was lost to follow up. (G36)

## OBSERVATION DATA:

21. Four eyes had central islands, six eyes had defects in one centrocecal area, 16 eyes had extensive arcuate defects. (E15)

## DEMOGRAPHICS:

22. Thirty-four patients (63%) were women. (G36)

## SUMMATED OBSERVATIONS:

23. In the negative control group, all parameters were stable over time (Table 3). (G10)

A common feature for these topoi is quantifiers: numerals, generalisation, and circumscription expressions (e.g. *all*, *at least*, *average*, *maximum*, *none*). (Generalisation expressions are more typical of SUMMATED OBSERVATIONS and less so of OBSERVATION DATA.) All data topoi also typically include participant lexis or the word *group*. Participant lexis usually comes here with numeric attributes. In addition to these common features, three of the data topoi have distinctive lexical features. INTERVENTION DATA includes clinical practice or examination lexis, PARTICIPATION DATA includes designation expressions or topos-specific lexis (e.g. *drop out*, *enrol*, *withdraw*), and DEMOGRAPHICS is also signalled with its own topos-specific lexis (e.g. *age*, *Asians*, *white*, *women*).

In their problem-solving inferences, the writers make comparisons, talk about found links and intervention effects, as well as about the statistical significance of such results:

## COMPARISON:

24. Angle a in the NTG group (35.1 (20.0)°) was significantly smaller than that of the POAG group (45.9 (21.9)°) ( $p=0.02$ ), while angle b in the NTG group (49.0 (31.9)°) was significantly larger than that of the POAG group (33.1 (23.9)°) ( $p=0.01$ ) (Fig 3). (G21)

## FOUND ASSOCIATIONS/CORRELATIONS:

25. As in previous studies . . . , we found that the higher the baseline IOP, the greater the IOP reduction, and that a statistically significant IOP reduction is more likely to occur at pre-treatment IOP levels of over 15 mmHg. (G6)

## FOUND CAUSES/EFFECTS:

26. Interestingly, age of patient had a significant effect on response to latanoprost. (G16)

## STATISTICAL SIGNIFICANCE:

27. There was no significant difference between these figures. (G31)

As in the data topoi above, quantifiers are frequent here, but not essential. Instead, a distinctive feature of two topoi in this group are comparison and juxtaposition. In the COMPARISON topos, the operation of comparison may be expressed with lexicalised expressions, such as *difference*, *similarity*, *compared to* . . . , *by comparison* . . . , or implicitly with comparative and superlative adjectives and adverbs or with comparative syntactic constructions. In addition, typographical comparison symbols, such as  $<$ , are a frequent shorthand for comparison expressions. Comparative degrees are also typical of the FOUND ASSOCIATIONS/CORRELATIONS topos. Here, however, they are used to express distinctiveness, rather than comparison *per se*, so they are interchangeable with other expressions of distinctiveness, such as *increased* or *decreased*. For a complete statement of association or correlation, the distinctiveness is typically complemented with a phrase like *for eyes with* . . . *in patients with* . . . or *patients with* . . . *showed* . . . Another

typical syntactic pattern for FOUND ASSOCIATIONS/CORRELATIONS is shown in example 25: ‘*the + comparative degree + the + comparative degree*’. Both COMPARISON and FOUND ASSOCIATIONS/CORRELATIONS topoi have one lexical feature in common: participant lexis or the word *group*. In FOUND ASSOCIATIONS/CORRELATIONS, however, such expressions are interchangeable with abstractions from the fields of clinical practice, analysis, observation, knowledge, or reasoning. Two other lexical features of this topos are association or enablement and possibility expressions.

The other two topoi from the group of problem-solving inferences, FOUND CAUSES/EFFECTS and STATISTICAL SIGNIFICANCE, have less intricate stylistic profiles than COMPARISON and FOUND ASSOCIATIONS/CORRELATIONS. Apart from the past tense in the main clause, STATISTICAL SIGNIFICANCE also typically includes the lexeme *significan\** and an abstract noun in the subject position. The latter two features may collocate in the abstraction *significance*. Two more abstractions, which may mark up this topos alongside the past tense and *significan\** are *difference* and *correlation*. The FOUND CAUSES/EFFECTS configuration includes cause and effect expressions and clinical practice, observation, or knowledge abstractions.

#### 4.2. Decision-making topoi

Decision-making is more complex than problem-solving, and the degree of the authors’ involvement in the outcomes of this reasoning mode is quite high. In this reasoning mode, the authors interpret their studies in practical terms and generally make them meaningful for their readers. While the generation of primary results is a matter of technique, decision-making induction (Kneale, 1949’s secondary induction) involves numerous choices. To a great extent, these choices account for the theoretical frameworks and methodologies applied to the problem. Other important functions of decision-making are the formulation of study objectives and the interpretation and evaluation of findings. One more important outcome of decision-making induction is higher-level analysis of theory and practice, which cannot be based on primary results alone. To arrive at such interpretations, the authors juxtapose and synthesise their methods and results with relevant information from their sources.

##### 4.2.1. Framing and cohesion in the arguments

Much of the distance from the title to major findings and recommendations is covered with the help of transformation and translation of ideas. In our corpus, these procedures are enacted with the following group of topoi:

OPEN ISSUES  
DISPARITY/UNCERTAINTY  
NEW APPROACHES  
THEME/PURPOSE  
HYPOTHESIS  
RECOMMENDATIONS

Such statements are typically categorised as framing and transition devices. Some of their essential functions are knowledge translation, as well as textual and discursive coherence and cohesion.

The exigencies or opportunities motivating the study are formulated early in introductions and discussions. They may figure as generalisations about the OPEN ISSUES in the domain:

28a. However, the correlation of the AVP to the degree of glaucomatous field damage has not yet been examined. (G34)

- 28b. However, few reports are available upon the long-term IOP-lowering efficacy of latanoprost in NTG. (G6)

This topos is used to point to lacking knowledge or insufficient literature on the issue pursued in the paper. Such lack or insufficiency is typically conveyed with negative diction (*not, few*). The stylistic features of OPEN ISSUES also include data acquisition or epistemic expressions (*examined, reports, unknown*) combined with phenomena and attributes, cause and effect, or association expressions (*correlation, long-term, lowering, efficacy*). Another characteristic feature of such statements is frequent abstract nouns from all lexical groups: clinical practice, argument and discourse organisation, observation, and knowledge (*correlation, AVP, degree, field damage, reports, IOP, efficacy, NTG*).

Like OPEN ISSUES, DISPARITY/UNCERTAINTY<sup>14</sup> statements usually appear in the introduction or at the start of the discussion. They typically point to a contradiction within the field's theory that needs to be resolved:

- 29a. The incidence of this pathology varies considerably in the studies that have been carried out, ranging from 5% of all types of glaucoma for some authors . . . to up to 15% of cases of POAG for other authors . . . (E3)

Alternatively, the authors can point out a contradiction between the theory and the demands of clinical practice:

- 29b. A more important outcome after filtering surgery is the prevention of further visual field deterioration; however, the detection of 'real' progression needs to be differentiated from the inherent 'noise' in visual field testing. (G31)

A distinctive feature of such statements are topos-specific expressions of variation, uncertainty, or incongruence (*varies, ranging*). Their other feature is juxtaposition signals, such as disjunctive and concessive connectives (*however*) and syntactic and lexical cohesive links within the statements (the two affiliated pairs of expressions *5% of all types . . . 15% of cases* and *for some authors . . . for other authors* in 29a and the contrasting pair '*real*' . . . '*noise*' in 29b). Such juxtaposition signals may be combined with disparity, variation, uncertainty, and incongruence lexis or act in its place.

