UNIVERSITY OF ALBERTA

Contextual Associations of Thinking Verbs:

A Corpus-Based Investigation of English

BY

Mildred Lau

A THESIS

SUBMITTED TO THE FACULTY OF ARTS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

BACHELOR OF ARTS

DEPARTMENT OF LINGUISTICS

EDMONTON, ALBERTA

April 2013

Revised Summer 2013

UNIVERSITY OF ALBERTA

FACULTY OF ARTS

The undersigned certify that they have read and recommend to the Faculty of Arts for acceptance, a thesis entitled **Contextual Associations of Thinking Verbs: A Corpus-Based Investigation of English**, submitted by Mildred Lau in partial fulfillment of the requirements for the degree of Bachelor of Arts.

…….

Dr. Antti Arppe

…….

For the Department

In memory of my grandmother

鍾少英
Chung Siu-Ying

1911 — 2013

ABSTRACT

In this study, the linguistic contexts of the English verb THINK and its near-synonyms CONSIDER, REFLECT, and PONDER will be investigated from a corpus linguistic perspective in order to determine what, if any, contextual factors are preferentially associated with the usage of each verb. Underlying this approach is Harris' (1954) distributional hypothesis that words with similar contexts have similar meanings, and its corollary that differences in contexts suggest differences in meaning. The near-synonyms are selected on the basis of their frequency in the British National Corpus and their dictionary meaning overlap within words expressing the activity of *thinking*.

Data extracted from the British National Corpus is annotated for morphological, syntactic, and semantic factors following the behavioural profiling principles compiled by Divjak & Gries (2006), and a statistical analysis, following Arppe (2008), is performed on the annotated information to reveal the contextual associations of the verbs. These results are used to construct semantic characterizations of the four verbs and the characterizations are then compared cross-linguistically to a selection of verbs meaning *think* in Finnish, which have been previously studied by Arppe (2008). More generally, the results provide evidence that statistical methods can be used to better understand in what ways a semantic field like *thinking* is lexically divided similarly or differently across languages.

ACKNOWLEDGEMENTS

"'Tulin, näin ja hävisin, mutta siksi koska itse tahdoin niin.
Kiitos teille, rakkaat, että elää sain, kuten halusin.'"

'I came, I saw, and I lost—but that is because I wanted it to be so.
It is thanks to you, my loved ones, that I got to live how I wanted to.'

— Valtteri Tynkkynen, "Vastavirtaan"

First, I would like to thank Dr. Antti Arppe, my supervisor, for the focus, inspiration, and advice throughout this project. If it was not for you I would not even know what to study. Also, thank you for the *ylioppilaslakki* ('for entertainment purposes only').

Second, I want to thank Dr. John Newman, for assistance with using the corpus tools and teaching me what corpus linguistics is all about.

I would like also to express my appreciation to the following people, in semi-alphabetical order, because you all played important roles in helping me see this project to the end:

Philip Dilts, for revealing to me the universe of linguistics.

Tiina Hohn, for your compassion and support in the face of injustice.

Kari Karsten, for your (mostly) patient and loving companionship.

Diane McKen, for believing in me all the way.

Dr. Sally Rice, interim Chair of the Department, for lending me a desk in the Masters students cluster—this entire thesis was written on the computers in that room.

Dr. Anne-Michelle Tessier, for urging me to face the tough questions I needed to ask myself.

Claudia Heinrich, Kailen Shantz, Yasemin Tulpar, and Richard Yanaky, for the socializing needed to clear my head after hours of panicking (sorry about the noise level!).

The rest of the Canada contingent: Anna Chen and the Graduate Recital Choir, Rebecca Cooper and the Undergrad Linguistics Club, Eva Guenther, Lilla Lesko and my coworkers at Rutherford Library, Daniel Lim, Danielle Klassen… I hope I didn't miss anybody.

(The) *Ulkomaalaiset*: Jari 'Huijari' Huilla, Antti Kangas, Samuel Kohonen, Norman Langerak, Patrick Newman, Sarah Pasay, Julia Simon, Sami 'Alajärvelä' Singh, Dr. Amy Wiles.

And finally to my immediate family, without whose love, patience, and material support I would never have come this far. I'm sorry that it took so long.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER I. Introduction

1.1 What is Synonymy?

Meaning is context-driven and context-dependent, or, as famously stated by Firth, "You shall know a word by the company that it keeps" (1957:11). The precision and subtlety of a word's meaning is made up of the sum total of the many contexts that it "participates" in, creating a set of patterns that uniquely pinpoint that word out of a field of candidate words (Hanks 1996). Partington (1998:27) asserts that "every lexical item in the language has its own individual and unique pattern of behaviour." Nowhere is this clearer in the lexicon than in the study of synonymy, for which the distributional hypothesis proposed by Harris (1954) posits that words occurring in similar contexts tend to have similar meanings.

Synonymy is an area in which the interests of lexical semanticists, computational linguists, and lexicographers meet (Divjak & Gries 2006). Despite how fundamental synonymy seems to be to the lexicon, polysemy appears to be much better studied. Edmonds & Hirst (2002) suggest that one reason for this is because researchers treat synonymy as a "non-problem": either synonyms are exactly the same in meaning (absolute synonyms), so there is little to say about them; or that there is no such thing as synonymy, so each word may be dealt with equally as individual items.

Most empirical research on synonymy has been motivated by automatic synonym identification and thesaurus generation, as stated by Divjak & Gries (2006), and these and lexical taxonomies designed for computational use such as WordNet (Miller et al. 1990) tend to treat semantically similar items as absolute synonyms (Edmonds & Hirst 2002). Lexicographers, however, have been acutely aware that in practice, absolute synonymy does not exist; what philosophers call and treat as *synonymy* is, in linguistics and practical usage, actually likeness of meaning rather than identity of meaning. What lexicographers are concerned about is: how much should meanings of two words overlap before the words can be considered sufficiently synonymous (Edmonds & Hirst 2002), that they should be mentioned in the same dictionary or thesaurus entry? And which parts of those meanings are permitted to differ while retaining synonymy? Individual lexicographers frequently differ on these measures (see Church et al. 1994 and Divjak & Gries 2006 for examples from English and Russian, respectively). Miller & Charles (1991) recognize that a semantic similarity continuum exists rather than a synonym / non-synonym dichotomy.

Is synonymy a relation between only two words? Traditional definitions of it are frequently phrased in terms of binaries. Synonym studies do tend to consider two words at a time, as though binary synonymy is the norm, but this is due to the relative ease of setting up tests for synonymy (Arppe 2008:11). As Murphy (2003:160) notes, "this is a limitation of the testing process, not a limit to the number of synonyms a word can have." Researchers have advocated studying synonym groups of more than two words in order to arrive at more complete semantic and contextual description (Atkins & Levin 1995; Divjak & Gries 2006; Arppe 2008:11; Gries & Otani 2010).

1.2 Judging Synonymy

The importance of meaning distinction between words having been established, we must define what (*near*-)*synonymy* actually means. Most philosophically-informed accounts of synonymy focus on similarity (Murphy 2003:141 collects a number of these), consistent with the distributional hypothesis. Cruse asserts that "synonyms must not only manifest a high degree of semantic overlap, they must also have a low degree of implicit contrastiveness" (1986:266). That is, near-synonyms are characterized by the differences between them. These differences may "involve concepts that relate roles and aspects of the situation" at hand (Edmonds & Hirst 2002), and may also include linguistic contexts. In contrast, antonym pairs, which have "negative similarity" or high contrastiveness, seem frequently interchangeable but experimental evidence suggests that differences in context between antonyms are actually greater than between synonyms (Miller & Charles 1991).

The tendency of philosophical semanticists to consider synonymy in terms of linguistic expressions that have the same meaning or truth conditions when one word is substituted by its synonym (Cruse 1986:88; Miller & Charles 1991), a tradition originating from Leibnizian philosophy, suggests that synonymy relations are external to the lexemes in question and bound to the context (the remainder of the expression) instead. Even so-called "neutral contexts" such as thesauri or synonym dictionaries are themselves contexts, which inescapably affect human judgments of semantic similarity (Murphy 2003:138).

Experimental evidence shows that language users have knowledge about collocation patterns that help them in choosing between near-synonyms (Miller & Charles 1991; Arppe & Järvikivi 2007; Dąbrowska 2009). These judgments invoke *contextual representations*, cognitive structures posited by Miller & Charles (1991) to consist of information regarding the conditions in which a lexeme may appropriately be used, including collocational, syntactic, semantic, pragmatic and stylistic considerations.

The underlying question behind determining whether a set of lexemes are synonymous is then: how can we determine the circumstances under which the lexemes can be interchanged without substantial change to meaning (Miller et al. 1990; Partington 1998:33)? An early attempt by Rubenstein & Goodenough (1965, cited in Miller & Charles 1991) to test the distributional hypothesis using subject-generated sentences, ran on the supposition that contextual similarity could be based on co-occurrence of words in the lexemes' contexts. However, the results could not comfortably account for anything less than high degrees of synonymy. But since context consists of the components included in the aforementioned *contextual representations*, not only co-occurring words, we can add constructional similarity or overlap to the equation and use collocation and constructional similarity as a proxy for semantic distance in order to determine sense relations like synonymy and antonymy (Church et al. 1994; Partington 1998:33; Murphy 2003:159; Divjak & Gries 2006; Mohammed & Hirst 2006).

1.3 Synonymy and Corpus Lexicography

One of the shortcomings of the traditional lexicographic approach of gathering citations and using introspection to compose word definitions and senses is that lexicographers historically

have likely not had sufficient evidence to confidently make generalizations about context features associated with word senses (Hanks 1996)—or indeed, between synonyms—and have had to resort to using *sometimes* or *usually* in their definitions to hedge claims about usage contexts (Edmonds & Hirst 2002). The advent of computing technology, access to massive amounts of digitized data, and the creation and curating of electronic linguistic corpora has been exploited by lexicographers (for example, the COBUILD project founded by John Sinclair) and linguists alike.

Early corpus studies of synonymy were simple and focused primarily on differences in collocations within a limited "window" of words surrounding the lexeme under investigation, for example Church et al.'s investigation of ASK *(for),* REQUEST, and DEMAND (1994), Biber et al.'s look at BIG, LARGE, and GREAT (1998:43–53), and Partington's study of SHEER, PURE, COMPLETE, and ABSOLUTE (1998:33–46). (The genesis of corpus linguistics predates these examples by several decades but they are representative of the sort of work.)

With the recognition that near-synonyms differ in syntactic context in addition to lexical collocation (Atkins & Levin 1995; Divjak & Gries 2006) came the realization that the corpus data needs to be analyzed at the sentence or clause level, rather than some arbitrary window, in order to account for structural considerations in other languages which do not have the relatively fixed word order of English. Edmonds & Hirst (2002) observe that structural variations between near-synonyms are a result of restrictions on their use that come from other parts of the clause or utterance and vice versa, and that these restrictions are not part of the meanings denoted by the words themselves.

Such structural considerations can comprise many factors. They can be collocational, syntactic, or even dependent on other words or concepts present in the utterance: Atkins & Levin (1995) finds differences in the semantic classes of nouns in the subject position of a number of verbs meaning SHAKE, QUAKE or QUIVER; Biber et al. (1998:95–100) finds differences in the preferred transitivity, types of agents and adverbial modifiers between BEGIN and START; Divjak & Gries (2006) find that Russian verbs for *try* do not prefer identical sets of adverbs; Arppe (2002) finds differing preferences for inflectional morphology in two Finnish verbs for *think*; just to list a few.

