

**Collocation as the Determinant of Meaning:
from Context of Situation
to Corpus-Derived Subtext**

by

William Ernest LOUW

Thesis submitted for the degree of

Doctor of Philosophy

University of South Brittany

June 9th 2014

Specialisation: English, Corpus Linguistics, Corpus Stylistics

Approved by the jury, composed of:

Professor Marina Bondi, Università degli studi di Modena and Reggio Emilia
Professor Geoffrey Williams, Université de Bretagne-Sud
Directors of the thesis

Reviewers and External Examiners

Professor Elena Tognini-Bonelli, Università degli studi di Siena
Professor Wolfgang Teubert, University of Birmingham

**UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA**

Dottorato di ricerca in Scienze umanistiche

nell'ambito della Scuola di dottorato in Scienze umanistiche

Ciclo XXV

**Collocation as the Determinant of Meaning:
From Context of Situation to Corpus-derived Subtext.**

**La collocazione come determinante del significato:
dal contesto situazionale al sottotesto derivato dal corpus.**

in co-tutela con l'Université de Bretagne-Sud

Candidato William Ernest Louw (firma) _____

Relatore italiano (Tutor): Prof. Marina Bondi

Relatore francese (Tutor): Prof. Geoffrey Williams

Coordinatore del Dottorato: Prof. Stefano Calabrese

Direttore della Scuola di dottorato: Prof. Stefano Calabrese

Contents

Contents	3
List of Figures	7
List of Tables	7
Acknowledgements.....	8
Abstract.....	9
Résumé.....	10
Riassunto analitico.....	11
CHAPTER 1	12
1.0 Introduction.....	12
1.1 The Scope and Nature of the Approach in this Research	13
1.2 Research Problem	16
1.3 Objectives of the Study.....	23
1.3.1 To prompt the move from analogue reading to reading by collocation.....	23
1.3.2 To prompt questioning.....	24
1.3.3 The provision of certainty.....	24
1.3.4 To export collocation into philosophy	25
1.3.5 To improve all ‘applications’ of language study, such as stylistics.....	26
1.4 Hypotheses.....	27
1.4.1 General Hypotheses	27
1.4.2 Hypotheses Relating To Software And Searches	28
1.5 Justification for the Research.....	28
1.6 Conclusion	31
CHAPTER 2: REVIEW OF THE LITERATURE.....	33
2.0 Introduction.....	33
2.1 The adequacy of corpus empiricism involving co-selection in order to bolster or satisfy a philosophical standpoint.....	35

2.2	The validity of working definitions of empiricism.....	36
2.3	Review of the literature relating to re-presentation as it subsists within philosophy and natural language and within a Firthian view	37
2.3.1	The first beginnings of reference corpora and their potential for becoming a scientific instrument.....	37
2.4	Bertrand Russell.....	42
2.5	Sense-data and their collocational aspect	48
2.6	General Guidelines for the inclusion of references in this review of the literature	58
2.6.1	An intention to proceed from philosophy to collocation as a priority and whenever possible.....	60
2.6.2	A review of the underpinning theory within Philosophy and its inaccessibility in Linguistics.....	62
2.6.3	Examples of the reasoning processes and corpus-based verifications set in motion by prompts	67
CHAPTER 3: COLLOCATION, CONTEXT AND CO-TEXT		73
3.0	Introduction.....	73
3.0.1	Trust the Subtext Using the Dialling Method: Summary	74
3.1	Context and Collocation as seen from a Philosophical Perspective	77
3.1.1	Introduction.....	77
3.1.2	Are there any early rivals to collocation within philosophy?	78
3.1.3	Section Two of Wittgenstein's Tractatus as States of Affairs and Context of Situation	79
3.1.4	Wittgenstein's thinking foresees something like KWIC concordances.....	81
3.1.5	Collocation and Wittgenstein's Picture Theory of Meaning compared	84
3.1.6	Developments within Philosophy between 1914 and 1950	91
3.1.7	Towards the provision of proofs of hypotheses	97
3.1.8	Examples of earlier proofs	101
3.2	Conclusion	106

CHAPTER 4: CORPUS-DERIVED SUBTEXT.....	107
4.0 Introduction.....	107
4.1 Subtext and the Notion of Variables.....	109
4.1.1 Variables: from the frequent to the hapax legomena.....	109
4.1.2 Variables and Quasi-Propositional Variables.....	111
4.2 The <i>A Priori</i> as Collocation.....	113
4.3 Subtext as a practical application.....	114
4.3.1 Intuitive opacity and the uncertain ‘destinations’ of searches and their use of entry points for interpretation.....	115
4.3.2 The method revolutionises stylistic investigation.....	117
4.3.3 The Fate of Insincerity Twenty Years on.....	122
4.4 Intuition: Its Roles, Opacity and Deficit.....	122
4.5 The Intuitive Deficit: Dilution and Qualification.....	124
4.5.1 The Intuitive Deficit as Dilution.....	124
4.5.2 The Intuitive Deficit Quantified.....	126
4.6 The Argument for Collocation as Instrumentation within Subtext.....	128
4.6.1 Introduction.....	128
4.6.2 The Method for determining Collocation’s Status as Instrumentation.....	131
4.6.3 The special characteristics of collocation as instrumentation.....	133
4.6.4 Some specific examples of collocation’s spontaneous autonomy.....	135
4.7 Collocation and The Cognitive.....	137
4.8 Developments in the Research of Louw.....	139
4.8.1 Proof of Hypotheses 4 and 6.....	140
4.8.2 Discussion of Hypothesis 6.....	143
4.9 Conclusion.....	154
CHAPTER 5: APPLICATIONS OF COLLOCATION: CONTEXT OF SITUATION TO CORPUS-DERIVED SUBTEXT.....	156
5.0 Introduction.....	156

5.1	The mind-set behind early searches.....	156
5.2	Delexicalisation and Relexicalisation.....	158
5.3	Semantic Prosody	160
5.4	Contextual Prosodic Theory [CPT]	161
5.4.1	Co-selection as a Contextualising Tool	165
5.5	Corpus-Derived Subtext and the Use of its Products in Literary Appreciation 166	
5.6	Subtext as a Practical Method.....	167
5.7	Collocation finds the <i>a priori</i>	168
5.8	The Logical Fallacy Involved in Text Mining.....	169
5.9	Political Spin and Collocation: Intriguing but <i>non-p</i>	172
5.10	Negotiating.....	173
5.11	Branding and Advertising.....	176
5.12	Conclusion	180
CHAPTER 6: CONCLUSIONS		182
6.0	Introduction.....	182
6.1	Simultaneous Paradigm Shifts within the Research of Language itself and within Applications.....	182
6.2	The Value of Intuitive Opacity as a Popperian Strenuous Test.....	186
References.....		190

List of Figures

Figure 3.1: Trust the Subtext using the Dialling Method	75
--	-----------

List of Tables

Table 2.1: Showing Search Procedures	34
Table 2.2: Firth's and Frege's interests compared	41
Table 2.3: Showing a Summary of Main Philosophical Prompts for collocational verification	63
Table 3.1: Philosophy and Linguistics	78
Table 3.2: Wittgenstein in <i>Tractatus</i>	80
Table 3.3: Wittgenstein foresees some of the advantages that were to come with the use of KWIC concordances	82
Table 3.4a: Wittgenstein's <i>Tractatus</i>, section 3	90
Table 3.4b: Wittgenstein's <i>Tractatus</i>, section 4	90
Table 3.5: Empiricism in Urmson	92
Table 4.1: Examples of subtextual searches and their 'destinations'	116
Table 5.1: Practical Applications of Collocation	181

Acknowledgements

I wish to express my gratitude to my supervisors, Professor Marina Bondi and Professor Geoffrey Williams, for their assistance and encouragement, especially under difficult circumstances. These included the death, during the period of this research, of my father, Arnold Louw, as well as the peculiar conditions that obtain in Third-World countries, such as interruptions to the supply of electricity and connectivity, which often made the writing, typing and production of the manuscript especially difficult.

I wish to thank my colleague and friend, Carmela Chateau Smith, for her unfailing encouragement of me to continue with the research and the writing, notwithstanding my many difficulties and for her substantial and unstinting assistance in gradually bringing the manuscript into a better state of preparation and order. Without her assistance, this project may have had to be abandoned.

My thanks also go to Sylvie Lohezic, from the Doctoral School at the University of South Brittany, for her help with many complex administrative matters.

I also wish to thank my typist, Lesley Paterson, in Zimbabwe, for the large amount work that was involved in converting an entirely hand-written text into a typescript. This was arguably one of the greatest challenges that she has ever faced not, only in terms of its scale, but often because the writing took place under artificial light and on battery power when the availability of electrical power fell below 4 hours per day. Typographical errors and the sending of incorrect files were inevitable under such conditions. She often made international telephone calls at her own expense in order to verify that files had been received.

All other shortcomings, apart from the unexpected termination of my employment and salary, are my own. Sadly, the degraded state of internet connectivity in Zimbabwe prevented access to a variety of corpora that may have been appropriate in the case of this piece of work.

Finally, I must thank my mother, Blanche Agnes Louw for her constant encouragement and prayers.

Abstract

The use of ‘prompts’ from philosophy that are capable of verifying meaning through collocation within natural language provides us with highly innovative methods of investigation. This technique involves the provision of increasingly strenuous Popperian tests (1) born of alterations to normal investigative procedure; (2) entirely shielded from bias because they operate with and through human intuitive opacity to the logic of natural language; (3) yielding insights into meaning unobtainable through the intuitive methods that have always dominated the discipline. The method begins as a computational alteration to the ways in which lexical collocation has dominated all discussion of the subject. It moves from relexicalisation through semantic prosody and contextual prosodic theory towards the use of co-selected collocates to establish states of affairs. This process prospects what is likely to occur in texts and also deals with emotions. Its plane is syntagmatic. Subtext involves the paradigmatic plane. Its power is the provision of empiricism for the interpretation by collocation of a text that shares the logical form of other texts, which cannot be seen by human intuition. This process works by splitting logic from metaphysics in natural language, in breach of a central tenet of the Vienna Circle, which ironically invites this situation through its own maxim: the meaning of a proposition is the method of its verification. As part of this process we can expect falsifications to take place within both linguistic and philosophical theory as well as improvements such as finding the *a priori* and the nature of what early analytic philosophers called structured variables.

Résumé

L'utilisation des *amorces* de la philosophie, qui sont en mesure de vérifier le sens par collocation dans le langage naturel, peut fournir des méthodes d'investigation innovantes. Cela implique la mise à disposition de tests popperiens rigoureux : (1) nés de modifications à la procédure d'enquête normale ; (2) sans biais parce qu'ils fonctionnent avec et par l'opacité intuitive humaine à la logique du langage naturel ; (3) donnant un aperçu de sens impossible à obtenir par les méthodes intuitives, qui ont toujours dominé la discipline. Le procédé commence par une modification informatisée de la manière dont la collocation lexicale a dominé toutes les discussions sur le sujet. Il part de la relexicalisation, la prosodie sémantique et la théorie prosodique contextuelle, vers l'utilisation de co-occurrents co-sélectionnés, pour établir des états de choses. Ce processus pressent ce qui est susceptible de se produire dans les textes et traite également des émotions. Son plan est syntagmatique. Le *sous-texte* implique le plan paradigmatique. Sa puissance est de fournir l'empirisme pour l'interprétation par la collocation d'un texte qui partage la forme logique d'autres textes, qui ne peut être vue d'emblée par l'intuition humaine. Ce processus fonctionne en scindant en deux la logique du langage naturel et la métaphysique, détournant la maxime du Cercle de Vienne : la signification d'une proposition est la méthode de sa vérification. Des falsifications se produiront au sein des théories linguistiques et philosophiques, ainsi que des améliorations telles que la recherche d'*a priori* et la nature de ce que les philosophes analytiques appelaient les variables structurées.

Riassunto analitico

Sulla base di proposte filosofiche che suggeriscono di verificare il significato attraverso la collocazione all'interno del linguaggio naturale, si offrono metodi di indagine altamente innovativi. Questa tecnica comporta la faticosa raccolta di tests popperiani (1) derivanti da alterazioni della normale procedura di indagine; (2) del tutto al riparo da pregiudizi perché operano con e attraverso l'opacità intuitiva umana nei confronti della logica del linguaggio naturale; (3) che generano intuizioni sul significato introvabile attraverso i metodi intuitivi che da sempre hanno dominato la disciplina. Il metodo inizia come un'alterazione computazionale ai modi in cui la collocazione lessicale ha dominato ogni discussione sull'argomento. Parte dalla rilessicalizzazione attraverso la prosodia semantica e la teoria prosodica contestuale per andare verso l'uso di collocati co-selezionati al fine di stabilire "stati di cose". Questo processo prospetta ciò che è probabile che si verifichi nei testi e si occupa anche delle emozioni, muovendosi su un piano sintagmatico. Il sottotesto implica invece il piano paradigmatico e consente una base empirica che utilizza la collocazione per l'interpretazione di un testo (che condivide la forma logica di altri testi). Questo processo funziona separando la logica dalla metafisica nel linguaggio naturale: in violazione di un principio centrale del Circolo di Vienna, che invita ironicamente a questa situazione attraverso la propria massima: il significato di una proposizione è il metodo della sua verifica. Come parte di questo processo, si possono prospettare sia falsificazioni della teoria linguistica sia filosofica, sia miglioramenti, come ad esempio una definizione di ciò che i primi filosofi analitici chiamavano variabili strutturate.

CHAPTER 1

1.0 Introduction

Collocation is sometimes defined as ‘*The way that some words occur regularly, whenever another word is used.*’ (COBUILD English Dictionary, 2nd edition, p 309)

I have divided collocation into an analogue period and a digital period. The digital period begins with the OSTI Report in 1970 by Sinclair et al. This report was later republished with a most illuminating interview of John Sinclair, by his successor to the Chair of Corpus Linguistics at the University of Birmingham, Professor Wolfgang Teubert (Krishnamurthy, 2004).

Technically, the work of J.R. Firth on collocation forms part of the analogue period of the subject. However, we need to realise that Firth was an extremely gifted man who had insights during the analogue period that were worthy of the digital period, e.g.,

One of the meanings of night is its collocability with dark, and of dark, of course, its collocability with night.

(Firth, 1957: 196)

Only a fraction of Firth’s work on collocation was published (personal communication in the form of lectures by Professor Frank Palmer in 1975-6 at the University of Reading).

The difference between the analogue and digital periods of collocation is that the former period was characterised by facile, intuitive examples that are in the nature of fixed expressions, such as *salt and pepper* or *bread and butter*. Examples from the digital period demonstrate an altogether deeper sense of autonomy and surprise. It is with the latter period that this thesis is concerned. Sinclair (1991), more than any other investigator of collocation, put in huge efforts to explain not only the complexity of digital collocation, but also aspects of its opacity to intuition.

The early history of collocation is altogether better explained than its current potential in computational terms, for there is far less discussion of the possible applications of collocation after 2010 than at any other time. Even the 2013 volume by Barnbrook, Mason and Krishnamurthy devotes more attention to collocation's past than it offers regarding its potential for future use.

This study sets out to reverse that trend. It is concerned with where collocation can and will take us, rather than where collocation has been in the past, whether distant or recent.

1.1 The Scope and Nature of the Approach in this Research

The nature of the approach to research in this thesis is unusual. It proceeds from philosophy and is applied to linguistics. In this way, the method of philosophy as an ancient discipline assists us in assessing linguistic models of the greatest merit, with the least reliance upon mentalism and intuition. The approach is *bold* in Popperian terms, as will be made plain in Chapter 6.

Furthermore, the arguments of philosophy are tested inside natural language with all of the empiricism of the corpus at our disposal.

For the purposes of this research I am referring to the method of arguing from philosophy as a *logical heuristic*. In fact, a better name would be *quasi-logical heuristic*, because the huge benefits that flow from the use of this type of method within natural language have no direct consequences for philosophical logic, except in the view of Wittgenstein, who enjoins his readers at the outset and in the preface to the *Tractatus Logico-Philosophicus* to realise that:

The book deals with the problems of philosophy, and shows, I believe, that the reasons why these problems are posed is that the logic of our language is misunderstood.

(Wittgenstein, 1922:3; Emphasis added)

Two other great philosophers have also provided inspiration for this work, just as they inspired Wittgenstein almost a hundred years ago, as he recognises a few lines later in the preface:

...I am indebted to Frege's great works and to the writings of my friend Mr Bertrand Russell for much of the stimulation of my thoughts.

(Wittgenstein, 1922: 3).

In fact, Wittgenstein saw more value flowing for philosophy from its treatment of natural language than the other way around. And, with the benefit of hindsight, I now concur with his position.

And so, at the very outset, it may be worth saying why this is the case. The fact is that the interface between linguistic computing and philosophical hypotheses takes the place of those sustained bouts of thinking that so exhausted Russell that he had to sleep for 14 hours after half an hour of deep thought. For it was those bouts of thought that were, at the time, and without recourse to computers, the only method available for the verification of hypotheses. In fact, radio interviews with Russell, now published under the relexicalising title *Bertrand Russell Speaks His Mind*, make it plain just how much disappointment flowed for him from the cumbersome process of verificational thought (Wyatt, 1960).

However, with the benefits of linguistic computing, the disappointments are only now appreciated. Philosophers other than Russell were often 'talked out' of major discoveries far too easily. For example, Rudolf Carnap was effectively talked out of the notion of subtext, which he called the remembrance of similarity, referring to variables in natural language (Mautner, 2000: 86).

I made the decision to argue an approach to research in my own way, (1) as much as possible from philosophy and (2) to verify philosophical principles computationally as quickly as possible, using collocation and a reference corpus. This was not really feasible when working with the earlier mainframe computers, but technological advances since the latter half of the twentieth century have rendered this method viable.

For example, if we take the advice of those who taught as literature at school, we find that their analogue routine was easy to state and remember. I recall a teacher of mine, whose name was Archibald Hatswell, extracting information of a literary critical kind from me by a technique which is now known as 'doing a Paxo' named after the BBC Newsnight interviewer, Jeremy Paxman. Mr Hatswell would find one detail in the text and then single me out with the words: "Now! Get as much as you can out of it! Yes!

More!” until I and the whole class were exhausted. But, after the discomfort of that interrogation and as I slumped back into my seat, I would always *understand* the text better.

The computational version of this is hugely powerful. Confronted with the line from a poem by W.B. Yeats, ‘*Now that my ladder’s gone*’, in order to ‘get as much as you can out of it’, you find yourself asking why the dictionary has no entry for ‘*now that my*’. And then you search for it in the BNC and, finally, in Google Books, but with an understanding also of how poor our intuition is at dealing with such lines. Further reading then discloses that what Hatswell was doing in the poetry class all those years ago was extremely close to what Russell did for most of his life, referring to himself as ‘a Logical Atomist’.

Of course, from there on, and with a computer, it was just a short step to being able to accomplish a *tour de force*: a literature lesson that never once needed to debate vocabulary. This method is now described in detail in *The Cambridge Handbook of Stylistics* (Louw and Milojkovic, 2014a).

But notice that it all hinges upon the *directionality* of argumentation and a prompt from within philosophy that “a logically perfect language” will have “only syntax but no vocabulary whatsoever” (Russell, 1918).

Those few words by Russell bear upon the separability of logic and metaphysics and they lead to the final dismantling of the Vienna Circle using the logical argumentation of a method of which the *Wiener Kreis* itself would, paradoxically, have approved; but all of it modelled in *our* natural language.

The method is fully explained in Chapter 5. Of course, it induces a sense of shock for authors. I well recall the surprise of the delightful Welsh poet, Herbert Williams, when he was shown in an email from me that the subtext of:

A heart made bleak by sacrifice

was

An opportunity made possible by

He waived the payment of all royalties payable by me for the use of his Collected Works, with the proviso that I set out his authorial and publication details in full (Williams, H. 2007). The poem was about a dishonest and opportunistic woman in charge of an infirmary.

Examples of this kind led to further insights into both linguistics and philosophy. With the discovery of the *a priori* now virtually settled, this approach will continue to revolutionise other applications, such as stylistics, even though the philosophers themselves at this time may still prefer, for sentimental reasons, to preserve Kant from too much forensic transparency.

1.2 Research Problem

The research problem with anything related to collocation is unique. The reason for this is that Firth, notwithstanding the signal absence of computers, had stated the primary function of collocation with greater accuracy than might have been the case even if computers had been available. I offer a proof for this in Chapter 5.

The implications of Firth's saying that the mere presence of a collocate is capable, reliably, of singling out one meaning of a stated form simply by means of its proximity with another term was inspiration enough to prompt Sinclair to set about providing proof. Even in science, insights of such high quality are rare.

Even more exciting was the fact that the proof sought was finally to be found at the highest degree of empiricism through the nascent technology of computing. Replicability was to be high. The findings, even by today's standards, cannot be faulted, and it is for this reason that the study in this thesis is unashamedly *qualitative*.

Sinclair's computing had been completed by 1970 with the 'publication' in written form of the OSTI Report (Krishnamurthy, 2004), with no more backup than could be provided at the time by means of photocopiers and carbon paper. Even the British Library 'lost its only copy' (Sinclair, personal communication). In fact, the only remaining copy was offered to Ramesh Krishnamurthy as a form of light reading. His reaction (personal communication) was to tell John Sinclair: "*This is gold. Would I have your permission to publish it?*"

In this way, the real qualitative investigation of the capabilities of collocation could flow out of its main achievement in Cobuild and lexicography, and *back* into the excellent interview of John Sinclair by Wolfgang Teubert (Krishnamurthy, 2004). And what becomes plain from a close reading of that interview is the huge and generous candour of Sinclair, as he tries to recall what was involved in ‘trusting the text’ in those early days. Even the extent to which there was scope for description or even doubt was dispelled by the power of collocation itself.

The fact that the main qualitative principle involved in meaning by collocation, namely that ‘one meaning of X is its collocability with Y’, *restricts* meaning rather than expands it, was at the time, even with the power of collocation in evidence, something that Sinclair never took up with either Firth or Halliday.

This thesis therefore comes in the wake of the quantitative proof of the power of collocation, as well as of its qualitative proof, both within the research of Firth, some of which has been lost (Palmer, 1968), and with the diagnostic proofs and practical applications recorded in my own work, from relexicalisation in 1989 to 1991, through semantic prosody in 1993 and Corpus Prosodic Theory in 2000, to deeper forensic treatments of corpus-attested terminology and subtext, from 2008 to 2014. These latter developments followed on naturally from the huge successes of the Cobuild dictionaries that constitute a paradigm shift within lexicography. My recent research was inspired also by John Sinclair’s largely inexplicable disappointments: the non-appearance of a dictionary of collocation; the stalled Project LUCID and the now stalled CAIN Scottish Schools corpus-based search engine, designed expressly in order to ‘chunk’ contexts of situation and put an end to the more intuitive aspects of word lists (see Philip, 2011), just as Gottlob Frege had foreseen, as the pre-Firthian founder of contextual studies within philosophical logic.

This current study is timely because linguistics stands at the crossroads of the admission of *subtext*, which will revolutionise the stance of the discipline: the power of collocation will grow as intuitive opacity and the intuitive deficit impel the new generation of linguists to seek verification *from* corpora rather than the invasion *of* corpora by scholars desirous of preserving a cognitive that is incapable of ‘cognising’ the lexical variables of grammar strings.

The impending revolution has now moved into philosophy, with the use of collocation (amidst many painful falsifications within linguistics) to discover Kant's synthetic *a priori*. It is taking hold also in stylistics, as subtext is now capable of providing through the Russellian argument of logical atomism, forms of Carnapian 'necessary meaning' that relieve the critic of all unnecessary debate about the special meanings involved in the 'fictional world' of a poem as a 'possible world', both in and as part of the external world. This process is likely to continue, as further applications, notably in NLP, take the objectives of the research further, as explained in Chapter 5.

There is also the ancillary problem of what will be lost to the world and to the benefits of science and literacy if collocation in a post-Cobuild world is simply packed up and thrown back into the toy-box, its task deemed to have been fulfilled.

Serious delusions remain to be dispelled. An anti-Popperian stance needs to be removed from within corpus linguistics (McEnery and Hardie, 2012:14) and its applications, such as corpus stylistics. We are told by the proponents of this stance that meaning is pluralistic and that theories may be selected in a spirit of eclecticism without any recourse to a deep consideration of their merits. This may be a good moment for seeing a domino-effect sweep through linguistics. Just as John Sinclair predicted (personal communication) that semantic prosody would 'run and run', so we need to foresee that the current edifice of linguistics may finally become the history of linguistics, with subtextual linguistics as the discipline's new quark. Contributions to theory are always rare and often unwelcome, but subtext offers us unique insights that are crucial for the expansion of theory and cannot be gainsaid.

A major concern of collocation research by computational means surrounds the notion that collocation always exercises some degree of *hidden* meaning, even where the collocates appear to be monolithically constant. For example, one of the very first issues to become apparent in the case of a settled semantic prosody such as *cause* was that even the *exceptions* to the prosody were found to bifurcate into forms of *irony* and *insincerity*. Between them, they occupied a mere 3% of a concordance (Louw, 1993) and, but for the use of the computer, may not have been noticed.

However, irrespective of the particular phenomenon using collocation, the moment at which its function is carried out takes place largely *subtextually*, but always *verifiably*. The fact that the ‘logic’ or argument of any sentence will dominate the vocabulary that occurs in the sentence presents us with a new and compelling argument for creating lexicographic access to grammar strings and their lexical collocates.

It is worth noting that the collocates in question in this category are essentially *all* to do with lexis. As far back as 1966, Michael Halliday urged his readers in a volume entitled *In Memory of J.R. Firth* to disregard the collocates of grammar words or grammar strings (Bazell *et al.*, 1966: 153). It was only within philosophy that the logic of grammar words was declared to be empirical (Dummett, 1976). This development assisted in the process of uncovering *subtext*, using a reference corpus and with some theoretical assistance from Bertrand Russell and Ludwig Wittgenstein. Neither of these philosophers lived to see the advent of linguistic computing. Only Quine may have lived long enough to see computer-generated concordances.

Several final pieces of inspiration responsible for shaping this thesis came from the one scholar who had never heard of collocation (even though he had met J. R. Firth). Roman Jakobson’s influence on my research is dealt with in an article that arose from a keynote paper I delivered at the University of Bologna for a conference entitled *Verbal Art Re-Visited*. The conference was held in honour of the contribution to stylistics of Ruqaiya Hasan. A substantial portion of my keynote address was devoted to the potential for using collocation to improve the practical aspect of Jakobson’s Poetic Function (1960; 1968) by and through the use of collocation (Louw, 2007b).

My analysis will not be repeated here, but I do now wish to record my gratitude for Jakobson’s inspiration in *three* crucial aspects:

1. Jakobson placed on record, in a number of scholarly works, his belief that poetry embodies the very best ‘discovery procedure’ for linguistic science;
2. With his Poetic Function, Jakobson used the paradigmatic axis to store up lexical forms for later projection into a poetic work;
3. The fact that *projection* in a very short poem, such as a Japanese *haiku*, is factually impossible in syntactic terms, but easily accommodated by a collocational model confirmed my perception that Jakobson’s two axes,

syntagmatic and paradigmatic, ought, with the use of appropriate software, to form the basis of my treatment of collocation in this thesis (Louw, 2007b: 154).

However, for this to be set in place, a further Firthian condition for the nature of collocation would need to be defined in corpus-attested terms (Louw, 2008). And this is Firth's notion that collocation is '*abstracted* at the level of syntax' (Firth, 1957). The expression caused a great deal of difficulty for a long time for the obvious reason that something that is averred to subsist ought to be recoverable in an observable target text. But Jakobson agrees that certain forms are (1) suitable for later *projection*; (2) stored up in readiness for such projection and (3) remain only putatively likely unless and until they *are* projected. For most readers they are not yet regarded as physically present and although their position in the text remains probabilistic, they are effectively not *there*. The corpus, with the arrival of computers, reveals that such collocates are both available and likely to appear. But they may not yet have done so. It is the solution to this dilemma that Firth terms *abstracted* (in Latin, 'drawn away from'). And it applies perfectly cogently to our computational context that, where the un-actualised is, it remains *hidden*.

However, it is the direction of abstraction that is crucial. For the Poetic Function, the terms to be projected are largely lexical collocates and the pool from which they find themselves projected into text will in all likelihood resemble what used to be termed a *lexical set*. Their role is often to settle the equilibrium that obtains between *context* and co-text. The direction of their drawing away or abstraction feels determinedly lateral or horizontal. They may even be of roughly the same frequency. The direction seems horizontal, like actors waiting to walk on to the stage.

And so it is that this thesis is structured, with the assistance of philosophy, around two complementary and verificatory forms of software within corpus linguistics: *co-selection* as context of situation and *wild-carding* as the quasi-logical subtextual underpinning of argument as the functional commentary on states of affairs in the external world and in its fictional counterparts.

A final difficulty needs to be mentioned even though the method of this study banishes it: authority is restored to the text that John Sinclair enjoined us to trust (2004). This also means that the old notion of monitor corpora now goes to the wall. Language is its own instrumentation and the extent of intuitive opacity is far too daunting for us to risk retaining unreservedly the right to monitor.

The phenomenon of subtext is fast turning into theory in its own right. Paradoxically its evolution was prompted in large measure by a combination of factors such as distance, poor connectivity and the restrictive price of subscriptions to large corpora. When all one has access to is a 44.5-million-word newspaper corpus made especially for me by the late Tim Johns (Times 1995 Corpus, Johns, 2005), one finds oneself asking questions like: “*Suppose all the logic in text is empirical?*” or “*Will the logic in text have lexical collocates and, if so, what will they look like?*” Questions of this kind led me to abandon the notion of ‘horses for courses’ corpora (see Chapter 5) and demonstrated that when it comes to subtext the corpus will both act as its own monitor and accept monitoring even from target texts, as the brief example below demonstrates. The discovery procedure behind it was, of course, Jakobsonian.

I have begun a project to determine whether the index of first lines in poetry can determine through prospection the issues that are likely to arise within the rest of any poem’s fictional world.

If we wild-card the first line from the poem by Philip Larkin that follows, using the MicroConcord concordancer¹, we find that its *a priori* ‘persons, personalities’ (Firth, 1957) are *trees* and a *songbird*. The argument in Fregean terms is that nature will respond to the seasons year upon year. However, one delexical ironic use in the Times 1995 Corpus² (Johns, 2005) suggests that team players in rugby or football games will follow the arrival of some unstated season and start to play dishonestly. Nature is programmed to do God’s will. Football players do not need putative seasons and their purported effects to prompt dishonesty. They have free will. Note how the poet foresees the question of a *trick*. (See also Louw, in Golinelli, 2008: 391). The *a priori* outnumbers the delexical and shows it the door. A poem monitors a corpus.

¹ The MicroConcord software program is a concordancer created by Tim Johns and Mike Scott, which operates under DOS. It is freely available to download as a zip file from Mike Scott’s website. <<http://lexically.net/software/index.htm>>.

² A 44.5-million-word newspaper corpus, created by Tim Johns in 2005.

The Trees

The trees are coming into leaf
Like something almost being said;
The recent buds relax and spread,
Their greenness is a kind of grief.

Is it that they are born again
And we grow old? No, they die too.
Their yearly trick of looking new
Is written down in rings of grain.

Yet still the unresting castles thresh
In fullgrown thickness every May.
Last year is dead, they seem to say,
Begin afresh, afresh, afresh.

Philip Larkin³

MicroConcord search SW: **the *s are *ing into**
80 characters per entry Sort : 1R/SW unshifted.

1 ard. Along the country lanes, the yellowhammers are coming into full voice calli
2 ide by the back lines or when the forwards are coming into rucks and mauls from
3 mission spokesman said. But the Germans are going into tomorrow's showdown mee
4 is moving out of its torpor. The pieces are moving into place for an economic l

Source: Times 1995 Corpus

The four concordance lines above share the same logical form and their relationship to one another is subtextual. The closest match to the target line: the trees are coming into leaf, is line 1 which has the *a priori* sense of marking the seasons and, in this case, the season of spring. However, note how the import of the other lines is, by comparison with the *a priori*, strongly delexical, imparting an ironic sense that aspects of sport, including cheating, are purportedly seasonal also.

³ A recording of Larkin reading this poem is online at <<http://www.poetryarchive.org/poem/trees>>. The poem was first published in 1974, in the book *High Windows*.

1.3 Objectives of the Study

By comparison with philosophy, linguistics is a relatively ‘young’ discipline and in many ways a largely shapeless one. Its phonology, morphology and grammar are well established. It is weakest at the level of meaning; but at that level, collocation is, in my respectful view, a science that is as hard and as settled as physics. My objectives for the research, therefore, are as follows:

1.3.1 To prompt the move from analogue reading to reading by collocation

This thesis sets out to regard collocation as the digital counterpart of analogue reading and literacy.

This will necessarily imply a *comparison* between the almost entirely *mentalist* act of ordinary reading and its new computer-assisted and data-assisted counterpart. The corpus is instrumental in this process; its integrity is vital.

The benefits of the digital approach cannot be stated unless and until we can quantify through analysis the *intuitive deficit*, which notion I shall expand upon later in this thesis. As we are now able to quantify what can be lost as a result of turning our backs on progress in linguistic theory, we should draw this loss to public attention.

And so, if we are to provide what is essentially a new form of literacy, we need to be certain that collocation thrives in all applications of literacy mentioned in this thesis (see Chapter 6). However, if we are setting our new and broad objectives and *uses* for literacy, we can probably do no better by way of capturing those objectives than to take the advice of a renowned Jesuit, Walter J. Ong, honoured for his work in the field of literacy and technology, particularly his 1982 book, *Orality and Literacy: The Technologizing of the Word*.

It is worth noting Ong’s view that successive forms of literacy become *consumed* by their successors (Ong, 1982). However, all is not as lost as Ong supposes, because, as we shall see in Chapter 6, *subtext* restores to us the primacy of the past; what Professor Vita Fortunati refers to as its *primitivismo* (Golinelli, 2008: 22), not from the Far East or the Polynesia of Gaugin, but universally within our own settings, even if they are today swamped by world capitalism and poverty. It is the conversation that Firth sees as the

precursor to language study. If such conversations were, in the past, unable to disclose the ideologies and the dynamics of power, they can today be brought to light through the use of collocation. The *a priori* has primacy over the modern because *nature* is fully lexicalised and the pure and the delexical are involved in the daily efforts of mankind and truly delexical albeit in terms of *primitivismo* as Ayer informs us (Golinelli, 2008: 395). Golinelli provides us with unique insights into the *metamorfosi* in which *il primitivismo* often lurks undetected. The collocates of subtextual argument will flush them out into the open. In this way, they will not be lost to the form of new literacies or in the manner and suddenness of their arrival.

1.3.2 To prompt questioning

A second objective is to prompt a questioning of what we take for granted. The most celebrated example within philosophy is the way in which a straight stick looks bent the moment it is immersed in water. Our new literacy involving collocation needs to prompt us to trust our senses and our intuition less and our instrumentation more; no matter how certain we may feel about being right. A rather expensive bent stick may be bought dearly as we use our intuition to unpack the meaning of *natural justice* in court, only to find later that part of our ‘prayer’ was to *ask* for natural justice when a simple concordance would have shown that its collocates make it untrustworthy.

In our society, there are many institutional ‘bent sticks’ and it is, or ought to be, the task of literacy to find them and, where possible, to show what would be involved in making them straight.

1.3.3 The provision of certainty

A third objective is to provide a reliable source of certainty. Earlier forms of literacy depended upon the reader’s stored prior knowledge for their interpretation (however limited that may be). However, the literacy involved in a digital collocational approach allows for a single utterance to be read by so large a reference corpus that it is, in effect, *allowing a sample of the world to read a single instance of the word*. It allows, for example, a single and apparently unique use of the term *natural justice*, in a case in an

authentic court setting to be read by the collocates of all its instances in a reference corpus. This example will be further discussed in Chapter 2.

Of course, the normal checks and balances for certainty apply, but they do so with evidence, as one encounters Wittgenstein's caveat below from paragraph number three of his book *On Certainty*, published posthumously in 1969:

- ‘3. *If, e.g., someone says “I don’t know if there’s a hand here” he might be told “Look closer”. This possibility of satisfying oneself is part of the language-game. It is one of its essential features.*’

(Wittgenstein, 1969: 2c; Emphasis added)

However, with or without a corpus, there are some problems that arithmetic cannot solve, but algebra can.

1.3.4 To export collocation into philosophy

It is plain that this study of collocation involves a spin-off in two directions and for two disciplines: the flawed recent discipline of linguistics and the ancient, refined discipline of philosophy.