The OPEN ISSUES and DISPARITY/UNCERTAINTY topoi can be seen as a type of implicit negative motivation for research. On the other hand, statements about NEW APPROACHES function as implicit positive motivation:

- 30a. A new concept has been proposed by Davanger, who accounted for the prevalence of NTG on the basis of the overlapping distribution of IOP in population and the pressure vulnerability of the optic nerve head . . . (E10)
- 30b. Latanoprost is a 'new generation' drug that has recently been evaluated as a potential therapy for patients with NTG in short term studies . . . (G16)

Such statements have strong affective appeals coming from lexicalised expressions of novelty or promise (*new, proposed, potential*) or recency and currency adverbials (*recently*), as well as the present perfect or indefinite tense in the main clause.

The THEME/PURPOSE<sup>15</sup> statements move the argument from general exigencies to specific research objectives. They typically occur at the end of introductions between literature reviews and methods sections:

- 31a. This study is aimed at assessing the effects of therapy on POBF and functional parameters in patients with NTG. (E3)



- 31b. In this study, we therefore wanted to investigate a possible coincidence between NTG and progressive sensorineural hearing loss (PSHL) and the association to APSA. (G14)

Such statements contain frequent expressions of purpose or volition (*aimed at, wanted to*), examination, analysis, or research lexis (*study, assessing, investigate*), autoreferential deixis (*this, we*), and various abstract nouns (*effects, parameters, coincidence, association*).

HYPOTHESIS<sup>16</sup> statements are typically viewed as a type of questions that motivate and organise inquiries (Stannard, 1965). This, indeed, is the most typical function of this topos in our corpus. In this role, it is similar to THEME/PURPOSE. In addition, however, HYPOTHESES can be used to draw up tentative conclusions at the end of the text. This topos is often signalled with its metalinguistic tag (e.g. *hypothesis*) but can also be signalled grammatically, by conditional sentences and the subjunctive mood:

32. If the pattern of RNFL loss in NTG has relationship with IOP, the mechanisms of optic nerve damage in NTG might be similar to its in POAG. (E10)

More typical in conclusions than HYPOTHESES are RECOMMENDATIONS,<sup>17</sup> which formulate the practical implications of the research (and may be a distinguishing topos for clinical genres over more purely observational or experimental genres in the same argument domain):

- 33a. In patients with progressive LTG, the normal IOP is relatively too high and a reduction to lower levels by means of filtering surgery is in our opinion indicated to improve the capillary perfusion pressure, resulting in a better oxygenation of the optic nerve head. (E15)
- 33b. Circulatory changes should be considered in the treatment regimen when the cascade of events leading to loss of visual function is most amenable to being interrupted. (G34)

The stylistic configurations of such statements include clinical practice lexis (*LTG, IOP, filtering surgery, etc.*), as well as topos-specific recommendation lexis (*is indicated*) or deontic modality (*should*), along with clinical practice, examination, phenomena and attributes abstractions (*LTG, IOP, reduction, levels, perfusion pressure, oxygenation, changes, etc.*).

A common feature of all framing topoi are their numerous coreference ties with the titles of their papers. These ties are created with terms from the titles, their synonyms, hyponyms, or hypernyms (cf. Halliday, 1994, Ch. 9). For example, every term from the title of the E15 paper ('Results of a Filtering Procedure in Low Tension Glaucoma') is invoked at least once in its RECOMMENDATIONS statement quoted above (*LTG, reduction, filtering surgery, pressure, resulting*).

#### 4.2.2. Secondary information

A big part of decision-making induction, utilising secondary information, is background reviews, mostly incorporated into introductions. Secondary information is also frequently cited in discussion sections. Such content is realised with the following topoi:

- KNOWN CAUSES/EFFECTS
- KNOWN ASSOCIATIONS/CORRELATIONS
- AVAILABLE CONCEPTS/CLASSIFICATIONS
- PREVIOUS FINDINGS
- AVAILABLE TREATMENT/RESEARCH

Statements from this group talk about the relevant state of knowledge and art: data, relations, and patterns reported in the literature, existing theory, treatment methods, and earlier research.

The first three of these topoi refer to well-established theory and practice. The similarity of their content is accentuated by their shared time and aspect features. They typically have present-tense predicates (with simple infinitives, if any) in the main clause:

34. KNOWN CAUSES/EFFECTS

Factors other than IOP also seem to be involved in the development of glaucomatous optic neuropathy in at least some eyes with NTG. (G12)

35. KNOWN ASSOCIATIONS/CORRELATIONS:

The more severe the defect the earlier visual field loss develops in the ‘second eye’. (G36)

36. AVAILABLE CONCEPTS/CLASSIFICATIONS:

Some authors have identified two distinct categories within this nosological form: a non-progressive and a progressive form. (E3)

Aside from the similarities, the distinctive content types of these topoi are apparent from their phrasing and forms. KNOWN CAUSES/EFFECTS includes cause and effect expressions (*factors*) combined with clinical practice, observation, or knowledge abstractions (*development, neuropathy*). KNOWN ASSOCIATIONS/CORRELATIONS often has juxtaposition expressions (*The more severe . . . the earlier*). Alternatively, such statements may contain association, comparison, or congruity and consistency expressions. Another part of their configuration is participant lexis (*eye*) or the noun *group*, as well as clinical practice, observation, or knowledge abstractions (*defect, field loss*). Finally, this configuration features expressions of identity (*distinct, categories, form*) or designation (*identified*).

In contrast to the previous three topoi, PREVIOUS FINDINGS is usually written in past indefinite. This represents its propositional content as unrepeated or atypical, or as especially relevant or interesting for the study. In addition to these time and aspect features, such statements are signalled with frequent bibliographical references<sup>18</sup> or epistemic lexis (*series, reported, contrast*) combined with participant lexis or *group* or expressions of clinical practice, observation, and/or knowledge (*LTG, 40%, progression, change, alter, etc.*):

37a. Of a group of LTG patients, only 40% showed progression in this series. In those patients that progressed, there was a 60% change of repeated periods of progression. (E15)

37b. McKibbin and Menage reported an average POBF increase of 21% in NTG eyes receiving latanoprost . . . (G24)

37c. In contrast, brinzolamide did not to alter ocular perfusion . . . (G10)

The first of these statements refers to observations from earlier research, the second one cites an association, and the third talks of a causal link that an earlier investigation failed to establish.

The last topoi in the secondary information group, AVAILABLE TREATMENT/RESEARCH,<sup>19</sup> provides a summative outlook on trends and developments in the field:

38a. Treatment for NTG has therefore concentrated on lowering IOP. (G31)

38b. In recent years, the understanding of development and progression of glaucomatous optic nerve damage has changed. (G14)

It has present tense verbs in the main clause and/or recency and currency adverbials (*in recent years*). The idea of a birds-eye-view of the domain is expressed with abstract nouns from the fields of clinical practice (*treatment, IOP*), reasoning (*concentrated, understanding*), phenomena and attributes (*development and progression, damage*), or cause and effect (*lowering*).

#### 4.2.3. *Commentary on findings and methods*

Appraisal and interpretation of primary results are indispensable for their synthesis with secondary information. Such synthesis may take various forms:

RESULTS CONSISTENCY  
 METHODOLOGICAL CONSISTENCY  
 QUALIFICATIONS  
 RESULTS RELIABILITY  
 LOCAL FACTORS

These topoi are primarily used in results and discussion sections but also occasionally occur in introductions.

The RESULTS CONSISTENCY<sup>20</sup> topos is used for commentary on how well the study results align with current knowledge:

- 39a. Concerning the postoperative complications, our data are consistent with those of others ... (E15)
- 39b. This is a considerably larger value than has been mentioned in previous reports of disc size in normal and NTG eyes ... (E1)

We can recognise this category by comparison or congruity and consistency expressions (*consistent, larger ... than ...*) combined with participant lexis (*eyes*), clinical practice, phenomena and attributes, or knowledge abstractions (*complications, value, size*) and frequent epistemic expressions.