Studies of synonymy such as the above tend to look at the lexical (semantic) level or the morphosyntactic level, but not so often both levels together (Arppe 2008:11; Gries 2010). The areas of morphological, syntactic, and semantic annotation come together in the contextual profiling approach. This approach is the combination of the ID tagging of data (Atkins 1987, cited in Divjak & Gries, 2006) and their use to form a "behavioural profile" of a word (Hanks 1996). Its procedural principles are articulated by Divjak & Gries (2006) as follows:

1. Manual annotation of the data set is necessary for precision in analysis. Automated tagging is good to start with, but it should be manually checked.
2. By restricting the annotation (mostly) to objectively identifiable properties consistently, intuition by the researchers will not creep into the analysis.
3. Annotating the data for a large number and variety of properties increases the comprehensiveness of the analysis.

In order to analyze corpus data in this way, it is necessary to devise some kind of systematic and objective method to annotate each concordance line under scrutiny. The ambiguity and vagueness of the differences between synonyms mean that methods used for studying polysemy are not sufficient (Divjak & Gries 2006). As for morphological annotation, part-of-speech and inflectional tagging can be automated by use of a parser. Likewise, syntactic annotation may be computer-assisted. But by and large, semantic annotation must be performed manually—for if we have already solved the problem of automatic semantic tagging, corpus study of semantic relations would be obsolete!

Hanks (1996) recognizes that, to analyze semantics of verbs, groupings of typical subjects, objects, and adverbials need to be generated and individual lexical items categorized into them. These categories may be determined organically based on intuition as they emerge (Hanks 1996), or they may be based on semantic taxonomies like semantic primitives (Divjak & Gries 2006) and WordNet (Arppe 2008:31).

Divjak & Gries (2006) use the annotated data set to generate a summary of the frequency of each tag value for each studied lexeme across the entire data set, and then use these frequencies for a clustering analysis that lumps their nine words for *try* into subgroups. Arppe (2008) takes a statistical modelling approach to the tag frequency data, using multivariate methods such as polytomous logistic regression to discover relationships between contextual feature variables and lexical choice, and to compute the probability that a particular lexeme out of the synonym set will be selected given a number of contextual features.

1.4 Possibilities and Limitations of Corpus Lexicography

One variable not yet considered in conjunction with the aforementioned profiling and statistical analysis or modelling approach is that of speaker or writer genre, register and style (Divjak & Gries 2006). Arppe 2008 considers genre to a limited extent, his research corpus consisting of newspaper and Internet discussion group texts. It would be worthwhile to consider and study further the usage differences among various genres, registers, and styles. Fortunately, the approach is expandable to include this or to even study diachronic changes. Its objectivity and consistent annotation scheme allows comparative study across varieties of a language or even across entire languages.

Translators would find the results of cross-linguistic synonym analysis of use to assist them in choosing the word with the best translation equivalence in their work (Partington 1998:51). Determining and quantifying the differences of usage contexts and meanings within synonym sets is also valuable in machine translation (Edmonds & Hirst 2002). If only the process could be automated and contextual feature description and preferences be easily generated for many words, along with automatic selection of corpus sentences containing the most prototypical usages of each, the lexicographer's task would be greatly simplified and dictionaries more representative of actual language use (Arppe 2008:258).

One must keep in mind, though, that corpora cannot write dictionaries by themselves, as they are but collections of words used in their natural contexts. The existence and use of corpora does not change the fact that word meanings and the relations between them need to be inferred from the

distribution of the words' formal elements, the latter being actually possible to extract from corpus data (Gries & Otani 2010).

An inconvenience of relying on distributional data is the fact that distributional information can also bring up evidence of relations other than synonymy (Church et al. 1994). Such semantic relations may be antonymy (high lexical contrast), hyponymy (type to token, or as Murphy (2003:229) calls it, *asymmetrical lexical contrast*), meronymy (part to whole), troponymy (relation by manner (Fellbaum 1990)), or entailment. That this occurs does not violate the distributional hypothesis; it is only that the similarities of meaning lie deeper than the mere similarity of denotation. This is another reason why human intervention is necessary to hone in on the lexemes fitting the desired relationship.

More importantly, word meanings and the differences between semantically similar words are fuzzy. Corpus analysis can suggest to us the contextual features that predispose the typical usage of a word among a selected synonym group, but cannot tell us when it *will* be used, for at least three reasons: 1. Corpus analysis only gives us *distributional* information (Church et al. 1994; Hanks 1996; Edmonds & Hirst 2002) from which we must infer meaning, per the distributional hypothesis; 2. No contextual feature or set of them can categorically characterize even corpus instances of any of those words (Arppe 2008:248); because 3. Use of language is human and infinitely innovative.

1.5 The Present Study

This study uses a corpus-driven approach to investigate the contextual semantic and morphosyntactic factors associated with a selection of English near-synonymous verbs expressing the act of thinking or cogitation, namely THINK, CONSIDER, REFLECT, and PONDER. The data will be prepared following Divjak & Gries' (2006) guidelines for the contextual profiling process and analyzed statistically via the method developed by Arppe (2008).

While the present study will be limited to the univariate analysis of the contextual variables surrounding the selected thinking verbs, it is hoped that the data compiled will form a basis for bivariate and multivariate analyses in the near future. This study is intended as the basis for further work that is, in essence, a replication of Arppe's (2008) study of Finnish thinking verbs, and, once completed, should result in data that may be compared cross-linguistically.

CHAPTER II. Data Extraction and Preparation

2.1 The Corpus

The British National Corpus (BNC) consists of approximately 100 million words of written and spoken British English, from a range of language use, mostly from the years 1985–1993 (Gries & Newman, in press). About 10.3% of this corpus represents transcribed speech—either conversational or non-conversational, such as lectures—and the remainder represents written language—mainly from published sources but also some unpublished texts such as letters and student essays. The great variety of types of text sampled in the BNC has made it one of the most heavily used and studied corpora, and this is also because it can be publicly accessed by using at least three Internet-based interfaces (Lee 2010).

In an age during which one prevailing philosophy is "bigger is better," why not use a 'mega-corpus,' such as the Corpus of Contemporary American English (COCA), containing 450 million words (and counting)? I selected the BNC not only for its ease of use via the BNCweb interface, but also because of its broader coverage of genres, media, and styles. In particular, COCA's sources do not include those of a private or unpublished nature—the private conversations, letters, and schoolwork that can grant insight into how the language is used when the speakers are not under the public eye (COCA's unscripted conversation section is based on transcripts from TV and radio programming).

2.2 Selection of Lexemes

In order to determine which near-synonyms of THINK were to be considered in this study, first, the online interface to the BNC by Mark Davies <http://byu.corpus.edu/bnc> was used to search for synonyms of THINK. After excluding from the results those verbs which are synonyms for THINK in "opinion" or "epistemic" contexts (the former, especially, is peculiar to English, *q.v.* Goddard 2003; Goddard & Karlsson 2008), the verbs in Table 1 remain. Each verb is accompanied by the number of tokens in the BNC (as calculated by BNCweb) and an estimated proportion of these used in the appropriate sense of cogitation based on random samples of 100 concordance lines containing that verb (for those verbs with polysemy).

Table 1. Frequency of thinking verbs in the BNC

Verb	**Frequency in BNC**	**'Cogitation' occurrences in 100 line sample**	**Adjusted frequency**
THINK	145438	29	42177
CONSIDER	28603	69	19736
REFLECT	11114	19	2112
CONTEMPLATE	1607	100	1607
PONDER	634	100	634
REASON	711	61	433
MEDITATE	161	62	100
DELIBERATE	182	54	98
RUMINATE	48	44 (out of 48 lines)	44
COGITATE	6	6 (out of 6 lines)	6

Because the Oxford English Dictionary does not use a standardized subset of vocabulary for writing its definitions, it is useful to consider the words used by its authors to define the above thinking verbs as a proxy measure of meaning overlap between them. If one thinking verb shares a majority of the key terms in its definition with another thinking verb, the two thinking verbs can be considered to be more overlapping in meaning than two thinking verbs that do not share words in their definitions to the same extent.

The relevant senses in the OED Online (2013) definitions for the above verbs were then selected, and the key terms contained in the definitions were extracted. (For the original definitions, see Appendix A.) Each possible pair of verbs was compared and the number of key terms they have in common (including the headword itself, if applicable) were counted. Each headword's score across all pairings was averaged and divided by the total number of key terms in its definition plus itself, in order to normalize for the differing numbers of key terms in each word definition. The resulting figure, which I shall call the *co-definition index* (CDI), is a rating of how well the headword's definition (and thus the headword) encapsulates the idea of "continued thinking, reflection, contemplation" as all of these words are semantically classified by the OED's Historical Thesaurus (OED Online 2013). A worked example for CONSIDER is presented in Table 2 and Table 3.

Table 2. Selection of key terms for CONSIDER

Headword	**OED definition (first occurrence of each key term in bold)**[1]	**Count**
CONSIDER	**3a.** *trans.* To **contemplate** mentally, fix the mind upon; to **think** over, **meditate** or **reflect** on, bestow attentive thought upon, **give heed** to, **take note** of. **4.** with *obj. clause*: To think, reflect, take note. **5a.** *intr.* To think deliberately, bethink oneself, reflect.	7

The CDI values for all thinking verbs considered are shown in Table 4. It may be said that, by and large, all of the candidate words do a decent job of describing thinking, since they range only from 0.389 to 0.571. But since I wanted to choose only four verbs to consider in this study, the measure is a way to rank the candidates into some semantically informed order.

While the five most frequent of this set in the BNC are all rated highly by the above evaluation, I chose to omit CONTEMPLATE for the reason that lexemes ending in *-ate* are derived directly from Latin words ending in *-ātus* rather than via French or Old English; speakers are more likely to consciously restrict their usage to formal, prepared contexts where one's choice of words is carefully and deliberately chosen for articulable reasons (Antti Arppe, personal communication). Though ideally, the frequency of the selected verbs would be on roughly the same order of magnitude, REFLECT, THINK, CONSIDER, and PONDER all are estimated to have more than 500

[1] Some verbs were not considered key terms on the basis of frequency in the BNC: *fix [the/one's] mind (up)on* (4 occurrences); *bestow … thought (up)on* (0 occurrences); *bethink [one]self* (1 occurrence). The markedness of these expressions suggests that their contributions to the overall meaning of the headword are, to the speakers documented by the BNC, very small. **N.B.** That no verbs are highlighted in senses 4 and 5a does not mean that these senses are excluded from the analysis, only that they otherwise present no further unique key terms.

relevant occurrences in the BNC, which is adequately frequent to consider in the present investigation.[2]

Table 3. Calculation of co-definition index for CONSIDER

Key terms in OED definitions (those in common with CONSIDER underlined)		**Count**
COGITATE	think, reflect, ponder, meditate	3
CONTEMPLATE	view mentally, consider, meditate, ponder, study	3
DELIBERATE	think, consider, examine	2
MEDITATE	think, contemplate, muse, reflect, consider, study, ponder, turn over in the mind	5
PONDER	weigh mentally, think, consider, reflect, wonder, meditate, contemplate, muse	5
REASON	think	1
REFLECT	consider, meditate, think, ponder, mull over, contemplate	5
RUMINATE	turn over in the mind, meditate, contemplate, consider, muse, ponder	3
THINK	turn over in the mind, meditate, ponder, consider, cogitate	3
Σ	3 + 3 + 2 + 5 + 5 + 1 + 5 + 3 + 3	30
Average	30 / 9	3.33
Co-definition index	3.33 / 7	0.476

Table 4. Co-definition indices of thinking verbs (selected lexemes for study in **bold**)