Preaching the power of collocation to linguists is largely a lost endeavour, as will be stated repeatedly elsewhere in this thesis. However, if one is able to convince philosophers of its value, it is the custom in philosophy for philosophy to hand over to *science* all findings that involve certainty. If this can be prompted, collocation may then be re-imported into a linguistics *that it then proceeds to falsify*, and collocation then achieves a goal unforeseen even by Popper of becoming *instrumentation*, i.e., a living document (Sinclair, 1991). Its functions would be akin to those of a monitor corpus, such as AVIATOR⁴ (Renouf, 2007) but operating largely subtextually to determine the arguments that underlie all utterances in authentic natural language (it would not apply to made-up examples, such as those of Chomsky). This move would bring into being Subtextual Linguistics. It would be different from previous movements. For example, it

⁴ The AVIATOR monitor corpus was a project developed by Antoinette Renouf in Birmingham from 1990 to 1993, using text from the Times newspaper dating back to 1988. AVIATOR stands for ‘The Analysis of Verbal Interaction and Text Retrieval’ (Renouf, 2007).

would *not* be a mentalist school of thought, but a scientific instrument. It would have no scions to purport to know that which is opaque to intuition. Instead there would be only a theory as a self-improving heuristic: a dynamic *p* in search of *non-p* in all language corpora and naturally occurring text. A major duty of philosophical truth is that *p* finds *non-p* and *identifies* it with a pointing finger and shows it the door.

1.3.5 To improve all ‘applications’ of language study, such as stylistics

A new instrumentation would be the welcome vector of falsification for all intuitively derived theories. It can already be proved that in the same way that Frege broadened Aristotelian logic, subtext will expand the reliability of corpus stylistics on a huge scale.

For example, in areas of corpus linguistics that see themselves as purveyors of pluralistic truth and eclecticism, subtext will identify and state the ‘intuitive deficit’. In other words, the extent to which theories identified for falsification can have their misleading element quantified, stated and corrected. It will usually lead to whole theories being discarded. All of the examples in a textbook like *Stylistics* (2012) by Jeffries and McIntyre would be re-appraised and their intuitive deficit stated. In books such as *Key Terms in Stylistics*, collocation would be used to demonstrate what would be required in order to make the key terminology they contain corpus-attested (Louw, 2008). Gradually, the new science will take shape.

It needs to be made clear that *no* duress would be involved. Falsification simply involves loss of credibility and this causes erstwhile supporters and enthusiasts to vote with their feet. Theories that are holed below the water-line quickly ship more water than can willingly be bailed out. Truth is also very converting. Nobody is forced to be truthful. The trick in life is to hide the truth and offer a beguiling substitute; and in the case of very old theories that pre-date the instrumentation that is capable of falsifying them, *sentimentality* is often the main reason for their retention, even in the face of compelling evidence of their inadequacy. It may well be that this phenomenon is even more ingrained in academe than anywhere else. A disturbingly high number of academics will often prefer face collegiality to truth.

Some of the objectives above may take a long time to be brought about in practice; but Max Planck once offered the observation that change that appears impossible amongst

adults, especially the elderly, comes into being effortlessly and without argument or ideology amongst the youth. They are our great hope for the emancipatory use of collocation in order to benefit mankind. The only obstacle now is that pedagogy remains in the hands of teachers who may be prepared to subject science to ‘damage-control’.

As Frege once memorably put it:

‘A proposition no more ceases to be true when I stop thinking of it than the sun ceases to exist when I shut my eyes.’

(Kenny, 2000:52)

1.4 Hypotheses

1.4.1 General Hypotheses

1. That there will always be a direct correspondence between Firth’s taxonomy (arrived at in consultation with Malinowski) (1957) and what collocation reveals both in particular states of affairs and in their outcomes.
2. That all literary ‘devices’, including humour, use the mechanism of collocation as relexicalisation.
3. That collocation in action confirms Sinclair’s observation that collocation reveals the function of utterances; and that this is especially the case in two circumstances: (1) Semantic Prosody, including Contextual Prosodic Theory, and (2) Forensically (Russell, 1948).
4. That subtext derived from reference corpora using methods inferred from Russell and Wittgenstein will always and with accuracy bear upon the state of affairs of which the target string is authentically a part.

1.4.2 Hypotheses Relating To Software And Searches

5. That, using *co-selection*, collocates within the span of collocative power (Sinclair, 1991) will both *spontaneously* (1) *chunk* contexts of situation, and (2) *relexicalise* in order to do so.
6. That, using the *wild-carding* of all vocabulary items, strings that have had vocabulary so wild-carded, will (1) *reveal* from reference corpora quasi-logical arguments that relate directly to the concerns of the states of affairs from which the target string has been drawn; (2) *operate* at levels of *depth* that are unable to be determined using *intuition* and (3) are not bound, in their own occurrence, by the genre and provenance of the original, and (4) may, through any lexical collocates that may be accumulated during the process of searching, reveal both *cultural* norms, as well as insights into the origins of philosophical phenomena, and (5) may falsify established linguistic classifications (such as *synchronic* and *diachronic*) and (6) have *primacy* in the sense that they are close to the origins of argument and that such argument confirms the status as ‘pure’ metaphysics of the majority of lexical items that were wild-carded in order to retrieve it. This latter development has already been applied in corpus stylistics and in the philosophy of language.

1.5 Justification for the Research

If, as I have averred in many publications and conference papers since 2004, collocation is both a scientific fact (Louw, 2010a) and instrumentation for meaning (Louw, 2008; and in the title of this thesis), then we ought to be able to find in literature that is fully up to date, reference to the very latest material on the *applications* of collocation.

However, a search of the relevant literature is almost totally and inexplicably silent on the matter. I opened the very latest book on collocation (Barnbrook et al, 2013) only to find the usual references, with nothing up to date. My work on subtext was published in 2010, albeit fairly obscurely, and yet by 2013 nothing had registered with the authors of the latest study on collocation. Stewart (2010) deals with semantic prosody as though all contributions are of equal value to the point that the science of semantic prosody and contextual prosodic theory looks as though its truth must be as pluralistic and eclectic as

McEnergy and Hardie (2012) make it look, and Jeffries and McIntyre (2010) make both eclecticism and pluralism in stylistics mandatory, with no proof offered and with no reference to subtext.

As a result, there are deficits in the treatment of collocation as science that this thesis seeks both to identify and to correct.

To be fair as well as critical, Barnbrook et al (2013) not only offer too few applications of collocation but their book contains a paucity of modern or even up-to-date references. My contribution to corpus stylistics is hardly ever mentioned in their text and not at all in McEnergy and Hardie (2012), or in Jeffries and McIntyre (2010). I started Corpus Stylistics using Cobuild reference corpora of 21 million words in April 1987. It was launched at St Hilda's College, Oxford in the presence of Ron Carter. I was the only stylistician who had access to more than one million words of running text at that time.

Here is a list of the potential uses and practical applications of the corpus approach that I have written about:

- 1991 Relexicalisation and literary devices (see Appendix, Paper 1, 1991)
- 1993 Semantic prosody, irony and insincerity (see Appendix, Paper 2, 1993)
- 1997 The contribution of corpus linguistics to critical literary appreciation
- 1999 Corpora and metaphor (see Appendix, Paper 3, 2000)
- 2000 Contextual Prosodic Theory including a proof of the Irony/Insincerity hypotheses (see Appendix, Paper 4, 2000)
- 2003 A falsification of the Truth and Reconciliation Commission of South Africa (see Appendix, Paper 5, 2003)
- 2004 Collocation as the determinant of Verbal Art. Keynote in Bologna. (see Appendix, Paper 7, 2007a).
- 2004 Closing Keynote for TALC6 in Granada to introduce the use of philosophy in collocation studies (see Appendix, Paper 8, 2007b)
- 2004 Keynote at Oxford Text Archive on Collocation

- 2004 Keynote presentation on Collocation in Primitivismo (Bologna)
- 2006 Applications of collocation to spin and negotiation theory (see Appendix, Paper 6, 2006)
- 2008 Article that urges that the terminology of corpus stylistics be corpus attested (see Appendix, Paper 9, 2008)
- 2008 Subtext launched at Queen's University, Belfast
- 2009 Application of collocation to Humour Studies (Keynote). Lucca and Collodi, Tuscany (see Appendix, Paper 10, 2009)
- 2009 Corpus 09, applying philosophical postulates to events (see Appendix, Paper 11, 2009)
- 2009 Contribution to the Bootcamp Debate
- 2010 First book chapter on subtext (see Appendix, Paper 12, 2010a)
- 2010 Subtext. Genoa PALA. ELR 50 Years since Firth's death (see Appendix, Paper 13, 2010b)
- 2011 Corpus11. Subtext (Birmingham)
- 2012 Subtext in the work of Mtshali (Namibia PALA)
- 2013 Collocation falsifies linguistics (Spanish Corpus Conference, Alicante)
- 2013 PALA Heidelberg: Collocation and subtext falsify the Cognitive
- 2013 Moscow NGIMO Conference: Collocation and the *a priori* (see Appendix, Paper 14, 2014)
- 2014 Dortmund. 'Collocation finds the *a priori*'
- 2014 ICAME. Nottingham. Whither the monitor corpus?
- 2014 CILC. Spanish Corpus Conference. Las Palmas. On new corpus pedagogy.
- 2014 IGEL. Turin, July. On Empiricism.
- 2014 4th International Convention on Humour. Milan. On dark humour in subtext.

An article co-authored with Marija Milojkovic in the *Cambridge Handbook of Stylistics* published in 2014 and edited by Peter Stockwell and Sarah Whiteley.

A forthcoming book (in press, autumn 2014) co-authored with M. Milojkovic entitled *Contextual Prosodic Theory* published by John Benjamins, Amsterdam.

The applications of collocation and the role of philosophy in refining our thinking in the area of collocation, and the advent of subtext with its huge potential and myriad applications, *must*, between them contribute substantially to the justification for this study, even *before* we take into account any contribution to theory.

1.6 Conclusion

Although the method which will be used in this research is sometimes difficult and often abstruse, it is not entirely unobserved within philosophy as Broad (1963: 65) observes:

*'Suppose that the results of taking a synoptic view of a number of different mutually relevant departments of knowledge or belief were to show that they all obviously fit together without difficulty into a single coherent whole. Then there would be little or no occasion there for philosophy. But actually, as illustrated by my examples, this is often not the case. It often happens that each of the several mutually relevant regions, which we habitually contemplate and react to separately, gives rise to its own system of concepts and principles; that each such set seems fairly satisfactory and coherent in isolation; but that, when we contemplate the various departments together, we find that the several sets of concepts and principles seem **prima facie** to conflict with each other. It is synopsis, revealing **prima facie** incoherence, which is the main motive to philosophical activity.'*

When we are faced with apparent conflict between different sets of propositions which, taken separately, seem to be satisfactory and evidently true, the most obvious first step is to try to analyse the terms in them and to formulate the propositions themselves more carefully. Such a process is an indispensable step towards deciding whether the inconsistency is real or only apparent, and towards formulating it precisely if it is real. And this is a precondition of any reasonable attempt to deal with the difficulty.'

This synoptic method has been modified with a view to testing prompts that arise from the practice of analytic philosophy against evidence provided by collocation from reference corpora. In the next chapter, the method is elaborated more fully and in Chapters 3, 4 and 5 collocation's contextual and subtextual aspects are tested more strenuously against philosophical prompts.

CHAPTER 2: REVIEW OF THE LITERATURE

2.0 Introduction

This thesis deals with the extent to which collocation and the philosophy of language are capable both in principle and in practice of evolving a form of *instrumentation* for meaning that will be acceptable as a form of science. The main purpose of the proposed method will be to examine the ability of collocation to re-present experience *digitally* or *computationally*. The analogue models of *knowledge* or *human* knowledge are, in large measure, those set out by Russell throughout his long career, starting with *Our Knowledge of the External World* (Russell, 1914) and ending with *Human Knowledge: Its Scope and Limits* (Russell, 1948). Part of the scientific approach evolved by Russell relates to Logical Atomism and our ability to isolate ‘atomic facts’ as stated by Russell in his preface to Wittgenstein’s *Tractatus Logico-Philosophicus* (1922) and as amended by him throughout his long career.

This study will devote more attention to Russell’s work after 1918 than it will to his work undertaken during the years of the First World War (1914-1918). Although Russell spent part of the war in prison as a conscientious objector, this is not the main reason for concentrating on the work that he began in 1918.

The main reason for dealing with Russell’s work *after* the war was that his thinking moved away from the fairly naïve aspects of sense-data to be found in Russell (1914) and towards a greater concentration on *events*. Events are of greater importance because of their deeper connection with the philosophy of Gottlob Frege, on the one hand, and the notion of Context of Situation which is crucial to all of contextual linguistics as set out by J.R. Firth in a paper entitled, ‘Personality and Language in Society’ and published in 1950 in *The Sociological Review*, XIII and reprinted in Firth (1957).

The crucial aspect of this paper is that it provides us with a taxonomy of Context of Situation which is neatly tabular and allows us to determine at each of its levels A-C, the extent to which collocation is an indispensable part of any attempt to deal with all aspects of situational context, from ‘the features of participants’ to the outcomes of communicative events that Firth and Malinowski call ‘the effect of the verbal action’.

In treating this taxonomy as a blueprint for automating John Sinclair’s desire to ‘trust the text’, we need also to add the criterion that all intuitive procedures be gradually excluded from the methods that are used in order to evolve our instrumentation. Hence, by the time we are dealing with subtext in Chapter 4, our reliance upon the empiricism of grammar strings will be operating within a situation of almost total intuitive opacity. This objective springs from Firth’s famous dictum that collocation does not concern itself with the *ideas* meaning of words.

Furthermore, this method has had to undergo considerable refinement during both research and writing. For example, it was assumed at an early stage of the work that *co-selection* and *subtext* were largely separate entities, the former operating at a level of abstraction from syntax as proposed by Firth (1957: 181) and the latter entirely intra-sentential (Louw, 2010b). Subdivisions of a kind that could not be anticipated arose and these were often so compelling as to force upon the investigator new taxonomic distinctions that could not be foreseen at the time. For example, the lexical collocations of grammar words and strings acting as *subtext*, turn out to be lexical forms that have been imbued *by grammar word collocates* with a form of protection from semantic change. They became impossible to ignore because they were only recently found to be capable of falsifying Saussure’s distinction between diachronic and synchronic and Halliday’s (Bazell *et al.*, 1966) assertion that the grammar word collocates of lexis were perhaps best disregarded. As findings of this kind are assimilated into the research, they begin to initiate and accomplish new directions and goals fairly quickly.

A tentative and probably still incomplete list of criteria for the role of collocation in the re-presentation of empirically recoverable experience or *events* is set out in the table below. As fresh discoveries are made, the list will, of course, need to be altered, often radically.

Table 2.1: Showing Search Procedures

1.	Co-selection as a chunking mechanism for situational as well as lexicographic meaning.
2.	Wild-carding of lexis as the main step in discovering and recovering subtextual empiricism, entirely independently of all <i>intuitive</i> ratiocination.

It is not a major concern of this research to present the philosophical work of the English empiricists. Our interest begins where we see manifest proof of a desire for empiricism that can only be met by the eventual arrival of computers and an ability to establish the variables that are *known* to exist but which because of intuitive opacity, simply cannot be seen or determined. However, as one begins to approach the period of analytic philosophy, it will be of great importance to point out how exactly the vision of these modern philosophers may well have been improved by linguistic computing if it had been available.

A good example may be provided to demonstrate the extent to which tautologies occur or fail to occur in natural language. I am indebted to Professor Francois Rastier (personal communication) who assures me that his searches for tautologies using the computer always met briefly with apparent success, followed by the realisation that the forms of wording found *all* turned out to be *states of affairs*. A case in point is the expression ‘business is business’. As a semantic prosody, the expression identifies the point at which, and using the exact words as a fixed expression, a person who is being taken advantage of financially makes his/her unhappiness plain to the person who is benefiting from the situation. That person finally declares that, considerations of friendship aside, he/she can no longer endure the financial hardship incurred by offering discounts that cripple their business as an enterprise.

2.1 The adequacy of corpus empiricism involving co-selection in order to bolster or satisfy a philosophical standpoint

It is important to note that almost all pre-computational approaches to empiricism that are experienced by a reader/hearer for the first time, will, even today, be found to rely extensively on explication by means of *concepts*. In fact, concepts are said to have a key role to play in the explication of meaning as late as McGinn (1981: 90), who comments favourably on Quine’s provision of the Third Dogma of Empiricism (1981). She also addresses Donald Davidson’s view that if Kant’s First and Second Dogma were removed along with Quine’s Third Dogma that sets out to replace them,

‘...it is not clear that there is anything distinctive left to call empiricism.’

(Davidson, 1973: 11)

McGinn refers to the fact that there is broad consensus also among the following scholars of the day: Richard Rorty (1980), Michael Williams (1977), David Wiggins (1980).

All of these studies are pre-computational and, as such, rely upon a technique for approaching meaning not only as

‘... independent of the mind’s cognitive activity...’
(McGinn, 1981: 89)

but dependent upon the mind’s interpretation of the given - the *conceptual scheme* – by means of which the mind puts a particular construction on what is given (McGinn, 1981: 90). Such a construction is in breach of Firth’s requirement that collocation has nothing to do with the ‘ideas’ meaning of words (1957:181).

Of course, McGinn was writing in 1981 and by that date computers were available, but we were still five years away from ‘large’ corpora such as those that were used to produce the first edition of the *Collins Cobuild English Language Dictionary* (1987). And even today, in March 2014, at a conference on empiricism in philosophy and linguistics held at the Technical University of Dortmund, Germany, only one paper out of six has any reference to a corpus and its use, even then, is naïve: use of a corpus for a traditional task in distributional semantics and that with no reference to collocation even though Firth is mentioned.

2.2 The validity of working definitions of empiricism

*‘In its ordinary use, the term “empiricism” (from the Greek “**empeiria**” = “experience”) means the employment of methods based on practical experience, rather than on an accepted body of theory. But in philosophy the word is used only in a **quite different, and technical sense, to refer to the philosophical theory that all knowledge is derived from experience.**’*

(Urmson, 1960:122)

It is providential that corpora enforce both of these emphases upon the investigator, because the concerns of this thesis are, on the one hand, not only of relevance to all human beings and to the languages they speak; but they are also of huge importance to the concerns of philosophy, a discipline that may well be likely to benefit from the role

of collocation. If philosophy does adopt collocation for use within logic, it is quite likely to hand collocation as instrumentation for meaning over to science (Russell, 1960: 3). The established practice outlined by Russell is that philosophy deals with matters of investigation that are not yet *settled* and once these become areas of settled knowledge, they are ‘handed over to science’.

Hence, the experience offered by regarding *human nature* and its collocability with *being what it is* has both experiential relevance in natural language as well as theoretical relevance within philosophy. Equally, the concordance for *business is business* qualifies as *Sachverhalt* in terms of the opening section of Wittgenstein’s *Tractatus* as well as occupying the position in logic as a failed tautology. Progress will be made within the interface between philosophy and corpus studies, but it is likely to be slow.

2.3 Review of the literature relating to re-presentation as it subsists within philosophy and natural language and within a Firthian view

2.3.1 The first beginnings of reference corpora and their potential for becoming a scientific instrument

A number of issues came together both from Philosophy and from Corpus Linguistics in order to create the rapprochement necessary for collocation to fulfil computationally the requirements of Firth’s ‘meaning by collocation’ on the one hand and to test issues in the area of empiricism on the other.

Most of the early empiricism, in the absence of the computer, pushed philosophy and semantics in the direction of conceptual models. Even as late as the 1980s, but before the advent of the first Cobuild Dictionary (1987), we see McGinn (1981) preparing to agree to Quine’s Third Dogma based upon concepts. And yet, Firth could not be deterred from his view that collocates between them regulate and contextualise meaning without any recourse to *concepts*. This insight was hugely ahead of its time.

At least one other Firthian concurred rather vehemently (F.R Palmer, 1981). However, as Frank Palmer points out, all that was needed for collocation to consolidate its position, was the removal of the *raison d’être* for concepts. Concepts *explain* data.

Where data is empirically plentiful, it not only explains concepts, but more often than not, falsifies them.

The extent to which the work of the early British empiricist philosophers has any bearing upon corpora of collocation is extremely limited. However, the empiricist tradition is valuable as part of this study for the reason that it prepared the ground for the detailed work of the analytic philosophers from Gottlob Frege to W. V. O. Quine that spanned more than one hundred years.

The work of the early British empiricists is also valuable because it established the basis of nationalism and its interaction with the cognitive, both of which came under attack within the British philosophical tradition as analytic philosophy sought to deal with other means available to philosophers for explaining data.

It would be fair to say that American philosophers adhered more assiduously to the mentalist aspects of philosophy and that linguistics in the USA began to consolidate mentalism in both disciplines. However, by way of contrast, mentalism was subjected to greater crucial scrutiny in Britain, and especially in Cambridge, where Russell began to translate and disseminate the writings of Frege to an extent that was without parallel anywhere else in the world.

In the case of Wittgenstein, there can be little doubt that his own writings after *The Tractatus*, ironically, began slightly to mirror an inversion of the second line of *The Tractatus*: he was forced to provide so many *ad hoc* and often made-up examples that the fabric of his later work looks a little like an attempt to replicate the ‘the totality of things’ rather than ‘the ‘totality of facts’ (Wittgenstein, 1922: 7).

Of course, only those developments that impinge upon the nature of collocation both before the First World War and between the World Wars (see Urmson, 1960) will feature below as part of the review of the literature for those periods.

Inevitably, such a review must begin with the research of Frege of whom Ayer famously said that he was not only the best logician since Aristotle, but that he worked on a much larger area of logic than Aristotle, with more inclusive implications for logic, *philosophy and language*.

Chapman and Routledge (2005) state Frege’s principle succinctly as follows:

'A view commonly attributed to Frege is that the reference of a complex expression is a function of (is determined uniquely by) the reference of its component parts...For example, if we replace 'Hesperus' in 'Hesperus is Phosphorus' by co-referential expressions 'Venus' or 'The Morning Star', we obtain a sentence which must have the same reference.'

(Chapman & Routledge, 2005:95)

(See also Frege's letter to Russell, in Beaney 1997:255)

Now, within natural language, the collocates of 'The Morning Star' will of course be admitted as part of the sentence's *argument*, which effectively allows those collocates to collocate with both the *sense* (*Sinn*) and *reference* (*Bedeutung*) of the term. Effectively, and possibly unwittingly, this sets up the earliest form of pre-Firthian contextual linguistics. It also, one hundred years later, prompts us (see Chapter 5) to find lexical collocates for grammatical strings that have been searched for with the lexical forms of an authentic sentence wild-carded and, hence, effectively *absent* from the search line. Some of the following primary works of Frege set out by Chapman and Routledge (2005:97) are of importance within the debate on empiricism to which Frege's writings gave rise. Note that only some of his works bear upon forms that computers may now be capable of finding as a result of simple search instructions.

Taken together, all of the work of Frege emphasises the dismantling of all reliance on the cognitive and Frege places all of his emphasis and his rules on contextualisation in the external world, as well as within the formulation of logical propositions. In a section entitled *Frege's Anti-Psychologism*, Scott-Kakures et al. offer three of Frege's principles for distinguishing the logical from the *Psychological*, as follows:

1. Always separate sharply the psychological from the logical, the subjective from the objective.
2. Never ask for the meaning of a word in isolation, but only in the context of sentences.
3. Never lose sight of the distinction between concept and object ...

In the beginning of his late essay 'Thoughts', for example, Frege writes:

'From the laws of truth there follow prescriptions about asserting, thinking, judging, inferring. And we may very well speak of laws of thought in this way too. But there is at once a danger of confusing different things. People may very well interpret the expression "law of thought" by analogy with "law of nature" and then have in their mind general features of thinking as mental occurrence. A law of thought in this sense would be a psychological law. And so they might come to believe that logic deals with the process of thinking and with psychological laws in accordance with which it takes place. That would be misunderstanding the task of logic, for truth has not been given its proper place.'

(Scott-Kakures et al., 1993:375) (Emphasis added)

Of course, such vicissitudes all occur in reference corpora and are easily recoverable within concordances.

The publications of Frege are of great importance for the computational study of his principles. Furthermore, Frege's influence upon Russell and Wittgenstein and their undergraduate and postgraduate teaching at Trinity College, Cambridge throughout the period between the two World Wars, and beyond them, was so thoroughgoing that it may well warrant a study of its own as far as collocational research is concerned. Dummett's interpretation of Frege (1981) may repay detailed analysis through corpora.

Many of Frege's publications are available, at least in part, in *The Frege Reader* by Michael Beaney (1997). This work is set out in the list of publications presented elsewhere as potential 'prompts' for collocational verification. The main concern of this study is to cross-refer the concerns of the analytic philosophers with those of collocation studies.

At first the task appears daunting. However, if we simply compare those issues that look unlikely to yield results from such comparisons, we find that the exercise is well worth the effort. For example, Firth's famous statement, which formed the basis for Sinclair's OSTI study matches almost point for point the entire career interests of Gottlob Frege as set out in *The Shorter Routledge Dictionary of Philosophy*:

'These three guiding ideas – lingua-centrism, the primacy of the sentence and anti-psychologism exercised a commanding influence on the early analytic philosophers, such as Wittgenstein, Russell and Carnap. Through them, these ideas have been spread far and wide, and they have come to create and shape analytic philosophy, with whose fathering Frege, more than anyone else must be credited.'

(Craig, 2005: 299) (Emphasis added)

As this is developed further in the research, we can see that Frege’s attempt to study the interplay between mathematics and logic allows us to appreciate the scientific power of digital collocation encompassing as it does through co-selection all aspects of the states of affairs that it chunks (see Glock, 1996: 116, and Louw, 2010c: 351, as part of the Bootcamp Debate).

The direct matches are astonishing as we see if we tabulate them.

Table 2.2: Firth’s and Frege’s interests compared

FIRTH	FREGE
Not the ‘ideas’ meaning of words	Anti-psychologism
Abstracted at the level of syntax	<i>Sinn</i> (sense) always preferred to <i>Bedeutung</i> (reference)
Internal to the language	Lingua-centrism

In addition to this, Frege galvanised the other analytic philosophers into *finding* or at least (pre-computationally) *speculating* about the variables that they could not see. If computers and corpora had been available to the analytic philosophers the falsification of the cognitive would today have been more than 120 years in the past. The mathematical contribution of A. N. Whitehead must also be considered. He and Russell falsified Plato’s axioms. Among the commentators, J. O. Urmson singles out the role of *variables* in Frege’s work for special mention.

‘He drew for the first time the distinction between axioms and rules of inference and introduced the device which distinguishes modern logic from its predecessors and makes it superior to them, the use of variables and (nested) quantifiers.’

(Urmson, 1960:147; Emphasis added)

The attempt to find these variables drove all of these philosophers as far *below* the surface of sentences as their *intuition* would allow. It inspired the return to subtext in a computational form more than 50 years later (Louw, 2010a; 2010b; 2014; Louw & Milojkovic 2014a; 2014b in press).

Each of the remaining analytic philosophers has a particular profile for the automation of their work through collocation. They will be dealt with individually now. Several

philosophers have been added to this section because of the perceived special significance of their work as it may finally relate to digital collocation studies.

Because it was as difficult for Frege to see all of the variables involved in his analyses as it was for the man in the street to see them; we find that although Frege falsifies psychology he does little more to follow up his victory than complain. However, complaints seldom prevent progress.

2.4 Bertrand Russell

Russell, from the very beginning of his writings is concerned with the external world and the ability of language to contain it and express its meanings. It is plain that he was bowled over by *The Tractatus* and used it to have Wittgenstein appointed to the staff at Trinity College. Russell wrote the preface to *The Tractatus* and in it he explores with great objectivity and fairness the limits of expression set out by Wittgenstein. By 1948, Russell was still exploring the limits of human knowledge. Luckily both areas are today enlightened by corpora.

Areas relating to perception and ontology that are described by Russell in his early writings are easily dealt with and verified by means of corpora and matters such as the existence of Piccadilly need not be dealt with. However, a major area of Russell's work has to do with Logical Atomism and the extent to which it can assist in the creation of instrumentation will continue to be of lively concern. Griffin (in Craig, 2005:923) refers to what must be, although Griffin does not say so, the growing influence in Russell's thinking of the views of Wittgenstein's distinction between *facts* and things.

'In 1914 he [Russell] proposed a solution to the problem of the external world by constructing matter out of sensibilia. After 1918 he proposed to construct both mind and matter out of events.'

(Griffin, in Craig, 2005: 923) (Emphasis added)

Again, the signal absence of the computer limited Russell's belief in and dependence on empiricism. Griffin points out that Russell '*... did not think that empirical evidence*

alone would be sufficient for the constructions...' (*ibid.*) he needed to make, but regarded science as the best source of truth.

Particular attention needs to be paid to the following works that relate to *qualia* during Russell's perception phase, and to his definitions of events, culminating in 1948, in his Five Postulates (Russell, 1948). The experimental aspects of Logical Atomism as transferred to corpora are also likely to be of great importance.

For the purposes of our investigation, the work of Bertrand Russell offers both advantages and limitations. It may be best to dispose of the limitations first and also to stress that they were not entirely of Russell's making. I mentioned earlier that he found that the absence of empiricism hampered his scientific work. As it turned out, this was not assisted by the fact that he spent a period before and during the First World War in prison as a conscientious objector. The solitude of incarceration had an immediate and influential effect on his research by moving it into the area of perception. His work involved classifying events in relation to conditions created by, for example, time of day and the quality of light (see his reference to the colour of furniture under particular lighting conditions in his most influential early work, *The Problems of Philosophy* (Russell, 1912).

It is likely that in solitary confinement the visualisation of the external world as *events* comprised of *qualia* began to move in the direction of events as *states of affairs* as visualised by Wittgenstein in the opening lines of *The Tractatus*. It is of interest that a copy of this work, written by Wittgenstein in the trenches, apparently reached Russell during the war, even though it was only published in 1921 in German and in English a year later. Wittgenstein began his notes as early as 1914.

From 1918 onwards we find that Russell, still starved of the empiricism that he needed in order to view events as scientifically verifiable entities, became less reliant on expressing them in relation to the physical characteristics of the *things* that they undoubtedly contained. This represents a key issue for relating collocation to events. It takes us down a path that leads from mere *qualia* in:

'One meaning of the word 'snow' is its collocability with 'ski' or 'melt''

into a new dimension that homes in upon specific forms of *truth*. For example,

'One meaning of St Gerard is his collocability with 'bilocation''

or

'One meaning of truth is its failure to collocate with reconciliation.'

The absence of the computer meant that Russell was still setting up his five *postulates* for events as late as 1948. However, this turned out to be a work of great sophistication and it is today largely capable of becoming computer-assisted.

Collocation refers to *both* types of event and may be used conclusively to deal with the classification of both. The parameter of *time* bears upon both and this needs to be factored into the review of the literature that needs a line of best fit with collocation.

Russell was an extremely careful researcher and, to his credit, he often stated his theories more strongly than he explained them. This prevented them from being lost while he waited for the supply of empiricism to improve. Several sources assist in putting together some of the loose ends that are created during this process. In particular, Ongley and Carey (2013) is a valuable recent source in this regard.

An astonishing feature of Russell's perception is the amazing clarity with which he saw the periods of silence *between* examples of similar events. Both he and A. J. Ayer visualise them as being almost like the *spaces* between concordance lines.

Valuable material on events from a Russellian perspective is to be found in his *History of Western Philosophy* (1946: 834), a work for which in large measure, he won the Nobel Prize for Literature in 1950.

Major commentators on Russell include David Pears (1985), A.C. Grayling (1996) and a seminal work by Landini (2011) whose style is especially refreshing as we see in the following example.

Think how odd it would be to imagine a new scientific breakthrough reported in the headlines: 'New concept discovered: Time. Henceforth, please consecutively order your thoughts in terms of before and after.' This is not just odd; it is unintelligible. Our thoughts are given consciousness with a temporal order.
(Landini, 2011:227)

It seems that the absence of appropriate levels of empiricism led Russell to be less vocal than Carnap (1928) with regard to what the latter called "*Ahnlichkeitserinnerung*" (recollection of similarity) , a definite precursor of subtext.

Landini (1998) had previously written a book entitled *Russell's Hidden Substitutional Theory*. All such substitutions are, of course, accessible through corpora, and *a fortiori* since subtext was made available at conferences from 2008 and in published form in 2010 (Louw, 2010a; 2010b).

The work of Grayling (1996) is especially useful because he shares Russell's pessimism about the practical value of the study of events in any analysis that bears upon the concerns of the external world.

In the run-up to Russell's book, *Human Knowledge: Its Scope and Limits* (1948), Dorothy Emmett (1945) shows something of the ways in which ingrained but now quaint thinking still needed to be disrupted by computing

'the use of the notion of probability, as in wave mechanics, emphasises that it is concepts with which we are treating, not entities, since probability is an attribute not of events but of our knowledge.'

(Emmett, 1945:78)

A major area in which Russell knew that he was on the brink of a major but still unrealised breakthrough was Logical Atomism. He was so deprived of empiricism that he was left unable to describe it in full, but so enthusiastic about it that he was unable to stop boasting about it. Even several years before his death, he continued to refer to himself as a logical atomist.

Both Russell and Wittgenstein had volumes devoted to the subject. Russell's *The Philosophy of Logical Atomism* was republished in 1985, with an Introduction by David Pears, where, with characteristic flair, he describes the book as follows:

The marriage arranged by Russell between logic and empiricism gives The Philosophy of Logical Atomism its special character. Most of his emphasis is on the empirical argument for logical atomism: if we think of a word and begin to analyze it by substituting definitions, we soon reach a point at which the supply of definitions runs out and the analysis terminates. The empirical argument claims that this shows that we have reached logical atoms.

(Russell, in Pears, 1985: 30)

However, because of the paucity of empiricism, the notion as set out above is, yet again, end-stopped at the *mind*. It leaves us, as Pears says ‘...*distinguishing between absolute atomicity and atomicity relativised to our present analytical achievements.*’ (1985: 31). And so we see that a corpus-based approach leaves the chaos that Pears describes next, looking almost fully settled.

‘So, to pick up the beginning of the argument again, if the analysis of ordinary descriptive sentences never terminated on logical atoms, their meaning would always depend on further factual truths. But that would not be acceptable, because it would make meaning indeterminate. The meaning of any descriptive sentence would depend on a truth, whose meaning would depend on a further truth, and so on ad infinitum.’

(Pears, 1985: 31)

The dialogic aspect of finding the atoms *intuitively* is converted into the inductive reasoning that unpacks concordances. Concordances specialise in the provision of further truth.

‘There are two main novelties in the modern way of studying language. One novelty is the amount of it that is available. Before the computer, linguists could only study small samples of language at a time because of the limitations of their powers of observation and their memories. Even scholars who relentlessly collected instances of usage all their lives only had a few examples of any particular pattern, and there was no way of telling what they had missed. The situation is almost the opposite with corpora, especially large corpora, where techniques have to be devised to control the often massive amounts of evidence.’

(Sinclair, 2003: ix)

It falls to Griffin (1964) to present Wittgenstein’s account of logical atomism. As Pears points out, Wittgenstein is committed to an account that leaves nothing putative unexplained.

2.4.1 Russell and the implications of his link between events and perception

In terms of Firth’s (1957) taxonomy for context of situation, the events of 1914 may well be found to reside within what Firth called ‘relevant persons, personalities’ and relevant objects. Those aspects of collocation that are suited to the representation of qualia, such as moods or appearance, would be relevant here; as would all references to

the 'relevant objects' and indicators of time. The latter will need to be added to the Firth taxonomy in the light of the research.

However, a key difference introduced in 1918 and culminating in Russell (1948) refers to *outcomes* in the external world and, in the Firthian taxonomy of 1957, as the 'results of the verbal action'.

In the absence of computing, Russell would have felt more comfortable with this latter view, and, in particular, with its diagnostic aspect. But we need to realise that even in the early period of his work he unwittingly left scope for improvement by computing by creating space for improved thinking.

He relies on common sense (possibly influenced by a career spent working with Moore at Trinity College) and relying upon records and memory until some better form of empiricism presented itself. Pears identifies this attitude. The computer rather than a person was to provide it long after Russell's death.

However, the result of his work up to 1918 created a climate for the investigation of perception, as well as a near reliance upon scepticism and doubt that the representation of physical objects could ever be reliably identified within a theory of realism. Tables that appeared brown in certain forms of light might be described as green or grey in other forms of light. Sticks that are manifestly straight one moment are perceived as bent when immersed in water. There was also a sense in which objects went unnoticed or were deemed to have disappeared unless they were referred to. All of these dimensions will need attention in terms of their recovery as empiricism within reference corpora.