In our corpus, the authors often comment on the METHODOLOGICAL CONSISTENCY of their studies with the field's practices:

- 40a. Diagnostic criteria for NPG were very similar to that of previous studies ... (G33)
- 40b. Our analysis varies from previous studies in that we looked at IOP updated for each 6 month period postoperatively as a risk factor for visual field progression. (G31)
- 40c. In the present study, we modified the standard for classification of nondippers, dippers, and overdippers, which was adopted in a previous study ... (G5)

This topos shares many features with RESULTS CONSISTENCY. Yet while in the RESULTS CONSISTENCY statements the object of scrutiny is findings, in METHODOLOGICAL CONSISTENCY statements scrutiny shifts to data acquisition (*criteria, analysis, classification*).

QUALIFICATIONS<sup>21</sup> statements allow the authors to focus on their own studies, commenting on their distinctive features, contingencies, benefits, and limitations:

- 41a. Given the retrospective nature of our analysis and the small number of patients investigated, our results must be interpreted with caution. (G36)
- 41b. The influence of the tested compounds on perfusion of the entire eye cannot be answered by the present study. (G10)

By explicating the assumptions and methodological decisions that affected the study results and findings, such statements delimit the scope or force of the conclusions. They may be signalled with topos-specific metalanguage (*caution*), but more often we can recognise them by more subtle signals, such as diminutive or negative diction (*small, cannot*) combined with deontic modality (*must*) or enablement and possibility expressions (*cannot*). As a class of methodological information, such statements are also characterised by data acquisition and epistemic lexis (*retrospective, investigated, interpreted, tested, answered*) combined with participant lexis, the

noun *group*, or abstractions from the fields of data acquisition, argument and discourse organisation, or phenomena and attributes (*nature, analysis, number, results, influence, perfusion, study*). Their self-reflexive nature is highlighted with autoreferential deixis (*our, present*).

Many authors go beyond simple commentary on results and methods. They also use statistical methods and combine multiple analytic methods to probe into their RESULTS RELIABILITY<sup>22</sup>:

- 42a. After adjusting for the factor of IOP, the increase of mean POBF associated with both regimens no longer reached statistical significance ( $p = 0.424$  and  $p = 0.345$ , respectively). To avoid the possible effects of systemic cardiovascular medication on POBF, data were further analysed after excluding patients with such medications. The results remained similar. (G24)
- 42b. Age, gender, ocular laterality, lens status, and the time of day of IOP measurement were evaluated as potential confounding factors, but were not found to be significantly associated with IOP reduction. (G6)

This topos often takes the form of multi-clause sequences explaining, in the past tense, how the authors tested their results for possible errors and artefacts. Such probing may involve adjusting analysis parameters and values or inquiring into received procedures. This is reflected in the topos-specific lexis, such as *adjust for ... factor, reach statistical significance, avoid ... effects, evaluate ... confounding factors, significantly*. The probing may also involve comparison of results obtained by various methods. This may be signalled either with concession/disjunction expressions (*but*) or with aspect diction (*no longer, further, remained*), which is often combined with comparison or congruity and consistency expressions (*similar*). One more typical feature is frequent clinical practice, data acquisition, observation, or knowledge abstractions (e.g. *IOP, regimens, effects, measurement*).

Apart from getting acknowledged, analysed, and reckoned with, research contingencies can work as a type of evidence. Some authors take them into account when interpreting their results. Such interpretations take the form of LOCAL FACTORS statements:

- 43a. A significant difference was observed between the two eyes because of the selection criteria. (E1)
- 43b. This may be related to the greater visual acuity loss and visual field progression also seen in this group. (G31)

One obvious feature of this topos is the presence of cause and effect or association expressions (*because, related*). The circumstances of the study may also be invoked with explicit topos-specific lexis, such as *selection criteria* and *confounding factors*. In addition, LOCAL FACTORS statements have high instances of participant lexis or the word *group*, clinical practice, data acquisition, phenomena and attributes, or knowledge abstractions.

#### 4.2.4. *Synthesis of primary and secondary information*

The next group of topoi serves the purposes of interpreting primary and secondary observations and drawing up the major outcomes of the study:

EXTRAPOLATIONS  
KEY FINDINGS  
FUTURE RESEARCH

In our corpus, we found most of such statements in discussion sections.

EXTRAPOLATIONS<sup>23</sup> are theoretical inferences where the authors relate their own or their sources' observations to underlying phenomena, causal links, and identities:

- 44a. These findings suggest that there may be two different types of NTG, the affected eyes differing in optic disc size and in other ocular characteristics. (E1)
- 44b. These differences suggest that optic nerve compression by ICA may be one of the possible causes or may be a risk factor for optic nerve damage of NTG in some patients. (G12)

The hedging of these statements shows that the *EXTRAPOLATIONS* do not follow directly from either primary or secondary evidence. Other features of this statement type include present tense verbs throughout, along with reasoning, research lexis or modals in the main clause combined with association, cause and effect, identity, or presence and appearance expressions. Like most *topoi* in their group, *EXTRAPOLATIONS* are also likely to have observation and knowledge abstractions.

As their *KEY FINDINGS*,<sup>24</sup> the authors may choose to present information of any level of abstraction, from plain data to causative inferences:

- 45a. In this study, latanoprost 0.005% administered once daily significantly reduced the IOP in NTG patients, and maintained this IOP reduction for up to 12 months. (G6)
- 45b. In conclusion, a high percentage of patients with NTG had marked nocturnal BP reduction. (G5)

The G6 statement (45a) points to the significance of a precedent (positive and sustained effects of the interventions), while 45b stresses the importance of a discovered association (between *NTG* and *nocturnal BP reduction*). Despite the different subject matter, both statements perform the same argumentative function: tell the readers what the authors consider to be the most important outcomes of their studies. This function is signalled with topos-specific metadiscourse (*in conclusion*) or autoreferential deixis (*in this study*) used at the beginning of the sentence, in addition to the stylistic features of the content that the authors choose to highlight as their key findings.

In their *FUTURE RESEARCH*<sup>25</sup> statements authors appeal to readers to advance the field in specific ways or directions:

- 46a. Hence, our suggestion that NTG may be subgrouped should be confirmed by further investigations on a larger number of subjects. (E1)
- 46b. Whether antihypertensive treatment has beneficial effect on CMF by flattening circadian BP fluctuation or not could be another subject for future research. (G5)
- 46c. Whether an increase in POBF is beneficial to the optic disc remains debatable. (G24)

This topos is written in the present indefinite tense. Its other definitive features are explicit time and aspect expressions (*further*, *future*, *remain*) and deontic modality (*should*), both of which orient it towards the future. *FUTURE RESEARCH* statements are also indexed by epistemic lexis (*suggestion*, *confirmed*, *investigations*, *subject*, *research*, *debatable*) combined with clinical practice, observation, or knowledge expressions (*NTG*, *subgrouped*, *antihypertensive treatment*, *effect*, *CMF*, *flattening*, *circadian BP fluctuation*, *increase*, *POBF*, *beneficial*).

### 4.3. Interpersonal *topoi*

Interpersonal argumentation can be seen as a way of translating the writers' categories into the readers' in a bid to influence their perceptions and behaviours. *Topoi* used in this reasoning mode have comparatively weak links with the problem-solving and decision-making operations. They seem to be brought in for the sake of delivering the results to the readers and engaging the community in their interpretation, dissemination, and integration into current theory and practice. In our study, we followed Spinoza (*Ethics*, 1677/2007) in dividing interpersonal argumentation

into affective and logical dimensions, rather than follow the more popular ethical-emotional-logical division.