Headword (# of terms in definition, including self)	**# Key terms in common with:**										**Average**	**CDI**	**Rank (greatest to least CDI)**
	COGITATE	CONSIDER	CONTEMPLAT	DELIBERATE	MEDITATE	PONDER	REASON	REFLECT	RUMINATE	THINK			
COGITATE (5)	–	3	2	1	4	4	1	4	2	4	2.78	0.556	2
CONSIDER (7)	**3**	–	**3**	**2**	**5**	**5**	**1**	**5**	**3**	**3**	**3.33**	**0.476**	**6**
CONTEMPLATE (6)	2	3	–	1	5	4	0	4	4	3	2.89	0.481	5
DELIBERATE (4)	1	2	1	–	2	2	1	2	1	2	1.56	0.389	9
MEDITATE (9)	4	5	5	2	–	6	1	6	6	5	4.44	0.494	4
PONDER (9)	**4**	**5**	**4**	**2**	**6**	–	**1**	**6**	**5**	**4**	**4.11**	**0.457**	**8**
REASON (2)	1	1	0	1	1	1	–	1	0	1	0.78	0.389	9
REFLECT (7)	**4**	**5**	**4**	**2**	**6**	**6**	**1**	–	**4**	**4**	**4.00**	**0.571**	**1**
RUMINATE (7)	2	3	4	1	6	5	0	4	–	4	3.22	0.460	7
THINK (6)	**4**	**3**	**3**	**2**	**5**	**4**	**1**	**4**	**4**	–	**3.33**	**0.556**	**2**

In contrast, Arppe (2008:20–27) used an initial selection of verbs exemplifying various domains of cognition to seed his process of finding candidate verbs for study. He extracted the definitions

[2] Word frequency may be thought of to correlate with possible contexts—if a word can be used in many different contexts, then, assuming all contexts are present equally, that word would be more frequent than another with fewer possible contexts. The implication is that words with fewer contexts are not only rarer, but are thus more marked (see Footnote 1).

from a dictionary for each verb in this selection, and collected all of the verbs among them that appear as a one-word definition. Taking this new set of verbs, the process was repeated. The result was a broad selection of verbs whose senses can in at least one way be used to describe cognition. A cluster analysis was then performed on this last selection of verbs using their one-word definitions and overlaps in definitions to find a group of candidate verbs that describe thinking (as opposed to understanding or feeling, for example). Finally, the frequencies of the candidate verbs were determined by searching in a corpus and the four most frequent selected for study, which were all on the same order of magnitude in frequency.

2.3 Linguistic Annotation

To begin, up to 2000 random lines for each lexeme (where possible) were retrieved from the corpus using the BNCWeb interface, where each line represents an entire sentence. These lines were pasted into a spreadsheet, and each line was inspected individually to weed out those where the lexeme was not used in the selected "cogitation" senses, until there were approximately 250 lines for each lexeme (THINK = 262 lines; CONSIDER = 259; PONDER = 250; REFLECT = 253; total = 1024).

Each line was then coded for a number of variables regarding the context of the thinking verb in it, including semantic categories of its agent, theme, adjunctive modifiers, and modality; and morphosyntactic factors such as the person of the agent, grammatical aspect, tense, voice, and mood. Also coded was whether the verb was used in a quotative manner, that is, used to indicate an expressed thought (*verbum dicendi*). A list of all of the tags used in the annotation process is provided in Appendix B.

Particular attention was required for the semantic factors, where the semantic groupings used as the headers for WordNet's "lexicographer files" (Princeton University 2010) were used as a starting point for annotating the agents and themes. Infrequent categories (< 30 instances) were merged into a single OTHER category, and amid the category of COGNITION for themes, a subcategory of cognitions regarding choices or decision-making emerged, to which I applied the tag COGNITION.CHOICE. The annotation of adjunctive modifiers proceeded entirely via a principle of emergent categories (Hanks 1996; and as used for annotating this kind of information in Arppe 2008:31).

At the syntactic level, annotation was relatively straightforward. Some special circumstances were treated as follows:

1. The aforementioned merging of infrequent categories (< 30 instances) was maintained throughout, except for the ADJUNCTS_NEGATION categories, where the threshold for inclusion as a separate category was 15 instances.
2. Implicit second-person arguments (as in imperatives) were annotated for second-person, but annotated as no semantic category for agent.
3. If the theme was anaphoric, it was annotated in its own category, THEME.PRONOUN. The rest of the line was inspected to look for its referent, and if necessary, I returned to the corpus to look at the sentence in context in order to determine the appropriate semantic category for that theme.

4. Where more than one type of adjunctive modifier (including negation) was present, all of them were noted. This is because, unlike the other variables, not only are they optional, but there can be more than one modifying the verb at the same time.
5. Where it could be considered that there was more than one possible category for modality and aspect, the OTHER tag was assigned. This results in perfect progressive aspect being lumped together with other aspects (such as inchoative) into the OTHER category.
6. If the thinking verb appeared as a complement of another verb, including auxiliary and modal verbs, tense and aspect were annotated based on the first (inflected) verb in the sequence. (Modal verbs are invariably non-past tense.) The "future tense" in English, as it is not a true inflectional tense, was considered a modality and treated separately from tense.
7. Non-complement non-finite forms, not forming part of a verb sequence, were tagged as either *to*-infinitive or -*ing*-infinitive.

To best illustrate the annotation process, some examples are given in Appendix C.

Once all 1024 lines were annotated in this manner, the resulting table of tags was exported as a comma-separated value (CSV) file which was read into the *R* statistical programming environment (R Development Core Team 2013) as a data-frame in order to perform the statistical analyses below.

CHAPTER III. Statistical Analysis

3.1 Procedure

The statistical analysis of each variable will undergo the same procedure, using functions available in *R* itself as well as some from the *polytomous* package (Arppe 2012). In outline the procedure involves the following steps: cross-tabulation; chi-squared test for independence and significance; examining the strength of association between lexeme and variable; and finally identifying the specific categories preferred or dispreferred by each lexeme and vice versa.

By way of example, I will focus on the variable representing the semantic category of the AGENT. Shown in Table 5 are the raw frequencies for each value in the variable AGENT, broken down into frequencies for each lexeme. Within the Other category for AGENT are agents that would be classed as animals (n=3), communications (n=7), and cognitions, states, and artifacts (n=1 each). These instances of thinking verbs may be understood as being used with anthropomorphized agents (Arppe 2008:94).

Table 5. Cross-tabulation of semantic AGENT across lexemes (frequency)

Lexeme \ Agent	**None**	**Individual**	**Group**	**Generic**	**Other**	**Σ**
THINK	39	199	18	4	2	262
CONSIDER	111	53	82	8	5	259
REFLECT	27	169	36	17	4	253
PONDER	15	181	36	16	2	250
Σ	192	602	172	45	13	1024

On first inspection, we can note that all of the verbs occur with every AGENT type. If AGENT is not a contextual factor in the selection of lexeme, then the occurrences of each AGENT type would be evenly distributed across the above table. Clearly this is not the case. It appears that across the entire data set, INDIVIDUALs are by and large the most frequent agents of these thinking verbs. Yet, CONSIDER much less frequently has an INDIVIDUAL agent compared to the other lexemes and is more frequently encountered with GROUP agents or without an explicit agent. But can we quantify these observations?

Pearson's chi-squared test will be used to test for the probability that the observed values in the cross-tabulation above could have happened by chance, assuming that the data falls evenly under a chi-squared distribution (Arppe 2008:78). This test can be performed with the function `chisq.test()` in R. Running this function on the data in Table 5 reveals that the observed distribution has a likelihood of $P(\chi^2 = 265.9638, df = 12) < 2.2\times10^{-16}$, that is, smaller than the software considers relevant to display, which means that the lexeme and semantic agent are clearly interdependent. In the behavioural sciences, the threshold probability below which we can consider a distribution significantly heterogeneous is $P = 0.05$, and here the *p*-value is so small that the criterion has been indisputably met.

Now that we are certain that the data distribution is heterogeneous, meaning that there is a relationship between the lexeme and its agent, it would be useful to find out how much the interdependence of the two variables can tell us about either of them. One statistical measure that

can do this is Theil's Uncertainty Coefficient (UC), described in Arppe (2008:90–91). The asymmetric form of UC is a measure of how much our uncertainty about one variable is reduced by knowing the value of the other. Part of the output of the function `associations()` in *polytomous* for R consists of the UC measures in both directions. This statistic shows that knowing the semantic grouping of AGENT reduces uncertainty about the lexeme used by $UC_{LEXEME|AGENT}$ = 0.096, but knowing the lexeme reduces uncertainty about what the AGENT might be by $UC_{AGENT|LEXEME}$ = 0.119. We did intuit this to some extent from our inspection of Table 5 earlier, where we saw that if we knew that the lexeme was CONSIDER, this effectively narrowed the likely AGENT down to either NONE or GROUP, yet if we knew that the AGENT was an INDIVIDUAL, this did not narrow down the lexeme as much, since it would likely be any of THINK, REFLECT, or PONDER.

Finally, we want to know exactly which co-occurrences of lexeme and AGENT are of greatest interest in terms of their contribution to the significant heterogeneity of the sample. For this, the standardized Pearson residual e_{ij} (Agresti 2002, cited in Arppe 2008:83) can show whether each cell in the cross-tabulation is significantly higher or lower in value than expected given a chi-squared distribution. The `chisq.posthoc()` function in *polytomous* does just this, in part producing a chart of the residuals as shown in Table 6.

The threshold values for distributional divergence using this measure is $e_{ij} > 2$ or $e_i < 2$, so if $|e_{ij}| \leq 2$, then the divergence is too small to be of note (Agresti 2002, cited in Arppe 2008:83). The `chisq.posthoc()`output conveniently also generates a table replacing significant divergences with a – or + symbol depending on the direction of the divergence, and a 0 where it is not noteworthy, as in Table 7. This simplified symbolic table facilitates human interpretation of the results.

Table 6. Standardized Pearson residuals for the distribution of AGENT across lexemes

Lexeme \ Agent	None	Individual	Group	Generic	Other
THINK	–1.857824	6.543587	–4.982435	–2.625285	–0.8483486
CONSIDER	11.500125	–14.497903	7.402902	–1.186108	1.0992692
REFLECT	–3.793831	2.982814	–1.259015	2.079112	0.5100219
PONDER	–5.940833	5.029016	–1.166033	1.779376	–0.7627228

Table 7. Signed preferences for the distribution of AGENT across lexemes

Lexeme \ Agent	None	Individual	Group	Generic	Other
THINK	0	+	–	–	0
CONSIDER	+	–	+	0	0
REFLECT	–	+	0	+	0
PONDER	–	+	0	0	0

From **Error! Reference source not found.**, we can summarize the results for lexemes in terms of semantic AGENT as follows:

1. No explicit agent: significant preference for CONSIDER; significant dispreference for REFLECT and PONDER

2. Individual agent: significant preferences for THINK, REFLECT, and PONDER; significant dispreference for CONSIDER
3. Group agent: significant preference for CONSIDER; significant dispreference for THINK
4. Generic agent: significant preference for REFLECT; significant dispreference for THINK
5. Other agents: no significant preference or dispreferences

Note how no two groupings of semantic AGENT have identical profiles, and indeed even when the preferences and dispreferences are regarded from the point of view of the lexemes, the profiles remain unique:

1. THINK: significant preference for individual agents; significant dispreference for group and generic agents
2. CONSIDER: significant preference for no explicit agent and group agents; dispreference for individual agents
3. REFLECT: significant preference for individual and generic agents; significant dispreference for implicit agents
4. PONDER: significant preference for individual agents; significant dispreference for implicit agents

This completes the walk-through of the statistical analysis procedure for the variable AGENT in relation to the lexemes. In the next section I will go directly to the discussion of the signed standard Pearson residual tables for each variable, after first noting any values of note among the $P(\chi^2, df)$ and UC evaluations. A full presentation of all of the cross-tabulations and relevant statistical figures resulting are provided in Appendix D.

3.2 Results

For all contextual features under consideration, the p-values fall under the $P < 0.05$ threshold, which means that their distributions are significantly heterogeneous. However, it is of note that the p-value for the feature NON_FINITE is quite large ($P(\chi^2)_{NON\_FINITE} = 1.159\times10^{-2}$)—a factor of 1000 greater than the next greatest p-value ($P(\chi^2)_{ASPECT} = 6.741\times10^{-5}$). So, while significant, the results for the categories NON_FINITE and ASPECT are perhaps not as distinguishing between the lexemes as the other categories are.