These aspects of perception in relation to collocation are dealt with in the literature which follows. A separate section deals with marked aspects of events and their being subsumed into five postulates is dealt with in the section that follows this one. For the most part, Russell's move from qualia to markedness is borne out within the literature over this long period from 1914 to 1948, although most of it is fairly lacklustre in character. This dissertation will concentrate upon those aspects where philosophy can focus upon collocation and improve its use and *vice versa*. As elsewhere in this research, there can be no doubt that benefits emerge for both philosophy and linguistics.

A particular issue that might benefit philosophy is the fact that where Firth's main principle and the focus of the OSTI Report is stated, the main beneficiary within philosophy will usually be *necessary meaning*. A 'dark night' will have 'no moon' as a necessary condition; 'melting' snow has a different texture to 'falling' snow and the collocates will always have overtones of causality that philosophy has always welcomed within single propositions, but with which it may well be less satisfied within narratives that move towards *relating* one proposition to another.

Furthermore, both approaches to Russell's events are often run together by commentators such as Landini (2011), Grayling (1996) and Ongley and Carey (2013). The fact that the two approaches *occupy* different parts of the Firth-Malinowski taxonomy must be allowed to decide the issue on the basis of collocation in the context of this research.

It is worth noting that of the early analytic philosophers, only Wittgenstein looked beyond sense-data in the direction of logic and the *a priori* for forms of empiricism that, as it turned out, computers were eventually able to provide. And only as recently as this research, his intuition that logic in natural language would be empirical, has proved correct (Wittgenstein, 1929). This view was shared by Dummett (1976), but never as explicitly as Wittgenstein characteristically set it out. In fact, Wittgenstein's short but seminal article 'Remarks on Logical Form' (1929) reads as though it has, even today, been written as an article in corpus linguistics. Only Frege begins to rival the predictive lucidity of Wittgenstein at his best.

After exploring sense data in the following section, we will eventually return to Russell and Wittgenstein, to consider the matter of logical atomism, an area that only poets such as Yeats were exploring, totally unwittingly, during the life-time of the analytic philosophers.

2.5 Sense-data and their collocational aspect

From the point of view of the Firthian perspective upon which our study is based, the most attractive aspect of sense-data is to be found in their responsiveness to one another

within context of situation. The statement *dark night* quickly disposes through necessary meaning of all of the probabilistic qualities that aside from this relationship might have been expected confidently to have occurred. While Alston (1996: 87) asks if we can cognise anything other than judgments, his view of sense-data is at once too broad and too narrow. He appreciates the need for a 'sense-data' language, but his link between sense-data and logic flows less easily than it ought to. No mention is made of the role of logic in such a language and interpretation appears to rely a good deal on metaphysics alone.

Jonathan Dancy (1985: 1995) doubts that sense-data can be treated as *facts* in Wittgenstein's sense of the term. The reason for this is that his theory appears to have no way of chunking facts into states of affairs in the same way that collocation combined with logical form would. This lends further impetus to the research.

Barratt (1995: xii) refers to the view that literary fiction is necessarily untrue, but makes the point along with it that great literature seldom warps its worlds to the point that they become incomprehensible. Collocation may be a crucial aspect to unravelling the warped minds of characters. In John Fowles' *The Collector*, all of the captured female victims are referred to as *guests*.

A special place needs to be found in this review of the literature for the work of a follower of Russell and Wittgenstein whose special credentials name him as the only British member of the Vienna Circle. Alfred Jules Ayer was a pupil of Gilbert Ryle at Oxford. Ryle was keen to have someone reliable keep track of the Vienna Circle and he sent a very youthful Ayer to Vienna to join the *Wiener Kreis*.

He seems to have found G.E. Moore's commitment to 'common sense' worryingly naïve when it is placed side by side with Russell's view that science would always somehow save the day. Ayer quotes this line with obvious warmth and sympathy for someone in search of large amounts of empiricism.

"Science is at no moment quite right, but it is seldom quite wrong, and has, as a rule, a better chance of being right than the theories of the unscientific. It is, therefore, rational to accept it hypothetically."

(Russell, 1959, quoted in Ayer, 1982: 116) (Emphasis added)

An extraordinary aspect of empiricism within philosophy is the automatic assumption, independent of events in the world of technology, that all experience flows from ‘the given’, with the given hardly ever visualised as *re-presented* rather than *presented*. Lexicographers alone, probably, who witnessed the gradual acceptance of John Sinclair’s methods, are likely to appreciate fully just how gradual the process of acceptance was. The fact remains that in 1987, the year in which the first edition of the Cobuild Dictionary was published, very few of those most vehemently opposed to change in lexicography, had ever seen a KWIC concordance or the output of a collocator.

Of course, dictionaries are a commercial product, and there were mutterings at Cobuild among the more elderly and staid staff that they were harming scholarship in order to please scrabble-players! But it is hard to imagine how much worse matters were in philosophy, where the academic staff felt an even deeper duty of care towards a discipline that might not recognise a computer even if it ever saw one.

And so, we are struck by a sense that the ancient discipline of philosophy was being practised in the midst of a computer age that could furnish more empiricism than would ever be needed by the careful best practices of the philosophers themselves. And yet nobody had made the potential link with computing explicit.

It is for this reason that Ayer is especially important. He had written extensively about the alternatives to empiricism such as the *a priori* and volubly about probability and evidence in a volume with those words as we see in its title below, and he responded at last, with characteristic charm, to the concerns of this very special volume late in his career (Macdonald, 1979). The book was entitled *Perception and Identity: Essays Presented to A. J. Ayer, with his Replies to Them*. It was edited by G.F. Macdonald and published by Macmillan.

The replies of Ayer offer us the greatest insight into how collocation *may* have been received *if* the OSTI Report that had been completed nine years earlier (but was apparently lost), had been made available to the community of analytic philosophers. Here are just five examples from the contents page.

Common Sense and Physics by Michael Dummett

Perception and Its Objects by P.F. Strawson

Ayer on Metaphysics by Stephan Kolner

A Defence of Induction by J.C. Mackie

A Comparison between Ayer's Views about the Privileges of Sense Datum Statements and the Views of Russell and Austin by David Pears

Here is an example of the calibre of argument from the contribution of Michael Dummett. Ayer (after Russell) had set up a scientific safety mechanism to preserve the value of his research for science as and when *empiricism* caught up. Ayer referred to it as a 'second system'. Dummett questions its existence as follows:

'When instrumentation is in this way applied to every scientific statement that rises above the level of simple generalisation, it becomes utterly implausible, and it is no wonder that it is resisted; we have here another example of the frequent phenomenon that realism scores too easy a victory, because the alternative to it has been incorrectly formulated. As I have tried to argue, a statement such as that light travels in straight lines is virtually forced on us in the attempt to make sense of our everyday experience; and many other scientific laws are arrived at by a process essentially the same in character as that by which, in ordinary life, we form a picture of the world we live in. Hence, if instrumentalism is to have any claim on our consideration, it must be maintained that the critical line, beyond which we begin to deal in theoretical statements properly so called, is crossed not as soon as anything recognisable as a scientific theory is framed, but at some much later stage in the development of science. Just when this occurs will be very hard to say. It is a problem in the theory of meaning; and the theory of meaning is not yet in a sufficiently well developed state for us to be able to give a ready answer to the question of when we may be considered to have conferred determinate truth conditions upon any given range of statements; that is, indeed, the central question over which the participants in a dispute about whether statements of a given class admit a realistic interpretation disagree with one another. It will be the task of scientific realism to maintain that there are no theoretical statements in the required sense – that, in other words, no line can be drawn: the progress of science is, on a realist view, everywhere continuous with the kind of attempt we make, even before science begins to understand our world and arrive at a description of it as it is in itself. Earlier in this essay, I did indeed express such a view, but I do not really want to commit myself to it: I wish to remain agnostic about whether it is possible to delineate a range of theoretical statements for which the instrumentalist thesis is plausible. The theses for which I wish to contend are the following; that instrumentalism, if correct at all, is so only for the most rarefied statements of scientific theory; that there is no conflict in principle between scientific truth and a common-sense view of the

world, and, therefore, no problem of reconciling them or necessity to opt for one object or the other; that there is no legitimate notion of knowing an object, by immediate awareness, as it really is, when this is taken as involving more than that we have not made a mistake; but that there is a legitimate notion of a description of an object as it is in itself, and that we cannot but view science, at least before it transcends some critical level of abstraction, as attempting to arrive at such descriptions.'

(Macdonald, 1979: 38-39)

Meaning by collocation (Firth, 1957) may well have settled more than it disrupted if philosophy had remained in closer contact with linguistics than is even now the case.

But even the appeal to the given (Ross, 1970), if made computational, would have required more alterations than common sense might have been willing to entertain:

1. the given presented on many separate occasions;
2. the given as a Firthian equation such as 'one meaning of *self* is its collocability with *poor*' in a poem by G.M. Hopkins.

As we continue to review the literature of Russell's later approach to events, it is almost convenient to begin at the end with the apex of this movement, with his *Human Knowledge: Its Scope and Limits* (Russell, 1948).

This is a far broader treatise than any other that deals with Russell's later view of events. Its main strength is the way in which it looks forward to being capable of handling huge amounts of data, even though, at the time of writing, there was not yet even the slightest hint of the advent of computer technology. Despite that, Russell's approach is far and away the most universal and scientifically inductive study.

Other important, strictly scientific work by Russell that survives to this day and is capable of computational proof, falls within the area that he referred to as Logical Atomism (Russell, 1918). It falls together, with small but significant variation, in the logical atomism of Wittgenstein. Judged intuitively, but as early as 1929, both

Wittgenstein and Frege were arguably in a better state of readiness for computerisation of their work than Russell. But Russell had a better vision of the future and more tenacity to adhere to empiricism than both Frege and Wittgenstein put together.

Inadvertently, Wittgenstein set up the parameters for work on logical atomism in his now famous statement in the *Tractatus* (1922) that what cannot be expressed must be passed over in silence. And because we are engaged in philosophy, the passing over of the unexpressed and the inexpressible only ever takes place after exhaustive processes of reasoning, and even then, finally, with great reluctance. Russell likens this process to exhaustive questioning as a form of refining *truth*. Pears (1987: 63) refers to it as the doctrine of forced acquaintance. Hence, one knows when there are no questions left to ask and that what remains as the *given* can only be a ‘logical atom’.

It would be true to say that for Russell the process of interrogation that leads us to a logical atom takes place purely *dialogically*. For Wittgenstein it is almost entirely *mathematical*! But no matter how we view it, the stumbling block for both Russell and Wittgenstein was the fact that their *intuition* was incapable of *furnishing* them with the *variables* of the vocabulary items within propositions. Russell offers us the formula for exploring subtext more than 40 years before we chose to use it; but for Wittgenstein, the grammar was always the centre-piece. As Carver (1996:139) puts it in an article entitled, *Philosophy as Grammar*.

‘It is one of the wonderful paradoxes of our time that the greatest and most stimulating philosopher of the century should identify his work with the stodgiest and dullest of schools’ subjects. It is nonetheless the case that for the last twenty years of his life, the years of his greatest productivity and his profoundest work, Wittgenstein identifies what he was doing, and what other philosophers really have been doing and should be doing, with grammar. This perspective is as carefully considered as it is puzzling. It emerged out of earnest and ongoing work ...’

(Carver, 1996:139) (Emphasis added)

It is this sort of ‘work’ that only corpora can unravel. Wittgenstein’s logical atomism is exhaustively dealt with by Glock (1996). Its best and most concise elaboration by Wittgenstein is in his article⁵ ‘*Some Remarks on Logical Form.*’ (1929).

⁵ N. B. Wittgenstein wrote only two published articles.

It is instructive if we place Firth's assertion 'one meaning of *night* is its collocability with *dark*' (1957) next to Wittgenstein's criteria for elementary propositions:

If 'p' entails 'q', its sense contains that of 'q', that is, analysis must reveal 'q' to be one of the truth-fundamental components of 'p'.

(Glock, 1996:103) (Emphasis added)

Wittgenstein's *Tractatus* offers what is arguably the most precise line of best fit between *facts* and the power of *collocation* to chunk them. Glock's exhaustive account of *facts* can be examined and re-examined until a line of best fit for their analysis computationally is reached on a basis that benefits *both* philosophy immediately and linguistics only tangentially. The reason for the tangential limitation in respect of the latter is that collocation is already capable of the wholesale falsification of linguistic theory to a greater rather than a lesser extent.

This brings us to a point where there is a rich cross-over between logic and collocation. It is only by working on logical atomism that one finds that single piece of luck that operates as the mainspring of this research: collocation is instrumentation for meaning because its operation mirrors almost exactly the main philosophical discovery of the 20th Century. The logic of Aristotle was based upon *objects*, in Chomsky's terms *subjects* and *properties*, in grammarian term, *predicates*, or $S \rightarrow NP + P$. But, as Ray Monk says, Russell believed ...that as well as *objects* and *properties*, the world (that Wittgenstein said is 'all that is the case') contained *relations* (Monk, 2005:12). We see how close this is to Firth's statement that Sinclair made the centre-piece of the OSTI Report.

One meaning of X is its collocability with Y

Note how this begins to gel as instrumentation as we read on and follow as Monk sets out the depth of the Russell-Frege breakthrough.

'Propositions such as 'John is taller than his father' and 'Four is the square root of sixteen' must be understood, not as the predication of a property to an object, but as the assertion of a relation between two objects. But, to understand them as such requires one to break out of the subject-predicate straitjacket imposed upon propositions by Aristotelian logic. The logic of Frege and Russell, which treats propositions as functions, achieves this breakthrough. The word 'function' here is used with a conscious allusion to its origins in

mathematics. Thus, just as the mathematical function x^2 has the value 16 when $x = 4$, and 25 when $x = 5$, so the propositional function 'x is taller than y' has the 'truth value' true when $x = \text{John}$ and $y = \text{his father}$ and false when $x = \text{his father}$ and $y = \text{John}$.'

(Monk, 2005: 12)

Now the relations referred to by Monk such as *tallness* or *square roots* are the stock-in-trade typical of philosophers' examples and of their intuition, especially as it bears upon the *a priori*; but we recall that Frege provides not only *propositional function* but also its extension or *argument*; and collocation provided from a reference corpus is capable of indicating that the extension, in Fregean terms the *argument* is capable of inferring from the external world a logical clash that falsifies the *function*. In other words, to paraphrase Firth (1957), one meaning of natural justice is its collocability with breached or denied.

The extension identifies *natural justice* as a member of a *class* that we may, say, call *j* [*justice*] determines its status empirically as *non-p* and indirectly shows it the door, because litigants in the external world who abandon *conceptual* reading in favour of *digital collocational* reading will begin to refuse to include the phrase in their stated prayer (*i.e.*, requested order) in court. Hence, *propositional function*, extension by *argument* and *class* membership, between them, will result in an outcome that is approved by both Firth and Wittgenstein. Monk concludes:

'For him [Wittgenstein] philosophy was the activity of solving philosophical problems. He used to say his father had been a businessman, so he took a businesslike attitude to philosophy. He wanted to clear up philosophical problems like a businessman clears his desk.'

(Monk, 2005:14)

The question of proof that collocation is instrumentation for meaning is not a 'hard sell': *all* of the sense disambiguation needed for two editions of the Cobuild dictionaries was carried out using software designed by Jeremy Clear, to provide "a computational approach to collocational analysis" (Clear, 1993: 271).

As we turn our attention to Rudolph Carnap, we become aware of two main issues. The first issue relates to *Der logische Aufbau der Welt* (1928), Carnap's attempt to project the world as a logical construct and the precise ways in which it inspired the American

writer Henry L. Goodman (1951) to produce a sequel to it in English entitled *The Structure of Appearance*. Both works appear to have some substantial relevance to collocation studies. However, a much more crucial point is made by Carnap himself. Rather than see events as causality of a logical aetiology entirely, Carnap argues for a largely forgotten form of *immediate* processing. It is almost as though he realises that although logical aspects will follow a projected path determined by reason, there will also be point instants at which we become aware of multiple, present and not necessarily competing forms of richness in meaning. No scholar addresses this issue within collocation, apart from Geoffrey Williams (2008), who presents diagrams with nodes that look like circular spiked valency units. Although Carnap used no diagrammatic aids, his desire to see *unitary* relationships within process are pleasingly Firthian in nature.

What Carnap has in mind is not far off from the notion of valency. So, in Firthian terms, once *night* and *dark* have been positioned on a node of the kind that Williams proposes, Carnapian necessary meaning will make any appearance of moon (apart from its negative statement, e.g., *there was no moon*) a *necessary* impossibility; and all of this within a Russellian point instant. We see in the following statement by Carnap something of this dilemma. He is accustomed to referring to *sense-data* (plural), but their very plurality pushes into time and space. The situation can only be saved if their presence is visualised as a node of the type that Williams proposes; a simultaneous point instant that, although it is up for revision as the world and the text advance, still provides a reliable statement worthy of being termed an *instrument* at the moment that the node itself acts as a ‘snapshot’. It is this form of *stasis* that also creates a fit with variables of logical form and, hence, subtext. Carnap shows us an analogue form of subtext.

Therefore, I took as elements total instantaneous experiences (Elementarerlebnisse) rather than single sense-data. I developed a method called ‘quasi-analysis’, which leads, on the basis of the similarity-relation among experiences, to the logical construction of those entities which are usually conceived as components.

(Carnap, in Schilpp, 1963: 16)

The term valency is not used in the fixed elemental sense of chemistry but in order to muster a constellation of collocates for even so natural event as a change in the weather,

non-fictional, or fictional and authorially re-presented. Of course, such nodes would be capable of being determined *in medias res*, as it were, by computational means. Although Sinclair (1991) alludes to a nine word window of collocative power, he nowhere makes the obvious point that it is of course, a *moving* window as text is produced and processed. The use of collocational nodes presupposes correctly that texts are more collocationally dense at certain points than at others.

Our review now takes a different direction to accommodate the answering of a major research question that was stated earlier. It is the question of *intuitive opacity*. This will involve a combination of Wittgenstein's 'Some Remarks on Logical Form' (1929) and Karl Popper's approach to *falsification* (see Chapter 5 for detailed examples that establish the falsification of intuitively-derived theories).

Popper turns into a requirement something that linguistics only talks about but never implements: *strenuous tests*. It is common cause that mere argument is unlikely to induce a sense of falsification in those who realise that their own adherence to a false theory is, as a position, scientifically misguided. If this were not the case, Frege's falsification using mathematics as logic would have leapt over the fence from philosophy into epistemology (linguistics had not yet been born) and completed the task of falsification. Practical applications in the form of falsifications of stylistic findings will be set out in Chapter 4.

However, unless philosophy shows its teeth (unusual for a gentle discipline) and disproves deeply cherished false theory, there will be nothing to hand over to science and nothing substantial enough to make science grateful for, let alone, glad of it.

And so it is that our strenuous test has a *strategy* (Greek *strategos*: a military general). The process involves at least three steps.

1. The soft proofs of Chapters 3 and 4, co-selection and the production for inspection of the empiricism of 'chunked' states of affairs.
2. Proof that logic is empirical (Dummett, 1976).

3. Proof that logic which is at the highest point of intuitive opacity (and its opacity is *easy* for the cognitivist to *admit*) and achieves a better practical result than any amount of toil using vocabulary alone. The profound weakness of the cognitive is that that which is paramount, logic, is simultaneously that which is opaque to intuition and in a position to dominate vocabulary.

Kitcher (1993) implies that where falsification is achieved it is more easily recognised in the *behaviour* of those who have been defeated than in any verbal admission.

A major finding of this research is that logic accommodates itself to context and co-text and it coverts its own collocates into part of its instrumentation for dealing with the human condition. And so we see that Teubert (2010) was right to reject corpora as *prima facie* 'inadequate' for dealing with the complexities and nuances of everyday life; but his anxieties will be relieved as we pursue the manner in which the logic of natural language takes unto itself the collocates of cultural argument and coverts them into a lived logic by removing from them the impact of what used to be termed the diachronic. The vocabulary of cultural institutions becomes inextricably bonded to the Fregean arguments that represent those social and cultural institutions throughout time.

2.6 General Guidelines for the inclusion of references in this review of the literature

The inclusion of references will generally be based upon the ability of phenomena within the philosophy of language to illuminate or contribute to the discussion of collocation in respect to two of its main functions: (1) the provision of *context of situation* and *co-text*; or (2) the provision of corpus-assisted access to *subtext*. The *directionality* issue is of particular value in this regard and contributes to what makes this study unique: the starting point will for the most part begin within *philosophy* rather than in *linguistics*. A tangential effect of this criterion will mean that there will be fewer references, but that they, although fewer, will be of greater relevance to the study. References that are unsupported by the relationship of philosophy to collocation studies will, in general, be rare. So, for example, general notions, such as the ability of collocation to provide insights into *qualitative* aspects of a description of a phenomenon

such as, say, *snow*, will always take their lead from philosophy. Hence, reference to *snow* may enter the discussion because of its observed *qualia* and its coldness and texture as *sense-data*. This may mean that reference to *snow* might identify a link between *collocation* and the research on sense-data of Bertrand Russell between the years 1914 and 1918. In the same way, a reference to snow after 1918 may concentrate primarily upon aspects of the way in which it concentrates upon the *event* of snowing.

The same general principle will be applied to references to theoretical aspects of collocation. Those that have no bearing upon the philosophy of language may, *ipso facto*, be found to have been excluded because the link between philosophy and its potential use both *in* and *as* collocation was not the starting point for or the aetiological starting point for the reference. The same criterion applies also in reverse. For example, where collocation manifestly has a significant bearing upon the philosophy of language and may well contribute to the way in which it is studied or regarded in the future, it will qualify for inclusion in this review of the literature. A good example may be the way in which collocation deals with experiential aspects of the empiricism involved in processing logical aspects of human experience into ‘bundles’, e.g., with *all* of a bundle, linked by the same quasi-logical or grammatical form, such as *being + what + it + is* or in the plural, *being + what + they + are*.

The criteria themselves are established within chapters 3, 4 and 5: *context, subtext* and *applications*. A division along these lines has acted to streamline the references and avoids the *circularity* of reference referred to by even the most modern commentators, such as Garvey and Stangroom (2012), who make sweeping statements to the effect that clarity in establishing positions in philosophy will never be established, or that (and this appears as a cover picture!) we may need to settle for a intuitive assessment of optimism or pessimism to describe its overall trends. It would be inappropriate for a subject that views itself as the hand-maiden of science to view itself as a glass that is half-full or half empty. It is precisely because studies of the kind that are represented in this thesis are likely to improve even philosophy itself, that we need to preserve an open mind rather than impose a *symbol* that leads the reader to suppose that the work may never be finished. Such reasoning plays into the hands of those who claim that *truth* is pluralistic.

Of course, reference works in philosophy will apply for inclusion as well as the primary texts of major analytic philosophers and their commentators. Such references will make

up the bulk not only of this review of the literature, but also may attract quotation within later chapters.

History is a general guide to the references and the unfolding of trends within philosophy on the one hand and the thought of J.R Firth and J.M. Sinclair on the other. Reference to my own interaction with both philosophers and academics has been kept to a minimum within this thesis.

Finally, there must be a note about the commonplace. The impact of this thesis would be markedly diminished if an inane correspondence between collocation and qualia or sense-data were reported. It should not be assumed that the author failed to note their existence on a large scale. A Cobuild CD of such collocations was released at a time when they were believed to be of great significance and when collocation itself was assumed to be both classifiable and marketable in a static form, rather than called upon as a dynamic phenomenon. The decision to provide examples that gain impact because of their counter-intuitive nature was entirely my own.

2.6.1 An intention to proceed from philosophy to collocation as a priority and whenever possible

Unless one prioritises along these lines, hours will be lost to a review of collocation studies that are either mentalist or syntactic or both. We are fortunate that John Sinclair trod this path during the OSTI research in the late 1960s. Firth disposes of two parameters that between them allow us to dispense with a review of mentalist approaches that *are* all to do with the *ideas* meaning of words; and approaches that continue to see collocation as a syntactic phenomenon.

Furthermore, it has become plain, largely as a result of the academic community having been denied a Cobuild Dictionary of Collocation, that we need to add a criterion here that refers to those approaches to collocation that will take it into the future: collocation has to be *dynamic* rather than *static*. This is because its role is never purely a reference role. It is dynamic because the nine-word window of its operation *advances* through the text during the act of reading, attracting new collocations and dropping others as it goes along its linear path.

Very few approaches, and possibly only *one* approach makes provision for this in both practical and symbolic terms. The approach is that of Geoffrey Williams, as described in Williams and Millon (2009). Let us begin with the symbolic aspect. Williams and Millon adopt a crown-cork symbol in order to demonstrate collocational dynamism. The tendrils can pick up collocates during the 'live' period of their show in the nine-word window and drop them once their re-lexicalising task has ended (but often in a particular text resumes later). In 2005 in a seminar to the Oxford Tape Archive I would have benefited from the crown-cork method. The audience asked me for a symbol to represent the interaction within the nine-word window. I offered the notion of a crucible that moves at the same pace as the text, drawing all text through its heat but *allowing* collocates to meld with one another briefly as, in Wittgenstein's terms, they become part of one another, in particular contexts.

However, it seems better to me, with the benefit of hindsight, to make provision for the sampling of collocational networks while the text is in progress. The emblems that Williams proposes will be better for the task of tracing the development of necessary meaning as narrative advances. Anyone who doubts the validity of such dynamic approaches to collocation need only compare dynamic methods with those that Sinclair found deplorably inadequate in *The BBI Combinatory Dictionary of English* (Benson et al., 1986), which he felt was set up as a poor rival to the dictionary that he had planned.

The serial comparison of collocational networks seems to be a more feasible model than that of melded meaning. Networks would have the advantage, in due course, of being computer-generated and rather like the single frames of an old-fashioned movie. Their subtexts would assist in obtaining a line of best fit.

A further advantage of always moving from philosophy to collocation is that all experimentation then springs from observations in philosophy. Subtext was born of one of these and the decision to split the logic-metaphysics atom of the Vienna Circle arose in part from descriptions of logical atomism by Russell and Wittgenstein. The main disadvantage of dealing with linguists is their poor theoretical background and shaky grasp of science. What does a mentalist linguist care about Firth's view that collocation liberates the investigator from the use of concepts? Philosophers will always be more respectful of empiricism, because it has been so hard-won.

Russell said of Wittgenstein:

“Most startling was his belief that you can read off the deepest structure of reality from the study of the deepest structure of language.”

(Robinson and Groves, 2002: 72)

2.6.2 A review of the underpinning theory within Philosophy and its inaccessibility in Linguistics

A second quotation from Wittgenstein offers us the task of this thesis:

“To analyse language in depth is to do metaphysics – and at the same time, to do away with it.”

(Robinson and Groves, *ibid.*)

The collocational evidence is that only a special form of lexical collocates adheres to the logical elements of ordinary language.

This type of evidence is best summarised in the form of a table (see below) of the most influential philosophers before we proceed further. This table must be read with the main prompting references of (1) the philosophers themselves; (2) general reference works in philosophy; (3) general reference works in specific areas of the discipline and (4) general reference works in Firthian-Sinclairian Corpus Linguistics. This four-point criterion will be further discussed after the presentation of the table. Several examples of prompting references will also be provided, before the chapter is brought to a conclusion.

Table 2.3: Showing a Summary of Main Philosophical Prompts for Collocational Verification

PHILOSOPHERS	THEORETICAL CONSTRUCTS	METHOD FOR VERIFICATION
<p>Gottlob FREGE (1848-1925)</p>	<ul style="list-style-type: none"> • The addition of contexts to mathematical functions 	<ul style="list-style-type: none"> • Collocation as variables
	<ul style="list-style-type: none"> • The enlargement of logic beyond any limitation visualised by Aristotle 	<ul style="list-style-type: none"> • Consequences in linguistics for a refusal to abandon metaphysics
	<ul style="list-style-type: none"> • Objectivity in science 	<ul style="list-style-type: none"> • Consequences for cognition and mentalism
	<ul style="list-style-type: none"> • The banning of lists of words 	<ul style="list-style-type: none"> • Relexicalisation of items on lists
	<ul style="list-style-type: none"> • Failure to appreciate limits of class membership 	<ul style="list-style-type: none"> • Falls victim to Russell's Paradox
<p>Bertrand RUSSELL (1872-1970)</p>	<ul style="list-style-type: none"> • Sense-data and qualia 	<ul style="list-style-type: none"> • Verified by collocation, e.g., qualia of <u>snow</u>
	<ul style="list-style-type: none"> • Event structure 	<ul style="list-style-type: none"> • Mirrored in Firth, 1935-1957
	<ul style="list-style-type: none"> • Criteria for Natural Language Logic 	<ul style="list-style-type: none"> • Results when automated by collocation in the production of subtext.

Bertrand RUSSELL (1872-1970)	<ul style="list-style-type: none"> • Logical Atomism (starved of empiricism during Russell's lifetime) 	<ul style="list-style-type: none"> • If pursued by collocation, the doctrine of forced acquaintance is borne out.
	<ul style="list-style-type: none"> • Taxonomy of Events in Five Postulates (1948) 	<ul style="list-style-type: none"> • If followed using collocation and corpora, results in an accurate taxonomy of events.
Ludwig WITTGENSTEIN (1889-1951)	<ul style="list-style-type: none"> • Analysis of the world as the totality of facts, not of things 	<ul style="list-style-type: none"> • A perfect link with collocation for chunking situational contexts and states of affairs (<i>Sachverhalte</i>).
	<ul style="list-style-type: none"> • Logical form takes precedence over metaphysics 	<ul style="list-style-type: none"> • Improves the heuristic of corpus-derived subtext
	<ul style="list-style-type: none"> • Predicts the instrumental power of grammar strings 	<ul style="list-style-type: none"> • As a result of computing, a precise logical picture can be derived.
	<ul style="list-style-type: none"> • Admits the philosophy of psychology more readily than was necessary 	<ul style="list-style-type: none"> • Contributes to the notion of intuitive deficit
The Vienna Circle <i>(Wiener Kreis)</i> (1924-1936)	<ul style="list-style-type: none"> • Sets up robust criteria and tests for scientific work 	<ul style="list-style-type: none"> • Logic and metaphysics must be kept integral.

<p>The Vienna Circle (<i>Wiener Kreis</i>) (1924-1936)</p>	<ul style="list-style-type: none"> • Its most powerful test is qualitative: “<i>The meaning of a proposition is the method of its verification.</i>” 	<ul style="list-style-type: none"> • Ironically, only collocation satisfies this principle as part of its falsification of the inseparability principle for logic and metaphysics.
<p>Rudolf CARNAP (1891-1970) (A member of the Vienna Circle)</p>	<ul style="list-style-type: none"> • Models the logical construction of the world 	<ul style="list-style-type: none"> • Suitable for collocational automation
	<ul style="list-style-type: none"> • Proposes on intuitive version of subtext; and was talked out of it by colleagues 	<ul style="list-style-type: none"> • A perfect model for automation
<p>Sir Karl POPPER (1902-1994)</p>	<ul style="list-style-type: none"> • His view of science was that it advances by serial falsification. 	<ul style="list-style-type: none"> • Collocation achieves such falsifications in spite of many admonitions from mentalist linguists and stylisticians to the contrary.
	<ul style="list-style-type: none"> • Wittgenstein often foresaw the potential <i>telos</i> of instrumentation; Popper opted for ever faster and better falsification. Only Frege would have sensed that the rectification of and intuitive deficit would become necessary. 	<ul style="list-style-type: none"> • Corpus-derived subtext may well signal the beginning of instrumentation for meaning. Intuitive opacity will guarantee that it has a long career ahead of it.
<p>Alfred Jules AYER * (1910-1989)</p>	<ul style="list-style-type: none"> • Exponent of the <i>a priori</i> 	<ul style="list-style-type: none"> • A powerful sounding board for potential links with collocation.
<p>* Ayer was the only English philosopher within the Vienna Circle</p>		

Michael DUMMETT (1925-2011)	<ul style="list-style-type: none"> Creates the first interventionist link between philosophy and the empirical evidence of natural language. 	<ul style="list-style-type: none"> Sadly, there is no evidence that Dummett encountered collocation.
	<ul style="list-style-type: none"> His system of 'warrants' for speaking up allows empiricism a better chance for identifying and eliminating <i>non-p</i>. 	<ul style="list-style-type: none"> The future of his approach may lay with his students.
Willard Van Orman QUINE (1908-2000)	<ul style="list-style-type: none"> An analytic philosopher in the USA whose best known contribution to the question of empiricism was a seminal article entitled, 'Two dogmas of empiricism'. 	<ul style="list-style-type: none"> Firth and his followers, especially Sinclair, took the view that collocation had nothing to do with the 'ideas' meaning of words. Instead of dealing with concepts to explain data, it is best for collocates to read and inform one another's meaning.
	<ul style="list-style-type: none"> He seeks to preserve a cognitive approach by insisting upon the schematic, conceptual structure of the given. 	<ul style="list-style-type: none"> The cognitive is destined to fall as a result of intuitive opacity that vitiates method.

2.6.3 Examples of the reasoning processes and corpus-based verifications set in motion by prompts

2.6.3.1 Example 1: Variables

The extended table above (Table 2.3) attempts to summarise something of the process involved in dealing with a ‘prompt’ that has been found in philosophy and taking it through the *four* stages identified on page 61. This process is often fairly complicated and now needs to be elaborated and exemplified in a prose account, rather than in tabular form. The matter becomes intricate even down to the fact that across the four points it is very often the case that philosophy and linguistics differ with regard to key terms that refer essentially to the *same* phenomenon. A good example to begin with applies in the case of Frege as we link the four points. The key term is *variables*.

Even before this term, which is used by all of the analytic philosophers, can be linked with Firthian linguistics, its precise meaning in philosophy needs to be differentiated from the mathematical and logical dimensions of the discipline. For example, we need to ask whether the categories that *contain* variables are applicable to natural language philosophy or not; and the question is far from trivial. Russell was aware of the problem to the point of being able to use it to leave Frege both baffled and incapable, for eleven years, of taking his research further: it lies at the very heart of Russell’s Paradox. Frege had succeeded in creating a new version of philosophy that was based around the very interface between logic and mathematics: a philosophy based upon *variables*.

But, indirectly, Russell was asking Frege a question that bears upon the distinction between mere ‘sets’ in what mathematics calls Set Theory and the *quality* of the members or variables in each set: does the criterion for membership of a set qualify it for membership of itself? It is a question that not only lies at the heart of philosophy itself, but is at the interface between logic and metaphysics. Is the set of all sets a member of itself? If it is, it is not and if it is not, it is. It was, among other things, Russell’s way of focussing upon the qualitative nature of variables in natural language; and unless and until we deal with it, we cannot accept, let alone *argue* for, a link that bridges the four-point criterion stated on page 61.

Although Russell, to say the least, was set in his ways and thinking, he, perhaps more than any other philosopher, responded to prompts, most notably from Wittgenstein, but constantly from other philosophers. The prompt that lies behind his paradox is this: if philosophy is to be brought to bear upon natural language, its own variables must at least resemble and then accommodate those of language rather than arithmetic or mathematics. And this prompt came to him from outside of the German tradition. Russell met the distinguished Italian mathematician Giuseppe Peano (born in Turin) at the International Conference of Philosophy in Paris, in 1900 and, as he remembers in his Autobiography: "*I therefore got him to give me all his works, and as soon as the Congress was over I retired to Fernhurst to study quietly every word written by him and his disciples*" (Russell, 1967: 232).

It was from Peano that Russell developed the notion of structured variables. And *this* offers us our cue within corpus linguistics: are the most frequent 'lexical' variables of a grammar string more like grammar and logical argument, because they often accompany them, or more like the other members of a slot that they fill more frequently than mere lexical items?

This question dominated some of the sessions of the weekly English Language Research Seminar at Birmingham University. John Sinclair adopted the term 'subtechnical terms' in order to describe them. There was no sense at the time of their likely subtextual value: 'That is no *country* for *old men*.' vs. 'That is no *reason/excuse* for...'. The possibility for reducing linguistic levels from five or six to two, *logic* and *metaphysics*, was still almost thirty years away. In fact, this possibility operates at the same high level of abstractness as the Chomskyan formula for Generative Grammar, $S \rightarrow NP+VP$; or Hallidayan notions such as System. As a new theory of language, Subtextual Linguistics would look like this

$$QP \rightarrow QL \triangleright M,$$

where QP = Quasi-Proposition; QL = a grammar string of natural language that is not a made-up example, with its vocabulary items represented by *; and where M [for metaphysics] = vocabulary that subsists in the target sentence from natural language. The post-search results from a reference corpus contain a list of QPVs [Quasi-Propositional Variables], the most *frequent* of which share the quasi-logical form of the search line and constitute its ST or Subtext. The analysis which follows once ST is

revealed is in essence *dialogic* in Firth's sense of the term *conversation*. It acts as a warrant to question the search line.