#### 4.3.1. *Affective appeals*

Affective argumentation allows the authors to set up their credibility, project their characters, signal their memberships in discourse communities, and express their concern with the patients' and other stakeholders' interests. In our corpus, affective topoi produce these effects by relating arguments to professional and social contexts:

MOTIVATION  
PREVALENCE/INCIDENCE  
RESEARCH ETHICS

The first topoi in this group, MOTIVATION,<sup>26</sup> is used to stress the significance of the studies in terms of research and clinical practice; the second one, PREVALENCE /INCIDENCE, points to the social dimensions of the disease. The RESEARCH ETHICS topoi conveys information on standard research procedures that were followed in the investigation. A typical temporal feature of the former two is the present indefinite tense in the main clause, which marks them up as generalisations drawn from a shared pool of facts:

MOTIVATION:

- 47a. However, the disease of the small vessels supplying the optic disc is till now hardly accessible for direct therapy. (E15)
- 47b. The treatment of progressive NTG represents a therapeutic challenge. (G35)
- 47c. Therefore, POAG patients with unioocular field loss represent an ideal population in which to investigate factors influencing the onset of field loss over a period of time. (G36)

Such statements tend to contain expressions of clinical practice, phenomena and attributes, associations, or cause and effect lexis (*disease, therapy, treatment, NTG, therapeutic, POAG, field loss, factors, influencing, period of time*) combined with analysis, examination, or argument and discourse organisation lexis (*population, investigate*). Another distinctive feature of this topoi is the meaning of evaluation. It tends to have expressions with negative or positive connotations (*hardly, challenge, ideal*) often combined with enablement and possibility expressions (*accessible*). In addition to their suasive function, MOTIVATION topoi also act as framing devices. That is why they tend to have coreference ties with the paper titles.

PREVALENCE/INCIDENCE statements are used for appeals to even broader contexts than MOTIVATION statements, highlighting the social dimension of the disease:

- 48a. In Japan Shiose found a prevalence of NTG of about 2% of residents aged 40 years or older, accounting for about 57% of all types of glaucoma . . . (E3)
- 48b. Normal tension glaucoma (NTG) has a prevalence of 0.6% within white populations and is thought to account for 20–30% of primary open angle glaucoma . . . (G33)

In our corpus, PREVALENCE/INCIDENCE is the only introductory topoi with numeric information. Its other prominent feature is topoi-specific expressions, such as *prevalence* and *populations*. This combination of features gives PREVALENCE/INCIDENCE statements a semblance of secondary information. However, in all but two instances in our corpus their function is to quantify the general relevance of the problem rather than provide evidence for the decision-making reasoning. Like MOTIVATION and other framing topoi, such statements also often have coreference links with the title, or just cite general data regarding *NTG* or *glaucoma*.



The RESEARCH ETHICS topos is used to explain how the researchers controlled their biases and ensured that their study participants' rights were respected:

- 49a. The visual fields were analyzed by an external observer. (E15)
- 49b. The study was approved by the Norwich District ethics committee and all patients underwent informed consent. (G16)

This statement type is marked by topos-specific lexis: either expressions of impartiality (such as *external* or *masked*) or explicit references to organisations or procedures created for the protection of research participant rights (such as *Norwich District ethics committee* and *informed consent*). The local, transient, nature of such information is communicated with the past tense of the main clause verbs.

#### 4.3.2. Logical and mixed appeals

Logical topoi help walk readers through the argument, as it were. The authors use them to make their texts reader-friendly while at the same time managing the readers' responses to the paper content. They do so by providing clarifications and addressing the readers' likely questions:

RELEVANT DETAILS  
 RESEARCH TYPE  
 DATA PRESENTATION  
 RELEVANT LITERATURE/COMPANION PUBLICATIONS  
 MATERIALS

Such statements provide commonplace information on research techniques and clinical procedures. The RELEVANT DETAILS<sup>27</sup> topos communicates methodological and technical background explanations. These may refer to how the values and formulae represent the phenomena, how the data are collected and analysed, what are the rationale for the procedures, the meanings of the terms, or the purposes of the equipment:

- 50a. The change in POBF should exceed the variability resulting from measurement and physiological variation to be attributable to drug effects. (G24)
- 50b. Performing multiple statistical tests in a study necessitates the correction of the p-value to reach a significance level of 0.05. (G10)

Despite their diversity, such statements have three features in common. One is their numerous abstractions from the fields of clinical practice (e.g. *POBF*, *drug*), data acquisition (e.g. *tests*, *correction*, *value*, *significance*), phenomena and attributes, or knowledge (*change*, *variability*, *effects*, *level*). Their subject group uniformly contains a clinical practice or data acquisition abstraction, and the major clause verb is in the present indefinite tense.

The RESEARCH TYPE<sup>28</sup> topos clarifies the study design by specifying the nature of the investigation:

- 51a. This was a retrospective clinical study. (G21)
- 51b. The study was designed as an interventional, randomized, prospective, institutional, single-blinded, controlled, clinical trial. (G10)

It contains research jargon like *retrospective* or *randomized* combined with research abstractions. The few instances of this topos in our corpus are all written in past indefinite.

The next category of logical topoi refers the readers to additional information within the texts, in the literature or the media environment:

## DATA PRESENTATION:

- 52a. The morphometric characteristics of the optic discs are summarised in Table 2. (G36)
- 52b. Figure 2 shows the mean diurnal curves for both randomised groups at baseline and at follow up. (G16)
- 52c. All data are given as mean  $\pm$  standard error of means (SEM). (G10)

## RELEVANT LITERATURE/COMPANION PUBLICATIONS:

- 53a. The mode of progression of LTG patients has recently been described ... (E15)
- 53b. The effect of such a lowering in IOP is to be addressed in a companion paper ... (G35)
- 53c. The method has been presented in detail elsewhere ... (G34)

## MATERIALS:

- 54. Latanoprost (50  $\mu$ g/ml) was obtained from Pharmacia Pfizer (Karlsruhe, Germany) as Xalatan<sup>®</sup>. Further ingredients are benz alkonium chloride, sodium chloride, sodium dihydrogene phosphate 1H<sub>2</sub>O, sodium monohydrogene phosphate, and water. (G10)

Such statements are more detailed than parenthetical remarks, cross-references, and citations but essentially perform the same functions. The first of these topoi, DATA PRESENTATION, mostly has main clause verbs in the present indefinite tense and contains topos-specific lexis (e.g. *figure*, *table*), reasoning or presence and appearance lexis (*summarize*, *show*, *give*), analysis, phenomena and attributes, or knowledge abstractions (e.g. *characteristics*, *baseline*, *data*, *mean*, *standard error of means*). RELEVANT LITERATURE/COMPANION PUBLICATIONS can be identified by bibliographical references combined with research lexis. The MATERIALS topoi in our corpus is signalled by the names of medications and their ingredients, the registered trademark symbol, and the words *ingredients* and *contain*.

One interpersonal topos, METHOD/DESIGN JUSTIFICATION,<sup>29</sup> mixes rhetorical appeals. Such statements explain the benefits of the chosen treatment, research methods, or equipment:

- 55a. This approach was chosen because a direct comparison of glaucomatous visual field defect and corresponding retinal microcirculation is possible. (G34)
- 55b. Analysis using microdensitometry and scanning laser polarimetry has the advantage of evaluating the severity of NFL defects in three dimensional mode by computer system. (G21)

This type of authorial commentary has high frequencies of abstractions related to clinical practice or data acquisition (*approach*, *comparison*, *microdensitometry*, *polarimetry*). The methodological information provided in such statements is always presented in a positive light. The positive assessment, however, is seldom explicit, as the word *advantage*. More typically, it takes the form of enablement and possibility expressions (*possible*) or diction that has positive (or, occasionally, negative) connotations in biomedical research and beyond (*direct comparison*).

## 5. Summary of findings

We have presented forty-four epistemic topoi that form the argumentative superstructure of ophthalmic research papers in our corpus. These topoi are associated with three modes of reasoning in the texts: problem-solving, decision-making, and interpersonal. Problem solving is focused on answering the technical questions formulated for the study. In decision-making, the formulation of the questions, the selection of methods and procedures, and the interpretation of results take

centre stage. Finally, interpersonal warranting includes affect and logos, the former appealing to the readers' sensibilities, and the latter making the arguments reader-friendly and educational.