In terms of uncertainty coefficients, their values range from $UC_{LEXEME|TIME.FREQUENCY} = 0.003$ to $UC_{PREP|LEXEME} = 0.344$ (average = 0.055). Knowing the thinking verb substantially decreases our uncertainty about not only the AGENT (as described above) but also whether the theme is a clause ($UC_{THEME\_CLAUSE|LEXEME} = 0.109$). The PREPOSITION is of special note because not only is knowing the verb highly associated with knowing the preposition after it, but the reverse association is also very high ($UC_{LEXEME|PREP} = 0.262$). No other feature comes even close to having UC-values of this magnitude, let alone in both directions—the next greatest association is $UC_{QUOTATIVE|LEXEME} = 0.151$). Even the highest uncertainty coefficient for the parallel part of this analysis in Arppe (2008) is only $UC_{LEXEME|PATIENT} = 0.214$.[3]

[3] This figure in Arppe (2008:518) is for a grouped analysis of separate features representing the semantic and structural characteristics of what he calls the PATIENT argument. Its closest equivalent in my analysis would be a combination of my THEME, THEME_PRONOUN, and THEME_CLAUSE variables.

Table 8. Preferences for agent PERSON across lexemes

Lexeme \ Person	None	1. Person	2. Person	3. Person
THINK	–	+	+	–
CONSIDER	+	0	+	–
REFLECT	–	–	–	+
PONDER	–	0	–	+

The feature of PERSON codes the grammatical side of the agent argument. As shown in Table 8, in terms of the lexemes, THINK is preferentially associated with first and second-person agents and dispreferentially associated with third-person or no agent. CONSIDER also disprefers third-person agents, but instead prefers second-person or no agent. REFLECT and PONDER both prefer third-person agents and disprefer second-person or no agent, with REFLECT additionally exhibiting dispreference for first-person agents. In terms of the grammatical PERSON of the agent, the most distinguishing features are that of no agent with CONSIDER and first-person agent with THINK.

Table 9. Preferences for PREPOSITION across lexemes

Lexeme \ Preposition	None	*about*	*of*	*on*	*over*	*upon*	Other
THINK	–	+	+	–	0	–	+
CONSIDER	+	–	–	–	–	–	0
REFLECT	–	–	–	+	0	+	0
PONDER	+	–	–	0	+	0	0

Table 9 presents distributional preferences for the PREPOSITION that follows the verb (without or without intervening modifiers such as adverbs). CONSIDER categorically does not take a preposition (n = 0 for all categories other than no preposition), but the other lexemes do. THINK is positively associated with *about, of,* and other prepositions (such as *through, out,* and *along),* but is negatively associated with *on, upon,* and no preposition. REFLECT takes preferentially the preposition *on* and its variant *upon*, and disprefers *about, of,* and no preposition. Finally, PONDER (without preposition) and PONDER *over* are preferred but PONDER *about* is not. Looking in terms of the prepositions, *about* and *of* are associated positively with THINK, and indeed *of* occurs categorically with THINK among the lexemes studied. *On* and *upon* are positively associated solely with REFLECT, and *over* only with PONDER. These multiple categorical results are reflected in the high *UC* values in both directions which were presented at the beginning of this section.

In Table 10, we can see that the lexemes prefer different kinds of semantic THEMEs, even though each category of THEME appears at least once with every lexeme. We find that CONSIDER prefers acts, cognitions, choices, and communications but not times, other miscellaneous themes, or no theme; REFLECT prefers attributes, events, and states but not cognitions or no theme; and PONDER prefers cognitions but not acts. THINK is a catch-all verb that prefers the miscellaneous themes contained in the "other" category as well as the lack of a specified theme at all. These preferences are largely distinguishing between the four lexemes, though CONSIDER and PONDER are both positively associated with cognition themes.

Table 10. Preferences for semantic THEME across lexemes

Lexeme \ Theme	None	Act	Attribute	Cognition	Choice
THINK	+	0	0	–	–
CONSIDER	–	+	0	+	+
REFLECT	–	0	+	–-	0
PONDER	0	–	0	+	0
Lexeme \ Theme	**Communication**	**Event**	**State**	**Time**	**Other**
THINK	–	0	0	0	+
CONSIDER	+	0	0	–	–
REFLECT	0	+	+	0	0
PONDER	0	0	0	0	0

Anaphoric themes expressed by a pronoun, as tabulated in Table 11, occur preferentially with THINK (as in THINK *about it*, THINK *of someone*, and other phrases of this type), even though they also occur with the other lexemes. CONSIDER and PONDER disprefer pronominal themes, instead preferring non-anaphoric themes. REFLECT has no significant preference either way.

Table 11. Preferences for pronominal themes across lexemes

Lexeme \ Pronoun	None	Pronoun
THINK	–	+
CONSIDER	+	–
REFLECT	0	0
PONDER	+	–

Table 12 considers a special kind of theme: those which are CLAUSEs in their own right, in particular, either *wh*- or *that*-clauses. Typically these are either indirect questions or relative clauses.[4] Such clauses were relatively rare in the sample, but there is enough evidence to show that *that*-clauses are characteristically associated with REFLECT and characteristically not associated with the other lexemes. In addition, THINK is positively and REFLECT negatively associated with non-clausal themes.

Table 12. Preferences for clausal themes across lexemes

Lexeme \ Clause Type	N/A	*wh*-Clause	*that*-Clause
THINK	+	0	–
CONSIDER	0	0	–
REFLECT	–	0	+
PONDER	0	0	–

Now that we have considered the obligatory 'participants' in an expression of thinking, that is, agents and themes, I will move on to investigate the adjunctive modifiers variable. These

[4] Sentences in which a thinking verb coupled with a *that*-clause attributes a belief or opinion—for example "I do not ***think*** *that* this is a problem in the German-speaking world." (EBU 233)—have been excluded from this study as they, in my opinion, do not express the *activity* of thinking. In this respect, my selection criteria for inclusion are more conservative than those of Arppe (2008), which does include such sentences.

arguments require special treatment: they are not only optional for a verb, but there also can be more than one of them modifying the same verb. In other words, the presence of a modifier of one category does not entail the absence of a modifier in every other category. Because of this lack of mutual exclusivity between categories, statistical calculations had to be performed separately for each category, testing for each kind of modifier's presence versus its absence. In Table 13, the number of occurrences of each type of adjunct across the entire data set is presented in addition to the signs of the significant Pearson residuals.

Table 13. Preferences for adjunctive modifiers (semantically classed) and verb negation across lexemes

Modifier	**None**	**Degree**	**Frame**	**Location**	**Manner**	**Time**[5]
Freq. of Presence Across Sample (1024 sentences)	646	50	26	47	68	83
Lexeme						
THINK	–	+	0	–	0	–
CONSIDER	0	0	+	0	–	+
REFLECT	0	–	–	0	+	0
PONDER	+	0	–	0	–	–

Modifier	**Duration**	**Frequency**	**Opportunity**	**Negation**	**Other**
Freq. of Presence Across Sample (1024 sentences)	52	36	19	25	22
Lexeme					
THINK	0	+	0	+	+
CONSIDER	–	0	–	0	0
REFLECT	–	0	+	0	0
PONDER	+	0	0	–	–

In terms of the semantic categories of adjunctive and adverbial arguments that modify the thinking verb, THINK displays preference for the most categories: degree, frequency, negation, and other modifiers. THINK also displays dispreference for location, time, or zero modifiers. CONSIDER prefers specifying a frame (of mind, or of context), or a time at which to contemplate something, but not a manner or for how long (duration) to do it. As for REFLECT, it prefers to indicate some manner and the availability of an opportunity to perform reflection, and not a frame of mind or any kind of duration. PONDER, however, is negatively associated with frame, manner, time, and negation, but is characteristically associated with the presence of either no modifying argument or one of duration. It is also noteworthy that aside from locations, each category is only positively preferred by one lexeme.

The remaining variables under study return to treating all of the possible categories as mutually exclusive, as had been with our look at the agents, prepositions, and themes. Verb modality (Table 14) is a feature which is preferentially characteristic in few ways for these lexemes. One

[5] Expressions of time refer here to a moment or a point in time, in contrast with expressions of duration or frequency.

tends to be obliged in some way to CONSIDER something, likely in the future; while PONDER*ing* tends to take place without any modal arguments at all.

Table 14. Preferences for modal arguments across lexemes

Lexeme \ Modal	None	Ability	Future
THINK	0	0	0
CONSIDER	–	0	+
REFLECT	0	0	–
PONDER	+	–	–

Lexeme \ Modal	Obligation	Possibility	Volition	Other
THINK	0	–	0	0
CONSIDER	+	0	0	0
REFLECT	0	0	0	0
PONDER	–	0	0	0

Table 15. Preferences for aspectual arguments across lexemes

Lexeme \ Aspect	None	Perfective	Progressive	Other
THINK	0	0	0	0
CONSIDER	0	0	0	–
REFLECT	+	–	–	0
PONDER	–	0	+	+

ASPECT (Table 15) is a feature that was originally assigned categories on a semantic basis, but the resulting collapse of infrequent categories into an "other" category resulted in the significant categories matching with the morphologically inflected aspects in English: perfective and progressive. Here we find that THINK and CONSIDER have no special preference for either of these aspects. REFLECT prefers having no aspect marking over either perfective or progressive, while PONDER takes a preference for progressive and other non-perfective aspects.

Table 16. Preferences for verb TENSE across lexemes

Lexeme \ Tense	Non-Past	Past	N/A
THINK	0	+	–
CONSIDER	+	–	0
REFLECT	–	0	+
PONDER	–	0	0

Table 16 represents the verb TENSE, either past or non-past, associated with the lexemes in the sample. The "future tense" in English takes the same finite verb form as the "present tense", so I have considered "future" as a modal argument (as above) and the verb form itself as the non-past tense considered here. We see that THINK*ing* preferentially happens in the past, and CONSIDER*ing* preferentially happens in the non-past (present or future). Infinitive verbs lack tense, and REFLECT and THINK have positive and negative associations with this category, respectively.

Proceeding to the NON-FINITE verb forms (Table 17), these consist of infinitive (and gerund) forms that are functioning as nominal arguments (in verb chains, verbal features such as tense,

aspect, and mood were assigned based on the finite verb at the beginning of the chain). These results show a clear preference for PONDER functioning as a nominal.

Table 17. Preferences for NON-FINITE verb forms across lexemes

Lexeme \ Infinitive Form	***-ing* Infinitive**	***to*-Infinitive**	**N/A**
THINK	–	0	+
CONSIDER	0	0	0
REFLECT	0	0	–
PONDER	+	0	0

Of most significant note in Table 18 is the complementary preference for active and passive VOICE for THINK and CONSIDER, respectively. Neither REFLECT nor PONDER prefer active voice, but both strongly disprefer passive voice.

Table 18. Preferences for VOICE across lexemes

Lexeme \ Voice	**Active**	**Passive**	**N/A**
THINK	+	–	–
CONSIDER	–	+	0
REFLECT	0	–	0
PONDER	0	–	0

The grammatical MOODs with frequencies sufficient for our statistical analysis amounted to the indicative and the imperative moods (Table 19). THINK and CONSIDER are positively associated with the imperative mood, while REFLECT and PONDER are negatively associated with it. In fact, CONSIDER also has a negative association with the indicative mood. Infinitive verbs lack grammatical mood, and REFLECT and THINK have positive and negative associations with this category, respectively (again).