2.6.3.2 Example 2: Qualia and Events

The relationship between variables and possible worlds is arguably easier to show from linguistics rather than philosophy, but if we are to obtain what Karl Popper refers to as bold research and with dramatic findings, the prompts ought always to come from philosophy; however, Firth's formula for 'meaning by collocation' prepares us for any eventuality. Co-selection is arguably the last point on a cline within corpus linguistics at which intuition is involved in experimental design and is, nevertheless, often agreeably surprised. Note how the delexical forms of *thaw* below are almost as frequent as their fully lexicalised counterpart.

MicroConcord search SW: thaw* CW: snow*
80 characters per entry Sort : 1R/SW unshifted.

```
1 uises the fact that the snow will be thawing and on steep slopes this may greatl
2 l (6 miles) SNOW CONDITION: Upper, thawing; beginner, thawed. GLENCOE LOCA
3 CONDITION: Upper, thawing; beginner, thawed. GLENCOE LOCATION: Meall a'Buiri
4 aerial photographs taken as the snow thawed revealed the original 1680s layout o
5 the house in winter as the snow was thawing showed an outline of the parterres
6 left until the spring when the snow thaws." There had been avalanche warnings
```

MicroConcord search SW: thaw* CW: relation*
80 characters per entry Sort : 1R/SW unshifted.

```
1 he trip had been seen as a sign of a thaw in Anglo-Chinese relations. The Chin
2 yesterday. The visit marks a further thawing of relations more than 13 years aft
3 t step by Baghdad in its attempts to thaw relations with the West. The search
4 uperpower relations finally began to thaw. So too does the picture of the same t
```

2.6.3.3 Example 3: Variables as part of a cline identified by means of prompts from philosophy

Suppose we process two prompts from philosophy simultaneously: (1) Russell's prompt that a perfectly logical natural language will have a grammar but no vocabulary at all,

and (2) Wittgenstein's prompt that analytic philosophy proceeds from metaphysics, but simultaneously *does away with it*.

Taken *seriatim*, our follow-up to the first prompt, using a reference corpus that is not even particularly large, yields a list of vocabulary variables and their *frequencies*. Now, if any of these variables are to qualify as forms of quasi-logical reasoning, they will necessarily be the most frequent variables. Logic is empirical and the variables, even though they fall into the category of metaphysics (as a result of Russell's reasoning), will need to be capable of being treated as truth-constant logical operators, in Carnapian terms. In other words, their link with quasi-logic must be empirically plausible. Even though *country* and *old men* may be averred to subsist in some counterfactual state in relation to the purported but possible world of Yeats's Byzantium, their lack of frequency in a reference corpus singles them out as probable *non-sense*. They are essentially, in Ayer's terms, presented as *evidence* for the way in which Byzantium treats old men, albeit that the statement is insincere: an old man as the persona in a poem writes it or says it and renders it doubly counterfactual (it is already a form of *fiction*) by sailing to Byzantium, possibly oblivious that the subtext of the first line, *logically precludes such a journey by **this** particular speaker-writer*.

And so we see that apart from the fact that *country* and *old men* are variables, in logical terms, they might just as easily be *hapax legomena*, mere one-off fragments of purported evidence made meaningless by the illogical conduct of yet another irrational Gerontius. But conclusively, *reason* and *excuse* occupy the highest frequency, without any sense that one is to be placed above the other. They share the *argument* that they bring to sentences, *even when they are not present and actualised*; and *this* exemplifies the power of subtext.

2.6.3.4 Example 4: Counterfactuals in Natural Language

The final example to be presented at this stage deals with the fact that where we detect apparent counterfactuals in natural language, part of our warrant may involve a choice between *intuitive* as opposed to *instrumental* verification.

Suppose you read an advertisement placed in the media by a service provider that offers immediate connectivity '*just like that*'. You purchase it and find that once you have got it up and running, it suddenly fails. A friend asks you what has happened and you say: 'The screen went blank, just like that!'

As a result of the advertisement and the failure of the software, you now have a warrant for intervening. Was there something to which your intuition was effectively blind? When you check the expression using a concordancer, this is part of what you find.

MicroConcord search SW: just like that
80 characters per entry Sort : 1L/SW shifted 2 characters.

```
1 My father died a few years ago, just like that. From being a very vital,
2 Now he is ready to brush me aside just like that, which hurts. Frank has
3 ople feel that to brush him aside, just like that, is unnecessary." The king
4 70 years of brand promotion away just like that. They will try to use the
5 houted. "You mean he sent you back just like that? Aren't you a shameless
6 ad crashed: 'The screen went blank just like that!', 'The program died on me jus
7 and I can't spring it on the Board just like that. They wouldn't listen. I'
8 eak. Of course it broke. It broke, just like that. It's still broke, if you
9 to stop <F02> It just ceased just like that. Yes. <M01> Yes. Mm. Well
10 started shooting into the crowds, just like that, women with their
11 e former empire will not disappear just like that. It has even more and
12 erproof tape and they'll disappear just like that." <p> The prosecution ou
13 to make your problems disappear, just like that. SHE line 0898 200172 <p>
14 le to convey; and of disappearing, just like that. At times he loathed being
15 and told us we were done-- done. Just like that. She was turning over the
16 d as he opened his own front door. Just like that. No apparent reason. Then
17 ds. So he just stops selling dope, just like that." Horner drank his beer
18 sex with the Prince of her dreams `just like that". Women's bodies do seem t
19 xer points out. `And it didn't end just like that, things were coming in a c
20 s hoping he would give up, expire, just like that. But he was beginning to
21 ike that. <p> He closed his eyes Just like that, huh?" <p> It isn't `just
22 dden heart attack <F04> Mhm. <F01> just like that. <F04> Mm. <F01> Erm aged
23 to do?" Because you're finished. Just like that. You're out. All the bills
24 been taken to close it down first. Just like that. There's clearly no point
25 we say we're sorry we're forgiven, just like that. Help us to forgive <p>
26 about Justice would be forgotten. Just like that. <p> On the other hand,
27 I wasn't going to let that go just like that." <p> With the 155p-a-
28 person. And that all seemed to go, just like that -and, like, why? All that
29 ver it. Leo, for God's sake! Gone. Just like that. I always pictured the
30 Enron said. `Here and gone. Just like that." Very slippery people." (
```

Source: Times 1995 Corpus

So here we have a case in point where *just like that* must have some empiricism that suggests rapidity or the advertising agency would not have used it; but the semantic prosody of stasis and shock that precedes the form did not enter the copy-writer's mind.

Of course, this raises the ethical question as to whether the advertising agency ought to be paid a fee for offering a faulty phrase that inadvertently highlights the possibility that the product may be flawed.

Branding is often to blame for its intuitive deficit. I recall the case of an electronic surveillance device invented in Japan for monitoring the movements of toddlers. Very few units were sold in the UK. The device was trademarked as Little Bugger.

And so, these examples begin to create an argument for checking whether the intuitive claims made for a potential node in a search of a reference corpus are worth checking. Is it ever the case that we are faced with choices where our snap intuitive interpretations are likely to mislead us and that with grave consequences? Well, one frequent *modus operandi* in courts of law involves an accused person being asked to state their ‘prayer’. Counsel often replies on behalf of his/her client with the words *natural justice*. Our intuition tells us that it is only natural to want justice. But what do the collocates say? Would they save us from falling into a trap? This example will be discussed again, in greater detail, in Chapter 3.

CHAPTER 3: COLLOCATION, CONTEXT AND CO-TEXT

3.0 Introduction

In Chapter 1, the strategy was adopted that in order to establish rapprochement between philosophy and collocation studies, all argumentation ought ideally to commence within philosophy and all applications of *theory* that evolve in this way will, ideally, be carried out as collocation in *practice*. A major benefit of this would be an improved method.

However, as in the case of subtext, when we are confronted with the difference between ‘That is no *country* for *old men*’, and ‘That is no *reason/excuse*’ the impression, based upon both the *frequency* and *nature* of the latter collocates, is decidedly *downward*. They are not actors waiting to walk on to the stage so much as travelling by lift to reach stage level. They have been brought from the origins of argument to appear and they make all vocabulary variables look like the purely metaphysical proxies for argument that they undoubtedly are.

The surface phenomena that contextualise old men in a land that ignores them give way to the empiricism of argument that questions their very *raison d’être*. The subtext questions ‘... the most important reason for them existing in the way that they do.’ (The Collins Cobuild English Dictionary, 1995:1356). Yeats, of course, rebels against the apparent *stasis* implied by the subtext, and the persona in ‘Sailing to Byzantium’ travels into adversity in search of both recognition and its concomitant golden accolades.

And so it is that, in this chapter, we examine the extent to which the ways in which analytic philosophy deals with the external world. All information gathered using this method is then visualised as practical collocation as a *first step*, and in computational terms as a *second step*, and verified using reference corpora as a *third step*.

Equally, any collocation-based method that failed to benefit from philosophy might be used to improve philosophy as a discipline, simply by *reversing* the method of argumentation. Most improvements of this kind will flow from subtextual studies in Chapter 4. However, an example is in order at this time. It is often the case in

scholarship that large numbers of scholars may be found to be working in the *wrong* place when *change* arrives. A. J. Ayer says of Frege and others:

They didn't actually refute Aristotle. They showed Aristotle's work as a little corner of logic. And they developed a much more wide-reaching, far ranging logic ... a very powerful tool of analysis.

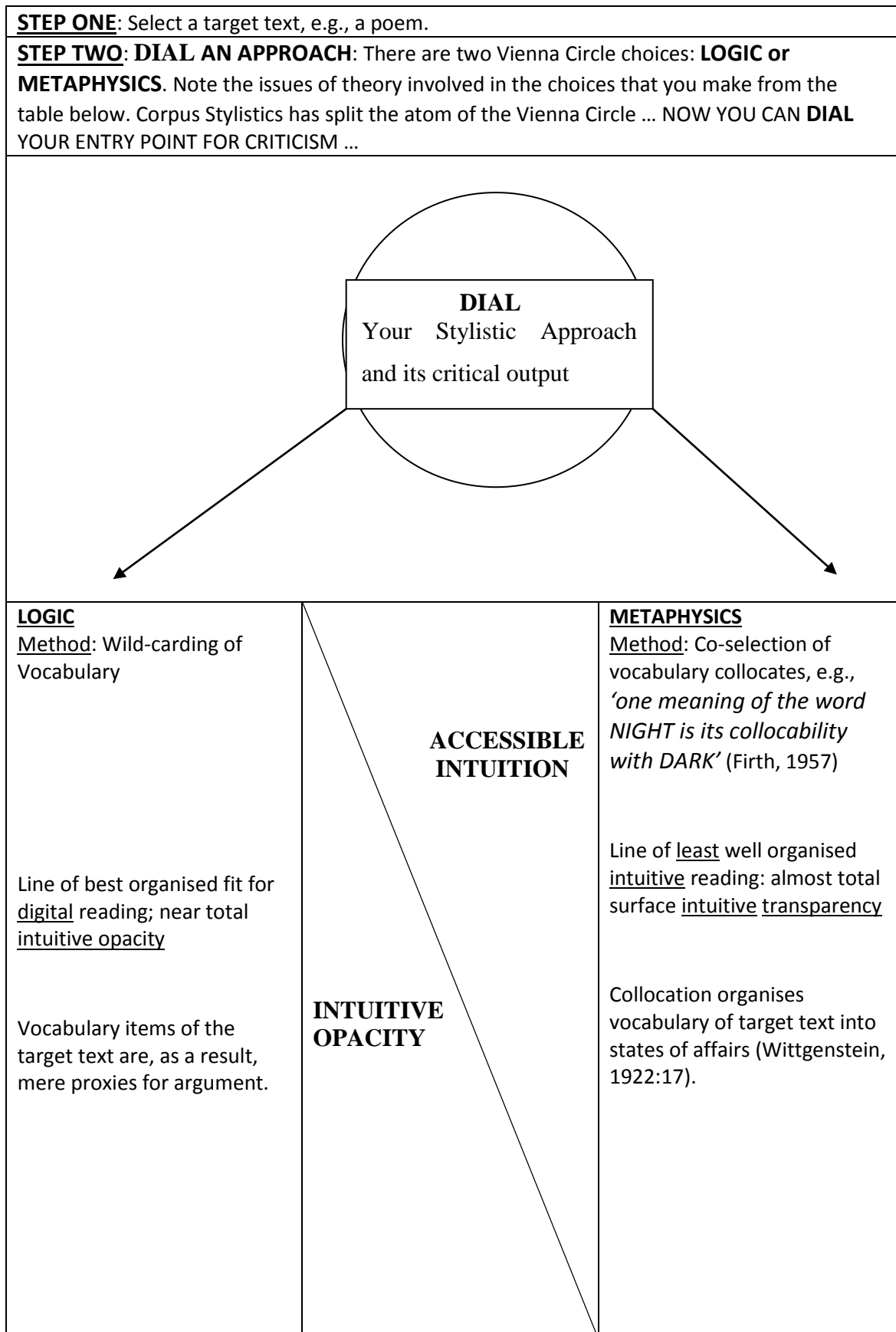
(Ayer, in Magee, 1978: 98)

3.0.1 Trust the Subtext Using the Dialling Method: Summary

Subtext may well become its equivalent in Stylistics, but it does both (1) falsify existing approaches and (2) show that they continue to operate in a 'little corner'. The little intuitive corner is revealed by the 'dialling' option I have evolved in Stylistics. For example, what follows is part of a summary for a presentation at a conference in Italy in July 2014. It is included here so that the simplicity of its method can be demonstrated.

Bertrand Russell states that a perfectly logical natural language will have a grammar but no vocabulary at all (Pears, 1985: 24). The method which follows is based on this remark and operates within 6 simple steps in Corpus Stylistic Analysis.

Figure 3.1: Trust the Subtext using the Dialling Method



<p>STEP 3: Run your choice using a simple concordancer and any reference corpus, such as The Bank of English or the British National Corpus [BNC], Google Fiction or Google Books etc.</p>
<p>STEP 4(1): IF you dialled LOGIC, THEN use the most frequent vocabulary quasi-propositional variables to interrogate the search line you drew from your target text, e.g.,</p> <p>Search Line: ‘That is no country for old men’</p> <p>Same Line Wild-carded: ‘That is no * for *’</p> <p>Result: <i>That is no reason/excuse</i></p> <p style="text-align: right;">(Louw, 2010b)</p>
<p>STEP 4(2): IF you dialled METAPHYSICS by selecting the line from a poem by G.M. Hopkins, e.g. ‘<i>Mine own <u>self</u> more let me have <u>pity</u> on’;</i></p> <p>THEN co-select <u>self</u> and <u>pity</u> (5:5) and use the output to interrogate the poem in order to assess the relevance of <u>any</u> or <u>all</u> of the <u>states of affairs</u> (Wittgenstein, 1922:7) found in the concordance lines. Do these states of affairs apply to its main concerns? Remember that Wittgenstein places LOGIC above METAPHYSICS (1914:93).</p>
<p>This leaves us with two steps to complete. They comment upon the quality of the heuristic gained, depending on the order in which we dial.</p>
<p>STEP 5: IF we dial METAPHYSICS, we automatically opt for the maximum use of our intuition. There are <u>opportunity costs</u> involved in this choice, because, at best, the fictional world of the ‘target’ text will <u>alter</u> symbolically the states of affairs provided from the reference corpus and, hence, from the external world. This computer-assisted aspect will simultaneously expand and amplify the <u>symbolism</u> and curb any excesses of cognitive response; but the resolution of the text’s meaning is swifter and more thorough-going once LOGIC is dialled.</p> <p>The logical reader’s response to having made the choice to begin with metaphysics is often: “<i>So all that stuff about X or Y was really just window dressing; while the main concern all along was Z.</i>” Of course, rationalisations assist the analysis, but they do not cut to the chase!</p>
<p>STEP 6: IF WE DIAL LOGIC FIRST, it is a little like turning to the last page of the novel. The empiricism of the grammar strings that our intuition cannot <u>cognise</u>, even if we try to call ourselves cognitivists, cannot be gainsaid. The search generates the very ‘<u>necessary meaning</u>’ that follows from it and <u>marches in step</u> with the logical construction of the world (Carnap, 1928). And the result of a logical reading is often filled with pathos: “<i>If only the ladder had been a real ladder, the persona would have been better off. But <u>as the ladder is Maud Gonne [gone], the bitter impasse of his life continues.</u></i>” In Aristotelian terms, the pathos creates what is arguably the fuller literary response. As Russell says, “... a language of that sort will show at a glance the logical structure of the facts asserted or denied.” (Pears 1985: 24)</p>

3.1 Context and Collocation as seen from a Philosophical Perspective

3.1.1 Introduction

It is only really in the language used by Firth to describe collocation that we begin to see that the expression ‘abstracted’ at the level of syntax is the main feature that distinguishes the analogue from its digital counterpart. A generalisation of this might look like this: The earlier the study of collocation, the more it will be sentence and syntax-bound. We see this in the understanding of the term within the work of investigators such as H.E. Palmer (1933) and Göran Kjellmer (1990). In fact, without the use of a computer it is easy to see how the terms *collocation* and *colligation* begin to be conflated. And yet, nothing could have been further from Firth’s mind if we consider the huge disparity between the *levels* of language that Firth ascribes to each of these terms (See Table in Louw, 2008: 248). Collocation is held in higher regard by Firth, who places it almost at the level of the context of culture in his taxonomy for levels of language arrived at through his collaboration with Malinowski. We may well reason that the *disparity* in the levels of collocation and colligation represents a primary instance of the way collocation is *abstracted* or drawn away from the notion that is similar to it in name alone. However, in philosophy, the term is redolent with meaning of a kind that does not apply to collocation, but begins to capture some of the nuances that Williams suggests for it in his hierarchical ‘crown’ diagrams (Williams: 2008; Williams and Millon, 2009). See Mautner’s entry below.

abstraction	
1.	the process of considering only some aspects of a whole.
2.	the result of this process.
3.	(in set theory) defining a set as the set of all objects that have a certain property.
4.	(in logic) an operation by which a predicate is formed from an open sentence. For instance, from the open sentence Fx , which can be read as ‘ x is F ’, one can form the expression $\hat{x}(Fx)$ which designates the property F . Similarly, $\hat{x}(Fx \ \& \ Gx)$ designates the property of being both F and G , $\hat{x}\hat{y}(Fxy)$ designates the two-place property, i.e. the relation F , etc.

(Mautner, 2000: 3)

3.1.2 Are there any early rivals to collocation within philosophy?

Given the well-researched history of the term collocation (Barnbrook et al: 2013: 32-51), it is worth using philosophy as the backdrop for asking questions that bear upon both context and collocation. This is something that has not hitherto been done. So, suppose we wish to consider what would constitute the most minimal form of context within (1) Philosophy and (2) Collocation Studies. We could tabulate our results as follows.

Table 3.1: Philosophy and Linguistics

PHILOSOPHY	LINGUISTICS
The Morning Star is not the Evening Star (Frege, 1884), within logic viewed as mathematics, or <i>Fa</i> .	One meaning of 'night' is its collocability with 'dark'. For example, in a single sentence (Firth, 1957:196).

Both of these scenarios visualise contexts of situation in the external world. And both allow for *logical* construction. Firth's example can be demonstrated to cause the disappearance of the moon or even the stars in computer-recovered contexts from a reference corpus. Even delexical forms appear, such as 'dark night of the soul', or 'dark night of Thatcherism' (Louw, 2007b: 157). Frege's formulation is mathematical. It allows for the fact that even though the referent in both cases is the same planet (Venus), the utterance itself and the conditions obtaining at the time the utterance is made and that refers to the same 'star', will be different. Its *Bedeutung* (Reference) remains constant but its *Sinn* (Sense) is governed by conditions in the external world that can penetrate even mathematics on the basis of its human logicity (Frege, 1892). It was this sort of innovation within the ancient discipline of logic that has allowed Frege to be termed the greatest logician since Aristotle. It also led to the accolade stated by Ayer referred to in the first section on this chapter, above.

Now, if we take this a step further, suppose we remove the collocates altogether. What do we achieve by doing so? In philosophy we create a situation of non-sense because *that -p* becomes too generalised to be considered. In the case of Firth's meaning by collocation *night* becomes generalised to the point of becoming a made-up example of

the type we encounter in the text books of TEFL. *Paradoxically*, and not unlike the situation in logic as expressed mathematically, the move to a *citation* form of the type we find on word-lists, *instead of having no collocates, now admits all collocates!* We perceive it in *intuitive* terms as fully lexicalised; something we never do purposefully in real life except as ‘defining language’ (Barnbrook, 2002). For Frege, the severance of the link with human life occurs and its situations and becomes *inexpressible*: *Fa* without *a* is non-sense. What becomes of word lists and tree-banks in such circumstances would not even rate a mention in philosophy and yet the Corpora List refers to them daily. This particular example of the relexicalisation of citation forms assisted Gill Philip (2011:118-9), in determining reference to colour terms in natural language, which was also the subject of her PhD at the University of Birmingham.

3.1.3 Section Two of Wittgenstein’s Tractatus as States of Affairs and Context of Situation

Now, if we examine the extent to which collocation might be said to subsist within philosophical approaches, we find that Wittgenstein writes with such clarity about his Picture Theory of Meaning that we have to admit this possibility. Look, for example at Section 2.012 of The *Tractatus* below:

‘In logic nothing is accidental: if a thing can occur in a state of affairs, the possibility of the state of affairs must be written into the thing itself.’

(Wittgenstein, 1922: 7; Emphasis added)

We see precisely the same sort of meaning (which philosophers call *necessary meaning*, see especially Carnap, 1928) that we saw set in place when *night* and *dark* are co-selected. Notice, though, how collocation studies benefit from Wittgenstein’s insight here. ‘*Written into the thing itself*’ has a number of meanings:

- Aspects that are written in are probabilistically most likely to surface and occur as part of the state of affairs and its habitual meaning.
- When they surface and occur, they will not and cannot be *accidental*.
- Wittgenstein appreciates the science that is involved in the null position (mentioned by Gill Philip).

His writings are presented in tabular form in the following table:

Table 3.2: Wittgenstein in *Tractatus*

SECTION NUMBER	TEXT OF <i>TRACTATUS</i>
2.0121	<p>It would seem to be a sort of accident, if it turned out that a situation would find a thing that could already exist entirely on its own.</p> <p>If things can occur in states of affairs, this possibility must be in them from the beginning.</p> <p>(Nothing in the province of logic can be merely possible. Logic deals with every possibility and all possibilities are its facts.)</p> <p>Just as we are quite unable to imagine spatial objects outside space or temporal objects outside time, so too there is no object that we can imagine excluded from the possibility of combining with others.</p> <p>If I can imagine objects combined in states of affairs, I cannot imagine them excluded from the possibility of such combinations.</p>
2.0122	<p>Things are independent in so far as they can occur in all possible situations, but this form of independence is a form of connexion with states of affairs, a form of dependence. (It is impossible for words to appear in two different roles: by themselves, and in propositions.)</p>
2.0123	<p>If I know an object I also know all its possible occurrences in states of affairs.</p> <p>(Every one of these possibilities must be part of the nature of the object.)</p> <p>A new possibility cannot be discovered later.</p>

This is precisely the point about monitor corpora. John Sinclair’s vision for the Bank of English was that it would be a ‘living document’ (Sinclair, 1991: 25), as a corpus that AVIATOR monitors, under the supervision of Antoinette Renouf at Birmingham City University, guards against Wittgenstein’s last line above:

‘A new possibility cannot be discovered later.’

(Wittgenstein, 1922: 9)

AVIATOR discloses a new possibility on the day on which it is first used. The term *ethnic cleansing* was detected by AVIATOR shortly after it had been used for the first time. A state of affairs that was too horrible to describe was no longer capable of being ‘passed over in silence’ (Wittgenstein, 1922: 3) and the inevitable obtrusiveness of its

birth into language was logged by the machine. It was printed out by that machine at the end of the day on which it first occurred. Today, it has its own collocates. They are recoverable. And the logic of the network that they form is recoverable as the valency-type of collocation visualised by the spiked circles used by Williams and Millon (2009) to signify the discovery of such networks. In fact, as the provision of funding for a collocation dictionary ‘failed’ in the mid-1990’s Sinclair (personal communication) kept the activities of AVIATOR separate from those that were involved in writing the dictionary. And as funding for updating the Bank of English dried up, Sinclair installed on his own computer in Florence, a Reuters corpus that was larger than the Bank, so that he could assess the deficit in terms of research that the loss of his living-document idea would cause. The value of AVIATOR is priceless in research terms.

3.1.4 Wittgenstein’s thinking foresees something like KWIC concordances

At the conclusion of my contribution to the Bootcamp Debate I included a concordance from a reference corpus for the co-selection of the terms *poor* and *self* (Louw, 2010c: 351). I made a point of stating that in doing so from a known source (the poem by G.M. Hopkins ‘My own heart let me more have pity on’), I was leaving it there for science to deal with as it catches up. That moment is now at hand.

If we consider one aspect of Firth’s taxonomy for *context of situation*, the whole of its second section is devoted to the occurrence of what he (and Malinowski) refer to as ‘*the relevant objects*’. The year of publication was 1957; but the article was written earlier than that, possibly as early as the late 1930s. At the time there was not a single computer concordance in sight. And yet, writing in the trenches in World War I, Wittgenstein offers the following insights that bear upon the arrival of concordances at least 50 years later. I have set some of them out below in tabular form with my own comments in the right hand column. Note also how many of his statements foresee the computational recovery of *subtext*. He was not the only philosopher to foresee something akin to concordances, but his commentary in ‘Some remarks on logical form’ (Wittgenstein, 1929: 162) is amazingly close to a perfect description of concordance lines (see also Monk, 2005: 55).

Table 3.3: Wittgenstein foresees some of the advantages that were to come with the use of KWIC concordances

SECTION NUMBER	TEXT OF THE SECTION <i>TRACTATUS</i>	COMMENTARY
2.01231	If I am to know an object, though I need not know its external properties, I must know all its internal properties.	<i>In other words, the <u>ladder</u> in Yeats' poem 'The Circus Animal's Desertion' is a <u>proxy</u> for the most frequent variables of 'Now that my * is gone'. (Louw and Milojkovic, 2014a)</i>
2.0124	If all objects are given, then at the same time all <u>possible</u> states of affairs are also given.	<i>A concordance will offer us a picture of these possibilities with their frequencies and, hence, their probabilities of occurrence. Wittgenstein was far ahead of Carnap in this regard.</i>
2.013	Each thing is, as it were, in a space of possible states of affairs. This space I can imagine empty, but I cannot imagine the thing without the space.	<i>Note how this foresees the object as NODE and the <u>span</u> as defining the <u>space</u>. Sinclair often treated <u>left</u> and <u>right</u> as spatial in terms of event sequences in the real world (see also Louw, 2010b: 83-87).</i>
2.014	Objects contain the possibility of all situations.	<i>But the hidden collocates of objects (see 2.01231, above) are a form of <u>dependency</u>. This coverts <u>possibility</u> into <u>probability</u>. It also makes all of Chomsky's early examples look very 'made up'.</i>
2.0141	The possibility of its occurring in states of affairs is the form of an object.	<i>Here, he summarises his position. It is usually stated in commentaries as 'meaning is use', but he says it is inside the object: he has no word to describe collocation.</i>
2.02	Objects are simple.	<i>But their composite aspects are obtainable by wild-carding. For example, the significance of the term <u>common</u> in the novel <u>Great Expectations</u> is displayed if one searches for *common*. Pip is ridiculed for being <u>uncommonly common</u>. By allowing co-selection to find all states of affairs, the forms <u>pity</u> and <u>self</u> between them psycho-analyse Hopkins more than 100 years after his death (Louw, 2010c: 351).</i>

2.0233	If two objects have the same logical form, the only distinction between them, apart from their external properties, is that they are different.	<i>Here, at last, is an example of intuitive opacity even in Wittgenstein. We need to substitute the word <u>different</u> for ‘...is that the more frequent one will be the <u>subtext of the other.</u>’ Its variables are quasi-propositional. This type of insight makes the research immensely exciting. It also bears out Wittgenstein’s point in 2.022 (below). This is a major finding (Louw, 2010b).</i>
2.022	It is obvious that an imagined world, however different it may be from the real one, must have <u>something</u> – a form – in common with it.	<i>This area of research will falsify and re-write all extant stylistics (Louw and Milojkovic, 2014a).</i>
2.05	The totality of existing states of affairs also determines which states of affairs do not exist.	<i>Of course, many fake states of affairs exist and continue to be relied upon. However, the forensic capability of the approach is to be welcomed. Louw (2003) demonstrated the unrelatedness of <u>truth</u> and <u>reconciliation</u> as collocates. As all legal institutions come into being as <u>precedent</u>, this lack of collocability falsifies the TRC. Nobody has accorded written recognition to that fact. The LIBOR rate has no correlation with the external world; but the ability to trace ‘negative facts’ using collocation must finally make a key contribution to the uses of literacy.</i>
2.062	From the existence or non-existence of one state of affairs it is impossible to infer the existence or non-existence of another.	<i>The key to making such inferences is stated clearly here. The forensic uses of collocation are still in their infancy.</i>

From this point forward, section two of the *Tractatus* moves into a description of The Picture Theory; and because states of affairs (Wittgenstein), sense-data (Russell) and context of situation (Firth) all have things in common, we need now to move on to consider the *manner* in which *change* takes place *during* any or all of them and the *extent* to which philosophy and collocation record those changes.

Firth covers himself very ably in this regard by building into his taxonomy for context of situation a functional boundary as well as a logical *telos*: His Section C refers to the 'Results of the Verbal Action'. Wittgenstein implies the importance of outcomes because he moves to The Picture Theory in response to having attended a legal trial in Paris that dealt with a motor accident and in which toy cars were used in court to reconstruct the accident. Russell, in the short space between 1914 and 1918 (some of it spent in prison) unwittingly moves towards the Firthian position: he begins to downplay *sense-data* in favour of *events*. In fact, Russell's thinking in this regard was ahead of its time. The earliest date of the diagram in Firth's papers was probably 1930 (Firth, 1957: 143).

As early as 1912, in his book, *The Problems of Philosophy*, Russell does take into account what collocation would be forced to examine (if he had ever encountered that term): the colour of the surface of a table alters along with the time of day. Issues of perception are dealt with as problems: straight sticks are immersed in water and look bent, but all of the enthusiasm for *qualia* begins to evaporate and, by 1918, events (which are also bound in time) move into a situation of greater prominence in Russell's thinking. This time they are there to stay and, although they do not reach finality during his lifetime, they are accorded Five Postulates in 1948, one of which is apparently borne out in Louw (2011:193).

3.1.5 Collocation and Wittgenstein's Picture Theory of Meaning compared

As we saw earlier, Wittgenstein believed that objects and the vocabulary that represents them contain the logic of necessary meaning somehow lodged *inside* them (Wittgenstein, 1922: 7). For Russell, *real* logic was more involved within the grammar of natural language than any other level. And although *objects* contained sense-data that were part of the key to those forms of simplification that were needed to attain logical atomism, for philosophers, vocabulary had a strong metaphysical component. It was difficult for philosophers to control it. In fact, although Russell was aware for most of his long life that better sources of empiricism were needed in order to bring his work on the external world to fruition, there was, even by 1914, very little difference between Russell's empiricism and that of Hume.

However, Wittgenstein was more fully aware that metaphysics was always infused to a degree by logical form. In fact, in his discussions with Russell on the notion of logical atomism, Wittgenstein was always keen to point out that one meaning of atomism was purely destructive and had its *reductio ad absurdum*: unless the process of breaking things down was fully logical at all times, it would simply revert to being a breaking down, ungoverned by logic. Only logic can make simplification scientific.

This is where pictures and collocation begin to look similar. Initially, Russell saw the components or sense-data as being capable of making a picture notwithstanding their having been simplified, but for Russell, the picture was largely a *still* picture. A picture of himself would, of course, need to contain his pipe (and the famous Canadian photographer, Karsh, took such a picture at Russell's request). But it took Russell longer to realise that the picture itself must be a component of a *class* of re-presentation that derived its meaning as much from its components as it did from having a purposeful beginning, a middle and an end. This realisation moved Russell's thinking away from snapshots to process and process offered a better characterisation than sense-data. By 1918, *events* were ready to replace sense-data. This posed a question.

How, then, were pictures to remain logically atomistic whilst simultaneously gaining in logic for the duration of that process? By 1918 he had settled upon the *name* for that larger class, and it was to have phases as well as components and these, it turned out, were to be compatible with *context of situation* as visualised by Firth and Malinowski more than twenty years later. Russell began to deal with *events* and his association with them was to last thirty years.

Of course, collocation could cope with the provision of the *qualia* of sense-data and that, *a fortiori* in modern times, from a reference corpus. But can collocation signal the transitions and phases of events? What are their location in time and space? How do we know when they are over? Books on negotiating and culture tell us that in Japan negotiations that we believed had been concluded decades ago are still felt to be current albeit 'paused'. In Japan, loyalty to companies that closed twenty years ago continues to subsist in the minds of 'former' employees. For this reason, collocation provides better empirical evidence for describing the process of events and especially their termination.

A good example of collocates that represent *phases* of an event of long duration may be seen in concordances from reference corpora for the co-selected nodes *position* and *untenable* (Louw, 2010a:84). Sudden transitions may collocate with *turn of events*. This concordance already appears in print (Louw, 2010a: 88) but is reprinted here because of the atmospheric nature of its collocates and hidden meanings that are characteristic of Contextual Prosodic Theory (Louw, 2000).

MicroConcord search SW: turn of events

80 characters per entry Sort: 1L/SW unshifted.

1 things suddenly got worse with a **turn of events** so improbable that I still can't
2 shareholders, be party to such a **turn of events**"? The answer is simple. There is
3 oosing his wicket to Salisbury, a **turn of events** not at all to the young man's ta
4 ecause, last week, in an amazing **turn of events**, I found myself nodding in agree
5 ment of defeat. This astonishing **turn of events** was the result of several factor
6 when Asprey unveiled the latest **turn of events** in its colourful history. The
7 yers were dismayed by the latest **turn of events**, which interrupted a critical ph
8 xpress my distress at the recent **turn of events** (report, February 13). I can u
9 s much frustration at the second **turn of events** as he did pleasure over the firs
10 o bowl his off-spin. A strange **turn of events**, which had everything to do with
11 re Italy invaded. The subsequent **turn of events** there meant that his pictures we
12 e UN forces is very upset at the **turn of events**, there being no intention to cau
13 wever, reasonably happy with the **turn of events** this season. They may not be dom
14 n Britain, are astonished at the **turn of events**. Cantona began this year in such
15 , to save the government. "The **turn of events** had left me in a distraught stat
16 We are frankly astonished at the **turn of events**. We were used to living rough an
17 opponent and were aghast at the **turn of events**. In America, from where the in
18 ters, are hugely relieved by the **turn of events**, keeping the player Ferguson des
19 Sanderson was devastated by the **turn of events** and later left Bovis, as the pro
20 Oxford's disappointment at this **turn of events** was allayed to some extent by Ca
21 Brian were full of hope and this **turn of events** was totally unexpected for them
22 in a gay disco. Crucially, this **turn of events** threatens the lifelong friendshi
23 them again after September. This **turn of events** will, I am sure, inspire Donald
24 st have been a further unnerving **turn of events** for Spring. Spring had led the L

Of course, there are two sides to every story. Some of the quarrels between Russell and Wittgenstein are today legendary. But while Wittgenstein was telling Russell that too much simplicity is destructive of logic, he had his own difficulties, as Glock remarks:

'The Tractatus's failure to provide examples of elementary propositions is due less to agnosticism than to difficulties Wittgenstein encountered (in The Notebooks) in trying to square his preconceptions about simplicity with his logical specifications.'

(Glock, 2005:105)

In fact, what Wittgenstein found, collocation might have provided more readily than the grid system that Wittgenstein evolved in 1929 for coping with the problem

(Wittgenstein, 1929; Monk, 2005: 54). Glock writes about Wittgenstein's method for resolving it. We see that it returns to Russell's position, but makes it more robust:

'Nevertheless, hints in the Tractatus ... indicate that analysis proceeds in the direction of the phenomenally given (sensory impressions).'