Our findings bear out the idea that argumentative organization is not signalled with isolated linguistic features but with their configurations. The elements of these configurations are not uniform. In our corpus they include lexico-grammatical and semantic relations, syntax, deixis, and coreference.

## 6. Related work and implications for argumentation

Our work extends naturally and richly into the area of computational argumentation. Grasso (2002) indicates that there are

three possible avenues for research should an AI scholar wish to undertake the task of creating a computational model of rhetorical argumentation (Crosswhite, 2000). The first is the exploitation of the argumentative schemata, of which literature in rhetoric provides a rich repository. The second is the exploitation of the figures of speech, and the ways they influence argumentation. The third is the explicit representation of the audience. (p. 59)

As Grasso (2002) explains, this figurative/topical line of exploration<sup>30</sup> 'relates, speaking in natural language generation terms, to the surface representation of the argumentative text' (p. 59). This focus is both prominent and promising in the field of computational argumentation.<sup>31</sup> Each analytic framework used in the field addresses a certain practical motivation and therefore emphasizes a certain facet of the complex phenomenon of natural argumentation (Liakata, Thompson, Waard, Nawaz, Maat, & Ananiadou, 2012). Some of the most popular frameworks are Toulmin modeling (e.g. Green, Dwight, Navoraphan, & Stadler, 2011), practical reasoning (e.g., Walton, 2009; Reed, 2010; Green, 2012), Rhetorical Structure Theory (e.g., Green, 2010; Green, Dwight, Navoraphan & Stadler, 2011), zoning analysis (e.g. Guo, Silins, Stenius, & Korhonen, 2013; Teufel, 1999), Swales's CARS model (e.g, Teufel, 2010, 2014), and rhetorical argumentation (e.g. Grasso, 2002). Our approach shares with this work the use of structured frameworks based on surface stylistic cues that correlate with specific argumentative meanings.

There is a key difference, however, between the stylistic configurations that we describe here and the more traditional loose collections of stylistic features (in what might be termed, somewhat facetiously, a "bag-of-features" approach). The features of stylistic configurations interact with one another and with their semantic and syntagmatic environments in rich *but regular* ways. We have found many benefits to configuration analysis. One benefit is the small unit size it can 'pick out.' Our model describes epistemic topoi at statement-level, but if necessary configuration analysis can go down to the level of clauses and even phrases. Another benefit is that stylistic configurations allow for analysis of topoi regardless of their location in the text. Our findings confirm that authors arrange their ideas with an eye to the writing conventions in their research fields. Yet for most topoi these conventions translate into fairly loose location predictors, not hard rules. Finally, stylistic configuration analysis allows for comprehensive modelling of argumentation. Rather than focus on one argumentation mode (such as problem-solving or interpersonal reasoning), our superstructure covers three modes: decision-making, problem-solving, and interpersonal reasoning. In this respect our superstructure is somewhat similar to Liakata, Thompson, Waard, Nawaz, Maat, and Ananiadou's 2012 model, which combines three schemes:

- "Core scientific concepts": *Hypothesis, Motivation, Background, Goal, Object-New, Object-New-Advantage, Object-New-Disadvantage, Method-New, Method-New-Advantage, Method-New-Disadvantage, Method-Old, Method-Old-Advantage, Method-Old-Disadvantage, Experiment, Model, Observation, Result, Conclusion*

- “Event Meta-knowledge”: *Investigation, Observation, Analysis, Fact, Method, and Other* (subdivided into three certainty levels and two source categories, which show whether the information comes from the current study or another source)
- “Discourse Segment Types”: *Fact, Hypothesis, Problem, Goal, Method, Result, Implication, Other-Hypothesis, and Regulatory-Hypothesis*.

In capturing these meanings, Liakata and her colleagues relied on a broad range of linguistic ‘clues,’ such as verbal forms and semantic classes, modality markers, deixis, syntactic structures, as well as combinations of these features. For example, they found that in their corpus “experimental goals are often given as a (mostly sentence-initial) clause with a to-infinitive . . .” often preceding a past-tense methods clause (p. 41). Liakata and co-authors established significant correspondences between the three schemes they used for annotating their three-paper corpus. Yet, from what we know, they have not yet developed a unified schema incorporating the full range of meanings and clues that they talk about in their 2012 papers. So it would be frivolous of us to compare our findings with theirs despite the significant similarities between our approaches.

Our future treatment of computational argumentation will complement existing methods of argument mining, such as, for example, Moens, Boiy, Palau, and Reed’s (2007) automatic detection of arguments in legal texts. This approach, which uses stylistic phenomena and Machine Learning algorithms to automatically detect and classify arguments, appears to be a very feasible approach for us to adopt in extending our analysis of epistemic topoi to the computational domain. Our obvious next step on this way will be to formalize our taxonomy for the purposes of annotation and verify it in terms of inter-annotator agreement against a corpus of NTG articles. We hope to recruit domain experts for our annotator team. Our annotated corpus and annotation guidelines will be made publically available for other researchers interested to test or advance our taxonomy, or to adjust it for other publication domains.

We would expect the specific topoi we have found to remain very robust in other NTG arguments, and probably ophthalmic arguments more generally, but some of them will likely be more consistently present than others. The lexical fields should also prove fairly robust, with some variation of specific lexemes across research genres and argument domains. We would also expect that some topoi would be robust across argument fields but within genres (such as clinical trials).

As a phenomenon of human collective reasoning, argumentation is not a simple object to study, and it will not yield to simple computational tools. We have found a tractable conceptual instrument for computational argumentation, combining semantic, structural, and relational attributes. We are confident that this work can add new dimensions to argument mining. For example, more intelligent systems could extract not just basic propositional content from certain parts of documents (such as in the text, in the title, or among keywords) but also from specific statement types (such as the HYPOTHESIS, INTERVENTIONS, or STATISTICAL SIGNIFICANCE statements). More research is needed to realize these possibilities, and such research requires more sophisticated theories of text and discourse organization. Our work also strongly supports research teams of highly trained linguists, computer scientists, and domain experts with an abiding interest in rhetorical argumentation, and we would welcome such collaborations.

### Acknowledgements

This publication was made possible with financial support from the University of Waterloo. The authors gratefully acknowledge John Swales’ and Olga Vechtomova’s feedback to the research materials on which this paper is based. We are also wish to thank *Argument and Computation*’s anonymous reviewers’ rich and helpful commentary on earlier versions of the paper. We are especially grateful for their insights into the current state of argumentation analysis and mining.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes

1. Also refer to the websites of the Computational Models of Natural Argument (CMNA) workshops (<http://www.cmna.info/>) and the Computational Models of Argument (COMMA) conference (<http://www.comma-conf.org/>) for recent developments in computational modelling of argumentation.
2. See Aristotle's *Rhetoric* (1924), especially 1358. There are many debates concerning *topoi* and many interpretations of Aristotle's text, which we will not enter. We are content with the classical insight, aligned with the term *topoi*, that (1) ways of arguing have structural signatures and (2) some of those ways of arguing are 'universal' while others are local to particular argument fields.
3. The term *syntagm* means units of linear organization in text and discourse, such as words, phrases, sentences, or text sections (De Beaugrande, 1997, p. 354).
4. *IMRD* stands for a now popular format of empirical publications, which includes four main sections: introduction, methods, results, and discussion.
5. We believe visual annotation to be a very important tool for text analysis, but do not have sufficient room and scope to include details about this method here. We will publish a justification and illustration of visual annotation elsewhere. In the meantime, Gladkova's dissertation is the best source (2010, pp. 88–93).
6. In all examples the emphasis is ours.
7. To go along with STUDY DESIGN, other composite *topoi* that we identified in our corpus are GENERAL RELEVANCE, STATE OF THE ART, PRESENT SERIES, RESEARCH PROCEDURES, COMPOSITE DATA, COMPOSITE FINDINGS, COMPOSITE COMMENTARY, and DISPARITY/ SIMILARITY ANALYSIS.
8. Cf. Trawiński's (1989) 'preliminary activities' and Liddy's (1991) 'subjects'.
9. Cf. Trawiński's (1989) 'schedule of testing method', 'place where testing was carried out', 'time of testing', and 'specification of procedures employed in testing', and Salager-Meyer's (1994) 'describe the process which led to the obtaining of the data'.
10. Cf. Trawiński's (1989) 'specification of objects used in testing'.
11. Cf. Trawiński's (1989) 'specification of equipment used', 'source of objects', and 'source of equipment'.
12. Cf. Trawiński's (1989) 'model used' and 'data reductions, calculations'.
13. Cf. Aristotle's 'definition' (Huseman, 1994), Trawiński's (1989) 'evaluation criteria used', Liddy's (1991) 'new terms defined', and Swales's (2004) 'definitional clarifications'.
14. Cf. Salager-Meyer's (1994) 'justify the reason for the investigation' and Swales's (2004) 'indicating a gap'.
15. Cf. Trawiński's (1989) 'idea of testing method', Liddy's (1991) 'research questions' and 'research topic', Myers's (1992) 'self-referential introductory statements', and Swales's (2004) 'announcing present research descriptively and/or purposively'.
16. Cf. Liddy's (1991) 'hypothesis' and Swales's (2004) 'presenting research questions or hypotheses'.
17. Cf. Trawiński's (1989) 'possible ways of improving solution', Liddy's (1991) 'practical applications' and Salager-Meyer's (1994) 'make suggestions'.
18. In all examples parenthetical citations are represented with ellipses.
19. Cf. Aristotle's 'existing decisions' (Huseman, 1994) and Liddy's (1991) 'relation to other research'.
20. Cf. Aristotle's 'proportional results' and 'identical results' (Huseman, 1994), Trawiński's (1989) 'comparison with results obtained by other authors', and Thompson's (1993) 'statements citing external consistency'.
21. Cf. Trawiński's (1989) 'evaluation of data completeness' and 'analysis of possible errors' and Thompson's (1993) 'evaluative comments on the quality of experimental data'.
22. Cf. Trawiński's (1989) 'evaluation of data precision' and Liddy's (1991) 'reliability'.
23. Cf. Trawiński's (1989) 'explanation of results obtained'.
24. Cf. Harmsze and Kircz's (1998) 'findings' and Swales's (2004) 'announcing principal outcomes'.

25. Cf. Liddy's (1991) 'future research needs' and Salager-Meyer's (1994) 'propose further questions', as well as Trawiński's (1989) 'new problems encountered during research', Thompson's (1993) 'calls for further research in the results section', and Harmsze and Kircz's (1998) 'new problems'.
26. Cf. Aristotle's 'the expediency or the harmfulness' (*Rhetoric*, I.3.1358b), Trawiński's (1989) 'possible usage areas in practice' and 'possible usage areas in science', Salager-Meyer's (1994) 'motivate the study', and Swales's (2004) 'stating the value of the present research'.
27. Cf. Trawiński's (1989) 'characteristics' content elements.
28. Cf. Salager-Meyer's (1994) 'describe the process of manipulating the data obtained during the experimental stage'.
29. Cf. Aristotle's 'incentives and deterrents' (Huseman, 1994), Trawiński's (1989) 'justification' content elements, and Thompson's (1993) 'justifications for methodological selections'.
30. While our work here is not concerned with figures of speech, but it is entirely consonant with that approach. Fahnestock (1999, 23–24, et passim) argues convincingly that figures of speech epitomize topoi; conversely, topoi are elaborations of the argumentative structures that figures can crystallise. The computational detection and plotting of figures is a fine-grained approach that reveals much (Harris & DiMarco, 2009; Gawryjolek, DiMarco, & Harris 2009), especially in stylistically rich argument discourses like political speeches and opinion pieces. But it also misses larger units of argument structure, particularly in the texts of authors not given to crystalline phrasing. The computational detection and plotting of topoi would operate at a mid-grained level, and should prove especially profitable for scientific and technical argumentation.
31. Based on our review of past CMNA workshops, COMMA conferences, and the First Workshop on Argumentation Mining, held at the 2014 Association for Computational Linguistics Conference.
32. Unlike BASIC TOPOI, which are monads indivisible into other recurrent topoi in our corpus, COMPOSITE TOPOI incorporate the BASIC ones. In this paper, we do not expand on COMPOSITE TOPOI. For a fuller delineation of the superstructure, with definitions and examples, see Gladkova (2010).

## References

- Aristotle. (1924). *Rhetorica* (W. R. Roberts, Trans.). In W. D. Ross (Ed.), *The works of Aristotle* (Vol. IX). Oxford: Clarendon Press. Retrieved from Scholars Portal, University of Waterloo. (Original work published in the 4th cent. BCE.)
- Association for Computational Linguistics. (2014). *Proceedings of the First Workshop on Argumentation Mining* (N. Green, K. Ashley, D. Litman, C. Reed, & V. Walker, Org.). Baltimore, MD. Retrieved from <http://acl2014.org/acl2014/W14-21/index.html>
- Berry, R. (2005). Making the most of metalanguage. *Language Awareness*, 14(1), 3–20.
- Biber, D., Csomay, E., Jones, J. K., & Keck, C. (2004). A corpus linguistics investigation of Vocabulary-based Discourse Units in university registers. In U. Connor & T. A. Upton (Eds.), *Applied corpus linguistics: A multidimensional approach* (pp. 53–72). Amsterdam, NY: Rodopi.
- Brown, J. D. (2004). Research methods for applied linguistics: Scope, characteristics, and standards. In A. Davies & C. Elder (Eds. & Intr.), *The handbook of applied linguistics* (pp. 476–500). Oxford: Blackwell Publishing.
- Channell, J. (1990). Precise and vague quantities in academic writing. In W. Nash (Ed. & Intr.), *The writing scholar: Studies in academic discourse* (pp. 95–117). Newbury Park, CA: Sage.
- Connelly, D. P., & Johnson, P. E. (1980). The medical problem solving process. *Human Pathology*, 11(5), 412–419.
- Crismore, A., & Farnsworth, R. (1990). Metadiscourse in popular and professional science discourse. In W. Nash (Ed.), *The writing scholar: Studies in academic discourse* (pp. 118–136). Newbury Park, CA: Sage.
- Crosswhite, J. (2000). Rhetoric and computation. In C.A. Reed and T. Norman (Eds.), *Symposium on Argument and Computation: position papers*. Retrieved from <http://www.csd.abdn.ac.uk/~tnorman/sac/> (As cited in Grasso, 2002).
- De Beaugrande, R. (1997). *Linguistic theory: The discourse of fundamental works*. New York: Longman Routledge.