Table 19. Preferences for MOOD across lexemes

Lexeme \ Mood	**N/A**	**Indicative**	**Imperative**	**Other**
THINK	–	0	+	0
CONSIDER	0	–	+	0
REFLECT	+	0	–	0
PONDER	0	0	–	0

Table 20. Preference for QUOTATIVE function across lexemes

Lexeme \ Quotative	**None**	**Quotative**
THINK	–	+
CONSIDER	+	–
REFLECT	–	+
PONDER	+	–

Finally, usage in a QUOTATIVE function is shown in Table 20 to be preferentially associated with think and reflect, and complementarily dispreferentially associated with CONSIDER and PONDER.

In the next section, I will consider what these results mean for the semantic description of the lexemes and form generalizations for them.

CHAPTER IV. Discussion

4.1 Characterizations of the Studied Thinking Verbs

The category-by-category explication of the results above may seem daunting and its implications for the meanings of the studied thinking verbs not obvious. When we group together all of the preferences for each lexeme, though, it is possible to get a sense of their basic meanings and to construct characterizations or provisional definitions of them (Arppe 2008:160). Table 21 presents the positive preferences for the four lexemes in a condensed form.

Table 21. Grouped positive contextual preferences

**\ Lexeme Domain **	THINK	CONSIDER	REFLECT	PONDER
Agent	Individual 1st person 2nd person	Group 2nd person	Individual Generic 3rd person	Individual 3rd person
Preposition	*about* *of* [others]	[none]	*on* *upon*	*over*
Theme	None Pronoun [others]	Act Choice Cognition Communication	Attribute Event State *that*-clause	Cognition
Modifiers	Degree Frequency Negation	Frame Time	Manner Opportunity	Duration [none]
Inflection	Past tense Imperative mood Active voice	Non-past tense Imperative mood Passive voice	–	Progressive aspect *-ing* infinitive
Modality	–	Future Obligation	–	–
Function	Quotative	–	Quotative	–

4.1.1 THINK

First, I would characterize THINK as the overarching hypernymic lexeme for the act of cogitation. Each of the studied lexemes has a preference for one basic preposition, with the exception of THINK, which shows preference for a variety of prepositions, so its meaning seems to be more generic. Thinking tends to be undertaken by individuals, in particular the self, who can also generically just THINK about nothing in particular, allowing thoughts to wander from one theme to another. The preference for anaphoric and a variety of 'other' themes goes well with THINK's character as a 'catch-all' verb for cogitation as well.

However, THINK prefers expressions of frequency and degree, taking place in the past, implying that the thinking is momentary rather than sustained, and one perhaps returns to it once and again. The momentariness of THINK is also reflected in its usage in a quotative function to

articulate a person's thoughts at a snapshot in time. The association of negation with THINK suggests that an individual thought may or may not be consciously on someone's mind at a given point in time (for example, when setting aside a thought for the time being), and that revisiting the topic has the possibility of causing a change of heart. Indeed, we frequently use the imperative *think about it!* to ask somebody to reconsider an opinion.

Selected examples from the corpus that illustrate these preferences:

- *The curious thing about industry is that if we **think** for ten minutes and draw a picture of the kind of organization we would least like to work in, and hence the one where we are least likely to be effective, we often look about and see just such an environment around us*. (EA8 764) [first-person agent; no explicit theme]
- *Always **think** how the passage would look if the bass were written at its true pitch instead of an octave higher*. (GVS 23) [individual, second-person agent (implied); modifier expressing frequency; imperative mood]
- *The forced brightness in Marjorie's voice disturbed him but he did not want to **think** about it*. (HH9 1564) [individual agent; pronominal theme; *about*; negation; past tense]

4.1.2 CONSIDER

In contrast to THINK, CONSIDER concerns actions, including choosing between alternatives or whether to take action on something. There is a sense of urgency associated with CONSIDER, as evidenced by its preference for modality expressing obligation and a time (in the future) for when the action needs or ought to be considered. These actions and decisions are thought about by or as a group. To get a group to think collectively and come to a consensus, it is often necessary to suggest a frame of mind or context within which the thinking should be constrained.

When a group thinks about or decides on a course of action, the matter itself or the result of the process is often more important than the identities of the decision-makers, which suits well the preference for the depersonalized passive voice in the context of CONSIDER. Frequently, topics to be considered by a group of people are presented in the form of a medium or a product of communication (such as an agenda at a meeting, a paragraph to interpret, or a film at a screening) or an abstract notion (that is, a cognition). Urgency is also reflected in the preference for the imperative mood and second-person agent (the addressee), that is, one is urged to think about a topic when directed to by CONSIDER.

Selected examples from the corpus:

- *It is usual for applications to be **considered** in total rather than by elements within them, so the full application will be approved or refused through the national competition*. (HD2 700) [communication theme; modifier expressing frame ('in total'); passive voice]
- *We will quickly **consider** these, before exploring in more depth the development of those policies that are in fact locally generated*. (B2L 1102) [group agent; cognition theme; modifier expressing time; future modality; non-past tense]
- *Of course, the COSHH regulations do not apply offshore, and the Minister ought to **consider** whether it is time to provide such protection for those who work in the offshore industry*. (HHX 19430) [decision toward action as theme; obligation modality; non-past tense]

4.1.3 REFLECT

REFLECT is an action that is taken by individuals or generic types of people (such as citizens, pupils, etc.) concerning abstract qualities and conditions (categorized in the data as attributes and states, respectively), or events experienced by the agent, which may or may not be expressed in terms of a *that*-clause. Reflection is a process that takes time and focused deliberation; rather than expressing a period or length of time during which it takes place, what is more critical is that the individual *finds* time or the opportunity to REFLECT.

These preferences combined suggest that reflection is a deliberate, conscious action that one directs toward the self, in order to learn from an experience or an observation. The observatory nature of REFLECT is reflected (pun unavoidable) in its preference for quotative usage, offering a glimpse into the thinker's mind.

Selected examples from the corpus:

- *Given the significance of regular heroin use, the psychological and physical dependency often associated with the drug, and the all-embracing nature of the lifestyle usually required to secure funds and supplies, users tend to live from moment to moment and find little time for* ***reflecting*** *upon their situation.* (EDC 906) [generic, third-person agent; *upon*; state theme ('situation'); modifier expressing opportunity]
- *The fact that most of the players in the Olympic match were from England seems to have borne this out, and our Blundellian correspondent sadly* ***reflects****: "A cricketer in France is a stranger in a strange land looked upon with mingled awe and contempt by the average Frenchman."* (CU1 1009) [individual, third-person agent; attribute theme; modifier expressing manner, quotative function]
- *As he walked he* ***reflected*** *that he had one other item to add to the profile of each boy.* (K8V 3697) [individual, third-person agent; *that*-clausal theme]

4.1.4 PONDER

Lastly, PONDER concerns individuals other than the self, thinking about abstract notions (cognitions). It frequently takes some time, such that its duration is often specified. But pondering does not take place in isolation—preference for *-ing* forms as in the progressive aspect and gerundive suggests that one *ponders* things while something else is happening within the frame of the situational context, whether it is done by the ponderer or someone else.

Selected examples from the corpus:

- *Roy Lacey confronts the weeds, and* ***ponders*** *the lessons of his first season with the labour-saving plot.* (A0G 1269) [individual, third-person agent; cognition theme]
- *Can the INSPIRAL CARPETS, who tear round Finland comparing testicles,* ***pondering*** *piss-drinking and Japanese prostitutes, really be the same band who make such brooding, impassioned pop?* (CAE 456) [third-person agent; gerundive *-ing* form; simultaneous action (same agent)]
- *As Ronni continued to* ***ponder*** *for a moment, Jeff leaned across the table.* (JXT 1293) [individual, third-person agent; modifier expressing duration; simultaneous action (different agent)]

4.2 Comparison to Finnish

Characterizations of each of the studied lexemes having been inferred, I will next attempt to draw comparisons and contrasts with a group of near-synonymous lexemes also expressing thinking, in a genetically unrelated language, namely Finnish, as studied by Arppe (2008). While Arppe's discussion centres mainly on the multivariate statistical analysis not performed in my study, and though his annotation scheme differs from my own at the level of detail encoded, his semantic characterizations of the verbs AJATELLA, MIETTIÄ, POHTIA, and HARKITA based on the univariate results (2008:160-162) can be compared with my above results for English verbs.

Since it is generally accepted that it is rare to have completely equivalent words across languages (Partington 1998:49), whenever a translator embarks upon a translation task, he or she must select the word in the target language that best suits the meaning and purpose of the translation. Even between closely related languages, the senses of near-synonyms across the languages will overlap imperfectly (Partington 1998:63). It is reasonable to expect that the denotative particulars of the verbs (as expressed by their contextual preferences) differ between the languages, as Edmonds and Hirst (2002) demonstrate with their conceptual model for synonyms of *untruth* in English and French. Finnish and English are no exceptions to this when it comes to their thinking verbs, as discussed below.

The Finnish lexeme that appears to share the most contextual preferences with THINK is MIETTIÄ, having in common preferences for individual agents of first- or second-persons, the imperative mood, and expressions of frequency. In addition, the preference of THINK in quotative context echoes the preference for direct quotes and indirect questions that MIETTIÄ displays. THINK also has a sense of being momentary and mutable, which MIETTIÄ does not have, and MIETTIÄ has a preference for communicative themes that THINK disprefers.

THINK's possibility of revisiting or reconsidering a thought is found not in MIETTIÄ but in the Finnish verb HARKITA, which expresses this possibility via a preference for 'again' as frequency, and arguments which describe a reason for contemplating something and conditions for coming to a conclusion. The emphasis on reasoning and argumentation matches best in English with CONSIDER, which shares with HARKITA a concern for actions and decision-making.

The impersonal passive voice associated with CONSIDER is not preferred by HARKITA but instead by POHTIA. POHTIA also parallels CONSIDER in other ways through its preference for collective (group) agents and thinking about abstract notions (cognitions). It is also notable that passive voice and group agents are only characteristic of these two verbs. CONSIDER's preference for arguments expressing a frame of mind is in Finnish preferentially associated with AJATELLA.

The deliberateness and preference for *that*-clause themes suggested by AJATELLA seem to match similar preferences in REFLECT. Indeed, while *that*-clauses representing a belief or opinion were excluded from consideration under THINK (and I hazard to say that it is by far an extremely frequent context of usage for this verb), such clauses also appeared frequently with REFLECT. It seems that while AJATELLA handles beliefs and opinions mostly on its own in Finnish, English

shares this burden between THINK and REFLECT.[6] In contrast, though, REFLECT has the denotation of being a very personal activity as MIETTIÄ does, which in the latter is the preference for being performed 'alone', and in the former is expressed as the need to find the opportunity to REFLECT (conceivably, some time to be alone).

Finally, PONDER, too, shares the individual character of AJATELLA, but also the preference for abstract themes of (its false cognate) POHTIA,[7] as well as the preference for duration being important of MIETTIÄ.

Table 22 summarizes the above overlaps in positive contextual associations between the considered sets of English and Finnish thinking verbs, based on the annotations and analysis of corpus data.

Table 22. Common positive associations between English and Finnish

**\ English Finnish **	THINK	CONSIDER	REFLECT	PONDER
AJATELLA	Agent: Individual Agent: 1st person Modifier: Negation	Agent: 2nd person Modifier: Frame	Agent: Individual Theme: *that*-clause Modifier: Manner	Agent: Individual
MIETTIÄ	Agent: Individual Agent: 1st person Agent: 2nd person Modifier: Frequency Mood: Imperative Function: Quotative	Agent: 2nd person Theme: Communication Mood: Imperative	Agent: Individual Function: Quotative	Modifier: Duration
POHTIA	–	Agent: Group Voice: Passive	Agent: 3rd person Theme: Attribute	Theme: Cognition
HARKITA	Modifier: Frequency	Theme: Act	–	–

[6] My intuition as a native speaker of English is that the difference between REFLECT and THINK "that [some belief/opinion]" is that REFLECT draws attention to the process of coming to that belief/opinion while THINK primarily asserts the belief/opinion. Of course, this would be worth investigating empirically.