(Glock, 2005: 105)

But look how cumbersome it all appears without collocation in 1929, even in Glock's sentence below that tries to make it sound easy.

'...take your visual field, flatten it and put a grid across. Elementary propositions use the coordinates to refer to a point in the visual field, and ascribe to it a shade of colour... Accordingly, objects are minima sensibilia ... such as shades of colours, tones and smells and simple spatial relations.'

(*ibid.*)

If computers had been available, a simple move to a probabilistic model would have supplanted the grid Wittgenstein produced in 1929.

And so we reach a point half way through Section Two of the *Tractatus* where the subject moves from states of affairs to The Picture Theory. Many commentators have indicated that the *Tractatus* is often more valuable for what it *hints* at than for its direct statements. States of affairs looked like a plausible approach to empiricism. But the transition into The Picture Theory is followed at the beginning of Section Three by the assertion that *'a picture is a thought'*. This shows us that we are suddenly in danger of departing radically from Firth's brief: collocation has nothing to do with the *ideas* meaning of words.

But if we use this point of transition in Section Two of the *Tractatus* to take stock we ought to do two things: (1) return to Firth's principle through Sinclair's statement during Wolfgang Teubert's interview with him as part of the preface to the OSTI Report (Krishnamurthy, 2004); and (2) to study the *hints* in the *Tractatus* for lines of thought similar to those of Sinclair.

In the interview with Professor Teubert, Sinclair says this: *"It took another twenty years for this particular penny to drop."* (Krishnamurthy, 2004). His amplification of Firth involves a form of reasoning that ought now to be applied to Wittgenstein. Similar

pennies have taken at least as long to drop in Stylistics, as we shall see later. Here is the passage, the import of which took twenty years to be understood:

*Collocation, I reasoned restricts the meaning of the words involved, it does not enhance it. Now if collocation is a reduction in meaning, then dark night actually is not one of the meanings of night but a reduction of the meaning of night. This cannot be the answer. But it took a long time to give up the traditional concept of the word as the unit of meaning... So the adjective dark is not selecting among all possible nights, the dark ones, but is reinforcing the dark element **already in night**. I call this the 'focusing' element of an adjective, as against the 'selective' function that we are familiar with.'*

(Sinclair, in Krishnamurthy, 2004: xxi)(Emphasis added to 'already in')

But note that Wittgenstein has already processed the substance of what I have emphasised above:

'If a thing can occur in a state of affairs, the possibility of the state of affairs must be written into the thing itself.'

(Wittgenstein, 1922, *Tractatus*, 2.012)(Emphasis added)

However, this is only one of the many 'hints' that are to be found in the *Tractatus*. There is much *pathos* involved when we compare the up-beat bravado of the philosophers with the depression of having *no access to stored empiricism* of the type that computers were to offer in future. The more the speed with which they thought they were advancing, the more tragic were the wrong turns, especially in the case of Wittgenstein, into what both Frege and Russell knew were areas that were 'rotten with psychology'. To watch the brilliance of Section Two of the *Tractatus* degenerate into the 'totality of things' recorded in almost three volumes of the philosophy of psychology and the largely anecdotal volume *Philosophical Investigations* (Wittgenstein, 1953) are tragic developments in scholarship that were caused partly by the signal lack of the computer.

Now, pennies that take a long time to fall, aside, we need to capitalise on John Sinclair's reasoning by moving to a level of abstractness that has only been made possible by linguistic computing. We need to argue the *focussing function beyond* grammar and away from adjectives. We need to argue it on the basis of the *hints* of Wittgenstein, with the simplifications provided by Russell and Carnap and with the dismay that it might have caused within the Vienna Circle.

Our main problem has been that we have concentrated on collocation as a phenomenon that affects vocabulary rather than one that *relates* vocabulary to logic. This is a major finding of this research. And when we do that, we find that what looked like a fairly trivial phenomenon of focus and of balance turns into the main collocational phenomenon that is, in the views of both Wittgenstein and Sinclair, *inside* or ‘somehow inside’ vocabulary itself. Once we agree that grammar strings have collocates, we witness the focussing phenomenon operating as the logic of necessary meaning and we can dismiss all of our earlier naïve beliefs: for example, who in their right mind invented stop words? And the restrictions or reductions begin to bear upon limitations of *possibility* as a result of massively improved *probability*. We feel the uncomfortable situation of the persona in Yeats’s ‘*The Circus Animals’ Desertion*’ not because of the loss of a *ladder* but because the reversal to which I am referring *turns* all vocabulary into mere proxies of logic. It is at once bewildering and brilliant.

The choice of the vocabulary item *ladder* has useful and fitting associations which are the normal stock-in-trade of literary criticism, such as aspirations, phallic symbols etc., but the term *ladder* itself is swamped with the duress of the loss of a human family member whose loss forces new circumstances on the bereaved. This duress stems from the logic of the subtextual line confirmed in a reference corpus, ‘now that my *’s’ (Louw and Milojkovic, 2014a). Americans may say of Yeats’s line: “Some ladder!” based loosely on a speech in the USA by Winston Churchill commenting upon Adolf Hitler’s assertion that if America joined the war, its neck would be wrung like that of a chicken. With perfect timing Churchill, the orator, said: “Some chicken! Some neck!”

Ladder has a symbolic appeal, but the link with Maud Gonne is established because ladder stands as a substitute for a very limited range of lexical collocates of *now that my*. This fact is gratifying from the point of view that this thesis, at a high level of abstractness, has two main chapters: one drawn towards *metaphysics* and the other towards the *logic* of natural language. Their use in combination feels comfortable for the very reason why the Vienna Circle insisted that *logic* and *metaphysics* should never be parted. We part them during analysis and re-unite them during the process of literary criticism.

We now move to an examination of the *hints* that show what Wittgenstein believed, but was unable to prove, because intuitive opacity applied to him as much as it does to all of us. There are many hints, but for reasons both of impact and of brevity, only a limited number of subsections of the *Tractatus* will be cited below. But note how they point more in the direction of variables that cannot be seen: their drift is deeply subtextual (see Chapter 4).

Table 3.4a: Wittgenstein's *Tractatus*, section 3

SECTION NUMBER OF TRACTATUS	TEXT OF TRACTATUS
3.3411	So one could say that the real name of an object was what all symbols that signified it had in common. Thus, one by one, all kinds of composition would prove to be unessential to a name.
3.344	What signifies in a symbol is what is common to all the symbols that the rules of logical syntax allow us to substitute for it.
3.327	A sign does not determine a logical form unless it is taken together with its logico-syntactical employment.

Notice how close to the notion of a *concordance* the following is. Most concordance lines are fragmentary. The function of sorting shows us a form of logic without any necessity to read each entry aloud or to ‘say’ it.

Table 3.4b: Wittgenstein's *Tractatus*, section 4

SECTION NUMBER OF TRACTATUS	TEXT OF TRACTATUS
4.1211	<p>Thus one proposition ‘<i>fa</i>’ shows that the object <i>a</i> occurs in its sense, two propositions ‘<i>fa</i>’ and ‘<i>ga</i>’ show that the same object is mentioned in both of them.</p> <p>If two propositions contradict one another, then their structure shows it; the same is true if one of them follows from the other. And so on.</p>

The reasoning involved in the above sections of the *Tractatus* is largely different from what we encounter in linguistics. However, Sinclair is exceptional in this regard also. He devotes more attention than other corpus linguists to the manner in which the logic of reading concordances proceeds or ought to proceed (Sinclair, 2003: ix) and he is conscious of the interplay between the linearity of the text and the extent to which devices are used for speeding up the progress of the action in the text or the slowing of the action. In *'Taking a poem to pieces'* (Freeman, 1970:129), he offers the terms *arrest* and *release* for these phenomena and offers a much more authentic view of the process of metaphorisation in an unpublished work entitled *'To the left of 'of''*.

3.1.6 Developments within Philosophy between 1914 and 1950

The influence of Wittgenstein was perhaps most useful in respect of his refusal to join the Vienna Circle and, in particular, his insistence on the primacy of natural language philosophy. How does this bear upon collocation and truth?

The main issue is always truth. The logical positivists had huge difficulties dealing with verification, because the criteria that they insist upon for determining verification are always deeply embedded in theory and doctrine. A further complication was that their members came to philosophy from very diverse backgrounds. Our problem is one of empathy. We need to speculate as to how they would have reacted to being shown analogue or pre-computational collocation. We also need a sense of how they may have reacted if they were shown collocation as a reliable instrument for producing a reading, based on the inherent empiricism of reference corpora. We need also to empathise with what their reaction might have been to a demonstration of the relatedness of philosophical logic and what Wittgenstein calls the *logic* of our language, that is to say, of everyday language.

For an appreciation of what verification involves in the pure sense and for logical positivists themselves, we cannot do better than rely upon the writings of J.O. Urmson (1960) that deal with the Vienna Circle between the two World Wars. We see something of the battle to find a respectable source for a theory in what Urmson sets out here:

The verification principle can be stated as a consequence of the atomic hypothesis in metaphysics; Wittgenstein hinted at this without an explicit statement in the Tractatus. If the only genuine facts are atomic facts then every significant sentence must be analysable into pictures of atomic facts. Hence to understand a sentence is to know what atomic facts must obtain for it to be true; and to know this is to know how to verify it. But it is clearly undesirable to derive an anti-metaphysical principle from a metaphysical doctrine in this paradoxical way. Moreover, since the metaphysics is the picture of the world which seems to be implied by an empiricist epistemology in combination with Russellian logic, it is undesirable to present a major epistemological thesis as a consequence of metaphysics. So we must be able to give a direct epistemological justification of the verification principle: and to give a plausible one is not hard... We may perhaps have had no previous experience of rugby football and start going to watch games, where we overhear people's comments, but have no expert interpreter with us. At first we shall hear such remarks as 'He was off-side' or 'He missed his tackle' without understanding them at all. But if we go on watching, listening when these things are said, and noting when they are accepted or rejected by other experts, i.e. observing what counts as a verification of them and what a falsification, we shall get at first a rough, and later an exact, understanding of them. When, and only when, I have learnt what empirical circumstances verifies and what falsifies the statement 'He has missed his tackle', then I understand the statement.

(Urmson, 1960: 109-110)

And so we can see that the requirements for satisfying the verification principle change from the time of Hume to the period between the wars.

Table 3.5: Empiricism in Urmson

<u>Old empiricism</u> (Hume)	Metaphysical thesis:	<i>All objects are sense-given.</i>
<u>New empiricism</u> (logical positivism)	Metaphysical thesis:	<i>All facts are sense-given.</i>
<u>Old empiricism</u>	Epistemological thesis:	<i>All significant words are names of sense-given objects.</i>
<u>New empiricism</u>	Epistemological thesis:	<i>All significant sentences describe sensible fact.</i>

(Urmson, 1960: 108)

But for our preparation for further and better empiricism just before the dawn of computing in philosophy we need to turn to Michael Dummett (2004). Here, at last, we find a definition that dovetails with collocation and with the notions of time, space and event. The shoe pinches sufficiently for Mautner to distinguish Dummett's approach from its predecessors by referring to it as *verificationism*.

*'As a theory of meaning and truth, verificationism, usually associated with the name of Michael Dummett, is the view that to accept a statement as true is to have adequate grounds for asserting it. In the case of mathematics, we accept a statement as true if and only if we think that a proof is available. In other kinds of inquiry, we may not have strict proof, but other forms of warrant fulfil a similar function. Generally, the statements that we accept as true are those that we have some warrant for asserting. It can happen that both a given statement *p* and its denial, not *p*, lack warranted assertibility.'*

(Mautner, 2000: 590)

But here, at last, is a definition from a philosopher that allows us to use collocates, for example, those of *natural justice* in an entirely verificatory manner. The computer would get us past the need for a canonical form of verification.

'For convenience, I shall use the term "verify" to cover what falls short of verifying a statement in the strictest sense; I shall treat it as covering also possession of grounds sufficiently compelling to warrant asserting the statement. Such grounds would justify the assertion, which is why I am speaking of a "justificationist" rather than a "verificationist" account of meaning. It is of interest to seek to make this notion precise; but that is not what concerns us in the immediate context.'

(Dummett, 2004:45)

Dummett's method bears out my use of grammar strings to read natural language without the use of intuition. He sees logic in natural language not only as empirical, but suitable for use as repeatable logical events. He ascribes the idea to Frege:

'As Frege remarked, theorems and their proofs store up deductive routines which do not have to be gone over again, but can simply be appealed to.'

(Dummett, 2004:43)(Emphasis added)

And so, we no longer need to visit the courts in order to find canonical proof of the meaning of natural justice. A cluster of collocates is now sufficient 'warranted assertibility' for raising the matter. Notice also that Dummett refers to the criteria of the

metaphysics of time and space and that these are crucial for a proper interpretation of the collocates of *natural justice*. It is in the courts of law (as shown by newspaper corpora) that *breaches* and *denial* become frequent collocates and *not* in legal textbooks. It is, of course, also a question of *genre*.

Sadly, the death of Michael Dummett (on 27th December, 2011) interrupted the main project of his life's work. His view that truth is vested in a 'warrant to speak' is much closer to the *forensic* concerns of collocation than even his most distinguished commentators realise. It is not enough to argue that his 'warrant' theory is weak because it seems to provide little more than theories of truth based upon traditional rhetoric and discourse.

Dummett's theory is stronger than all of its predecessors for the very reason that traditional rhetoric and discourse are often used to debate *non-p*. What Dummett singles out for his later theory is the fact that where the empiricism *reveals non-p*, that automatically furnishes a warrant (a duty) to speak up. Such an intervention might only *ever* be properly prompted by *collocation*. This would make collocation part of the procedures used by a philosopher. It may take the form of a simple question, such as this one, with its accompanying warrant:

Why is it that references to *breach* and *denial* of *natural justice* only ever seem to come up in court and never in legal textbooks? They ought to be taken seriously if we consider that these collocates outweigh all other collocates in newspaper corpora.

That 'warrant' appears to have accounted for the term *natural justice*'s having attracted a 'D-Notice' as it no longer appears in newspaper reports. Here is the concordance for *natural justice* from the Times 1995 Corpus.

MicroConcord search SW: **natural justice**

80 characters per entry

Sort : 2L/SW unshifted.

1 g Japan. England 2 Japan 1 IF **NATURAL justice** had prevailed at a wet Wembley o
2 June 1995</Date> <Headline>No **natural justice** in cross-examination refusal</he
3 e her powers in accordance with **natural justice**: see R v Birmingham City Justice
4 tices to act in accordance with **natural justice** and that had not happened in the
5 al with them in accordance with **natural justice** and the public interest. Thursda
6 f Glasgow had not acted against **natural justice** in refusing Conor O'Reilly a pla
7 ce, it would seem an affront to **natural justice** for the Serious Fraud Office to
8 t Mr Errington had been "denied **natural justice**" because the JP, Elizabeth Wilso
9 laimed Hitchins had been denied **natural justice** by the "general inadequacy of th
10 sues, of their man being denied **natural justice**. His case has received massive s
11 held for life. I believe that **natural justice** demands the Corporation's reason
12 eems to me beyond the bounds of **natural justice**. I wonder how many of those with

13 lved an allegation of breach of **natural justice**, the secretary of state was not
 14 station is claiming a breach of **natural justice**. ScotFM claims Northsound could
 15 e university acted in breach of **natural justice** in denying him a place at medica
 16 onably and are not in breach of **natural justice**. Yet all too often it seems th
 17 to other soldiers, in breach of **natural justice** and their individual rights. T
 18 ner of that dog was a breach of **natural justice**. The Queen's Bench Divisional
 19 that there had been a breach of **natural justice** because Mrs Fanneran had not bee
 20 here there had been a breach of **natural justice**. His Lordship did not accept t
 21 here there had been a breach of **natural justice**. There were any number of theo
 22 ly occurred, but as cases where **natural justice** had been denied because the natu
 23 introduction of the concept of **natural justice** into a category of situations,
 24 decision appears to conform to **natural justice**. The letter sent to Dr Coker by
 25 procedures and, consistent with **natural justice**, no further announcement will be
 26 spondents had acted contrary to **natural justice** by failing to advise him that a
 27 it will be strongly contrary to **natural justice** for him to be given an opportuni
 28 been irregular and contrary to **natural justice**. Section 60 (4) provided: "The s
 29 giving evidence is contrary to **natural justice**. "The sums involved could be e
 30 it. Yet it would be a denial of **natural justice**, to say nothing of the obvious,
 31 he refusal had been a denial of **natural justice**. The decision concerned a large
 32 asons was a "blatant" denial of **natural justice**. He gave the Egyptian-born broth
 33 ngs, on the ground of denial of **natural justice**. In particular it was averred th
 34 ic evidence, it was a denial of **natural justice** for the magistrate to refuse to
 35 his occasion to the dictates of **natural justice**. What makes the RFU's decision
 36 rce cannot be at the expense of **natural justice**. In deciding on terms of a div
 37 trative decisions, fairness and **natural justice** required that the Court of Alder
 38 l reflection, that fairness and **natural justice** required that that decision shou
 39 e that warm feeling from seeing **natural justice** done. Here is my suggestion fo
 40 exonerate Scott from respecting **natural justice** and a sense of proportion. Par
 41 "However, it is fundamental to **natural justice**, which governs all Jockey Club d
 42 lf a sleazy one it goes against **natural justice**. These are serious accusations o
 43 95</Date> <Headline>Breach of **natural justice** finding sets Dempsey free</headl
 44 rules and the interpretation of **natural justice**". Mr Marriott, 48, was relieved
 45 w. If that is what it is doing, **natural justice** surely demands that ministers ex
 46 it. It is not proven. It is not **natural justice**. Fyfe argues that the alterna
 47 e just wants to air the lack of **natural justice**." </Group> </Story> </Art
 48 er legal footing in not denying **natural justice**. And unlike the present system i
 49 ldegrave and an offence against **natural justice**. Instead Sir Richard amazed Down
 50 he face of another principle of **natural justice**: nobody with a direct financial
 51 tors Act 1974 nor principles of **natural justice** required that to be done and a n
 52 d contrary to the principles of **natural justice** for the secretary of state to ne
 53 e notion that when the rules of **natural justice** had not been observed one could
 54 rwise in breach of the rules of **natural justice**. Solicitors: Cornell & Co, Leo
 55 cited claimed that the rules of **natural justice** required that not only was he to
 56 ured) would expect the rules of **natural justice** and a fair hearing to be availab
 57 (iii) in breach of the rules of **natural justice**, the inspector had regard to mat
 58 d been a breach of the rules of **natural justice**: see Fairmont Ltd v Secretary of
 59 e implications and the rules of **natural justice**, we must give him the chance to
 60 ld seem to require it. So would **natural justice**. Capital FM has invested £18 mil
 61 e relatively straightforward as **natural justice** would indicate that anyone holdi
 62 persecution must surely breach **natural justice**, the branch of law that upholds
 63 he expectation in any system of **natural justice** that there should be a right of
 64 to be conducted by the charity **Natural Justice**, follows a pilot project in Cumb
 65 it also would appear to distort **natural justice**. A small tour operator which org
 66 dify their structures to ensure **natural justice**. Sadly all this will divert scar
 67 cott inquiry are a "travesty of **natural justice**", as the heading says, how can w
 68 nquiry has become a travesty of **natural justice** John Major should set up an in
 69 plunged into a civil war about **natural justice**. THE Labour leader is partly a
 70 ealt with firmly, it was hardly **natural justice** to reward the victim with a free
 71 ween Mr Dewar and Mr Vaz "was a **natural justice** exercise" to enable both sides t
 72 Any such interference with the **natural justice** that this case so urgently deman
 73 Loudon feels that it would be "**natural justice**" for Niugini shareholders to hav

Source: Times 1995 Corpus

So what exactly needs to be revived if we are to take Dummett's project further?

It may be best to put this in point form, because Dummett was arguably the greatest exponent of the work of Frege and had by the time of this recent death succeeded in applying his 'warrant theory' to the language of mathematics. Barry Taylor (2007), the

commentator on Dummett's life and works believes that the automation of the empiricism of Frege's arguments would be highly 'non-trivial': a form of high praise within philosophical circles. As we see from the last paragraph below, Dummett intended to convert the mathematical model for use on ordinary language.

If he had accomplished that, the functional empiricism of collocation would have been embraced by all of the natural language school of philosophy. Collocation does, after all, do both of the main things required by logic: it identifies that p as well as $non-p$. Collocation's forensic interventions would have, and still can, Dummett's death notwithstanding, allow Frege to show $non-p$ the door. Below is shown the crucial part of the entry in *The Shorter Routledge Dictionary of Philosophy* (Taylor, 2007), that deals with the life of Michael Dummett yet, crucially, misses the opportunity that awaits philosophy in the form of collocation as the prompter of warrants.

The only hope for the survival of linguistics involves the recognition that *collocation* is scientific. Only philosophy has the ready mechanism for running collocation in parallel with an automation of Frege's language of mathematics in order to achieve this handover process. Collocation will need to be handed over to science by philosophy.

Once this has been carried out, collocation will return to take its place in subtextual linguistics, as a place from which collocation has been used to banish the cognitive in all of its forms. This was one of Frege's objectives.

What nobody appears to be saying is that the *warrant* is likely to deal with *exceptions* of a forensically recoverable kind. Louw (1993) would have been a case in point. The binarity of *irony* and *insincerity* would surely have constituted a 'warrant'. And so, with all due respect to Taylor's entry below, one is inclined to say that what he sees as the weaker truth, in his view, would be the stronger truth where it is supported by corpus data.

'Analysis of truth suggests we seek such verification in patterns of assertion. But scrutiny of those patterns shows that there is no distinction between use which manifests knowledge of classical truth-conditions and use which manifests knowledge of a weaker kind of truth - for example, one which holds whenever we possess a potential warrant for a statement.

Such considerations motivate reconstructing sense as given by conditions for this weaker kind of truth. But rejigging Fregean semantics in line with such a conception is highly nontrivial.'

(Taylor, 2007: 205)

And from here going forward, an appreciation of the science of collocation (which is already fairly complicated even in 1993) would be almost inevitable, especially if one recalls that my proof for the binarity principle based upon *inadvertency* and *advertency* for *insincerity* and *irony*, respectively, took seven years to appear in Louw (2000) and whose authenticity and durability have, to my knowledge, never been queried. It was based upon the inference that the delexical form *recipe for*, co-selected with *success*, would be insincere. It was, at the time, borne out by more than twenty instances in the Bank of English. However, it was not until 2004 that I began work on the philosophy of language. I only realised the full significance for collocation of Dummett's project in time for my presentation to the Spanish Corpus Conference in Alicante in March of 2013. My intervention, although more than 'warranted', received only slightly shocked approbation.

And so, the climate for interdisciplinary co-operation between Firthian linguistics and analytic philosophy on the matter of collocation as a supplier of Dummettian warrants now stands ready. It will be even more conducive as we move to the next chapter on corpus-derived subtext: collocation is active in the border-line between the 'logic' and 'metaphysics' of natural language. Its activity is not only capable of falsifying aspects of the Saussurean distinction between *synchronic* and *diachronic*. It is also able to uncover a better understanding of the *a priori*.

As we approach Chapter 4, the interest in its concerns would, for Jakobson, have been the paradigmatic axis; however, the approach through philosophy is strictly Fregean, with all the collocates *present* but almost entirely opaque to intuition.

3.1.7 Towards the provision of proofs of hypotheses

In Chapter One, a total of six hypotheses were set out: *four* general hypotheses and *two* hypotheses relating to software and searches. The latter two are based upon the Firthian-Sinclairian approach to meaning as collocation (Firth, 1957; Sinclair et al., 1970) and collocative power (Sinclair, 1991) as a nine-word window.

What follows will amount to attempts to prove the hypotheses. In this chapter, the following hypotheses will be considered: Three general hypotheses and one software-based hypothesis relating to co-selection:

A. General Hypotheses

1. That there will always be a direct correspondence between Firth's taxonomy (arrived at in consultation with Malinowski) (Firth, 1957) and what collocation reveals both in particular states of affairs and in their outcomes.
2. That all literary 'devices', including humour, use the mechanism of collocation as relexicalisation (Louw, 1991; Louw, 2009).
3. That collocation in action confirms Sinclair's observation that collocation reveals the function of utterances; and that this is especially the case in two circumstances: (1) Semantic Prosody, including Contextual Prosodic Theory, and (2) Forensically (Louw, 2003; Russell, 1948).
4.

B. Hypotheses Relating To Software And Searches

5. That, using *co-selection*, collocates within the span of collocative power (Sinclair, 1991) will both *spontaneously* (1) *chunk* contexts of situation, and (2) *relexicalise* in order to do so.
6. ...

However, it is worth noting that although the division that created Chapters 3 and 4 stems from Jakobson's distinction between the *syntagmatic* and the *paradigmatic*, because of Firth's assertion that *collocation* is *abstracted* at the level of syntax, there will inevitably be an area of overlap between the concerns of Chapters 3 and 4.

A brief but powerful example will be set out below in order to clear the way for this.

Attempted proofs for subtext will be set out in Chapter 4. They include only two hypotheses: a software hypothesis (No. 6) to demonstrate the power of wild-carding and a general hypothesis (No. 4).

And so, to begin the process, we consider the issue of overlap with examples from a recent paper (Louw and Milojkovic, 2014a).

Below is the product of a search for the first line of famous war poem by Wilfred Owen.

Move him into the sun

The wild-carded line (see Chapter 4) is * *him into the* *

The resulting concordance from the 44.5 million word reference corpus from *The Times* for 1995 actually contains the poet's line from a review of a sung performance to it.

MicroConcord search SW: * him into the *

80 characters per entry Sort : 1R/SW unshifted.

1 year contract, believed to take him into the £10,000-a-week bracket. Manchest
2 id not always endeavour to draw him into the action as much as they might. Hap
3 uffalo and a few pigs went with him into the afterlife. I stood in the rain and
4 child's arm as a lever to haul him into the air and snapped the limb. Seven mon
5 of coloured balloons that lift him into the air, is an intriguing spectacle, pr
6 l Berndt reached out and pulled him into the aircraft. The rescue operation to
7 the West Indian cause has taken him into the alpha class, he has already attract
8 looked quite upset as they put him into the ambulance," he said. "The gunman
9 a sports car with Odette beside him into the American lines, in the hope that he
10 allowed the ball to roll behind him into the area, where he tried to pick it up.
11 nother titan, if only to entice him into the arena where he can be destroyed.
12 urther off course by propelling him into the arms of a hard-riding matron whom o
13 a pistol in his ribs and forced him into the back of a car. He was driven across
14 eculatation: will this one launch him into the big time? The tribute album will ad
15 e head with the gun and bundled him into the boot of his car before capturing a
16 attle for democracy and invited him into the cabinet as minister of home affairs
17 th a shove as Mundee's run took him into the Cardiff box. Grainger lashed the pe
18 and sister. Last year they put him into the care of social services for two mon
19 how many of his clients whoosh him into the chaise d'amour in the VIP room, all
20 gained in this field which led him into the chamber yesterday afternoon to deba
21 ent with the system will propel him into the chancellorship. He may well be ri
22 d, but Mr Gramm's mother pushed him into the city's best school. He performed wr
23 e court heard. Malone had lured him into the club where she worked. The schoolbo
24 n. After the sport had booted him into the cold for his transgression, he chos
25 humacher may have tried to trap him into the collision by braking earlier than n
26 ise American boy, who initiates him into the comic confusion of an adult world.
27 health authority that released him into the community. David Hoare, a paranoi
28 acular turns of event then hurl him into the company of brigands, gypsies, Musli
29 nal Farmers' Union affairs took him into the complex business of agricultural po
30 found public mystery) catapults him into the cultural elite. Besides, inarticula
31 s Aeneid. As his talent propels him into the dangerous public spotlight, Virgil'
32 early stage. The company moved him into the design and drawing office where he
33 him in their pocket, they tread him into the dirt. </Group> </Story> </Ar
34 s to reconcile. He carries with him into the domain of faith the mental apparatus
35 ian shadow, was going to follow him into the dressing room. Throughout the first
36 on and his players will welcome him into the dressing-room as a unifying asset.
37 tehall was pouring poison about him into the ears of anybody prepared to listen.
38 owtine's first clear round took him into the eight horse jump-off, where he was
39 ot enough, apparently, to force him into the England XI. In his absence, Range
40 lent as a footballer had thrust him into the England schoolboy team in three dif
41 ront and his momentum just took him into the fence. With hindsight, I should hav
42 ntances and colleagues followed him into the firm which bore his name. There he
43 ther day (attempting to flatter him into the football team), they combined exper
44 was simply not enough to pitch him into the forefront of expert prognostication
45 her." As I have just followed him into the fourth decade, it strikes me that t
46 onies, where fortune soon flung him into the furnace of revolution. There he ser
47 next week's cock-up will bring him into the future." Mr Clarke has long trodd
48 to find touch, partly to bring him into the game and partly because all week he
49 ents later when Crompton popped him into the gap. The scrum half's chivvyng, al
50 of ants. Eventually, they took him into the garden, behind the building. They p

51 wife, a swimsuit model, brought him into the giantesses' fold, although he notes
52 ord Curzon during life followed him into the grave. He was laid to rest at Kedle
53 eels to lock up and had pitched him into the gravel trap. It was another piece o
54 for his "treachery", I followed him into the ground floor of the filthy building
55 ence of United Artists, who put him into the Guinness Book Of Records with a dea
56 le classes is prepared to elbow him into the gutter where he belongs and get on
57 at close range. Instinct leads him into the heaviest cover available. But becau
58 rom the wrecked fighter and get him into the helicopter. The doctor gave Leona
59 the sheikh stepped up to whisk him into the hot seat. As you would expect of an
60 coaster of Archer's career took him into the House of Commons at the age of 29.
61 tres clearance would have taken him into the javelin, the ninth event, with the
62 the ten points that he won took him into the lead in the drivers' championship f
63 l excesses as Jerry Bailey sent him into the lead rounding the home turn. "It's
64 il a late burst of scoring took him into the lead on 134, ten under par. Farry
65 trategy. Attacking golf had put him into the lead, and attacking golf was his mo
66 eight holes from the 5th thrust him into the lead. A second bogey on the last wa
67 llion, For Pleasure, which took him into the lead. Sloothaak, third on Weihaiwej
68 round of golf which catapulted him into the lead. "Now," he said, "I'm the se
69 motionless, as it whistled past him into the left-hand corner. SPARTAK VLADIKA
70 til luck took a hand and thrust him into the limelight before he received proper
71 er in early June had catapulted him into the limelight because he had called for
72 nge in the rules has catapulted him into the limelight. The International Hock
73 ctator, only served to catapult him into the limelight. It is a position with wh
74 h this year, Bailey transformed him into the most exciting hurdler seen for year
75 away from the team that drafted him into the most coveted seat in Formula One at
76 te as the Raith man surged past him into the Motherwell half, earning Roddie a y
77 harmer when Malcolm X recruited him into the Nation of Islam in 1956. To judge f
78 the Tory conference has thrust him into the national spotlight is being exploit
79 rt to help his stepson, getting him into the Navy. "He was a bad lad even then,
80 nds to the ball, it fell behind him into the net. So Swindon go seven points c
81 udsen let the ball slither past him into the net. Earlier, the Faeroe Islands,
82 21 and 28, June 4) did not take him into the North. The standard of culinary abi
83 rish Open win last week had led him into the novel situation of making the cut i
84 time in Lancashire has embedded him into the once proud, and for so long fallow,
85 f she was really trying to draw him into the open so that his psychopathic forme
86 f she was really trying to draw him into the open so that his psychopathic forme
87 f she was really trying to draw him into the open so that his psychopathic forme
88 ng the war. A later one brought him into the orbit of Home Counties Quarter Sess
89 t up. Unfortunately, we'd glued him into the pants some hours before, and the he
90 he did they jumped and dragged him into the park. He did not stand a chance." O
91 ocked driver's door and bundled him into the passenger seat. One of the assailan
92 get enough bat on Kumble to hit him into the pavilion for six. Thus did he thras
93 ant offence when Gillespie beat him into the penalty area, but no foul was given
94 nd the telephone button to link him into the person responsible for the next sec
95 oyal chauffeur, taking him with him into the political wilderness, during which
96 t square leg and McMillan lured him into the pull shot that had twice cost him h
97 Swing out of the stalls to get him into the race and get him travelling sweetly
98 a Sioux "wise woman", initiates him into the realm of herbal healing and occult
99 f victory, knowing it would get him into the right frame of mind for the US Open
100 nated his appalling family, got him into the right clubs and endlessly introduce
101 t no manner of beating will get him into the ring ... Last season at the Tetua
102 ine after three skinheads threw him into the river during the National Front ral
103 tended business trip and forced him into the role of nursemaid. His mother, a ma
104 ended business trip, and forced him into the role of nursemaid. His mother becom
105 ended business trip, and forced him into the role of nursemaid. His mother, a ma
106 ended business trip, and forced him into the role of nursemaid. His mother, a ma
107 on Browne's supervision, helped him into the Rolling Stone league with this ecce
108 me 12 months too late to propel him into the Scotland World Cup side of 1974, al
109 t was Lanark that truly blasted him into the Scottish firmament. The book made A
110 6-1, 4-6, 6-1. The result took him into the semi-finals of the eight-man exhibi
111 form with Newcastle will force him into the side at Shearer's expense. "My co
112 of the hapless Temple, forcing him into the sort of displays of scheming incomp
113 d a wine waiter. This brought him into the sort of social circles that appear
114 g son round the garage, lifting him into the spare car, like a father on a famil
115 rt overall. The crash propelled him into the spotlight, but his reaction, typica
116 ve us a wave!" before following him into the station concourse, singing "Walking
117 r woman stepped forward and led him into the strains of Jealousy. I no longer
118 mes with a handgun, then chased him into the street before killing him with a fi
119 bouncers. Mr Collymore followed him into the street and set about him with his f
120 ham and Isaac episode. His Move him into the sun was the finest I have heard him
121 daughter, both of whom followed him into the teaching profession. </Group>
122 gh-profile recruits will follow him into the team in the spring as they qualify.
123 LTA hierarchy who have invited him into the tent may live to regret it, while m
124 Illingworth had mentally marked him into the Test team. Russell now has 631 cham
125 re of the wound. "We followed him into the thicket with difficulty because the
126 on him. Hayles's victory lifted him into the top 20. </Group> </Story> </
127 to him this year. It will push him into the top bracket of Formula One earners

128 re respect, and a sigh followed him into the trap. A perfect splash-out to withi
 129 managerial skills to integrate him into the United team at short notice. With t
 130 nd his sociable nature, brought him into the university's academic life. He trie
 131 mpionship, which would have got him into the US Ryder Cup team on points,by the
 132 the present, as his train bore him into the vast drought belt yesterday. Mr.
 133 I would unquestionably throw him into the Welsh team at the next possible opp
 134 ing the people who hope to send him into the wilderness is another matter.
 135 the relationship which brought him into the world. Conor's father, "half drifte
 136 vague, who would try to coerce him into the wrong policies. He quickly learnt b

The subtext, as we can see involves a very descript semantic prosody (Louw, 1993; 2000), but the ‘destination’ of the more energetic verbs involves *light* in at least seven (7) cases. No apology is made for the use of co-selection of

* *light* *

in order to capture this subtextual relationship. By comparison with the *sun*, *spotlight* and the *limelight* are decidedly delexical, quasi-propositional, tungsten variables of an *a priori sun*.

```

MicroConcord search SW: * him into the * CW: light*
80 characters per entry
Sort : 1R/SW      unshifted.
  1 s Aeneid. As his talent propels him into the dangerous public spotlight, Virgil'
  2 til luck took a hand and thrust him into the limelight before he received proper
  3 er in early June had catapulted him into the limelight because he had called for
  4 ctator, only served to catapult him into the limelight. It is a position with wh
  5 nge in the rules has catapulted him into the limelight. The International Hock
  6 the Tory conference has thrust him into the national spotlight is being exploit
  7 rt overall. The crash propelled him into the spotlight, but his reaction, typica

```

3.1.8 Examples of earlier proofs

It took seven years before I found a way to prove the irony-insincerity binarity set out in my 1993 paper. It came as a surprise to me that nobody either in 2000 or since has commented upon or even referred to my proof, from the Bank of English, which hinged upon the terms *a recipe for* and *the recipe for*. There was not even a single counterfactual example. There was absolutely no reaction from anyone. At least my 1993 paper attracted detractors and at least one spoiler, but *proof* seemed to go down like a lead balloon.

Exactly the same phenomenon took place in Stylistics. For at least twenty years, the most avidly sought-after ‘Holy Grail’ was *the norm* from which the lines of poetry were

said confidently *to deviate*. At the first TaLC (Teaching and Language Corpora) Conference in Lancaster (Louw, 1997) and with Mick Short in the audience, I provided both norm and deviation. My victory was forgotten before I reached the car park!