- Fahnestock, J. (1999). *Rhetorical figures in scientific argumentation*. New York, NY: Oxford University Press.
- Gawryjolek, J., DiMarco, C., & Harris, R. A. (2009, July). *An annotation tool for automatically detecting rhetorical figures*. Paper presented at CMNA (Computational Models of Natural Argument), 13 July 2009, Pasadena, CA.
- Gladkova, O. (2010). *The identification of epistemic topoi in a corpus of biomedical research articles* (Unpublished dissertation). University of Waterloo. Retrieved from ResearchGate.
- Grasso, F. (2002, 4–6 September). Towards a framework for rhetorical argumentation. In J. Bos, M. E. Foster, & C. Matheson (Eds.), *EDILOG 2002 – Proceedings of the 6th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 53–60), Edinburgh, UK.
- Green, N. L. (2010). Representation of argumentation in text with rhetorical structure theory. *Argumentation*, 24(2), 181–196.
- Green, N. L. (2012). Argumentation and risk communication about genetic testing – Challenges for health-care consumers and implications for computer systems. *Journal of Argumentation in Context*, 1(1), 113–129.
- Green, N., Dwight, R., Navoraphan, K., & Stadler, B. (2011). Natural language generation of biomedical argumentation for lay audiences. *Argument and Computation*, 2(1), 23–50.
- Guo, Y., Silins, I., Stenius, U., & Korhonen, A. (2013). Active learning-based information structure analysis of full scientific articles and two applications for biomedical literature review. *Bioinformatics*, 29, 1440–1447.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Harmsze, F.-A. P., & Kircz, J. G. (1998, 23–25 April). Form and content in the electronic age. In C. S. Nielsen & J. R. Herkert (Eds.), *Socioeconomic dimensions of electronic publishing workshop proceedings: Meeting the needs of the engineering and scientific communities*, Santa Barbara, CA, USA (pp. 43–49). Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Harris, R. A., & DiMarco, C. (2009, 8 April). Constructing a rhetorical figuration ontology. In J. Masthoff & F. Grasso (Chairs), *AISB (Artificial Intelligence and Simulation of Behaviour)*, Edinburgh, Scotland (pp. 47–52). Symposium of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour, Edinburgh.
- Hersh, W. R. (2009). *Information retrieval: A health and biomedical perspective* (3rd ed.). New York, NY: Springer Verlag.
- Huseman, R. C. (1994). Aristotle's system of topics. In E. Schiappa (Ed.), *Landmark essays on classical Greek Rhetoric* (pp. 191–199). Davis, CA: Hermagoras.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.
- Ide, N., & Romary, L. (2004). International standard for a linguistic annotation framework. *Natural Language Engineering*, 10(3/4), 211–225.
- Kneale, W. C. (1949). *Probability and induction*. Oxford: Clarendon Press.
- Kumpf, E. P. (2000). Visual metadiscourse: Designing the considerate text. *Technical Communication Quarterly*, 9(4), 401–424.
- Lazaraton, A. (2002). Quantitative and qualitative approaches to discourse analysis. *Annual Review of Applied Linguistics*, 22, 32–51.
- Levene, R. (1980). Low tension glaucoma: A critical review and new material. *Survey of Ophthalmology*, 24(6), 621–664.
- Liakata, M., Thompson, P., de Waard, A., Nawaz, R., Maat, H. P., & Ananiadou, S. (2012). A three-way perspective on scientific discourse annotation for knowledge extraction. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse* (pp. 37–46). Association for Computational Linguistics.
- Liddy, E. D. (1991). The discourse-level structure of empirical abstracts: An exploratory study. *Information Processing and Management*, 27(1), 55–81.
- Litman, D. J. (1996). Cue phrase classification using machine learning. *Journal of Artificial Intelligence Research*, 5, 53–94.
- Malcolm, L. (1987). What rules govern tense usage in scientific articles? *English for Specific Purposes*, 6, 31–43.

- Moens, M.-F., Boiy, E., Palau, R. M., & Reed, C. (2007, 4–8 June). Automatic detection of arguments in legal texts. *The Eleventh International Conference on AI and Law*, Stanford, CA, USA, 4–8 June, 2007.
- Myers, G. (1992). ‘In this paper we report . . .’: Speech acts and scientific facts. *Journal of Pragmatics*, 17(4), 295–313.
- Paice, C. D. (1990). Constructing literature abstracts by computer: Techniques and prospects. *Information Processing and Management*, 26, 171–186.
- Palau, R. M., & Moens, M. (2009). Argumentation mining: The detection, classification and structure of arguments in text. In C. D. Hafner (Ed.), *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, Barcelona, Spain (pp. 98–107). New York: ACM Press.
- Reed, C. (2010). Walton’s theory of argument and its impact on computational models. In C. Reed and C. W. Tindale (Eds.), *Dialectics, dialogue and argumentation* (pp. 73–84). London: College Publications.
- Saint-Dizier, P. (2012). Processing natural language arguments with the < TextCoop > platform. *Argument and Computation*, 3(1), 49–82.
- Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*, 13(2), 149–170.
- Spinoza, B. (2007). Ethics demonstrated in geometrical order (J. F. Bennett, Trans.). Retrieved from <http://www.earlymoderntexts.com/sp.html> (Original work published in 1677.)
- Stannard, J. (1965). The presocratic origin of explanatory method. *The Philosophical Quarterly*, 15(60), 193–206.
- Stirling, L., Fletcher, J., Mushin, I., & Wales, R. (2001). Representational issues in annotation: Using the Australian map task corpus to relate prosody and discourse structure. *Speech Communication*, 33, 113–134.
- Strong, P. M. (1988). Minor courtesies and macro structures. In P. Drew & A. Wootton (Eds.), *Erving Goffman: Exploring the interaction order* (pp. 228–249). Oxford: Northeastern University Press.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University.
- Swales, J. (2004). *Research genres: Exploration and applications*. Cambridge: Cambridge University.
- Taboada, M. (2009). Implicit and explicit coherence relations. In J. Renkema (Ed.), *Discourse, of course: An overview of research in discourse studies* (pp. 127–140). Amsterdam: John Benjamins.
- Teufel, S. (1999). *Argumentative zoning: Information extraction from scientific articles* (Unpublished dissertation). University of Edinburgh. Retrieved January 20, 2007, from <http://www.cl.cam.ac.uk/~sht25/thesis/t.pdf>
- Teufel, S. (2010). *The Structure of scientific articles: Applications to citation indexing and summarization*. Stanford, CA: CSLI Publications.
- Teufel, S. (2014). Scientific argumentation detection as limited-domain intention recognition. In E. Cabrio, S. Villata, & A. Wyner (Eds.), *ArgNLP 2014: Frontiers and connections between argumentation theory and natural language processing*. Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing. CEUR-WS.
- Thompson, D. K. (1993). Arguing for experimental ‘facts’ in science: A study of research article results sections in biochemistry. *Written Communication*, 10(1), 106–128.
- Thompson, S. E. (2003). Text-structuring metadiscourse, intonation and the signalling of organisation in academic lectures. *Journal of English for Academic Purposes*, 2(1), 5–20.
- Trawiński, B. (1989). A methodology for writing problem-structured abstracts. *Information Processing and Management*, 25(6), 693–702.
- Van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction and cognition*. Hillsdale, NJ: Lawrence Erlbaum.
- Walton, D. (2009). Enthymemes and argumentation schemes in health product ads. *CMNA IX, Computational Models of Natural Argument Workshop*, Pasadena, 13 July 2009
- Wilbur, W. J., Rzhetsky, A., & Shatkay, H. (2006). New directions in biomedical text annotation: Definitions, guidelines and corpus construction. *BMC Bioinformatics*, 7, 356–365.
- Willard, C. (1989). *A theory of argumentation*. Tuscaloosa: University of Alabama Press.
- Wyner, A., Mochales-Palau, R., Moens, M., & Milward, D. (2010). Approaches to text mining arguments from legal cases. In E. Francesconi, S. Montemagni, W. Peters, & D. Tiscornia (Eds.), *Semantic processing of legal texts* (pp. 60–79). No. 6036 in Lecture notes in Computer Science. Berlin: Springer.