[7] Incidentally, English *ponder* does have a cognate in Finnish, *punnita*, both coming originally from Latin *pondus* 'weight' (Itkonen & Joki 1962). While *punnita* still retains its concrete sense 'to weigh' to some extent even in the present day (Arppe 2008:262), English adopted *ponder* from Middle French already with the abstract cognitive meaning partially developed; today, the concrete meaning has been lost (OED Online 2013). Further, Antti Arppe (personal communication) pointed out to me that there is another Finnish verb meaning 'to weigh, consider' derived from the same Latin root, namely, *puntaroida* [from L. *pondus* > Fin. *punta* ('weight') > *puntari* ('steelyard balance') > *puntaroida* (+verbal suffix *-oida*; 'to use a balance')].

4.3 Summary and Directions for Further Research

In this study I have taken the first steps toward a replication of Arppe's (2008) study of a selection of Finnish verbs meaning *think* using their English equivalents and data from the British National Corpus, with view to ultimately conduct a full cross-linguistic comparison between expressions of *thinking* between the two languages.

Since I have analyzed only the univariate statistics and drawn conclusions from them in this study, the natural extension of the current work would be to perform and analyze the data using the bivariate and multivariate methods developed and used in the remainder of Arppe (2008). The probabilistic results arrived at by this process may then be used as the basis for experimental work paralleling Arppe & Järvikivi (2007), which tested native speakers' lexical choices versus those evidenced by the corpus.

One great advantage of contextual profiling via tagging and analysis as was performed here is that it is extensible; and the analysis here can be easily extended to include genre, domain, and register, as this information is tagged in the BNC. This would allow for investigation into the pragmatic and stylistic components of Miller & Charles' (1991) contextual representations, which were not considered in my study.

It would be worthwhile to characterize the semantic differences between THINK *about*, THINK *of*, and THINK with other prepositions, via their collocations and colligations (as Partington (1998:80) does with prepositions following PERSIST). How do the different prepositions divide up the 'genericness' of THINK, and can or do these expressions of THINK+[preposition] match up with thinking verbs other than those studied here, such as CONTEMPLATE or MULL *(over)*?

What do the results so far tell us about thinking? I hypothesize that one dimension that characterizes the way English subdivides the domain of thinking is the distance of the purpose of thought from the ego. From the inward-directed and solitary nature of REFLECT, to the presence of another action or person in PONDER, to the other-directedness and generic character of THINK, to the impersonal distance of CONSIDER, there seems in each of these lexemes an implicit expression of how far the cogitation reaches out to and impacts the people around the thinker.

Finally, like the Finnish thinking verbs, the contextual associations of the studied English thinking verbs may be partially explained by etymology, from metaphorical extensions of verbs of perception and experience (Arppe 2008:48–49, 162–163; after Fortescue 2001). Are the etymologically-based associations culturally-dependent overlays atop universal contextual patterns across languages based on semantic primitives (Arppe 2008:262)? Indeed, does culture play a role in conditioning contextual associations even between varieties of the same language, such as British and American English? Further diachronic and cross-linguistic investigation using historical English corpora and corpora in other languages may open up the possibility of such hypotheses from a historical linguistics or typological perspective.

REFERENCES

Agresti, Alan 2002. *Categorical data analysis*, 2 edn. Hoboken: John Wiley & Sons.

Arppe, Antti. 2002. The usage patterns and selectional preferences of synonyms in a morphologically rich language. In Morin, Annie & Pascale Sébillot (eds.), *JADT–2002. 6th International Conference on Textual Data Statistical Analysis*, vol. 1, 21–32. Rennes: INRIA.

Arppe, Antti 2008. Univariate, bivariate and multivariate methods in corpus-based lexicography: A study of synonymy. Helsinki: University of Helsinki dissertation.

Arppe, Antti. 2012. Package 'polytomous': Polytomous logistic regression for fixed and mixed effects. Version 0.1.4. The R Project for Statistical Computing. URL: http://www.r-project.org/.

Arppe, Antti & Juhani Järvikivi. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2). 131–159.

Atkins, B. T. Sue. 1987. Semantic-ID tags: Corpus evidence for dictionary senses. *Third Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary.* 17–36.

Atkins, B. T. Sue & Beth Levin. 1995. Building on a corpus: A linguistic and lexicographical look at some near-synonyms. *International Journal of Lexicography* 8. 85–114.

Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/

Church, Kenneth W., William Gale, Patrick Hanks, Donald Hindle, & Rosamund Moon. 1994. Lexical substitutability. In B. T. Sue Atkins & Antonio Zampolli (eds.), *Computational approaches to the lexicon*, 153–177. Oxford: Oxford University Press.

Cruse, D. Alan. 1986. *Lexical semantics.* Cambridge: Cambridge University Press.

Dąbrowska, Ewa. 2009. Words as constructions. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New directions in cognitive linguistics*, 201–223. Amsterdam & New York: John Benjamins.

Divjak, Dagmar & Stefan Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1). 23–60.

Edmonds, Philip & Graeme Hirst. 2002. Near synonymy and lexical choice. *Computational Linguistics* 28(2). 105–144.

Fellbaum, Christiane. 1990. English verbs as a semantic net. *International Journal of Lexicography* 3(4). 278–301.

Firth, John R. 1957. *Papers in linguistics*. Oxford: Oxford University Press.

Fortescue, Michael. 2001. Thoughts about thought. *Cognitive Linguistics* 12(1). 15–45.

Goddard, Cliff. 2003. Thinking across languages and cultures: Six dimensions of variation. *Cognitive Linguistics* 14. 109–140.

Goddard, Cliff & Susanna Karlsson. 2008. Re-thinking THINK in contrastive perspective: Swedish vs. English. In Cliff Goddard (ed.), *Cross-linguistic semantics: Explications, metalanguage, grammar*, 225–240. Amsterdam & New York: John Benjamins.

Gries, Stefan Th. 2010. Behavioral profiles: a fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon* 5(3). 323–346.

Gries, Stefan Th. & John Newman. in press. Creating and using corpora. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*. Cambridge: Cambridge University Press.

Gries, Stefan Th. & Naoki Otani. 2010. Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal* 34. 121–150.

Hanks, Patrick. 1996. Contextual dependency and lexical sets. *International Journal of Corpus Linguistics* 1(1). 75–98.

Harris, Zellig. 1954. Distributional structure. *Word* 10(23). 146–162.

Itkonen, Erkki & Aulis J. Joki. 1962. *Suomen kielen etymologinen sanakirja III* (Lexica Societatis Fenno-Ugricae XII). Helsinki: Suomalais-Ugrilainen Seura.

Lee, David. Y. W. 2010. What corpora are available? In Anne O'Keeffe & Michael McCarthy (eds.), *The Routledge handbook of corpus linguistics*, 107–121. Oxon UK: Routledge.

Miller, George A. & Walter G. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6. 1–28.

Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, & Katherine J. Miller. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography* 3(4). 235–244.

Mohammad, Saif & Graeme Hirst. 2006. Distributional measures of semantic distance: A survey. Unpublished manuscript: arXiv:1203.1858v1 [cs.CL]

Murphy, M. Lynne. 2008. *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms*. Cambridge: Cambridge University Press.

Oxford English Dictionary Online. 2013. Oxford: Oxford University Press. URL: http://www.oed.com/ (accessed 16 March 2013)

Partington, Alan. 1998. *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam & New York: John Benjamins.

Princeton University. 2010. "About WordNet." *WordNet 3.1*. Princeton University. URL: http://wordnet.princeton.edu/

R Development Core Team. 2013. R: A language and environment for statistical computing. URL: http://www.R-project.org/

Rubenstein, Herbert & John B. Goodenough. 1965. Contextual correlates of synonymy. *Communications of the ACM* 8(10). 627–633.

APPENDIX A
Selections of Lexical Entries from OED Online (2013)

Lexeme	**Definition**
Cogitate	**1.** *intr.* To think, reflect, ponder, meditate; to exercise the thinking faculties.
Consider	**3a.** *trans.* To contemplate mentally, fix the mind upon; to think over, meditate or reflect on, bestow attentive thought upon, give heed to, take note of. **4.** with *obj. clause*: To think, reflect, take note. **5a.** *intr.* To think deliberately, bethink oneself, reflect.
Contemplate	**2.** To view mentally; to consider attentively, meditate upon, ponder, study.
Deliberate	**2a.** *intr.* To use consideration with a view to decision; to think carefully; to pause or take time for consideration. **2b.** Of a body of persons: To take counsel together, considering and examining the reasons for and against a proposal or course of action.
Meditate	**1.** *intr.* To exercise the mind in thought or reflection; (freq.) to engage the mind in religious or spiritual reflection, contemplation, or other discipline. **2a.** *trans.* To muse over or reflect upon; to consider, study, ponder. **3.** *trans.* To plan by turning over in the mind; to conceive or design mentally.
Ponder	**3a.** *trans.* To weigh (a matter, words, etc.) mentally, esp. before making a decision or reaching a conclusion; to think over, consider, or reflect on; to wonder about. Freq. with clause as object. **4.** *intr.* To think or reflect deeply, to meditate; to contemplate, to muse; to wonder. Freq. with *on, upon, over, about*.
Reason	**6.** *intr.* To think in a connected or logical manner; to employ the faculty of reason in forming conclusions.
Reflect	**7a.** *intr.* To engage in reflection; to consider, meditate. **7b.** *trans.* With clause as object: to think about, consider. **8a.** *intr.* With *on* or *upon*. To fix the thoughts or attention (back) on something, esp. a past event or experience; to think deeply or carefully about; to ponder, mull over, contemplate.
Ruminate	**1a.** *trans.* To revolve, turn over repeatedly in the mind; to meditate deeply upon **1b.** *trans.* To contemplate or consider (a plan, course, etc.) with a view to subsequent action. Now *rare*. **2a.** *intr.* To muse, meditate, ponder. **2b.** *intr.* With *about, of, on, upon, over*.
Think	**2.** *trans.* To turn over in the mind, meditate on, ponder over, consider. **3a.** *intr.* To exercise or occupy the mind, esp. the understanding, in any active way; to form connected ideas of any kind; to allow or cause a train of ideas to pass through the mind; to meditate, cogitate. With *about, of, on,* or *upon* (arch.).