And so, for the reasons stated above, I have entitled this part of my thesis ‘Towards the provision of proofs of hypothesis.’ At the very least, I have provided some wriggle-room in all of the areas examined, except for subtext, which is the most robust and where the proofs remain falsifyingly rigid and final.

Proofs are difficult things to work into shape. Here too, the assistance of philosophy has been useful. Philosophy often works with null positions or a *reductio ad absurdum*, and it is instructive to try to find such positions within literary texts, because they act as discovery procedures.

It is for this reason that I wish to begin my work on proofs with my hypothesis that all literary devices will contain within them one or more instances of relexicalisation (Louw, 1991). This is hypothesis number 1 in Chapter 1.

In my 1991 paper, published in the ELR Journal (and also included in the Appendix to this thesis), I set out to examine what is required for rare lexical forms to be kept in a fully ‘literal’ state. The hypothesis was born during a class given by John Sinclair to undergraduate students at the University of Birmingham in 1982. He invited me to sit in and he began his lecture with characteristic simplicity. He announced that every word in use as part of our vocabulary will have at least *two* meanings, one more literal than the other, even in the case of very rare words.

This statement set up a deep sense of resonance with me because I had been teaching a poem by Louis MacNeice entitled ‘Refugees’⁶. Its subject was the trauma experienced by Jewish refugees as they sail on board a ship into New York harbour. The words involved were *resurrected*, *disinterred* and *Lazaruses*.

I had begun to ‘concordance’ for them on microfiche at what was then the headquarters of Cobuild (‘under wraps’ as Sinclair described them) on the corner of the Bristol Road and Edgbaston Park Road.

⁶ The poem can be found in the full text of the 1991 paper, which is included in the Appendix to this thesis, pp. 3-30. The poem itself is presented on pp. 8-9.

MacNeice was keen to keep the literal sense of these three words in order to offer a composite meaning for people who looked like corpses that had been dug up in Europe, resurrected like Lazarus into a new life in America and received without any warmth by officialdom. Here is the poem. What I discovered from it was the fact that relexicalisation was able to be sustained *beyond* what, even then, John Sinclair was referring to as a ‘nine word’ window of collocative power (Sinclair, 1991).

The fact that the same words in the concordances looked more like *resurrected ideas*, lost items *disinterred* from piles of rubbish and business executives with nine lives being compared with Lazarus, simply bore out Sinclair’s hypothesis that there will always be at least *two* meanings, even in the case of words as rare as these.

This began in me a desire to work on the continuum between relexicalisation and delexicalisation, an achievement with which John Sinclair credits me in the last footnote of his book *Trust the Text* (Sinclair, 2004:198).

John Sinclair wrote very little about the nature of authorship and writing, but in the piece in which he comments on my research, he refers to collocation in the hands of writers and authors in this way:

*‘Such features as collocation are part of the **control mechanism** available to the writer.’*

(Sinclair, 2004: 195) (Emphasis added)

We have already examined the rare vocabulary in this poem by Louis MacNeice (Louw, 1991) and shown how that poet exercises a control mechanism by preventing these forms from being regarded as *delexical*. We now move to an attempt to use the control mechanism in order to curb the broader application of collocation within a nine-word window and to force it to apply primarily to a compound form of the poet’s own creation.

Our hypothesis refers to all devices; and, for example, alliteration would be one such device. But, of course, as a matter of authorial style, there may be stronger and weaker forms of alliteration. However, suppose for a moment, a poet wished to create a form of alliteration that simultaneously excluded as much spontaneous collocation as possible and also forced alliterating forms into exclusive relationships by removing their abstract

collocative subtleties and insisting upon alliteration and collocation by duress. That would take Sinclair's notion of a *control mechanism* to new heights.

A poet who uses or abuses collocation in this way is Gerard Manley Hopkins. The simple device is the hyphen: it is used in two ways. The first way is to bring about collocation by relating two issues that might not be obviously related in real life and the second way involves the same type of unification of issues, but in combination with a device such as assonance or alliteration or both. The examples below will illustrate this. They are all taken from a small corpus of the poetry of Gerard Manley Hopkins.

MicroConcord search SW: d*-d*

80 characters per entry

Sort : 1R/SW unshifted.

```
1 n three of the thunder-throne! Not a dooms-day dazzle in his coming nor dark as
2 elled shank-' That fledted else like day-dissolved dreams Head and foot, shoulde
3 ion, king dom of daylight's dauphin, dapple-dawn-drawn Falcon, in his riding Of
4 th melted to a rose of air; With the dainty-delicate fretted fringe of fingers T
5 For how to the heart's cheering' The down-dugged ground-hugged grey Hovers off,
6 these' Were vein'd and streak'd with dusk-deep lazuli, Plashes amidst the billow
```

MicroConcord search SW: m*-m*

80 characters per entry

Sort : 1R/SW unshifted.

```
1 d Of seamen whelm'd in chasms of the mid-main, 120 As poets sing; or that it is
2 poising palms were weighing the Thou martyr-master: in thy sight 23 Joy fall to
3 l with sighs on sighs. Look, look: a May-mess, like on orchard boughs!' Look! Ma
4 trem factam Salute thee, mother, the maid-month's Queen! At non inter omnes sem
5 ng of the strong sun, Thou matchless maiden-mother,' Tall sun's tingeing, or tre
6 50 Of us, the wimpledwater-dimpled, not-by-morning-matchv The flower of beauty,
```

MicroConcord search SW: t*-t*

80 characters per entry

Sort : 1R/SW unshifted.

```
1 cresseted east,' (aa2) THE WOODLARK Tatter-tangled and dingle-a-dangled The Woo
2 re I gather strength.... Between the tower-top and the ground Below me in the be
3 cover in their overflow, Of wringing tree-tops, chalky lanes, Marking the spot,
4 im forge, powerful amidst peers, His tear-tricked cheeks of flame For fond love
5 s; such a melting, a madrigal start! Before-time-taken, dearest prized and price
6 r detesting this' Self-trellises the touch-tree in live green twine And loops th
7 me,' Mid-numbered he in three of the thunder-throne! Not a dooms-day dazzle in h
8 re washing away: To the shrouds they took,-they shook in the hurling and A proph
9 I can teach him happiness Or take it thus-that the concurring stars That shall n
```

An example of a more general use of the hyphen by Hopkins as a control mechanism for collocation now follows.

MicroConcord search SW: *-*

80 characters per entry

Sort : 1L/SW unshifted.

1 e Escorial was built in the form of a gridiron,-the rectangular convent was the
2 play with a wide meadow; behind is a palm-tree. 4 fi Alluding to Raphael's `Lo
3 ght to me with fear. Divinity of air, fleet-feather'd gales, Ye river-heads, tho
4 r, Tresses dipp'd in rainbow fire, An olive-branch whence richly reek Earthless
5 e the waste Amongst Castilian barrens mountain-bound; A sombre length of grey; f
6 droop'd, himself reviled And bleeding saw.-Thus hung from room to room The skill
7 a cross of flowers, in purple bloom; j-He, where the crownals droop'd, himself
8 right, Down-splinter'd rocks crush'd cottages.-Drear sigh An endless round of d
9 iness Alien from yours as heaven from Nadir-fires: Fallen from my peers. IO You
10 clings to what is hers Or half-human, half-divine? Closer because faint morning
11 s brood That battled Gods for heaven; brilliant-hued,* With golden fillets and r
12 Fledged things do rustle near; Lines 88-127 Whispers of the mid-air stirring PR
13 founts and postur'd stone he mix'd. -Before the sepulchre there stood a gate, A
14 With massy pillars of the Doric mood Broad-fluted, nor with shafts acanthus-cro
15 when night clings to what is hers Or half-human, half-divine? Closer because fa
16 ther down the valley, left and right, Down-splinter'd rocks crush'd cottages.-Dr
17 ge of ready piety Legates might rush, zeal-rampant, fiery, 40 *Upon the stubborn
18 ic mood Broad-fluted, nor with shafts acanthus-crown'd Pourtray'd along the frie
19 in lone air stood the sun, Grant that close-folded peace that clad To sing scarc
20 on the mould'ring terraces amain; The Altar-tapers flar'd in gusts; in vain Loud
21 veil that covers mysteries; Under the cloud-festooned roof, And soul is subtle a
22 ough amber of dark eyes To quench the fine-drawn cataract; Went forth to compass
23 le near; Lines 88-127 Whispers of the mid-air stirring PROMETHEUS With light pul
24 g still its sunned girth; To hear the sheep-bells dimly die Touch me and purify,
25 the envied fruit with fatal smile To golden-girdled Cypris; Ceres there Raves t
26 And to the mystery of those Things To pause-then from his gurgling bill Let the
27 Then pass'd the wind, and sobb'd with mountain-ech 14 Next morn a peasant from t
28 my throbbing breast. That nest within close-barred doors, Or, if a sudden silver
29 ity of air, fleet-feather'd gales, Ye river-heads, thou billowy deep that laugh'

The corpus of Hopkins's poems, upon being searched for **-** offers the investigator a total 841 hyphenated forms.

The more memorable uses of hyphenated forms seem also to involve devices that are phonetic such as alliteration and assonance. However, in many ways, Hopkins uses the control mechanism for collocation by deploying a hyphen in order to create an illusion which is similar to the effect of a descending word frequency list. In other words, the poet has been able to restrict the use and appearance of the full range of normal

collocates and to prioritise a particular collocate, making it look like the most frequent form and freezing it in perpetuity with the use of the hyphen.

Similar forms of control are quite difficult to find in other poets and would make a very exciting research topic.

In the case of poets who use punctuation as devices, such as e. e. cummings, much more radical forms of collocational containment may well be found.

3.2 Conclusion

The search techniques involved in the co-selection of two lexical terms is open to at least some measure of intuitive commentary. However, as we shall see in the next chapter, the lexical collocates of grammar strings always come as a surprise when the grammar string is the search term. Grammar strings are hardly ever used as search terms for the reason that throughout the history of corpus linguistics, searches for a single term and co-selection of more than one term have always involved a search for lexical items intervening within a predetermined window, rather than grammar strings.

CHAPTER 4: CORPUS-DERIVED SUBTEXT

4.0 Introduction

There is, of course, an extent to which all potential candidates for filling a particular slot in a sentence are *potential* collocates, and because they are missing as we consider their candidature, they are no longer entirely putative, but are in reality recoverably *subtextual*. They are not putative because intuitive opacity prevents their being ‘considered’. However, for a form to be subtextual in the sense intended in this thesis, it needs, in order to qualify for that term ‘subtextual’, to be derived from the principle set out by Russell that a purely logical natural language will have a grammar and no vocabulary at all (Pears, 1985: 24). Hence, the term subtextual in corpus-based study really only begins when the paradigm-shift in the *nature* of the *search* of a reference corpus takes place (Louw, 2010a; 2010b). The target line is set out as ‘logical from’ (Wittgenstein, 1929) with *all* vocabulary terms wild-carded.

The two examples upon which this paradigm shift was initially founded occur in two separate poems that were being taught to the second year BA General students as part of a stylistics course at the University of Zimbabwe in 2009. In many ways, the first example, drawn from a poem by the distinguished Welsh poet, Herbert Williams, is the more striking for two reasons (Louw, 2010a). Firstly, it is more compelling because it reveals the *motive* of the central character in the poem. The poem is entitled ‘**Daughter of the House**’. The poet is at pains to show that (1) she has an unsullied reputation for self-sacrifice, but (2) is, in fact, a vicious opportunist. But we notice at once that the paradigm shift is in place: we are not searching on the basis of a vacant slot-filler, as we are trying to look *below* an *argument* for what Harold Pinter calls ‘the power behind the words’.

The result allows us to rest our case: The subtext (note how we have verified that old critical notion literally, see Cudden, 1980:665) of the poem is that the protagonist, through ‘a knack of lies’, has lived a comfortable life made possible by deception.’

Daughter of the House

It is not love that keeps her here, tending
The stubborn enterprise of age. Her hands
Are clinical expressions of a **heart**
Made bleak by sacrifice, her eyes
Neutralise her therapeutic smile.

Love is an easy master, but her guile
Springs from more terrible demands.
It is the blood's dictatorship, bending
Her uninvited kinship to the part,
Masking indifference with a knack of lies.

(Herbert Williams)

The subtext is recovered by wild-carding the vocabulary in the fragment in bold above:
heart; bleak; sacrifice.

MicroConcord search SW: **a * made * by**

80 characters per entry

Sort: 1R/SW unshifted.

1. requiring the giving of reasons, a point made explicitly by Megarry V.C. in McIn
2. d set up in Threadneedle Street, a move **made possible** by his marriage on 9 Janua
3. law, since the damage must be of a kind made likely by the characteristics known
4. ural rights even more important, a point made cogently by Lord Wilberforce in Ma
5. t was a profession in name only (a point made indirectly by my lecturer who stud
6. ee what it would cost me to have a tank made up by the local chap that most enth
7. d guilty. The youths were black, a fact made clear by the relatively unusual use
8. urseries and botanical gardens - a trip **made possible** by the award of the first
9. ting seated events last year, in a move **made possible** by the new developments in
10. ical equipment in the workplace, a task made compulsory by the UK's 1990 Electri
11. t he would have to do in 1962 is a leap made only by Gaullists predisposed to el
12. curing the reasons for doing so: a task made easier by his tortuous prose style.
13. Iranian revolution) on March 16, a loan **made possible** by the US decision to rema

A corpus-attested term for the most frequent variable is required (Louw, 2010b). The term *quasi-propositional variable* (QPV) is suggested. It is termed *quasi-* for the reason that we are not operating with philosophical logic but within the *logic of natural language*. The term *propositional implies* that this will be the most *frequent* of the variables. The use of a less frequent variable will take us further away from natural language logic in the direction of its metaphysics and that of the external world.

4.1 Subtext and the Notion of Variables

4.1.1 Variables: from the frequent to the hapax legomena

Now, because vocabulary items have been wild-carded, there will be a strong expectation that vocabulary items will replace those elements in the target text that were wild-carded. Why, then do I make it a pre-condition that only the most frequent variables uncovered in and through the process of wild-carding be regarded as the *subtext*? The reason is part of two further issues: of *depth* and of *age*. *Bleak* and *sacrifice* may not even collocate. The form may be a *hapax*. *Bleak* and *sacrifice* may be as obfuscatory as mere pieces of metaphysics are capable of being. After all, the form of fraud that they serve to mask is at the very least *rare*. However, the most frequent forms may **look** uninteresting, but the *logical* nature of the quasi-proposition that is set out in the grammar string remains related to *bleak* and *sacrifice*, because that grammar string deals with *causality* and the causality involved in being accepted by a board of trustees as a selfless person. On a one-to-one level this is logically no different in terms of its *outcome* from the causality brought into play by a *decision* to appoint, extracted from a board that has been too easily duped by the appearance or event-histrionics of the ‘daughter’. An *opportunity* was created and filled by a board in circumstances that demand *compassion*; and even the *appearance* of sacrifice had been enough to secure it. The individual ‘lucky’ appointment was, it turned out, sufficient to secure an impression of *bona fides* and the post was safely secured for life. The serendipity involved is reflected even in the words that go with other successful attempts from the reference corpus:

by his marriage
by the award of the first
by new developments
by the US decision

Each of these seems to represent an *actus novus interveniens*: some new factor that proves to be decisive. In her case, the therapeutic nature of her smile is said to have been neutralised by a sadness or a seriousness in her eyes. The fact that ruthlessness and guile are present are not always easy to spot in an interview. The unfortunate outcome is fairly rare also, but even the most frequent variables offer us a repeatable event that may be characterised as a breakthrough that is as sudden as the subtext that brings about its recognition.

The poem celebrates her *guile* and hints that her nature will have unfortunate consequence for any inmates under her control.

In terms of illocutionary force, the variables that are effectively *hapax legomena* lend great colour and tone to the contexts of situation from which they are drawn. They deepen the emphasis on the deception that gets her the job; but the logic of the grammar string is purely frank: jobs come to the just and the unjust alike. The chance of avoiding an inappropriate appointment is often fairly slim.

However, a major area in corpus stylistics that is just beginning to open up involves authorial sincerity (Milojkovic, 2013). It is either accessed via the binarity proposed by Louw (1993) or by authorial reaction. In the case of the poem **‘Daughter of the House’**, the BA class at the University of Zimbabwe decided to contact the poet and present Herbert Williams with the subtextual evidence that relates to the line set out in bold in the poem itself.

One difficulty that awaits the common man or woman, including poets, is that a line of poetry, stated with such certainty as this line

a heart made bleak by sacrifice

is found to be not only totally undermined by its subtext, but that the subtextual meaning *prospects* exactly the ease with which the daughter’s deception is carried out.

Carnap's explanation is correct, even in the absence of a computer. The empiricism of the invisible variables causes the reader/hearer such palpable unease that Carnap was prepared to call the phenomenon 'the remembrance of similarity'.

Mautner tells us that Carnap was talked out of the idea:

*Carnap proposed one framework constructed on a very slender basis consisting of one relation, 'remembrance of similarity', and basic data. The basic data are occurrences of **total immediate experiences**, and so-called sense-data are constructions based on those. A sensory quality is in turn defined on that basis. It did, however, become clear to critics, and to Carnap himself, that this construction project was too daring.*

(Mautner, 2000: 86; Emphasis added)

Slender or not, there may have been some pressure and heat from Carnap's colleagues to get him to drop the notion. In the huge retrospective on Carnap's work, edited by Paul Arthur Schilpp, no scholar takes up the notion by way of a tribute. All that we find about it is a brief reference made by Carnap himself in his own autobiographical introduction (Schilpp, 1963:26). Carnap explains:

*... I took elements as **total instantaneous experiences** (Elementarerlebnisse) rather than as single sense-data. I developed a method called 'quasi-analysis' which leads on the basis of the similarity-relation among experiences, to the **logical construction** of those entities which are usually conceived as components.*

(Carnap, in Schilpp, 1963: 17)

Carnap makes no further reference to it and one suspects that Mautner wanted to do better justice to it by including it in an entry in a dictionary of philosophy. Sadly, Carnap died in 1970, with no computer in sight or even in prospect. And it is only with the computing power of today that we discover that one crucial line in the poem of Herbert Williams is underpinned by huge amounts of subtextual empiricism.

4.1.2 Variables and Quasi-Propositional Variables

From the work of Frege forward into analytic philosophy proper, *variables* were an essential feature of all logical inquiry. This is arguably nowhere better understood than by Wittgenstein, but if we take an explanation of the term from an attempt to create A

Wittgenstein Dictionary (Glock, 1996: 215), we sense at once that two things are missing that collocation could immediately provide: (1) The conversion of *Sachverhalte* or States of Affairs into contexts of situation; and (2) the probability for the combining logically and plausibly of *objects*, as Glock painstakingly explains. Note how the notion of collocation would simplify the writing of propositions:

*The form of a name is what it has in common with all other names of the same logico-syntactical category, its combinatorial possibilities which are represented by the variables of which those names are values...An object has both internal ... and external properties. Its external properties consist in its being **combined** with whatever other objects it happens to be combined with. Its **internal** properties consist in the **possibility** of its combining with some objects to form **states of affairs**, and the **impossibility** of its combining with other objects. Each of these combinatorial possibilities is a form of an object, and essential to it. Their **totality** is **the** (logical) form of the object – the logical equivalent of the chemical valency, which determines the combinability of elements.*

(Glock, 1996: 215) (Emphasis added)

The exquisite complexity of this logical and syntactic phenomenon is in need of an empirically respectable phenomenon that is, as Firth says, *abstracted* at the level of syntax and in which combinability is either expected or disallowed on the basis of proximity as governed *probabilistically*. Collocation is such a phenomenon. Considerable support was expressed in Moscow in 2013 at the Moscow State Institute of Foreign Relations (MGIMO University) Conference on Innovation of Techniques and Technologies for Foreign Language Teaching (Louw, 2014). The audience agreed that the Copernican revolution demanded by Kant in order to disclose the *a priori* had now arrived. Collocation fits the description of such a revolution because it effectively drops the main feature that continues to keep philosophy hide-bound: syntax. It was Wittgenstein who said on more than one occasion: “Distrust of grammar is the first requisite of philosophising.” (Glock, 1996:212).

Of course, in computational terms, the difference between variables and propositional variables is less crucial and Louw (2010b) allows them to fall together as terms for the reason that propositional variables as a form of ‘concept-shopping’ would not apply to a written text but only conceivably to a text during the process of its composition. The composition of a fictional text is, of course, bound to an extent by necessary meaning.

My model was chosen from the work of Russell primarily for the reason that Wittgenstein during this period simply cannot settle upon a theory. There is no sense in going along with the major and dogmatic changes that he continued to embrace or repudiate. His later repudiations are often based upon mentalist assumptions. However, Wittgenstein's 1929 article 'Some remarks on logical form' reads like a modern work in corpus linguistics and I adopted it as my standard in obtaining a philosophical foundation for subtext. The best commentary on this article by Wittgenstein is to be found in Monk (2005:66).

4.2 The *A Priori* as Collocation

It is quite plain that where forms of meaning can be said to have been acquired without the sentient experience of the receiver, such forms of knowledge must at the very least, have been *accepted* by their recipients as a form of an eternally valid 'given'. There is very little scope for intuitive access to the *a priori*, for the reason that any intuitive appreciation of this area will lead to quarrels and disputes about what might have been experienced from the pre-natal to the 'second childhood'.

Only collocation and its as yet largely undiscovered rules will begin to show us such reliable and replicable findings as will place matters beyond dispute. It is gratifying to note that Sir Karl Popper, in replying to his critics in a now famous two-volume work, says as much:

All writers here quoted by Campbell, including myself, have pointed out that we may reinterpret Kant's a priori so as not to mean "objectively valid" but "prior to sense experience".

(Schilpp, 1974:1064)

It turns out that collocational evidence for all of this not only subsists, but, as a direct result of Popperian tests and their strenuous requirements, has become available through experimentation in stylistics. I had pointed out at the Spanish Corpus Conference in Alicante in March, 2013 and in Heidelberg (July, 2013) at the annual conference of the Poetics and Linguistics Association, that the lexical collocates of grammar strings (in the Vienna Circle's terms, forms of metaphysics related to expressions of quasi-logic in natural language) fall into a *special category of collocates*. They are spared all forms of

diachronic change, to the point that they effectively falsify Saussure's distinction between the synchronic and the diachronic. That theory subsisted for at least 100 years.

4.3 Subtext as a practical application

No two texts, except where one is the plagiarised version of the other, will ever be identical. This means that subtext where it is recovered will be relevant to the target text only insofar as (1) the brief stretch of logical form provided from a reference corpus underscores departures from the logic of argument that are inherent in the skeletal form of the search line and, (2) *a fortiori* in the oldest, i.e., most frequent variables of the form recovered by the search.

It may well turn out that, along lines established by A.A. Markov in his search for the longest sustainable matches within stochastic processes, subtexts may finally be shown to sustain themselves alongside a target text; but, for a considerable period as this sustainability is gradually uncovered, we will need to be content with the detailed matching and integration of subtexts to the lines that prompt them within a manageable and coherent space. As things stand, within this new line of discovery and investigation, no better discovery procedure exists than the stylistics of short and yet complete literary worlds (see Louw and Milojkovic, 2014a; 2014b, in press).

The idea for the use of poetry as a discovery procedure in linguistics comes from the early researches of Roman Jakobson during his work on what he called the 'Poetic Function'. There are compelling reasons for using the first few lines of a poem. These will become plain as I advance below the method I propose for the integration of subtextual findings into literary *interpretation* as a first step and literary *criticism* as the method becomes more fully established. From both a literary perspective as well as a philosophical one, this makes sense. If we split logic from metaphysics in order to allow argument to dominate the process of interpretation, we ought to re-unite logic and metaphysics so that the full extent to which the latter has acted as a mere distraction may be stated as part of our interpretation. Where metaphysics act as mere proxies of logic, we have a duty or a 'warrant', in Dummett's terms, to say so. Logic dominates the process of interpretation and the metaphysics altered by logic into devices works purely as non-sense, their old interpretation having been shown the door.

4.3.1 Intuitive opacity and the uncertain ‘destinations’ of searches and their use of entry points for interpretation

In July 2013 at the PALA Conference in Heidelberg, my paper was entitled “Subtext is deeper than genre: ‘The consequences for stylistics’”. My audience agreed that all of the fairly random-looking destinations for subtextual search were relevant to the text under discussion and accepted the view that this would be the case of all attempts to obtain a subtext. However, because as a search line they all involve total intuitive opacity, the function of the ‘destination’ imposes upon human induction a barrage of unexpected and yet highly original material, all of which finds an immediate and inevitable place within the process of interpretation. This is truly on the cusp between data-assisted and data-driven (Tognini-Bonelli, 2001). Some of these ‘destinations’ look highly irrelevant as they are disclosed by the computer; but all are, of course, related to one another at the highest level of abstractness. This is because the approach does not search for vocabulary, but instead for the empiricism of natural-language logic. They are all reliably found to be related to *argument* and *truth*. Arguments in the classroom that bear upon dictionary definitions of vocabulary are almost a thing of the past when we use this method. We are dealing with a situation in which scholarship *rarely*, if ever, finds itself: using this method, the ‘destination’ is always *terra incognita*, except for the joy and comfort that the logical form that is common to both target line and concordance line from the reference corpus will *always* relate to the *degree of veracity* of what the poet’s text is proposing.

So, while the surface forms fall in with the conventions of *genre* as a predeterminable act of composition, the subtext’s prominences that are allocated by frequency, and hence *de facto* always *old*, can take us beyond *genre* as part of a ratiocinative journey that is never predictable but is always relevant. It is abstraction without irrelevance; but with all of the apparent magic both logical and fully explicable, even though it may have remained hidden for aeons. This is likely to appeal to the popular imagination once the method has become as routine as ordinary reading. See the table below for examples of destination points.

Table 4.1: Examples of subtextual searches and their ‘destinations’

TARGET LINE	POEM	SEARCH LINE	FREQUENT QPVs (Quasi-Propositional Variables)	GENRE FIT (IF ANY)	PROSPECTION
“That is no country for old men”	W.B. Yeats, ‘Sailing to Byzantium’	“That is no * for **”	REASON EXCUSE	Rationalisation of causality	Persona’s lie In line 14 Ignores his own advice
Colloquial variant: “This is no country for old men”	As above	“This is no * for **”	TIME	Political slogan. This is no time for fools/ dithering	Political change will reward poets, etc.
“A heart made bleak by sacrifice”	Herbert Williams, ‘Daughter of The House’	“A * made * by **”	OPPORTUNITY POSSIBLE	Commentary on change (sudden) in newspaper	Power and ruthless control, but disguised.
“The earth is charged with the grandeur of God”	G.M. Hopkins, ‘God’s Grandeur’	“The * is *ed with”	SKY FILLED SMOKE	Man-made fires and pollution	Second stanza defilement of sin
“If ever any beauty I did see, Which I desir’d, and got, ‘twas but a dream of thee.”	John Donne, ‘The Good Morrow’	“was but a **”	DREAM, SHADOW, SHORT STEP	Favourable comparison; easy move to completion. Colloquial.	Happy state of love required.
“Move him into the sun”	Owen Thomas	** him into the **”	CATAPULT, LIMELIGHT, SPOTLIGHT	News media	SUDDEN UNEXPECTED FAME

Also notable are the number of insights that the method offers into the process of composition. In the poem ‘**God’s Grandeur**’ by Hopkins, the logical form of the first line may well be the poet’s response to a line containing the *sky is filled with smoke*, possibly read by the poet in the newspaper. Note the contrast: man’s sins fill the sky with smoke; this is countered by God’s storing up of punishment and reprisal in a world *charged* with the grandeur of God, ready to flame out from a sword wielded by the Creator Himself.

It is noteworthy that *earth* is wild-carded for the search and that the empiricism of the logic in the grammar words alone comes up (from the reference corpus) with the *a priori* collocate, *sky*! If computers had been invented when Carnap proposed a presentient explanation for Kant’s *a priori* as we saw earlier the task of philosophers would have been made easier. The *a priori* would today be less of a mystery. Note how the subtextual line prospects the defilement that we see in the second half of stanza one, in the poem printed below.

God's Grandeur

The world is charged with the grandeur of God.

It will flame out, like shining from shook foil;

It gathers to a greatness, like the ooze of oil

Crushed. Why do men then now not reck his rod?

Generations have trod, have trod, have trod;

And all is seared with trade; bleared, smeared with toil;

And wears man's smudge and shares man's smell: the soil

Is bare now, nor can foot feel, being shod.

And, for all this, nature is never spent;

There lives the dearest freshness deep down things;

And though the last lights off the black West went

Oh, morning, at the brown brink eastwards, springs –

Because the Holy Ghost over the bent

World broods with warm breast and with ah! bright wings.

(G. M. Hopkins)

4.3.2 The method revolutionises stylistic investigation

We see that the logical form of the first line or of any line in a poem alters the manner in which the stylistic analysis proceeds, as I say in a paper I recently delivered on the subject. As I explained in Chapter 3, it is now possible to select a slant: logic or metaphysics. The Vienna Circle would expect both to be chosen, if and only if it was agreed that the *results* from quasi-logic and quasi-metaphysics would always be reintegrated. However, the process of reintegration always takes place after the tricks of any relexicalised metaphysics have been outed.

We pause at this moment to consider what it is that makes this approach different. We may even attempt this in point form:

1. Because we intend to argue from the subtext of the first line as generalised into the rest of the poem we are attempting a paradigm shift. Most critical appreciation is accustomed to proceed from inferences that relate to *deviation from a norm* within the sphere of vocabulary.
2. By starting with a grammar string in which all vocabulary has been wild-carded, our new paradigm shift has chosen to rely upon *logic* as its starting point. This will focus our criticism of the poem upon argument rather than speculation: upon truth that is verifiable. This process is likely, in time, to become routine.
3. Hence, instead of deviance, we have made the decision to impose the logical form of the first line (Wittgenstein, 1929) upon any interpretation of what occurs later in the poem. It is easily foreseeable that this process may be automated computationally. It is also readily appreciated that his process would have taken place between the Two World Wars if computers had been available.
4. This results in the *saturation* of the poem and its world with the logic of the first line as its first objective, rather than any attempt to structure deviance at the level of vocabulary. Logic is preferred to metaphysics as part of that process. Piecemeal deviance is replaced by *systematic logical saturation*.
5. Having chosen this path, logic then proceeds *ex mero motu* to deal with any and all interlopers of the type *non-p* and shows them the door. Any refusal by them to leave then *writes* the poet/persona's motive (Louw, 1993). In '**Sailing to Byzantium**' the persona fails to follow his own advice in line 1, and by line 14 declares:

‘And so I have sailed the seas,
And come to the holy city of Byzantium.’

6. The suitability of the country for old men is unmasked as a ruse or a pretext because its subtextual variables show that they, on the surface of the text, are little more than a rationalisation.

7. The disqualification of the term *country* makes the search for *motive* all the more urgent and intensive. In fact, *country* as a motive turns out to be less tawdry than the *real* motive: recognition of one's *status* as *poet* by pagans!
8. Concentration upon logic and argument does two things:
 - 8.1 The vocabulary of the metaphysics of context may now be treated less seriously as mere intensification of reference, and;
 - 8.2 The most frequent vocabulary variables of the grammar string are seen to operate at such a high level of abstractness through their frequency, that they come to represent the moment at which they *cease to be vocabulary items and are subsumed often as 'subtechnical' terms into argument*. Such collocates are so deeply associated with logic that they finally become part of it, and in doing so, they become *immune* to the semantic change that affects the rest of the lexis. This will be a major finding. It will reverberate within natural language philosophy and improve philosophy generally, as the emphasis moves within the discipline.
9. In fact, when one examines enough lexical collocates of grammar words, one realises that they constitute a category whose presence has been averred for hundreds of years but never proven: *the synthetic a priori*.

And so, we find that the spin-off for our teaching is *objectivity*. The approach is based upon the computational analysis of material that is hugely *frequent* and simultaneously *opaque* to intuition.
10. Hence, the inability to *cognise* even the most frequent variables using intuition, leads, inevitably and through loss of face, as a direct result of the use of this method, to the *demise* of the *cognitive* with the inevitable recognition that *logic is empirical* (Dummett, 1976). The new method levels the difference in status between student and teacher.
11. It was Emmanuel Kant who declared that acceptance of the synthetic a priori would be tantamount to a scientific revolution '... something he claimed will take a Copernican revolution in the way we think about metaphysics and the nature of the mind.' (Garvey and Stangroom, 2012: 258). Up to now, it has been

only vocabulary that our intuition is confident with, but if we look at the demands that are being made by Kant, our insistence that the search for subtext *must begin at the point of greatest intuitive opacity* involves a leap of faith that can railroad the interpretation of a line by Yeats away from fake evaluation towards the *a priori* of reasoning itself. The new subtextual version of corpus linguistics will run and run until it has disclosed the very origins of argument. This can only be achieved if we accept the conditions: (1) We 'Trust the Subtext' as well as we 'Trust the Text', and (2) appreciate the fact that because the logical aspect is *finite*, we will eventually disclose the whole of argument and its logical collocates.

As can be seen from the brief table presented earlier, classification of this phenomenon may prove problematic. It may be accommodated perhaps as a hidden form of Keynes (Bondi and Scott, 2010).

In the fullness of time, this method is likely to amplify the scope of stylistics from the limited and haphazard discipline that it currently is, often hobbled by mentalism, into an instrument that is capable of grading the very literary canon on the one hand, and the provision of evidence of sloppy composition on the other. Allow me to offer a case in point by way of an example. In the following example by Sylvia Plath, we find that what she chooses to write as a first line is far too warm and cordial in its subtext by comparison with the lethal forms of test that await the person addressed in the poem. Even the lake is dragged for a corpse, presumably, within four lines of the *first* line.

The Manor Garden

The fountains are dry and the roses over.
Incense of death. Your day approaches.
The pears fatten like little buddhas.
A blue mist is dragging the lake.

You move through the era of fishes,
The smug centuries of the pig –
Head, toe and finger
Come clear of the shadow. History

Nourishes these broken flutings,
These crowns of acanthus,
And the crow settles her garments.
You inherit white heather, a bee's wing,

Two suicides, the family wolves,
Hours of blankness. Some hard stars
Already yellow the heavens.
The spider on its own string

Crosses the lake. The worms
Quit their usual habitations.
The small birds converge, converge
With their gifts to a difficult borning.

(Sylvia Plath)

MicroConcord search SW: the *s are * and the *s

80 characters per entry

Sort : 1R/SW unshifted.

1 oft, the atmosphere serene, **the cicadas are sleeping and the hippos** are doing th
2 s are particularly popular; **the names are evocative and the certificates** are att
3 Osborne is very beautiful, **the grounds are undulated and the views** very fine...
4 ertone-bodied GTV coupes of **the 1970s are classics, and the previous Spider,** im
5 ans, or mountain ash trees, **the leaves are crimson and the berries** scarlet. Lond
6 but Bacher is not fretting. **The vines are down and the grapes** will follow, wheth
7 ewhere. Perhaps that is why **the animals are lazy and the whites,** as Lawrence obs
8 ic talent to America, where **the rewards are huge and the public's** appetite appea
9 into the water; cook until **the whites are firm and the yolks** runny. Arrange thr
10 head oatmeal is from Kelso, **the onions are fresh, and the haggis** skins are real,
11 a method of cooking, where **the whites are hard and the yolks** are soft but not r
12 roup> THE SEASIDE is grim, **the towns are tacky and the cities** overpriced. It ha
13 tain times of the year when **the eggs are laid and the turtles** hatch. After the

Plath is definitely a poet who lives her tragedies and dies tragically, rather than one who can invent tragedies from a first line. If we omit the word *over* from our search, we obtain the fairly warm subtext above.

4.3.3 The Fate of Insincerity Twenty Years on

It has been twenty years since the article on irony and insincerity (Louw, 1993). In all of that time, apparently, no further research into insincerity has taken place. Subtext is likely to take this subject further fairly decisively and possibly briskly.

However, what will become immediately apparent, as my co-author and colleague, Marija Milojkovic points out, is that a major new dimension for study will open up in the area of *authorial* insincerity (Louw and Milojkovic, 2014b, in press, John Benjamins: Amsterdam). Marija Milojkovic is in the process of developing from my work a cline that has four descript and easily traced areas: banality, sincerity, inspired sincerity and inspired insincerity (Milojkovic, 2014).