## Appendix 1. Annotated NTG corpus

- Ang, A., Reddy, M. A., Shepstone, L., & Broadway, D. C. (2004). Long term effect of latanoprost on intraocular pressure in normal tension glaucoma. *British Journal of Ophthalmology*, 88(5), 30–34.
- Choi, J., Jeong, J., Cho, H. S., & Kook, M. S. (2006). Effect of nocturnal blood pressure reduction on circadian fluctuation of mean ocular perfusion pressure: A risk factor for normal tension glaucoma. *Investigative Ophthalmology & Visual Science*, 47(3), 831–836.
- De Jong, N., Greve, E. L., Hoyng, P. F., & Geijssen, H. C. (1989). Results of a filtering procedure in low tension glaucoma. *International Ophthalmology*, 13(1–2):131–138.
- Fontana, L., Armas, R., Garway-Heath, D. F., Bunce, C. V., Poinoosawmy, D., & Hitchings, R. A. (1999). Clinical factors influencing the visual prognosis of the fellow eyes of normal tension glaucoma patients with unilateral field loss. *British Journal of Ophthalmology*, 83(9), 1002–1005.
- Kremmer, S., Kreuzfelder, E., Bachor, E., Jahnke, K., Selbach, J. M., & Seidahmadi, S. (2004). Coincidence of normal tension glaucoma, progressive sensorineural hearing loss, and elevated antiphosphatidylserine antibodies. *British Journal of Ophthalmology*, 88(10), 1259–1262.
- Liu, C. J., Ko, Y. C., Cheng, C. Y., Chou, J. C., Hsu, W. M., & Liu, J. H. (2002). Effect of latanoprost 0.005% and brimonidine tartrate 0.2% on pulsatile ocular blood flow in normal tension glaucoma. *British Journal of Ophthalmology*, 86(11), 1236–1239.
- Membrey, W. L., Bunce, C., Poinoosawmy, D. P., Fitzke, F. W., & Hitchings, R. A. (2001). Glaucoma surgery with or without adjunctive antiproliferatives in normal tension glaucoma: 2 Visual field progression. *British Journal of Ophthalmology*, 85(6), 696–701.
- Membrey, W. L., Poinoosawmy, D. P., Bunce, C., Fitzke, F. W., & Hitchings, R. A. (2000). Comparison of visual field progression in patients with normal pressure glaucoma between eyes with and without visual field loss that threatens fixation. *British Journal of Ophthalmology*, 84(10), 1154–1158.
- Membrey, W. L., Poinoosawmy, D. P., Bunce, C., & Hitchings, R. A. (2000). Glaucoma surgery with or without adjunctive antiproliferatives in normal tension glaucoma: 1 intraocular pressure control and complications. *British Journal of Ophthalmology*, 84(6): 586–590.
- Ogata, N., Imaizumi, M., Kurokawa, H., Arichi, M., & Matsumura, M. (2005). Optic nerve compression by normal carotid artery in patients with normal tension glaucoma. *British Journal of Ophthalmology*, 89(2), 174–179.
- Oh, J. Y., & Park, K. H. (2005). The effect of latanoprost on intraocular pressure during 12 months of treatment for normal-tension glaucoma. *Korean Journal of Ophthalmology*, 19(4), 297–301.
- Ravalico, G., Pastori, G., Toffoli, G., & Croce, M. (1994). Visual and blood flow responses in low-tension glaucoma. *Survey of Ophthalmology*, 38(S1), 73–76.
- Tomita, G., Nyman, K., Raitta, C., & Kawamura, M. (1994). Interocular asymmetry of optic disc size and its relevance to visual field loss in normal-tension glaucoma. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 232(5), 290–296.
- Woo, S. J., Park, K. H., & Kim, D. M. (2003). Comparison of localised nerve fibre layer defects in normal tension glaucoma and primary open angle glaucoma. *British Journal of Ophthalmology*, 87(6), 695–698.
- Yamazaki, Y., Koide, C., Takahashi, F., & Yamada, H. (1992). Diffuse nerve fiber layer loss in normal tension glaucoma. *International Ophthalmology*, 16(4–5), 247–250.
- Zeitl, O., Matthiessen, E. T., Reuss, J., Wiermann, A., Wagenfeld, L., Galambos, P., Richard, G., & Klemm, M. (2005). Effects of glaucoma drugs on ocular hemodynamics in normal tension glaucoma: a randomized trial comparing bimatoprost and latanoprost with dorzolamide. *BMC Ophthalmology*, 5(6).

## Appendix 2. Abbreviations

APSA	Antiphosphatidylserine antibodies
AVP	Arteriovenous passage
BP	Blood pressure
BMC	BioMed Central
CARS	Create a Research Space model (Swales, 1990)
CMF	Circadian mean ocular perfusion pressure fluctuation
CDI	Colour Doppler imaging
HTG	High-tension glaucoma
ICA	Intracranial carotid artery

IMRD	The textual structure including an introduction, methods, results, and discussion sections
IOP	Intraocular pressure
MEDLINE	Medical Literature Analysis and Retrieval System Online
MOPP	Mean ocular perfusion pressure
NFL	Nerve fibre layer
NTG/NPG/LTG	Low/normal pressure/tension glaucoma
POAG	Primary open-angle glaucoma
POBF	Pulsatile ocular blood flow
PSHL	Progressive sensorineural hearing loss
RNFL	Retinal nerve fibre layer

**Appendix 3. Epistemic topoi in our corpus**

BASIC TOPOI fall into argumentation categories (like problem-solving and decision-making), within which they cluster into smaller topoi classes, like *METHOD* and *RESULTS NARRATIVES*.

**BASIC TOPOI**

---

Problem-solving argumentation

---

<i>METHOD NARRATIVES</i>	<i>RESULTS NARRATIVES</i>	<i>OUTCOMES</i>
--------------------------	---------------------------	-----------------

---

COHORT SCREENING	INTERVENTION DATA	COMPARISON
INTERVENTIONS	PARTICIPATION DATA	FOUND ASSOCIATIONS/CORRELATIONS
INFORMATION	OBSERVATION DATA	FOUND CAUSES/EFFECTS
DATA HANDLING	DEMOGRAPHICS	STATISTICAL SIGNIFICANCE
INSTRUMENTS	SUMMATED OBSERVATIONS	
DATA PROCESSING/ANALYSIS TOOLS		
STIPULATED CONCEPTS/CLASSIFICATIONS		

---



---

Decision-making argumentation

---

<i>FRAMING &amp; COHESION</i>	<i>SECONDARY INFORMATION</i>	<i>COMMENTARY</i>
-------------------------------	------------------------------	-------------------

---

OPEN ISSUES	KNOWN CAUSES/EFFECTS	RESULTS CONSISTENCY
DISPARITY/UNCERTAINTY	KNOWN ASSOCIATIONS/CORRELATIONS	METHODOLOGICAL CONSISTENCY
NEW APPROACHES	AVAILABLE CONCEPTS/CLASSIFICATIONS	NEW/REMAINING ISSUES
THEME/PURPOSE	PREVIOUS FINDINGS	QUALIFICATIONS
HYPOTHESIS	AVAILABLE TREATMENT/RESEARCH	RESULTS RELIABILITY
RECOMMENDATIONS		LOCAL FACTORS

---

*SYNTHESIS OF PRIMARY AND SECONDARY INFORMATION*

---

KEY FINDINGS		
FUTURE RESEARCH		
EXTRAPOLATIONS		

---



---

Interpersonal argumentation

---

<i>AFFECTIVE APPEALS</i>	<i>LOGICAL &amp; MIXED APPEALS</i>
--------------------------	------------------------------------

---

MOTIVATION	RELEVANT DETAILS
PREVALENCE /INCIDENCE	RESEARCH TYPE
RESEARCH ETHICS	DATA PRESENTATION
	RELEVANT LITERATURE/COMPANION PUBLICATIONS
	MATERIALS

---

COMPOSITE TOPOI<sup>32</sup>

---

GENERAL RELEVANCE  
STATE OF THE ART  
PRESENT SERIES  
STUDY DESIGN  
RESEARCH PROCEDURES  
COMPOSITE DATA  
COMPOSITE FINDINGS  
COMPOSITE COMMENTARY  
DISPARITY/SIMILARITY ANALYSIS

---