APPENDIX B
List and Description of Tags Used in Linguistic Annotation

AGENT – who/what is doing the action
AGT.GENERIC
AGT.GROUP
AGT.INDIVIDUAL
AGT.OTHER
NONE

PERSON (of Agent)
AGT.1P
AGT.2P (includes implicit 2nd person)
AGT.3P
NONE

PREPOSITION
PREP.ABOUT
PREP.OF
PREP.ON
PREP.OVER
PREP.UPON
PREP.OTHER
NONE

THEME – topic of cogitation
THEME.ACT
THEME.ATTRIBUTE – qualities of people and things
THEME.COGNITION – mental abstractions and constructs
THEME.COGNITION.CHOICE – choice between options or whether or not to perform some action
THEME.COMMUNICATION – media or products of communication
THEME.EVENT
THEME.STATE – natural states of affairs
THEME.TIME
THEME.OTHER
NONE

THEME (PRONOUN)
THEME.PRONOUN – theme is anaphoric
NONE

THEME (CLAUSE)
THEME.THAT – theme is a *that*-clause
THEME.WH – theme is a *wh*-clause (*who, what, when, where, why, whether, how*)
NONE

ADJUNCTS _NEGATION – modifiers to the verb; these are best illustrated by examples
ADJ.DEGREE – e.g. *just; only; a lot*
ADJ.FRAME – e.g. *in context of; in relation to; separately; like a child*
ADJ.LOCATION
ADJ.MANNER
ADJ.TIME
ADJ.TIME.DURATION – e.g. *for a while*
ADJ.TIME.FREQUENCY – e.g. *often; once*
ADJ.TIME.OPPORTUNITY – e.g. *have time to; have a chance to*
ADJ.OTHER
NEG – verb negation
NONE

MODALITY
MOD.ABILITY – *can; could*
MOD.FUTURE – *will; would; shall*
MOD.OBLIGATION – *must; have to; need; should; ought to*
MOD.POSSIBILITY – *may; might*
MOD.VOLITION – *force to; urge to; care to; bound to*
MOD.OTHER
NONE

ASPECT
ASP.PERF – perfective
ASP.PROG – progressive
OTHER – semantic aspects such as inchoative, durative, habitual, etc.
NONE

TENSE
Applied by considering the entire clause holistically based on the entire sequence of verbs including those expressing aspect or modality.
TENSE.NONPAST
TENSE.PAST
NONE

NON-FINITE FORMS
INF.ING – *-ing*-infinitive
INF.TO – *to*-infinitive
NONE

VOICE
VOICE.ACTIVE
VOICE.PASSIVE
NONE

MOOD
MOOD.IMPER – imperative
MOOD.INDIC – indicative
MOOD.OTHER
NONE

QUOTATIVE FUNCTION
QUOTATIVE – marks usage of verb to report expressed thought (i.e. *verbum dicendi*)
NONE

APPENDIX C
Examples of Linguistic Annotation

Line	Martha , who had decided to stop <<< thinking >>> about the inconvenience they were causing , asked Woodie not to stop at the boats ; they would like to go on to the New King 's Road . (H0R 1196)		
Factor	**AGENT**	**PERSON**	**PREPOSITION**
Value	INDIVIDUAL (Martha)	AGT.3P	ABOUT
Factor	**THEME**	**THEME_PRONOUN**	**THEME_CLAUSE**
Value	STATE (the inconvenience)	NONE	NONE
Factor	**ADJUNCTS_NEGATION**	**MODALITY**	**ASPECT**
Value	NONE	VOLITION (decided to)	PERFECT
Factor	**TENSE**	**NON_FINITE**	**VOICE**
Value	PAST	NONE	ACTIVE
Factor	**MOOD**	**QUOTATIVE**	
Value	INDICATIVE	NONE	

Line	<<< Consider >>> today whether you look upon your job as “work” or as an enjoyable way to earn a living . (C9Y 2861)		
Factor	**AGENT**	**PERSON**	**PREPOSITION**
Value	NONE	AGT.2P	NONE
Factor	**THEME**	**THEME_PRONOUN**	**THEME_CLAUSE**
Value	COGNITION.CHOICE (whether…)	NONE	THEME.WH (whether…)
Factor	**ADJUNCTS_NEGATION**	**MODALITY**	**ASPECT**
Value	ADJ.TIME (today)	NONE	NONE
Factor	**TENSE**	**NON_FINITE**	**VOICE**
Value	NONPAST	NONE	ACTIVE
Factor	**MOOD**	**QUOTATIVE**	
Value	IMPERATIVE	NONE	

Line	Despite his misfortunes Maskell once <<< reflected >>> : " I am the luckiest chap in the world . " (CBF 9305)		
Factor	**AGENT**	**PERSON**	**PREPOSITION**
Value	INDIVIDUAL (Maskell)	AGT.3P	NONE
Factor	**THEME**	**THEME_PRONOUN**	**THEME_CLAUSE**
Value	ATTRIBUTE ("I am the luckiest…")	NONE	NONE
Factor	**ADJUNCTS_NEGATION**	**MODALITY**	**ASPECT**
Value	TIME.FREQUENCY (once)	NONE	NONE
Factor	**TENSE**	**NON_FINITE**	**VOICE**
Value	PAST	NONE	ACTIVE
Factor	**MOOD**	**QUOTATIVE**	
Value	INDICATIVE	QUOTATIVE	

Line	Even though the NSDAP was to achieve its majority in the Danzig Volkstag with a very clear mandate from the electorate to do what it thought necessary , most Danzigers were prepared to reap the benefits of being on the winning side without <<< pondering >>> too deeply the significance or morality of their own personal support for a party they did not entirely trust or like . (BN2 1026)		
Factor	**AGENT**	**PERSON**	**PREPOSITION**
Value	GENERIC (Danzigers)	AGT.3P	NONE
Factor	**THEME**	**THEME_PRONOUN**	**THEME_CLAUSE**
Value	ATTRIBUTE (the significance or morality…)	NONE	NONE
Factor	**ADJUNCTS_NEGATION**	**MODALITY**	**ASPECT**
Value	NEGATION DEGREE (too deeply)	NONE	NONE
Factor	**TENSE**	**NON_FINITE**	**VOICE**
Value	NONE	INF.ING	NONE
Factor	**MOOD**	**QUOTATIVE**	
Value	NONE	NONE	

APPENDIX D
Statistical Data

D1. Overall Frequency Count of Each Tag

LEXEME		AGENT		PERSON		PREP	
consider	259	AGT.GENERIC	45	AGT.1P	177	NONE	704
ponder	250	AGT.GROUP	172	AGT.2P	97	OTHER	4
reflect	253	AGT.INDIVIDUAL	602	AGT.3P	609	PREP.ABOUT	91
think	262	NONE	192	NONE	141	PREP.OF	40
		OTHER	13			PREP.ON	144
						PREP.OVER	18
						PREP.UPON	23

THEME		THEME_PRONOUN		THEME_CLAUSE	
NONE	140	NONE	954	NONE	875
OTHER	207	THEME.PRONOUN	70	THEME.THAT	49
THEME.ACT	144			THEME.WH	100
THEME.ATTRIBUTE	102				
THEME.COGNITION	188				
THEME.COGNITION.CHOICE	34				
THEME.COMMUNICATION	95				
THEME.EVENT	40				
THEME.STATE	44				
THEME.TIME	30				

ADJUNCTS_NEGATION		MOD		ASP	
ADJ.DEGREE	50	MOD.ABILITY	41	ASP.PERF	35
ADJ.FRAME	26	MOD.FUTURE	35	ASP.PROG	74
ADJ.LOCATION	47	MOD.OBLIGATION	58	NONE	882
ADJ.MANNER	68	MOD.POSSIBILITY	25	OTHER	33
ADJ.TIME	83	MOD.VOLITION	26		
ADJ.TIME.DURATION	52	NONE	826		
ADJ.TIME.FREQUENCY	36	OTHER	13		
ADJ.TIME.OPPORTUNITY	19				
NEG	25				
NONE	646				
OTHER	22				

TENSE		NON_FINITE		VOICE		MOOD	
NONE	235	INF.ING	118	NONE	241	MOOD.IMPER	57
TENSE.NONPAST	411	INF.TO	117	VOICE.ACTIVE	718	MOOD.INDIC	708
TENSE.PAST	378	NONE	789	VOICE.PASSIVE	65	NONE	236
						OTHER	23

QUOTATIVE	
NONE	900
QUOTATIVE	124

D2. Results for Each Category

D2a. Agent

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$AGENT)
X-squared = 265.9638, df = 12, p-value < 2.2e-16

         AGT.INDIVIDUAL NONE AGT.GROUP AGT.GENERIC OTHER
think               199   39        18           4     2
consider             53  111        82           8     5
reflect             169   27        36          17     4
ponder              181   15        36          16     2


$cells$std.pearson.residuals
         AGT.INDIVIDUAL      NONE AGT.GROUP AGT.GENERIC      OTHER
think          6.543587 -1.857824 -4.982435   -2.625285 -0.8483486
consider     -14.497903 11.500125  7.402902   -1.186108  1.0992692
reflect        2.982814 -3.793831 -1.259015    2.079112  0.5100219
ponder         5.029016 -5.940833 -1.166033    1.779376 -0.7627228


$cells$std.pearson.residuals.sign
         AGT.INDIVIDUAL NONE AGT.GROUP AGT.GENERIC OTHER
think                 +    0         -           -     0
consider              -    +         +           0     0
reflect               +    -         0           +     0
ponder                +    -         0           0     0


$uc.RC
[1] 0.09607239

$uc.CR
[1] 0.1190517

$uc.sym
[1] 0.1063347
```

D2b. Person

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$PERSON)
X-squared = 185.4674, df = 9, p-value < 2.2e-16

         AGT.3P AGT.1P NONE AGT.2P
think       139     58   19     46
consider     95     44   84     36
reflect     188     33   25      7
ponder      187     42   13      8


$cells$std.pearson.residuals
            AGT.3P     AGT.1P      NONE    AGT.2P
think    -2.453425  2.4078963 -3.549134  5.180309
consider -8.644506 -0.1461221 10.084592  2.814808
reflect   5.539323 -2.0563279 -2.068385 -4.197702
ponder    5.677847 -0.2333473 -4.522896 -3.895608


$cells$std.pearson.residuals.sign
         AGT.3P AGT.1P NONE AGT.2P
think         -      +    -      +
consider      -      0    +      +
reflect       +      -    -      -
ponder        +      0    -      -


$uc.RC
[1] 0.06325076

$uc.CR
[1] 0.07907615

$uc.sym
[1] 0.07028364
```

D2c. Preposition

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$PREP)
X-squared     =     771.0247,     df     =     18,     p-value     <     2.2e-16

         NONE PREP.ON PREP.ABOUT PREP.OF PREP.UPON PREP.OVER OTHER
think     129       0         88      40         0         2     3
consider  259       0          0       0         0         0     0
reflect   116     113          2       0        21         1     0
ponder    200      31          1       0         2        15     1


$cells$std.pearson.residuals
              NONE    PREP.ON PREP.ABOUT   PREP.OF PREP.UPON
think    -7.899395 -7.5903718  16.288358 11.002953 -2.844265
consider 12.553297 -7.5319782  -5.814991 -3.754063 -2.822383
reflect  -9.056511 16.1362862  -5.215584 -3.695860  7.489729
ponder    4.414085 -0.8697382  -5.424101 -3.666755 -1.774863

         PREP.OVER       OTHER
think    -1.419948  2.26935414
consider -2.490614 -1.16600135
reflect  -1.900649 -1.14792353
ponder    5.870892  0.02733321


$cells$std.pearson.residuals.sign
         NONE PREP.ON PREP.ABOUT PREP.OF PREP.UPON PREP.OVER
think       -       -          +       +         -         0
consider    +       -          -       -         -         -
reflect     -       +          -       -         +         0
ponder      +       0          -       -         0         +

         OTHER
think        +
consider     0
reflect      0
ponder       0


$uc.RC
[1] 0.2617087

$uc.CR
[1] 0.3444384

$uc.sym
[1] 0.2974279
```

D2d. Theme (Semantic)

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$THEME)
X-squared = 230.5997, df = 27, p-value < 2.2e-16

         NONE OTHER THEME.ACT THEME.ATTRIBUTE THEME.COGNITION
consider    4    40        59              19              62
ponder     37    41        20              24              68
reflect    19    58        32              34              34
think      80    68        33              25              24

         THEME.COGNITION.CHOICE THEME.COMMUNICATION THEME.EVENT THEME.STATE
consider                     18                  41           8           7
ponder                       11                  23           9           6
reflect                       4                  22          18          21
think                         1                   9           5          10

         THEME.TIME
consider          1
ponder           11
reflect          11
think             7


$cells$std.pearson.residuals
             OTHER THEME.COGNITION  THEME.ACT       NONE THEME.ATTRIBUTE
think     2.681575       -4.458461 -0.7918708  9.2098868      -0.2624960
consider -2.211906        2.683068  4.6691156 -6.5727828      -1.6320665
reflect   1.236987       -2.329821 -0.7457537 -3.2878659       2.1287358
ponder   -1.727548        4.152889 -3.1716016  0.5971955      -0.2191876

         THEME.COMMUNICATION THEME.STATE THEME.EVENT THEME.COGNITION.CHOICE
think             -3.77860965  -0.4442202  -1.9349025              -3.077595
consider           4.20556648  -1.4637435  -0.7855993               3.771883
reflect           -0.36754128   3.6189756   3.0355719              -1.779495
ponder            -0.04848473  -1.7011742  -0.2874736               1.095949

         THEME.TIME
think    -0.2869949
consider -2.8084190
reflect   1.5415154
ponder    1.5856425
```

```
$cells$std.pearson.residuals.sign
         OTHER THEME.COGNITION THEME.ACT NONE THEME.ATTRIBUTE
think        +               -         0    +               0
consider     -               +         +    -               0
reflect      0               -         0    -               +
ponder       0               +         -    0               0

         THEME.COMMUNICATION THEME.STATE THEME.EVENT THEME.COGNITION.CHOICE
think                      -           0           0                      -
consider                   +           0           0                      +
reflect                    0           +           +                      0
ponder                     0           0           0                      0

         THEME.TIME
think             0
consider          -
reflect           0
ponder            0


$uc.RC
[1] 0.08437325

$uc.CR
[1] 0.05540111

$uc.sym
[1] 0.06688453
```