4.4 Intuition: Its Roles, Opacity and Deficit

My reference, earlier in this chapter, to ‘destinations’ of subtextual searches demonstrates that paying further attention to the undesirable influence of *intuition* has now become necessary.

Our best understanding of intuition in its primary pre-computational role, comes from Kant’s writings, most notably from *The Critique of Pure Reason* (1781) and later from

The Critique of Judgement (1790). The former sets out mechanisms for the use of intuition. The latter shows how certain aspects of intuition are abandoned in attempts to work towards the higher forms of feeling that are involved in aesthetic judgement.

In the latter case, the decision is made that tastes are impossible to debate: *de gustibus non est disputandum*. For Kant, the Thesis and Antithesis are:

(1) *Thesis. The judgement of taste is not based upon concepts; for otherwise it would admit of controversy (would be determinable by proofs).*

(2) *Antithesis. The judgement of taste is based on concepts; for otherwise, despite its diversity, we could not quarrel about it (we could not claim for our judgement the necessary assent of others).*

(Kant, translated by Bernard, 1914[1892])

The antithesis feels as though it is in greater need of resolution by means of corpora; the thesis senses the danger alluded to hundreds of years later by Wittgenstein, to the effect that it may not be capable of being expressed and, hence, best passed over in silence (Wittgenstein, 1922:3).

However, considering that intuition was all that Kant had to work from in 1781, we find all of the mechanisms that he provides for cognition are mental rather than empirical. The huge empiricism of corpora has combined with the power of collocation to provide evidence of an established convention of ratiocination that differs from Kant's proposal. It assists us in delimiting our post-computational take on the role of collocation as an instrument that covers all of the misgivings that pre-computational philosophers may have expressed concerning the future of human empiricism as these are mediated through data.

Kul-want and Klimowski set out the process that Kant uses to visualise the role of intuition in *The Critique*.

'Kant is concerned with sensibility or sensible intuition, for which the faculty of imagination is responsible. Sensibility is a passive power for receiving data. Kant is solely concerned with the form this data takes rather than with its matter. Unlike the tradition of metaphysics, he is not concerned with what the data might be and what it represents ...

Hence, Kant's task is twofold: (1) to isolate that which is specific to intuition (2) and paradoxically, to discover a relationship between understanding and that which is specific to intuition.'

(Kul-Want and Klimowski, 2011:54-55)

Kant offers *space* and *time* as ‘two pure forms of sensible intuition.’ These dimensions, the work of Saussure and *diachronic* and *synchronic* aside, are wholly under-utilised within corpus linguistics, except in very few studies, carried out by myself in order to determine the prevalence in the external world of *events* as described by Russell, between 1918 and 1948. I used his five postulates to determine the likelihood that (1) the removal by mercenaries from office as President of Haiti of Jean Bertrand Aristide and (2) the incarceration of Simon Mann in Harare, Zimbabwe, were *one* and the same event. The paper, which includes the fact that the same aeroplane is common to both purported events, still remains unpublished. I included a reference to the matter to be determined for posterity when the Office of Public Records releases details after 70 years (Louw, 2011: 193).

However, we must give Kant credit where it is due: the time and space of events as presented in the node of a concordance are *re-presented* in real time by the spaces between the lines of a concordance, especially as this can also be calculated in terms of the *genre* of newspaper corpora.

4.5 The Intuitive Deficit: Dilution and Qualification

4.5.1 The Intuitive Deficit as Dilution

While *intuitive opacity* is often easy to deny (I once called it 20:20 hindsight) (Louw, 1993: 173), the *intuitive deficit* is determined in the form of the quantum of knowledge that has been lost as a result of the *ideological protectionism* of cognitive approaches to language study. What do we mean by this term? Certain schools of thought cannot but water down meaning. Ideological protectionism takes a number of forms. Its most usual form results from a conscious decision to protect a perilously weak mentalist theory, notwithstanding the fact that its weaknesses make it both obviously and easily falsifiable. For example, bare-faced denial is a frequent and often successful technique for overcoming the threat of falsification. A good example is to be found in the writings

of Gottlob Frege, who, in 1884, declared that much of philosophy is ‘rotten with psychology’.

Frege developed a method for falsifying the cognitive and set it out in step form. Scott-Kakures *et al* set out its principles as follows and in point form:

- ‘1. Always separate sharply the psychological from the logical, the subjective from the objective.
2. Never ask for the meaning of a word in isolation, but only in the context of a sentence.
3. Never lose sight of the distinction between concept and object.

*...The problem with psychologism ... is that it in some way undermines, falsifies or otherwise fails to respect the distinction between what is **objective** and **subjective**. Its failure to respect this distinction is not distinct from its failure, as Frege puts it in other places, to do justice to the centrality of the notion of **truth** in logic.’*

(Scott-Kakures, *et al*, 1993: 375, emphasis added)

Subtext respects all of the criteria that Frege set out in 1884 in *Die Grundlagen der Arithmetik*. In fact, subtext as set out in Louw (2010b: 642) respects all of the above criteria in meticulous detail and *transcends* all of them by allowing no psychologism to enter the picture. If we match our approach point by point in terms of Frege’s three rules above, we obtain a view of how those points are to be transcended:

1. The psychological has no role to play because grammar strings are opaque to intuition; and vocabulary terms are the wild-carded variables of the logic of our own language. There is no contest there.
2. All subtextual searches are grounded within the sentence and potentially in *all* sentences that share its logical form (Monk, 2005).
3. Concepts are banished (Palmer, 1981) because the empiricism is so strong as to make them utterly surplus to requirements.

Furthermore, no *terminology* is used at all unless and until it has been ‘corpus attested’ (Louw, 2008: 243). This places subtext (Louw, 2010b) well out in the clear when it comes to the danger of conflation that Scott-Kakures and his colleagues then set out by way of summarising Frege’s position in the following long, but rewarding paragraph from pages 375-6 of their account of Frege’s work.

‘In the beginning of his late essay “Thoughts”, for example, Frege writes:

*“From the laws of truth there follow prescriptions about asserting, thinking, judging, inferring. And we may very well speak of laws of thought in this way too. But there is at once a danger of confusing different things. People may very well interpret the expression “law of thought” by analogy with “law of nature” and then have in their mind general features of thinking as mental occurrence. A law of thought in this sense would be a **psychological law**. And so they might come to believe that logic deals with process of thinking and with the psychological laws in accordance with which it takes place. That would be **misunderstanding the task of logic**, for **truth** has not been given its proper place.”*

*This passage reveals that at the core of psychologism, for Frege, is the tendency – either explicit or implicit – to identify **logic** with the science of human **thinking**, to view the laws of logic as, at bottom, those laws of nature in accordance with which human thinking takes place. This **conflation**, according to Frege, results from a failure to do justice to the place of truth in logic and, consequently, in an inability to sustain a coherent distinction between the subjective and the objective.’*

(Scott-Kakures et al, 1993:375)(Emphasis added)

All three of Frege’s rules are, of course, breached on a daily basis as the cognitive that he falsified decisively continues to thrive informally.

4.5.2 The Intuitive Deficit Quantified

The least popular form of intuitive deficit is that which is characterised as the calculable loss of knowledge and scholarship as a direct result of the failure to pursue more empirically respectable theories, notably in the face of intuitive opacity, but especially because of the threat that better theories pose for mentalist approaches.

The best way of proving the deficit is to quantify what is *lost* as a result of the use of what is now demonstrably the wrong theory. This method is often regarded as poor

sportsmanship within an ideology to protect all methods as if they were equal; however, once other theories have been falsified, the differences between what has been lost by using poor theories and gained by using better ones, creates a lasting ideological imperative for the rejection of empiricism for its own sake (see Jeffries and McIntyre, 2010:24) on this controversial but unavoidable subject. (See also Losee, 1972; 2005).

I was recently called to order by Vander Viana, editor of the volume *Perspectives on Corpus Linguistics* (2011), for declaring that linguists were not scientific in any objectively verifiable sense. And it is certainly the case in this current research that philosophers are far and away better and more conscientious scientists than we tend to find in linguistics, even at the very top of its hierarchy. In neither Chomsky nor Halliday do we see much respect for computers. When collocation made the move from analogue to digital, Halliday abandoned it almost totally even though it had formed the basis of his notion of *texture* within *cohesion* (Halliday and Hasan, 1976). The philosopher, Bryan Magee has this to say and what he says does not bode well for the reception among linguists of subtext.

*'The creative artist, like the philosopher, is fully committed to a truth-seeking activity, **trying to see below the surface of things** and acquire a deeper understanding of human experience ...'*

(Magee, 1982: 8) (Emphasis added)

Investigation is actively discouraged and is replaced warmly with *choice*: we are told that we may select a theory without any admonition being provided that we may need to study the merits of theories on offer. Of course, most theories are still intuitively derived and corpus-based approaches quickly and conveniently become the needle in the undisturbed haystack. Falsification is today only mentioned with a view to its being brushed aside. Where it is not mentioned, all reference to Karl Popper disappears. Popper's assistants are still loyal, but they have almost completely died out. But note how refreshing the passage below is:

'For the aim of science is simply truth; not some epistemically distinguished variety of truth, but truth alone. Or, more enterprisingly, more truth, as much truth as can be achieved. According to the simplest version of the falsificationist doctrine, a doctrine that is almost entirely due to Popper, that is all that we should try to achieve in science; it is on truth alone that the methods of science should be concentrated. This does not mean, of course, that the methods we adopt need guarantee in any degree the attainment of this goal;

only that they should not demonstrably frustrate it. It is in this sense that methodology should be minimal. As the appropriate method for the search for truth Popper proposes the method of conjectures and refutations.'

(Miller, 1980: 113)

But there is hope at last, because subtext is so impenetrably opaque to intuition that its replicable results *must*, at last, be taken seriously. And the results are often so 'off the wall' from an intuitive perspective that young students will welcome the approach. It will be a breath of fresh air that simply cannot be gainsaid or denied. Jeffries and McIntyre (2010: 24) select neither of these options. They choose to *ignore* subtext and do so in a book that is so keen to celebrate theoretical eclecticism that the omission of any reference to subtext stands out. The gloves are likely to come off when subtext begins to quantify what I am calling the *intuitive deficit*. So, which needs healing then, the CUP 'red book' that signally omits a major theory, or the empirically valid theory of subtext itself? Time will tell. But time can be given a tiny push, as we see when subtext begins to deal with some of the examples of Jeffries and McIntyre. And subtext, after all, lies beneath the surface of all that we read or say; so it is hardly trivial and no longer entirely unknown. The non-trivial fact that cannot be denied is that if two pieces of text share the same logical form, they are inevitably related to one another subtextually. Even Wittgenstein, as noted earlier, missed this point because he never had access to the analytical power of computers.

4.6 The Argument for Collocation as Instrumentation within Subtext

4.6.1 Introduction

What I have termed the *intuitive deficit* refers to those forms of analytic gain that are lost as a result of any decision to remain committed to an intuitively-derived theory. This process is openly encouraged by some of our best scholars in textbooks published by the best academic publishers.

Lesley Jeffries, former Chair of the Poetics and Linguistics Association (PALA), praises the free *eclecticism* involved in selecting a theory for use in researching stylistics. She implies that the truth of the discipline is pluralistic and obtainable from many different sources.

On the matter of falsifiability, McEnery and Hardie (2012:14) have a section entitled 'Total accountability, falsifiability and replicability'. Everything that they write on these subjects they state very dogmatically, but all of it needs to be completely reworked because of the **nature** of subtext. For example, they write:

If you approach a corpus with a specific theory in mind, it can be easy to unintentionally focus on and pull out only the examples from the corpus that support the theory (this is technically called a confirmation bias). But the theory can never be shown to be false by such an approach, even in principle. As such, this approach runs counter to one of the key features of the scientific method identified by Popper ([1934] 2006:18), namely, falsifiability. The principle of total accountability is, simply, that we must not select a favourable subset of the data in this way. When approaching the corpus with a hypothesis, one way of satisfying falsifiability is to use the entire corpus – and all the relevant evidence emerging from the corpus – to test the hypothesis.

(McEnery and Hardie, 2012: 14)

Almost everything that McEnery and Hardie have written, on the subjects of total accountability, falsifiability and replicability, will now need major revision, when applied to subtext. For example, the notion that the corpus is representative of the whole language needs to be completely reconsidered when one is dealing with the *empiricism* of natural language *logic* as opposed to the vocabulary of natural language. In respect of finding a total picture of all the empiricism of grammar strings, even quite a small corpus is likely to account for most of them, even though the vocabulary and subject content of the small corpus might be assumed correctly to be too limited to offer a sample from which a dictionary could be written. Their suggestion that the entire corpus be used is based on an intuitive understanding of the subject matter dealt with by such a corpus. But subtext deals with the totality of argumentation and logic in the corpus as a whole and the scope and scale of this is not even considered by McEnery and Hardie and even if it were, their intuition would be incapable of making any predictions at all on the matter of the logic of natural language in the whole of a corpus. The idea that we select a corpus for its *content* must be reviewed. I am calling this the 'horses for courses' fallacy. For example, if one states an intuitive need and creates a corpus to

meet it. This corpus will be based on the intuitive use of vocabulary as metaphysics. Such reasoning is standard practice in NLP.

For this reason, any decision that McEnery and Hardie make about the possibility and applicability of Popperian falsification simply fails to apply to subtext. Unless they are prepared to consider subtext (both of my 2010 papers were out when they wrote their book; and they even refer to one of them) their book will need a major overhaul in the very near future and as a matter of urgency. And on the matter of major overhauls, the same reasoning applies within all applications of subtext in sub-disciplines, such as stylistics.

It is especially the case within organisations, such as PALA, that insufficient attention has been paid to the relative merits of stylistic theories. That organisation has never recommended that the merits of the fifty or so intuitively-derived theories that are on offer be studied. The notion of untested eclecticism is propounded in Jeffries and McIntyre (2010) and any notion of Popperian falsification is strenuously discouraged. McEnery and Hardie (2012) in a volume on corpus linguistics in the same series also argue as though there are no issues in the domain of theory that are likely to advance corpus linguistics with any suddenness and take it in new directions. Both volumes see scholarship as choice and that always of an entirely mentalist and intuitive kind. The work on subtext (Louw, 2010a; 2010b) is well known to both sets of authors and the incoming Chair of PALA, Professor Paul Simpson, because subtext was first revealed at the The Queen's University of Belfast in a staff research Seminar in January, 2008. Ireland was deliberately selected for this because of the power of the example drawn from W.B. Yeats's poem **'Sailing to Byzantium'** and the fact that all of the works of Yeats effectively became a 'corpus' in 2009 as all of his works came out of copyright.

And so, the imperative for demonstrating the instrumental power of collocation, as well as the intuitive and intellectual deficit involved in making the decision to ignore corpus-derived subtext, will now need urgent attention. There have been other casualties in this process. Jeremy Clear, who used collocation as the main lexicographic computational tool for sense disambiguation (Baker et al, 1993: 271) in both editions of the Cobuild dictionary was the last 'Cobuilder' to leave the project on lexical computing when it

closed. He continues to pursue a career in primary school teaching and his collocation-based online consultancy has closed.

Collocation is a British discovery and no permission is required to pursue this aspect of the science of language. We all witnessed the power of collocation as the computer created the entries for the Cobuild Dictionaries unassisted and entirely on its own, by means of algorithms devised by John Sinclair. Is collocation to be thrown back into the toy box so that scholarship in linguistics can be capped forever at and within the margins of intuitive opacity? The extent of *bias* in our discipline needs now to be quantified. And there is no better yardstick for determining the claims of collocation to instrumentation than collocation *itself*, especially now through the use of subtext. If this is done at the highest level of intuitive opacity this will have the major tactical advantage of keeping the critics silent while we complete the work.

4.6.2 The Method for determining Collocation's Status as Instrumentation

The role of collocation in relexicalisation was established 22 years ago (Louw, 1991). At the time, no scholar, apart from Roman Jakobson (Sebeok, 1960) and his theory of projection as part of what he called the Poetic Function, came anywhere near what collocation was capable of achieving. Oddly, both J. R. Firth and Roman Jakobson attended a conference on Stylistics in the USA in 1958. The proceedings of this conference were finally printed and released in a volume entitled *Style in Language*, edited by Thomas Sebeok. Firth died in December 1960. Six years later a volume entitled *In Memory of J. R. Firth* and edited by Bazell *et al.* appeared.

In this volume, we find a tribute to Firth by Jakobson. Jakobson is warm and reminisces about his visit to Firth's home in the countryside; but we see no mention of *collocation*. We know that Jakobson would have picked up on collocation if it had been mentioned. For it is collocation alone that might have saved the poetic function from the main problem that it faces: how can we have *projection* in a very short text? We can only project at points of *choice* in the *chain*. Syntax governs Jakobson's account; but proximity or collocation allows the poetic function to thrive even within a mere *haiku* (Louw, 2007b). Jakobson's principle reads as follows:

'The poetic function projects the principle of equivalence from the axis of selection into the axis of combination.'

(Jakobson, in Weber, 1996: 17)

If we consider for a moment how close to collocation Jakobson comes, without knowing how to escape from syntax, we begin to sense why it was that he decided to write a tribute to Firth that was based on the life and work of one of Firth's mentors: Henry Sweet.

Sweet, we are told by Jakobson, was Jewish. He had changed his name from German into English and bore the distinction, along with Peirce, of being one of only two distinguished scholars of his day that were *not* awarded Chairs at Oxford University.

The impact of Sweet (the supposed model used by Shaw for Henry Higgins) on the life of Firth was huge, especially when it came to *context of situation* and to *conversation* and the now famous utterance in a pub:

"Ahng gunna gi' wun fer Bert." (I'm going to get one for Bert.)

(Firth, 1957: 182)

But is there still perhaps a hint of the 'curse' of collocation: it may have been handed down from Sweet to Firth, to Sinclair (signally avoided by Halliday after cohesion became computerised by L. John Chapman of the Open University in 1980) and more recently to others?

Well, we sense some dissatisfaction with syntax alone in the language Jakobson uses to describe the poetic function and it is gratifying to see where this leads him:

The selection is produced on the basis of equivalence, similarity and dissimilarity, synonymy and antonymy, while the combination, the build-up of the sequence, is based on contiguity.

(Jakobson, in Seboek, 1960; emphasis added)

Did Firth discover collocation too early or too late? Jakobson offers us a fairly broad hint to the effect that *innovation* in scholarship is signally unwelcome. This hint is offered by Jakobson almost six years after Firth's death. Jakobson seemingly dedicates to the memory of Firth a scholarly article about his own ostracised mentor. The passage

below repays close study and reveals something also of the internal politics of The Philological Society (an organisation that declined to publish contributions by John Sinclair, personal communication). Jakobson tells us that even those who chaired it, and Sweet was one, fared little better for being successful.

As I wrote in my afore-cited paper about Baudouin de Courtenay and his school, 'the proverb says that it is wrong to discover America too late, after Columbus, but also a too early discovery may be detrimental.' The great precursors of the modern science of language – John Hughling Jackson (1835-1911), Charles Sanders Pierce (1839-1914), Henry Sweet (1845-1912), Jan Baudouin de Courtenay (1845-1929), Jost Winteler (1846-1929), Mikolaj Kruszewski (1851-1887), and Ferdinand de Saussure (1857-1913) – each of them in his own way bears a stamp of tragedy on his whole life.

In the final Annual Address of the President, delivered by Henry Sweet at the Anniversary Meeting of the Philological Society, 7 May 1878, under the eloquent heading 'The Future of English Philology' it was stated:

"There is one form of charlatanry to which I will call your attention, and which is specially insidious and dangerous, veiling itself under the guise of conscientiousness and accuracy. It may be termed the mechanical view of language, and is based on the assumption that language ... is not governed by general laws, but consists merely of a mass of disconnected details."

The scholar who dared to look far ahead and to defy the creed of his time was proscribed to become le savant maudit.

(Jakobson, in Bazell *et al.*, 1966: 251)

4.6.3 The special characteristics of collocation as instrumentation

Would not Sir Karl Popper, notwithstanding his great distinction in the field of philosophy, perhaps feel excited by the new digital era of collocation, if he were still alive today?

The vindication of this new approach still contains some surprises

1. The capitulation of other theories is as immediate as it is sudden, especially on the part of those challenged to cognise. It is a clear choice. I recall an advertisement that Geoff Leech used to refer to. It was for a cleaning cloth. The line went: "If it doesn't have holes, it isn't a J-Cloth!" Similarly, "If you cannot cognise the vocabulary variables of these grammar strings, you are not a cognitivist!"

2. Philosophers are better prepared than non-scientists for sudden advances in the realm of science. Russell called himself a logical atomist for all of his life, trusting that the arrival of better empiricism would prove him correct. It has; but sadly, posthumously.

The first surprise that Popper would have been presented with is this (it will be dealt with in the next section): he would not have expected that collocation has a propensity for falsifying theories *unbidden*. Collocation will take control and carry out the falsification. It does this with Schema Theory, as I demonstrated at a keynote lecture in Granada in July 2004 at the TaLC6 Conference. At that event, Popper was vindicated by a metal statue of an antbear (*nhungu* in Shona, one of Zimbabwe's local African languages) from Zimbabwe. As Schema Theory fell, the antbear symbolically popped balloons, in deference to Popper.

3. The third point is *all* surprise. Popper was mysteriously silent on the subject of any *telos* or end-point to the process of serial falsification and the advancement of science. Most philosophers hedge their bets. Russell says memorably that when you think you have found a better theory, you are always very hopeful. You check carefully and in most cases you are disappointed (Russell, 1960: 3). Jeffries and McIntyre quote similar sentiments by Siobhan Chapman. I have placed Chapman's quotation in its full context (Jeffries and McIntyre, 2010: 25) in order to demonstrate its slight edge of hubris.

The result of this eclecticism in stylistics is the explosion of new theories and methods in recent years, drawing from cognitive linguistics, corpus linguistics, critical discourse analysis and functional linguistics. In addition, the practice of text analysis still depends extensively on those methods of language which developed in the formal linguistics of the early twentieth century and gave us detailed methods of description and analysis that are exemplified in Chapter 2.

Stylistics, then, takes on the new, but does not necessarily throw out the old in doing so. As Chapman (2006) says:

'Theorists in all fields generally accept that they are unlikely to come up with the definitive account of their subject matter that will be proved to be correct and will be universally accepted.'

(Chapman 2006: 22)

*One way to think of the question of theory/theories is to realise that we do not abandon theories wholesale **when they are proved wrong in some small way**. We either adapt them or we add another theory to explain those aspects of the world which our original theory didn't cover. By these means, **scientists and***

social scientists add to human understanding by accretion rather than revolution.

(Jeffries and McIntyre, 2010: 25; emphasis added)

At a recent PALA Conference in Malta, in answer to a question I asked, Chapman seemed much more hopeful that subtext could uncover aspects of the Irish political struggle than her statement from 2006, quoted above, suggests.

It would have been a huge surprise for Popper if collocation brought instrumentation to one last major field: **language itself**; and, especially, then only because linguists had succumbed to apparent pressure to keep its potential **hidden**.

We miss Popper sorely, now that we are so close; now that we are forced to travel to Russia to reveal that collocation has found the a priori and that in spite of Kant's prediction that finding it would take a scientific revolution of Copernican proportions (Garvey and Stangroom, 2012: 258).

And, of course, what we have with collocation is a scientific revolution that removes the *autonomy* that we *thought* we had. We have witnessed the end of intuitive decisions being relied upon for both interpretation and for directing corpus research.

4.6.4 Some specific examples of collocation's spontaneous autonomy

Two delexical expressions that are brought close to one another relexicalise spontaneously if and only if they each contain a lexical collocater. An example of this is the opening line of '**Tropic of Capricorn**' by Henry Miller:

'Once you have given up the *ghost*,
everything follows with *dead* certainty.'

I liken the effect to the twitching of a zombie. *Ghost* and *dead* had almost co-selected one another in a process that is easily automated. Proof of the relexicalisation can be provided by co-selection.

MicroConcord search SW: ghost CW: dead

80 characters per entry

Sort : 1R/SW unshifted.

1 through a footbridge. Every new-dead ghost Comes to that worn-out blood for its
2 re had introduced – contact with the ghost of the dead king – was not only the o
3 ed himself to be in contact with the ghost of the dead king, were complications
4 <p> Some say the Big Grey Man is the ghost of a long dead giant. Some Buddhists
5 sts! By using magic he can order the ghost of any dead person to be his servant
6 hrough the woods who was chasing the ghost of his dead father, Bradley neverthel
7 apparition? An earth-bound soul? The ghost of the dead Lady Eleanor? </p> <p> Fa
8 p> Yes, but you're talking about the ghost of a dead person, aren't you? This ma
9 saw that a servant was actually the GHOST of his recently dead neighbour. Resis
10 ected from an old singer in 1906 the ghost of the dead lover returns after a yea
11 <p> The Romans called the spirit or GHOST of a dead person his manes or mares,
12 e soundtrack of psycho prison movie, Ghost Of The Civil Dead. Therapy? then slam
13 been a witch when alive, she'd be a ghost once she was dead. </p> <p> Most adul
14 been a witch when alive, she'd be a ghost once she was dead. </p> <p> Most adul
15 like one returning from the dead, a ghost (the allusion is to the drowning of P
16 and distaste. Surely the dead man's ghost would object to this? Was his soul st

Source: British National Corpus

The concordance is the product of proximity. Collocation had done the rest by remote control in the form of Sinclair's now famous nine word widow. By 2008 I had published a paper in a book edited by Sonia Zyngier as a *Festschrift* for Willie van Peer. It argued for all terminology to be *corpus-attested*. It began as a Cinderella article, but has since coped robustly with many forms of obfuscation that have been directed at both Contextual Prosodic Theory (Louw, 2000) and subtext (Louw, 2010a; 2010b).

The notion of corpus-attested terminology has been useful in setting out the basis of subtext. Oddly, as the science involved in the methodology settled, so did the terminology. It amounts to fewer than five terms: frequency; quasi-propositional variables; wild-carding; logic and metaphysics. Russell (1948) had five postulates, Peano even fewer.

So today, we see a huge disparity of methods: one approach uses a shopping trolley and experimentation and concoction and another approach takes a familiar object and invests in ways of taking a closer look at it; a closer look than our senses and intuition as human beings are capable of providing. Whom do we remember more? Galileo or

Roffeni, the astronomer who relied on the use of the naked eye and ejected Galileo from Bologna University for showing the staff ‘more stars than existed’? (Kitcher, 1993: 231)

Ayer sums up for us something of Galileo’s method.

‘By looking through a microscope I may discover that some material thing has properties that I have not previously detected in it; but I should not express any fact of this kind by saying that I found some of my sense-data to have properties that they did not previously appear to have. For while the material thing remains the same whether or not I make use of a microscope to observe it, the sense-data do not. And the reason for this is that I choose to say, not that the use of the microscope enables people to detect new features of sense-data which they were already apprehending, but that it leads them to apprehend new sense-data. I adopt the convention that reinforces the distinction between sense-data and material things, in preference to one that would encourage us to confuse them.’

(Ayer, 1969: 130-131; Emphasis added)

4.7 Collocation and The Cognitive

The cognitive approach to stylistics purports to be able to cognise expressions and has been doing so for some time. Don Freeman’s work in the area has gone unchallenged for all of the years that he has worked with stylistics (Freeman, 1970: 3). His view is fairly literal; and along with Lakoff and Johnson (2003[1980]) the approach may be summarised in two parts. If we say that we *spend time* rather than *money*, then (1) *time is money* and (2) it is a metaphor we ‘live by’.

As John Sinclair (personal communication) used to say of cognitive argument: “It’s difficult to unpick.” He said this of Sam Whitsitt’s all-intuitive approach to semantic prosody in the only non-computational article ever to appear in IJCL, let alone as the lead article (Whitsitt, 2005:283).

But the matter is easily solved, even by analogue standards. Firth sees the form as ‘one meaning’. It is a fixed expression. If we co-select *time* and *spend** we find very few of the collocates of money, as one expects.

One meaning of time is its collocability with spend.

It is technically not even worth a mention in a book on stylistics in 2010; but when undeniably, a *disproportionate* amount of space is devoted to the cognitive, we need to

respond. Although there may be a good reason to look for a motive, it may be better simply to dispel the bias by means of ratiocination and science instead.

Jeffries and McIntyre (2010) make what at first looks like a plausible claim for eclecticism in stylistics when it comes to selecting a theory, but, if we examine their book more closely, it soon becomes apparent that mentalist approaches are very strongly favoured. Even where reference is made to corpus-based approaches, the text never offers even a single concordance and all of the references offer examples that themselves rely upon *intuition* alone.

(1) *How comforting it is, once or twice a year,
To get together and forget the old times.*

(‘A German Requiem’, Fenton 1980)

*The second line of this extract stands out because of Fenton’s use of the verb ‘forget’, which seems unusual in this context. Prototypically, we might expect ‘remember’ in its place. This is because the phrase ‘old times’ has a positive semantic prosody – that is, it collocates most often with positive concepts. We will discuss the concepts of collocation and **semantic prosody** in Chapter 7, but briefly, **collocation** refers to the fact that particular words frequently occur within close proximity to a restricted set of words, and semantic prosody refers to the meanings that delexicalised (i.e., relatively ‘empty’) words take on as a result of collocational patterns. Since ‘old times’ has a positive semantic prosody, it therefore seems appropriate for people to want to get together and remember rather than forget them. Moreover, ‘forgetting’ is not normally considered to be a conscious act, so it is odd to imagine a situation where a group of people would gather together in order to collectively and actively forget something – on the face of it a logical impossibility. We might describe the verb, technically, as changing its **transitivity** pattern, being used to describe an intentional action instead of an unintentional one, or **supervention**. Transitivity will also be more fully discussed in Chapter 3.’*

(Jeffries and McIntyre, 2010: 5)

Notice how, quite apart from the fact that my work on delexicalisation and semantic prosody is not acknowledged here, collocation is quickly dealt by these authors with in an *intuitive, analogue* definition and no concordance is offered. However, if we were to add concordance material, the glib and convincing text begins to suffer from the effects of too much evidence. In the concordance that follows there is one occurrence of ‘how convenient’. It refers to the communication skills of the Prince of Wales and if the

context is opened we see that the expression ‘how convenient’ is ironic because the writer says that Prince Charles communicates well with older people and wishes that he could do so with young people. Hence, the burden of the irony is not only as is felt intuitively on the term ‘forget’, but is also to be found in the expression ‘how convenient’.

MicroConcord search SW: how *ing it is

80 characters per entry

Sort : 1R/SW unshifted.

1 e myths and legends, you realize how amazing it is what people will do for gain.
2 e's whole business – do not know how testing it is of courage and nerve. </p> <p>
3 When you come to think about it, how amazing it is that a singer whose recording
4 arred by a long speech detailing how annoying it is that the price of eggs has g
5 perhaps fail to understand just how crushing it is to make beds and casseroles
6 d try to live with nothing- find how degrading it is to go to those places and a
7 ere you put us. You have no idea how boring it is to be a computer. Especially f
8 s really, really I can't believe how boring it is yes come on mum talk to me </p>
9 t. Let him assume that she knows how annoying it is and does it to annoy – if yo
10 ited rather than the exploiters. How cheering it is to read the assurance of tha
11 the ice by saying interestingly how confusing it is that the Odeon Mezzanine, t
12 he then proceeds to whinge about how confusing it is that the Odeon Mezzanine, t
13 ed to one of his lords. </p> <p> How amusing it is, he said to him, that an inse
14 furiously. Do you have any idea how embarrassing it is to be caught in flagrant
15 to knitters who share mistakes. How frustrating it is when other people's knitt
16 the season of goodwill. </p> <p> How refreshing it is that this year's number on
17 ncle, Lord Mountbatten. </p> <p> How comforting it is to know that Charles can c

Source: British National Corpus

If we study the concordance material, we realise that a change of attitude is now inevitable. The main reason rather looks to be the combination of intuitive opacity combined with the inevitability, sooner or later, of further and better falsifications of intuitively derived theories and the flawed nature of their products. Subtext is more satisfying than the accumulation of guesswork.

4.8 Developments in the Research of Louw

It was not my intention to provide an overview of my research from 1991 to the present day. However, it has become increasingly apparent that only two of my papers are ever

quoted. It has become almost a convention to mention only my 1993 paper on Semantic Prosody and occasionally my 2000 paper on Contextual Prosodic Theory. It was always the intention of John Sinclair to write the preface for a collection of twelve or so of my papers. The Appendix to this thesis includes fourteen papers, from 1991 to 2014.

My research has progressed through at least five distinct phases with publications (often during periods of sustained censorship) to back them up. A table appeared in Louw (2008: 258-259) to demonstrate the work undertaken in stylistics.

The five phases are as follows:

1. Relexicalisation.
2. Semantic Prosody.
3. Contextual Prosodic Theory.
4. Forensic Applications and Spin.
5. Corpus-Derived Subtext.

4.8.1 Proof of Hypotheses 4 and 6

Of all the hypotheses, these hypotheses are both easily proven and a joy to work with. It is especially satisfying for the reason that both Wittgenstein and Russell were unable, for reasons of intuitive opacity that apply to all of us, to state a subtextual hypothesis without the assistance of a computer. However, Russell has a slight edge over Wittgenstein in this regard. Far and away and of the greatest importance was Russell's mind and his ability to *simplify* matters of huge complexity.

It is worth setting out once more the power of his statement of the requirements for a perfectly logical natural language, so we can note the ease with which it lends itself to the wild-carding hypothesis. And as we read Russell's words, we can picture the scene as David Pears must have dealt with the difficulties of getting a book written in typescript to press. We reflect, as we read Russell's words, that the very machine that was soon to replace the typewriter, was also to be the source of stored empiricism as well as a resident source of collocation software and even corpora, that, although small, were

about to make the recovery of subtext and the inherent power of its logic more fully available for the first time.

There were other ironies also. The computer was still to be termed a ‘word processor’ and there was the problem for me also as a researcher of waiting for a very long time for the crucial penny to drop: we no longer need to begin with the text’s metaphysics; we need to trust the empiricism of the text’s logic even though we cannot see where it will take us.

Once we do this, we can watch as long grey beards of the old theories grow longer and whiter. The distant clatter of falling dominoes grows louder.

One feels tempted to say of the fourth hypothesis that it is a case of *res ipsa loquitur*, and we would be right, but for the purpose of scientific proof, more needs now to be said and written.

The instrumentation behind the wild-carding is so sensitive a ‘finder’ of the variables that eluded philosophers for so long, that it is well worth using an example that very generously provides us with more than we bargained for: it is ‘*Move him into the sun*’, the first line of ‘**Futility**’, Wilfred Owen’s poem from the First World War, mentioned and a concordance provided at the beginning of the confirmation in Chapter 3 of hypotheses 1-3 and hypothesis 5.

I am not going to dwell upon how a traditional or classroom reading might have proceeded. We have more important findings to unpack that will in future see to it that all corpus-assisted reading takes place in this way. The surprises are intensified by the *method* (Greek *metá* + *hodos* an *after-path*). They appear immediately and with the empirical force of a fire-hose.

1. The search for the line’s logical form *finds* the poet’s line even in the newspaper corpus of 44.5 million words of running text. This may *look* like luck, but it proves the empiricism of the grammar strings and the power of inter-textuality.
2. The term *sun* appears only once. The *a priori* experience of the sun takes *all* of us back to our infancy, *except* the dead soldier. The function of the sun is to awaken infants from their sleep. Wittgenstein’s ‘language as use’ principle

verifies that we did not ‘raise this clay’ so that he could be converted by war into a young corpse.

3. It confirms the link with the *a priori* as fully lexicalised and the link between the delexical and the tawdry (See my keynote lecture delivered at the conference on *Primitivismo* in Bologna in 2007, published in Golinelli, 2008: 391). The *spotlight* and the *limelight* are both man-made and have an unnatural colour-cast.
4. The link with semantic prosody is largely opaque to intuition. The prosody is one of sudden and fast-moving fame or fake reclame. We see it in verbs such as *catapulted*.
5. Above all, by proceeding from logic we not only get better focussed argumentation in our criticism, but the semantic prosody becomes *exophoric* and points outside the poem. The corpse of this ‘unknown soldier’ becomes the universal *cause celebre* for all of the *protest* poetry of World War One. Wilfred Owen’s struggle against the ideology of war is *catapulted into the spotlight* of the press and the popular imagination.
6. The issue of intuitive opacity is likely to make an impression in the debate of ‘keyness’ in *genre* studies (Bondi and Scott, 2010). Above all, we may safely aver that the target string in the examples set out earlier in a table, as well as the poem by Wilfred Owen, above, are primarily ‘logically’, as Hypothesis 4 states, part of the same state of affairs as all lines that are linked through their logical form. The unexpected nature of those strings bolsters our faith: we ‘trust the subtext.’