D2e. Theme (Pronoun)

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$THEME_PRONOUN)
X-squared = 39.4131, df = 3, p-value = 1.419e-08

         NONE THEME.PRONOUN
think     222            40
consider  249            10
reflect   242            11
ponder    241             9


$cells$std.pearson.residuals
              NONE THEME.PRONOUN
think    -6.268899      6.268899
consider  2.194943     -2.194943
reflect   1.807297     -1.807297
ponder    2.331988     -2.331988


$cells$std.pearson.residuals.sign
         NONE THEME.PRONOUN
think       -             +
consider    +             -
reflect     0             0
ponder      +             -


$uc.RC
[1] 0.01201493

$uc.CR
[1] 0.06678378

$uc.sym
[1] 0.02036587
```

D2f. Theme (Clause)

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$THEME_CLAUSE)
X-squared = 127.1068, df = 6, p-value < 2.2e-16

         NONE THEME.WH THEME.THAT
think     238       21          3
consider  231       28          0
reflect   183       25         45
ponder    223       26          1


$cells$std.pearson.residuals
              NONE    THEME.WH THEME.THAT
think     2.868488 -1.10640454  -3.199913
consider  1.974877  0.65558139  -4.174119
reflect  -6.819100  0.07150687  11.165366
ponder    1.934524  0.38865077  -3.736235


$cells$std.pearson.residuals.sign
         NONE THEME.WH THEME.THAT
think       +        0          -
consider    0        0          -
reflect     -        0          +
ponder      0        0          -


$uc.RC
[1] 0.03978924

$uc.CR
[1] 0.1087832

$uc.sym
[1] 0.05826654
```

D2g. Adjuncts / Negation

	NONE	DEGREE	FRAME	LOCATION
N	646	50	26	47
p(X2)	1.652E-02	1.710E-04	8.05E-05	0.007386
uc.RC	0.003600854	0.006299592	0.009179937	0.00571422
uc.CR	0.007579728	0.04476801	0.1075298	0.04252527
std.pearson.residuals				
think	-2.8620868	4.3889692	1.068819	-3.4310488
consider	0.3884351	-1.5499916	3.849700	1.0691908
reflect	0.2090836	-2.1360534	-2.037552	0.8266948
ponder	2.3041841	-0.7449952	-2.935424	1.5931470
think	-	+	0	-
consider	0	0	+	0
reflect	0	-	-	0
ponder	+	0	-	0

	MANNER	TIME	TIME.DURATION	TIME.FREQUENCY
N	68	83	52	36
p(X2)	1.630E-03	0.000153	1.641E-12	0.02603
uc.RC	0.005514469	0.007113416	0.01839054	0.002896993
uc.CR	0.0312957	0.03504708	0.1269441	0.02637802
std.pearson.residuals				
think	1.323560	-2.423747	-0.7517943	3.0288465
consider	-2.656054	3.689568	-3.3243049	-1.2121791
reflect	2.967878	1.192836	-3.2498294	-0.7453059
ponder	-1.636572	-2.469214	7.3904484	-1.1016356
think	0	-	0	+
consider	-	+	-	0
reflect	+	0	-	0
ponder	0	-	+	0

	TIME.OPPORTUNITY	NEGATION	OTHER
N	19	25	22
p(X2)	3.252E-04	2.120E-02	0.000912
uc.RC	0.006158906	0.003885199	0.005268126
uc.CR	0.0822755	0.04692998	0.07906343
std.pearson.residuals			
think	-1.5185519	2.6003370	4.1348409
consider	-2.0274024	0.3152462	-0.2798672
reflect	3.9224946	-0.5524549	-1.7167741
ponder	-0.3442924	-2.4056152	-2.1930795
think	0	+	+
consider	-	0	0
reflect	+	0	0
ponder	0	-	-

D2h. Modality

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$MOD)
X-squared = 77.6298, df = 18, p-value = 2.219e-09

         MOD.ABILITY MOD.FUTURE MOD.OBLIGATION MOD.POSSIBILITY MOD.VOLITION
consider          15         22             30                8            5
ponder             4          3              4                8            4
reflect           10          3             11                8            8
think             12          7             13                1            9

         NONE OTHER
consider  176     3
ponder    221     6
reflect   212     1
think     217     3


$cells$std.pearson.residuals
              NONE MOD.OBLIGATION MOD.ABILITY MOD.FUTURE MOD.VOLITION
think     1.026426     -0.5700332  0.55152176 -0.7706439    1.0688195
consider -5.992465      4.7677244  1.69773791  5.2021043   -0.7203125
reflect   1.452984     -1.0437929 -0.04800047 -2.2520786    0.7259624
ponder    3.562376     -3.1974759 -2.22996629 -2.2201009   -1.0856554

         MOD.POSSIBILITY      OTHER
think         -2.5042632 -0.2086513
consider       0.7810645 -0.1849882
reflect        0.8559612 -1.4314741
ponder         0.8939351  1.8363724


$cells$std.pearson.residuals.sign
         NONE MOD.OBLIGATION MOD.ABILITY MOD.FUTURE MOD.VOLITION
think       0              0           0          0            0
consider    -              +           0          +            0
reflect     0              0           0          -            0
ponder      +              -           -          -            0

         MOD.POSSIBILITY OTHER
think                  -     0
consider               0     0
reflect                0     0
ponder                 0     0


$uc.RC
[1] 0.02717372

$uc.CR
[1] 0.0459609

$uc.sym
[1] 0.03415424
```

D2i. Aspect

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$ASP)
X-squared = 34.6958, df = 9, p-value = 6.741e-05

         NONE ASP.PROG ASP.PERF OTHER
think     218       26        7    11
consider  228       15       13     3
reflect   238        7        3     5
ponder    198       26       12    14


$cells$std.pearson.residuals
              NONE  ASP.PROG   ASP.PERF     OTHER
think    -1.588999  1.954532 -0.7706439  1.036806
consider  1.022596 -1.031954  1.6410412 -2.176505
reflect   4.210502 -3.157310 -2.2520786 -1.293710
ponder   -3.648215  2.228957  1.3833598  2.448204


$cells$std.pearson.residuals.sign
         NONE ASP.PROG ASP.PERF OTHER
think       0        0        0     0
consider    0        0        0     -
reflect     +        -        -     0
ponder      -        +        0     +


$uc.RC
[1] 0.01322523

$uc.CR
[1] 0.03366461

$uc.sym
[1] 0.01899014
```

D2j. Tense

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$TENSE)
X-squared = 74.5949, df = 6, p-value = 4.651e-14

         TENSE.NONPAST TENSE.PAST NONE
think               93        125   44
consider           155         50   54
reflect             80        103   70
ponder              83        100   67


$cells$std.pearson.residuals
         TENSE.NONPAST TENSE.PAST       NONE
think        -1.776397   4.197766 -2.7466397
consider      7.486504  -6.794268 -0.9297671
reflect      -3.184754   1.442471  2.0570221
ponder       -2.573667   1.162986  1.6654280


$cells$std.pearson.residuals.sign
         TENSE.NONPAST TENSE.PAST NONE
think                0          +    -
consider             +          -    0
reflect              -          0    +
ponder               -          0    0


$uc.RC
[1] 0.02676981

$uc.CR
[1] 0.03461199

$uc.sym
[1] 0.03018994
```

D2k. Non-Finite Forms

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$NON_FINITE)
X-squared = 16.4369, df = 6, p-value = 0.01159

         NONE INF.ING INF.TO
think     218      17     27
consider  205      25     29
reflect   183      36     34
ponder    183      40     27


$cells$std.pearson.residuals
               NONE   INF.ING     INF.TO
think     2.7466397 -2.958747 -0.6608673
consider  0.9297671 -1.090991 -0.1339558
reflect  -2.0570221  1.553372  1.1598992
ponder   -1.6654280  2.549700 -0.3577465


$cells$std.pearson.residuals.sign
         NONE INF.ING INF.TO
think       +       -      0
consider    0       0      0
reflect     -       0      0
ponder      0       +      0


$uc.RC
[1] 0.00596748

$uc.CR
[1] 0.01185507

$uc.sym
[1] 0.007938807
```

D2l. Voice

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$VOICE)
X-squared = 115.5607, df = 6, p-value < 2.2e-16

           VOICE.ACTIVE NONE VOICE.PASSIVE
think             214   46              2
consider          150   58             51
reflect           177   70              6
ponder            177   67              6


$cells$std.pearson.residuals
         VOICE.ACTIVE       NONE VOICE.PASSIVE
think       4.7395892 -2.6441291     -4.297597
consider   -4.9634288 -0.5009477     10.189941
reflect    -0.0627575  1.7858334     -2.989344
ponder      0.2712864  1.3996597     -2.944576


$cells$std.pearson.residuals.sign
         VOICE.ACTIVE NONE VOICE.PASSIVE
think               +    -              -
consider            -    0              +
reflect             0    0              -
ponder              0    0              -


$uc.RC
[1] 0.03573874

$uc.CR
[1] 0.06480633

$uc.sym
[1] 0.04607081
```

D2m. Mood

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$MOOD)
X-squared = 54.4997, df = 9, p-value = 1.517e-08

         MOOD.INDIC NONE MOOD.IMPER OTHER
think           187   44         22     9
consider        166   55         30     8
reflect         177   70          3     3
ponder          178   67          2     3


$cells$std.pearson.residuals
         MOOD.INDIC       NONE MOOD.IMPER     OTHER
think     0.9072636 -2.7860644   2.316551  1.505676
consider -2.0348149 -0.8008538   4.886186  1.058927
reflect   0.3253545  2.0114541  -3.502430 -1.311717
ponder    0.8108205  1.6207774  -3.780853 -1.283923


$cells$std.pearson.residuals.sign
         MOOD.INDIC NONE MOOD.IMPER OTHER
think             0    -          +     0
consider          -    0          +     0
reflect           0    +          -     0
ponder            0    0          -     0


$uc.RC
[1] 0.02129636

$uc.CR
[1] 0.03516573

$uc.sym
[1] 0.02652761
```

D2n. Quotative Function

```
data:  table(THINK.BNC2$LEXEME, THINK.BNC2$QUOTATIVE)
X-squared = 93.5371, df = 3, p-value < 2.2e-16

         NONE QUOTATIVE
think     201        61
consider  258         1
reflect   202        51
ponder    239        11


$cells$std.pearson.residuals
              NONE QUOTATIVE
think    -6.426336  6.426336
consider  6.690922 -6.690922
reflect  -4.522496  4.522496
ponder    4.297704 -4.297704


$cells$std.pearson.residuals.sign
         NONE QUOTATIVE
think       -         +
consider    +         -
reflect     -         +
ponder      +         -

$uc.RC
[1] 0.04012669

$uc.CR
[1] 0.1506929

$uc.sym
[1] 0.06337722
```