We now move to a consideration of the software hypothesis, Number 6.

Probably the most gratifying aspect of the proofs of the hypotheses relating to subtext are to be found in the manner in which the proofs themselves contain semantic prosodies and bear out the irony/insincerity binarity (Louw, 1993) and as proven in Louw (2000).

4.8.2 Discussion of Hypothesis 6

This hypothesis is divided into *six* subsections and each of these will be dealt with separately below. For the reader's convenience, the whole of Hypothesis 6 will be restated at the beginning of the work on each subsection and that subsection will be highlighted in bold.

4.8.2.1 Hypothesis 6

That, using the *wild-carding* of all vocabulary items, strings that have had vocabulary so wild-carded, will

(1) *reveal from reference corpora quasi-logical arguments that relate directly to the concerns of the states of affairs from which the target string has been drawn;*

(2) *operate at levels of depth that are unable to be determined using intuition* and

(3) are not bound, in their own occurrence, by the genre and provenance of the original, and

(4) may, through any lexical collocates that may be accumulated during the process of searching, reveal both *cultural* norms, as well as insights into the origins of philosophical phenomena, and

(5) may falsify established linguistic classifications (such as *synchronic* and *diachronic*) and

(6) have *primacy* in the sense that they are close to the origins of argument (Louw, in Golinelli, 2008) and that such argument confirms the status as 'pure' metaphysics of the majority of lexical items that were wild-carded in order to retrieve it.

This latter development has already been applied in corpus stylistics (Louw and Milojkovic, 2014a) and in the philosophy of language.

4.8.2.2 Hypothesis 6.1 and 6.2

That, using the *wild-carding* of all vocabulary items, strings that have had vocabulary so wild-carded, will

(1) *reveal* from reference corpora quasi-logical arguments that relate directly to the concerns of the states of affairs from which the target string has been drawn;

(2) *operate* at levels of *depth* that are unable to be determined using *intuition*

A key part of all of these proofs is the fact that the most reliable part of the proofs will depend on the quasi-propositional variables that are *most frequent*. Although variables that are *hapax legomena* will be valid in terms of their logical form, they will be more influenced by *metaphysics* than *logic* in terms of the model here proposed.

For example, in the case of the search line derived from:

That is no *country* for *old men*

the terms to be wild-carded are clearly in a different semantic sphere by comparison with their frequent variables: **reason** and **excuse**. (See Chapter 2.)

But the most frequent variables *are* the *instrument* of quasi-logic and apply to the inferences that may be drawn from *That is no country for old men*.

Syllogistic logic applies. Yeats writes the poem in old age. The country is Byzantium.

The syllogism unfolds in the normal way as follows:

1. *I am an old man*
Ergo
2. *Byzantium is no country for me.*

Major inferences are:-

1. Therefore, I will not go there.
2. Therefore, I will, if I go anywhere, not go to Byzantium.

Reductio ad absurdum:

If I stay home, I will, ironically and paradoxically, be staying in the very place that the persona projects as a sort of Byzantium. Poets receive no golden accolades in Ireland.

Ironically, the poet disregards his own advice. The process of argumentation involved in doing so will cover such rationalisations as:

But that is no reason not to go

or

That is no excuse for not going

or

*That is no reason for staying away
etc.*

(Louw, 2010b)

The process of *inner debate* remains *hidden* and so does the *subtext*, but only until a reference corpus reveals it. The ratiocination involved in dealing with the subtext of the first line, makes the persona's self-contradiction in line 14 of the poem stand out.

‘And so I have sailed the seas
And come to the Holy City of Byzantium.’

Finally, *if* Byzantium is intended to represent Ireland as a place where old men are treated badly, we may confidently expect an eventual summary that alters the first word of the line from *That* to *This*, possibly as part of a political slogan and as *vox populi*.

When this occurs in the common parlance of the politically discontented, the most frequent Quasi-Propositional Variables change: *Reason* and *Excuse* exeunt and *Time* enters in the costume of Impatience.

MicroConcord search SW: this is no * for *

80 characters per entry

Sort : 1R/SW unshifted.

Source The British National Corpus

1 sec that I shock you a little. This is no way for a grandfather to talk, hey? Es
2 for corn circles. </p> <p> But this is no hunt for an extra-terrestrial creation
3 our more troubled communities, this is no time for any of us to strike tradition
4 reduce socialist consciousness this is no basis for automatically equating them:
5 and men who remain in prison. This is no time for complacency. We ask for help
6 convinced neither him nor us, this is no cause for concern. We might try to exp
7 lashed the horse on. </p> <p> This is no hour for conversation. I do not wish t
8 my mother put on a brave face. This is no time for crying, she told me firmly. V
9 y to steer him away from them. This is no time for delicacy and posing, it is ti
10 h, Hubert, McGann said softly. This is no time for divisive talk. We've got a fi
11 ould help and defend the RUC. This is no time for equivocation. </p> Historic d
12 at back on his heels. </p> <p> This is no time for false modesty, Paige. If we d
13 lves quietly and with dignity. This is no occasion for jubilation, certainly not
14 question in a broader context. This is no comfort for a patient who has to wait
15 he work by economic output but this is no argument for lack of careful managemen
16 in some specialities. However, this is no excuse for inaction, for a number of s
17 an seem quite bewildering. Yet this is no reason for ignoring what, taken as a w
18 emporary, says scientists, but this is no reason for complacency about ozone-dep
19 rough its indicators. However, this is no reason for despair since it is by exam
20 enerally increased every year. This is no victory for women, even though it is o
21 ployment, but he warns: All of this is no substitute for a coherent industrial p
22 take you to the top. </p> <p> This is no place for a short Sunday afternoon str
23 inspection at the school, but this is no substitute for attendance at meetings
24 f ocean to turn around in, but this is no argument for abandoning our commitment
25 s do appear in Hollywood film, this is no consolation for feminists. For the wom
26 es to hours, or even days; but this is no reason for abandoning sexual activity.
27 nice Master's little trickses. This is no excuse for Sam, but it shows maybe whe
28 p in worse conditions outside, this is no argument for making them sleep on conc
29 bing everything in a brief but this is no substitute for a profound knowledge an
30 he adventurers have backed off this is no problem for them, if they haven't (e.g
31 ous glance towards the salon. This is no place for a de Courmont, nor for any F
32 , Miss Abbott. Now in you get. This is no night for hanging about in the air. </
33 from a pretty woman. </p> <p> This is no time for jokes! </p> <p> It seems to m
34 yclops and returned to Ithaca. This is no time for modesty. You have achieved mo
35 of 460 feet. </p> <p> Clearly this is no place for novices but it is possible f
36 ussion of her needs and yours. This is no time for pretending that all will auto
37 ave sold at rising prices, but this is no basis for saying prices for that artis
38 gton, Hampshire S041 9WL. </p> THIS IS NO TIME FOR SEDIMENT! </h2> <p> Legionnai
39 m to produce absolute justice. This is no quest for some Utopia. But people do e
40 . God rest us merry Gentlemen, This is no time for sorrow Because ten thousand r
41 o up to my room, Your Majesty. This is no place for state secrets. </p> <p> The
42 always more lurking below, and this is no good for the production of root vegeta
43 be put off by the thought that this is no job for the novice – not everything in
44 . Even were the worst to come, this is no path for us. To put things in men's he
45 comfortable with. </p> <p> But this is no time for Yesterday's Men. </p> <p> It
46 at my teeth rattled – it said, This is no place for you, whoever you may be. <

The variable time down-steps reason and excuse and *where they are present* the nature of the states of affairs in which they appear are now suffused with greater urgency. *Time* has a semantic prosody of urgency. This is a new discovery about semantic prosody that will now need to be written up. Hitherto, we had blithely assumed that a semantic prosody was the metaphysics of its negative or positive collocates. This is only partly true. If we start our search with the grammar string and the vocabulary wild-carded, we find that the semantic prosody of the most frequent variable creates ‘family resemblances’ as Wittgenstein calls related forms (1958:17) that allow the context of situation to seep into the other variables. Wittgenstein says that we ‘crave for generality.’ All of this is very different from the advice of Michael Halliday (in Bazell *et al.*, 1966: 159).

For example, in the case of:

This is no * for *

The first slot is filled most frequently with TIME (see concordance above from the BNC). It creates a semantic prosody of hasty intolerance (**citations 8-12**). The terms REASON and EXCUSE are down-stepped, away from courses of action into *attitudes* that cause delays (see **citations 16-19**). This observation from a concordance satisfies hypothesis 6.2 also. Intuitive opacity is at its greatest in the study of subtext.

One of the reasons why discoveries like subtext come as a greater surprise than might have been the case is always the mind-set, even within science, that is created by intuitive and cognitive opacity and incapacity.

It is an extension of the phenomenon John Sinclair had to contend with in lexicography. The short-cut is always what we *think* we know and we will always move towards where intuition takes us; unless and until we *know* that we are dealing with instrumentation. The earliest example is especially striking. Sinclair had to depart from all lexicographic tradition in the first edition of the Cobuild dictionary (1987) in order to add a brief caveat for the reader that was inserted into the entry-section of the dictionary just before the huge entry appeared for the word TAKE. Here is part of that entry. Note that it has some similarities with the way in which we now need to clear away the ‘mind-set of metaphysics’ before we ‘dial’ LOGIC at an early stage in corpus stylistics.

'The most frequent use of take is in expressions where it does not have a very distinct meaning of its own, but where most of the meaning is in the noun that follows it...'

(Sinclair, 1987:1488)

From our point of view, 25 years later, Sinclair is really saying TAKE often behaves like a grammar word.

One might just as easily offer a caveat in the case of subtext, but notice that such a caveat would today be altogether more far-reaching, because there is no way forward for stylistic investigation *unless* one warns the person who dials LOGIC that they will have no choice but to 'trust the subtext' and that having done so, logic takes up its position *above* metaphysics anyway. It turns into the very short cut that we wanted, but is always *better* than the short cut of intuition because of the inherent heuristic aspect of the method. It is like Frege's logic, a short cut that offers truthful instrumentation on a scale that dwarfs Plato and Aristotle. The material reward for trusting the subtext is always rewarding as truth.

And it turned out that the dictionary sceptics that John Sinclair battled so hard to win over to his new method in 1987, *all* finally *adopted* that method. They did so *without acknowledgement* (for so it is with the naturalness of instrumentation). Subtext is likely to become an even more powerful resource, for the reason that it offers the early choice for selecting the logic of an argument that will dominate the world of the text. It does this to the point that its authors' best efforts to conceal the devices inherent in the circus animals of vocabulary will be uncovered as mere proxies of the logic that Wittgenstein says overpowers metaphysics, trounces its shortcomings and shows them the door. The days of 'sexed-up', dodgy dossiers are numbered. 'Sexing-up' is the thin and tawdry veneer of metaphysics; by starting with data-assisted (Tognini-Bonelli, 2001: 84) logic the veneer is stripped away before it can cause any damage, let alone start a war.

4.8.2.3 Hypothesis 6.3

Hypothesis 6.3

That, using the *wild-carding* of all vocabulary items, strings that have had vocabulary so wild-carded, will (2) *operate* at levels of *depth* that are unable to be determined using *intuition* and **(3) are not bound, in their own occurrence, by the genre and provenance of the original.**

The fact that the subtext of sentences extracted from a particular genre does not necessarily relate to that genre within the quasi-propositional variables disclosed by corpora means that the *depth* issue, even though we still know very little about it, will still have primacy over all matters of metaphysics.

This fact implies that a huge amount of research will need to be done as part of the objectives for future study that arise out of the findings of this thesis.

One fact that will be up for re-investigation is what the context of situation of the deepest subtextual strings will look like. We know that the subsets created by Frege will be well capable of the task, as will the terminology that is free from all mentalism that has been gradually set in place within corpus studies as these bear upon the philosophy of language.

4.8.2.4 Hypothesis 6.4

Hypothesis 6.4

That, using the *wild-carding* of all vocabulary items, strings that have had vocabulary so wild-carded, **(4) may, through any lexical collocates that may be accumulated during the process of searching, reveal both *cultural* norms, as well as insights into the origins of philosophical phenomena,**

This sub-hypothesis is arguably of the greatest importance of all in terms of the way in which disciplines are capable of visualising their own boundaries. We often find that they are discovered or set out as part of the work of mature scholars. Russell's title in 1948 was indicative of the way in which his thinking was beginning to push against existing boundaries. And even though a title like *Human Knowledge: its Scope and Limits* pushed the boundaries of its time as far as it could, we become aware that if

Russell had been granted a view of a single wild-carded grammar string in fulfilment of his own definition of a 'perfectly logical natural language', there would have flowed from him several books of even greater clarity than those that won him the Nobel Prize for Literature in 1950. And certainly, the insights that are gained from concordances of subtext will assist in re-booting a Linguistics in which all of the keynotes have begun to sound the same. The discipline has begun to respond to its laager mentality and draw its wagons into a circle, to defend its mentalism from science.

And, very often a surface meaning that is crafted to look inspirational is subtextually rapacious and war-like. My MBA students were set a subtextual task: find the subtext of your organisation's slogan.

A cellular service provider has the slogan: *just like that*. It suggests that you will be connected with the speed at which you can click your fingers. Far from producing the cultural norms of speedy connectivity suggested by its suasive appearance, the result from the BNC looks in part like this.

MicroConcord search SW: just like that

80 characters per entry

Sort : 1L/SW shifted 6 characters.

5 My father **died** a few years ago, just like that. From being a very vital,
12 Now he is ready to **brush me aside** just like that, which hurts. Frank has
13 People feel that to **brush him aside**, just like that, is unnecessary." The king
15 70 years of brand promotion **away** just like that. They will try to use the
16 shouted. "You mean he **sent you back** just like that? Aren't you a shameless
26 and I can't **spring it on** the Board just like that. They wouldn't listen. I'
29 break. Of course it broke. It **broke**, just like that. It's still broke, if you
31 to stop <F02> It just **ceased** just like that. Yes. <M01> Yes. Mm. Well
36 started **shooting** into the crowds, just like that, women with their
38 The former empire will not **disappear** just like that. It has even more and
39 waterproof tape and they'll **disappear** just like that." <p> The prosecution ou
40 to make your problems **disappear**, just like that. SHE line 0898 200172 <p>
41 able to convey; and of **disappearing**, just like that. At times he loathed being
42 and told us we were **done-- done**. Just like that. She was turning over the
43 kind as he opened his own front door. Just like that. No apparent reason. Then
44 yards. So he just **stops** selling dope, just like that." Horner drank his beer
45 sex with the Prince of her dreams `just like that". Women's bodies do seem t
48 Boxer points out. `And it didn't **end** just like that, things were coming in a c
50 was hoping he would **give up, expire**, just like that. But he was beginning to
51 st like that. <p> He **closed his eyes** Just like that, huh?" <p> It isn't `just
52 **Sudden heart attack** <F04> Mhm. <F01> just like that. <F04> Mm. <F01> Erm aged

54 to do?" Because you're **finished**. Just like that. You're out. All the bills
 56 been taken to **close it down** first. Just like that. There's clearly no point
 57 if we say we're sorry we're **forgiven**, just like that. Help us to forgive <p>
 58 about Justice would be **forgotten**. Just like that. <p> On the other hand,
 61 I wasn't going to **let** that **go** just like that." <p> With the 155p-a-
 62 person. And that all seemed to **go**, just like that -and, like, why? All that
 63 over it. Leo, for God's sake! **Gone**. Just like that. I always pictured the
 64 Enron said. `Here and **gone**. Just like that." Very slippery people."

Source: The Bank of English

The students spent some time in the complaints queue at their place of work, only to find that the slogan was used by clients to describe the *failure* of their computers. Intuitive opacity is a poor accessory in the world of branding.

The last part of this subsection, 6.6, relates to the link between philosophical phenomena and the corpus. We have already observed something of this in the first line of Wilfred Owen's poem 'Futility' (Move him into the *sun*), where the sun represents the *a priori* and the delexical aspects are the man-made light of *spotlight* and *limelight*. More detailed research along these lines is to be found in Italian (Louw, in Golinelli, 2008: 407). What we witness here is that we are so thoroughgoingly subjected to the reprehensible conduct of human beings in such a myriad variety of ways, that our consciousness of when we first experienced it is dulled as we gradually become inured to it. Each outrageous event may be collocated with *being + what + it + is* **or** we can subsume what occurs in the subset collocate *human nature*.

MicroConcord search SW: **being what it is**

80 characters per entry

Sort : 1L/SW shifted 2 characters.

1 1 the image of female aggression being what it is. Those howls of feminist
 2 2 r a new partner, what with AIDS being what it is and the chief buyer at W
 3 3 of irony. The Oxford Apollo being what it is a theatre the sense of It fa
 4 4 so sure but with arachnophobia being what it is, somehow I doubt we'll alum
 5 view The force of that attraction being what it is, out of all the degrees of ve
 6 5 Mm. <M02> and the school budget being what it is erm does not at the
 7 6 but <tc text=pause> bureaucracy being what it is # <M01> Yeah. <M02> I <M01> M
 8 7 because, the pop-music business being what it is, and me being as old as prim
 9 8 stralian schoolyard camaraderie being what it is, those taking part were is
 10 9 s stove - although the catering being what it is, it's not a bad idea to

11 is concerned, his known character being what it is. As all the duties which we d
12 0 always so. Female conditioning being what it is, I grew up assuming that
13 1 c of such dreamy confrontations being what it is, or rather isn't - that'
14 2 a mother. <p> And birth control being what it is, that means many of the <p> A
15 3 e the official rate. Corruption being what it is, the foreigner has no
16 4 ning. <p> Limited-overs cricket being what it is, one cannot yet be sure
17 5 ils to wash but celebrity crime being what it is, he will probably take exac
18 6 The Bank's closely-knit culture being what it is, that claim will be the end
19 7 WITH the cost of pellet cutters being what it is, a great saving may be th
20 8 up is easy and, army discipline being what it is, they are unlikely to confide
21 9 g stakes. Sexual egalitarianism being what it is, this state of affairs all e
22 0 y and in the USA. But, Euro-law being what it is, such live recording are is
23 1 end of the conflict, expediency being what it is, they were putting them
24 ore editions And their experience being what it is, they of course look with gre
25 The nature of the muscular fibre being what it is, the purposes of the animal c
26 rection of the ventricular fibres being what it is, the apex of the ventricle mu
27 2 hing started to take hold, film being what it is. It happened around group. W
28 3 t for some time." His following being what it is, there were probably des
29 4 e you from attacking a play for being what it is, in fact, supposed to t bu
30 5 will never forgive the PO for being what it is - a world class public our p
31 6 down the lane. <p> Fox-pressure being what it is, we were dismayed when like
32 7 ut vacuous) popsies. Friendship being what it is, Sir David had casually
33 8 sloth and two because, Glasgow being what it is, the place is pulsing </h> W
34 9 he Test surface. But Headingley being what it is, despite relaid pitches
35 CFoQmBYwBg" More editions Heaven being what it is, and mankind in a state of na
36 0 bottled water. <p> Hollywood being what it is, a few celebrity off down t
37 nts the venerable host, that host being what it is ? I fear not the displeasure
38 1 u wait on my son # <p> Humanity being what it is, the little tyke will country
39 2 ewer than Cantonese). Hong Kong being what it is, dozens of schools have r
40 3 and let's face it, Sod's Law being what it is, there will be nobody an i
41 4 olid on the road. <p> Sod's Law being what it is, it rained, or bucketed
42 5 ential"; but the art of leaking being what it is in Washington, it might has
43 6 o you to do. Live in life. Life being what it is, you'll encounter standards
44 7 <F01> Right # <M01> and life being what it is nobody's going to fork on th
45 8 nts to hear again </h> <p> Life being what it is, we aren't perfect, said t
46 nature of the mus- i cul.tr lihrc being what it is. the ... HYPERLINK "http://b
47 9 inspired move and, Tapie's luck being what it is, Alen Boksic came good of
48 0 aid: With the automative market being what it is at the moment we are
49 1 or not. And with the mass media being what it is today, no one can say, ` Am
50 2 esirable or not. And the media, being what it is, will try to draw a <p> Cadbu
51 till, however, the average of men being what it is, and it being neither possibl
52 1 view The record of our ministry being what it is in this respect, for the sake
53 itima, I have given it that name, being what it is usually called when in its ..
54 8 human # and political # nature being what it is, this is an unlikely to be a
55 9 o genuine loggers, human nature being what it is, there is always someone B
56 0 act is small. <p> Human nature being what it is, drivers will continue t
57 6 eir own # However, human nature being what it is, and male and female judgment
58 7 eir criticism. <p> Human nature being what it is,we should expect that a clas
59 5 st few years that, human nature being what it is, a backlash was and
60 2 ailing to do that, human nature being what it is, staffing rural day short-liv
61 3 76 on forward. And human nature being what it is, people will tend to to 49, b

62 4 artners in Europe. Human nature being what it is, I'm glad I'm not a 1976
63 5 cancer and burns. Human nature being what it is, will these warnings remaini
64 zgK" More editions -Human nature being what it is, it works in varying ways. 10
65 1 dbury, Greenbury # human nature being what it is, most of us prefer to eye. T
66 ted effect of which, human nature being what it is, cannot be otherwise than pow
67 2 tion, respect and, human nature being what it is, probably a touch of be
68 1 he said. <p> And human nature being what it is, maybe he will be caught Mind
69 0 ou, my perverse Scottish nature being what it is, the chances of my can he de
70 9 e spokesman said: `Human nature being what it is, they want to keep what litt
71 8 but 50 and above. Human nature being what it is, if we don't get this week.
72 3 tion, respect and, human nature being what it is, probably a touch of group.
73 6 towards it. Human nature being what it is, this more profound and fa
74 5 ecause otherwise, `human nature being what it is", we would have total with at
75 4 eeple. `Regrettably, owl nature being what it is, some do conform to the
76 3 h us. I felt that, human nature being what it is, they would be an to l
77 7 to live. However, human nature being what it is, we do not always live b
78 4 an keeping it. <p> Human nature being what it is, there must also be s tru
79 5 majority of MPs - human nature being what it is - say they find the them up
80 6 uy such arguments. Human nature being what it is, people don't always A m
81 7 it is, counselor, human nature being what it is, Dom started expanding m
82 8 V, but, of course, human nature being what it is, the diaries tend to get t
83 4 ards or otherwise, human nature being what it is, the whole structure t - t
84 0 e too easy on her. Human nature being what it is, I fell for her feminine
85 1 2> And er you know human nature being what it is if <tc text=pause> you
86 2 ight. <ZGY> <M02> Human nature being what it is. <M01> Yes. Yeah. Erm <M02
87 3 true but I suppose human nature being what it is I hope that er <ZF1> I s r
88 9 side? <p> However, human nature being what it is, you are obliged to open
89 6 week. And with the week's news being what it is, it seems especially th
90 7 tly that. But human originality being what it is (ie, non-existent for 98
91 8 ical boldness, the Labour Party being what it is; but ditching a view got con
92 9 erating to women but patriarchy being what it is, cultural practices grew sk
93 0 photographs of erect penises being what it is, Pauline has to vet each
94 1 l development of the guinea pig being what it is, one of the males was m not
95 2 op with the state. <p> Politics being what it is, the Queensland ALP and rock
96 3 nt here. <p> Education politics being what it is, no one can be wholly an auto
97 4 te and expect a reply, the post being what it is nowadays, so I sent a double
98 5 y appointing him as such. Pride being what it is, they inevitably will on th
99 6 animals." <p> The Turner Prize being what it is, Hirst's winning exhibit
100 7 that human nature is a problem. Being what it is, you're always going to
101 8 er inland, beach-front property being what it is. The dialect of Southern s
102 9 obiography although, publishing being what it is, it's only a matter of is lib
103 and situation of the royal pupil being what it is, any infirmity in his mind, e
104 0 be (pronounced, the Brazilian r being what it is, `hobby # Any short politica
105 1 ing what they are. British Rail being what it is, nobody has deemed it killing
106 2 e deliver? <p> His track record being what it is, and IRA divisions being Pal
107 ' ought to say, ' 'the recruiting being what it is, notwithstanding the operatio
108 3 t Hellman did. Public relations being what it is, though, she is still how
109 5 . With the special relationship being what it is, this very-high-level admirat
110 6 boats <p> Human resourcefulness being what it is, the survivors rebuilt way o
111 7 yself of. But with the schedule being what it is, the endurance, the further
112 8 ssification # computer security being what it is # could easily and in new

113 9 or is that self-delusion?) being what it is, teams from these shores in t
 114 0 ssador, but the Foreign Service being what it is, I may never be to write and
 115 1 learned. The Sunday bus service being what it is, I decided I'd walk at wit
 116 2 go about it now, the situation being what it is and her mind certainly
 117 re editions The temper of society being what it is, some small gang of cotton-de
 118 3 s out the door. Thespian spirit being what it is, I held out a glimmer of pea
 119 4 njoy `parity of esteem". Status being what it is, it took no time for logic
 120 ISEMENT. added, that, the subject being what it is, O 6 ... HYPERLINK "http://b
 121 5 lived - the power of suggestion being what it is, I fell each time within wasn
 122 PPENDIX A MicroConcord search SW: being what it is 80 characters per entry Sort
 123 6 onfused. China's justice system being what it is, such a mistake can take th
 124 7 a worry. THE CITY'S water table being what it is, the dead are buried in thei
 125 8 f spectacle. The rock gig thing being what it is, everyone resolutely the last
 126 9 newspaper files, the road toll being what it is, the deaths are so develop w
 127 0 nionable at pre-sent and, trade being what it is, I can't afford to put
 128 1 if you ask me, civil unrest being what it is. Is that all?" <p> Steepl
 129 2 ful means - the world around us being what it is - then we will do it want
 130 3 surprise, the publishing world being what it is, that he had a name for sexu

Source: The Bank of English

This example is only one of many that may now be seen to falsify Ferdinand de Saussure's binary distinction: *synchronic* and *diachronic*. The form 'human nature' is preserved from diachronic semantic change through and because of its *proximity*, as a *collocate*, to the quasi-logical form *being + what + it + is*. Quasi-logic gives its most frequent collocates immunity from change. The fact that the logic of natural language does not change and prevents change in the case of its collocates provides sufficient cause for Saussure's distinction to be reviewed. The term *being + what + it + is* acts as the creator of a proxy-subclass collocate: *human nature*, inside which the outrageous metaphysics of human behaviour is smoothed. Hence, we all wonder when we first noticed the similarities between different outrageous acts. We are so accustomed to them and their predictable recurrence that they become for us a form of our *a priori*.

4.9 Conclusion

Hypotheses 4 and 6 of this research now stand proven.

Hypothesis 6 is especially valuable because its sections will, between them, assist in creating the method and path for future research. There are several reasons for this.

The first reason inevitably involves the fact that **argument** is **dialogic**. The most frequent quasi-propositional variables (QPVs) will interact dialogically with the logical form of the target line. Initially, this may be carried out manually, but as this dialogic aspect of subtext becomes settled, algorithms may be generated computationally, in order to carry out the dialogic task, and reach a conclusion. In philosophy, **counterfactual** behaviour such as saying: "That is no country for old men.", and then actually going there, and spending time there, is likely to trigger dissenting warrants.

CHAPTER 5: APPLICATIONS OF COLLOCATION: CONTEXT OF SITUATION TO CORPUS-DERIVED SUBTEXT

5.0 Introduction

The notion of using a reference corpus in order to read a single line of text was introduced in 1987 when the reference corpus upon which the first Cobuild Dictionary reached 22 million words of running text, including what were known at the time as Reserve Corpora (Sinclair, 1987). The first occasion on which this method was used is written up in Louw (1989), although the session at which it took place occurred in April, 1987 at St Hilda's College, Oxford. At this time, the corpus itself was only accessible, even to lexicographers, in the form of microfiches. This event marked the beginning of corpus stylistics. Before this date, some 'corpora' of literary works were available to users of mainframe computers, among these were the Brown and LOB Corpora of 1 million words each. Louw (1993) began its life on microfiche, and was improved before publication, with eventual access running at 90 million words on several different machines.

Concordances were printed from microfiche and, where more sophisticated sorting of the output was required, an instruction in UNIX needed to be given to the machines by a technician.

5.1 The mind-set behind early searches

Because much of the concordance materials had been captured on microfiche there was a default sorted output of one word to the right of the node.

This was enough to get semantic prosody started because a line of poetry, such as line 2 of Philip Larkin's poem 'Days', was immediately useful to theory:

'Days are where we live ...'

The product of that search was photographed on the screen of a microfiche reader and re-typed. The results were both amazing and addictive (Louw, 1993: 162).

Where we live ... was in a state of tension with done, gone, over and past.

In 1993, the long cherished goal of finding literary deviance was suddenly available on microfiche. Personal computers were available, but the hard drives that they were supplied with could not hold even a fraction of a corpus. The prototype XT computer had only a 20mb hard drive. These early studies of Larkin have been much improved by Marija Milojkovic (2011: 127).

However, the goal that Donald Freeman (1970) had declared impossible in 1970 was already available in that year in the first version of the OSTI Report, albeit still only in mimeo (Sinclair, et al, 1970; Krishnamurthy, 2004). By 1993, Freeman's reservations were overturned and corpora were not only well established, but capable of dealing with *irony* and *insincerity*. But the most disappointing aspect, with the benefit of hindsight, was the fact that all of us were completely unprepared for the scientific aspect of the forms we were working with. This was, as John Sinclair was to say, to be another penny that took almost twenty years to drop. Instead of noting the role of time in what we were seeing and knowing that we needed to refer to matters such as time or predictability or causality in the work of the philosophers of language, we simply soldiered on, looking, initially at any rate, for *direct matches* between the search line and the target line. We did this routinely and without ever realising fully that we were finding variables that no philosopher had hitherto been able to see, contextualise or interpret.

So, to summarise the concerns of this chapter, beyond these early beginnings, what is to be dealt with? The following ten phenomena will be exemplified. They will be presented as exemplars of the hypotheses of this thesis. However, reference to the hypotheses will be made briefly only in the summary table at the end of this chapter.

1. Delexicalisation and Relexicalisation;
2. Semantic Prosody;
3. Contextual Prosodic Theory [CPT];
4. Corpus-Derived Subtext;
5. Subtext as a Practical Method;
6. Collocation finds the *a priori*;
7. The Logical Fallacy Involved in Text Mining;
8. Aspects of Political Spin;
9. Negotiating;
10. Branding and Advertising.

5.2 Delexicalisation and Relexicalisation

One of the main discoveries of the Cobuild project in lexical computing was *delexicalisation*. A good example of this was mentioned earlier. The Concise Oxford Dictionary has about six entries for the form *take*; the first and second editions of the Cobuild dictionaries run to almost 80 entries. This large number of entries was found by the computer, using collocation software written by Jeremy Clear (1993: 271) according to what he termed ‘Firth Principles’, such as

‘One meaning of *take* is its collocability with *The Guardian*’ (meaning *subscribe to*)

John Sinclair (personal communication) described the process of sense formation by delexicalisation as being like blue jeans. The more we wash blue jeans, the more they fade. The more we use a word like ‘take’, the more its meaning is washed out progressively into new senses. These senses will be recognised by computer and the computer will write a new dictionary entry for each new sense.

The concerns of this thesis work also with the converse of this principle. *Devices*, from weak to sophisticated humour will put the *blue* back into the *jeans*, as will all literary devices and others often used in advertising.

This re-blue-ing or relexicalisation is very common and recognition of it begins at an early age, as some of these ‘really silly jokes’ for the entry *zombies* will attest (Tibballs, 2011: 494-5).

“What happened when a vicar saw a zombie with nothing on his neck?”

“He made a bolt for it.”

“What did the zombie’s friend say when he introduced him to his girlfriend?”

“Good grief! Where did you dig her up from?”

“Why did the zombie go to hospital?”

“He wanted to learn a few sick jokes.”

“What did the depressed zombie say to his friend?”

“Mind if I pick your brain for a bit?”

“Why do zombies learn Latin and Ancient Greek?”

“Because they like dead languages.”

Note how in these examples relexicalisation takes place even where syntactic links are dubious. This vindicates Firth’s assertion that collocation is abstracted at the level of syntax and refutes Halliday’s claim only six years after Firth’s death to the effect that:

... lexis seems to require the recognition merely of linear co-occurrence together with some measure of significant proximity, either a scale or at least a cut-off point. It is this syntagmatic (sic.) relation which is referred to as ‘collocation’. The implication that degree of proximity is here the only variable does not of course imply how this is to be measured; moreover it relates only to statements internal to the lexical level...

(Halliday, in Bazell *et al.*, 1966: 152)

No mention is made the more Firthian requirements of context of situation or collocation’s ‘abstractness’ from syntax. In the above quotation Halliday creatively raises the question of the instantiation, durability and extent of the collocational span. Rather than say (as Firth does) that collocation operates at a level abstracted from mere

syntax, Halliday immediately reins collocation back into grammar and syntax, notwithstanding the inability of those categories to explain what Firth had already asserted. Yet another penny got stuck in the slot mechanism as a result of this move back into syntax.

In fact, my hypothesis that *all devices relexicalise* has turned out to be more robust than I thought it would be when I first offered it (Louw, 2008: 243).

Of course, the *quality* of the humour marches in step with the quality of the *empiricism* of the corpus and, in the case of humour, even cartoons are part of that empiricism. It can be demonstrated with forms of precision that Carnap (1928) had no access to, that it is possible to predict, in spite of all claims that determinism is impossible (Berovsky, 1971), jokes that are ‘waiting to happen’, as discussed in Louw (2009). Knowledge of such simultaneity is no longer impossible, and was prepared for by Russell (1948).

The cartoonist, Matt, of *The Telegraph*, is arguably the best of them. He observes society and recognises a humorous similarity between two *contexts of situation*. Matt realised in 2008 that once the salary increase for the police had been ‘viremented’ to the UK military in order to pay for the war in Iraq, the police were likely to be reduced to *begging* on the streets rather than *patrolling* them.

Now, as we see from the first three general hypotheses for this thesis, the claim is effectively made that *situational context* is itself a form of instrumentation. An elaboration of the 1993 definition of semantic prosody as an ‘aura’ is expanded in order to accommodate a definition of semantic prosody that turns upon any *fracture* of a context of situation.

5.3 Semantic Prosody

The section that follows this one discusses Contextual Prosodic Theory [CPT]. It is important that Semantic Prosody (1993) and CPT (2000) and their proofs be read together. The reason for this is that CPT was intended to amplify Semantic Prosody. This it did in that CPT produced proof of the binarity principle for *irony* and *insincerity*.

However, during this process it became plain that CPT provided a new dimension for the primal semantic prosodies such as those first noticed by Sinclair, such as *happen*.

The example which follows demonstrates this dimension as a *fractured* context of situation. This places semantic prosody at the heart of the structure of culture and society. Halliday's notion of grammatical metaphor is unsatisfying because it mistakes personification for fractured contexts of situation. Halliday mentions Malinowski very little in his work, but completely overlooks the extent to which fractured contexts of situation are instantiated in the very fabric of the language that depicts fractures.

If Halliday had read Russell's doctrine of forced acquaintance, he may have realised that fractured contexts of situation must show some emotional distress within the way that language copes with them.

5.4 Contextual Prosodic Theory [CPT]

The next example involves the use by the Welsh poet Edward Thomas of a fractured context of situation in order to provide an image of God's ubiquity: an ever present absence or an *ever apparently* absent presence. It is the logical tautology of the Greek *ego eimi* or of the Old Testament "I am that I am."

Here is the poem, 'Raptor'. The image is of a giant owl, ... *abroad in the shadows*...ever vigilant – sometimes punitive: to deliver us from evil denizens or to punish us for flagrant disobedience or disregard. The poem opens with the fashionable rationalisation that God can be reduced, for purposes of definition, as something made effable by science. The poem is altogether less sinister than some of its corpus-attested lines. The reason for this is that *we* cannot hide from God.

Raptor

You have made God small,
setting him astride
a pipette or a retort,
studying the bubbles,
absorbed in an experiment

that will come to nothing.
I think of him rather
as an enormous owl
abroad in the shadows,
brushing me sometimes
with his wing, so the blood
in my veins freezes, able

to find his way from one
soul to another because
he can see in the dark.
I have heard him crooning
to himself, so that almost
I could believe in angels,

those feathered overtones
in love's rafters. I have heard
him scream, too, fastening
his talons in his great
adversary, on in some lesser
denizen, maybe, like you or me.

(R.S. Thomas)

Now we are accustomed to the often uncontroversial nature of delexicalisation. A colleague who is on leave may be *abroad* in a specific place; often as the result of a long-cherished ambition: such absences might easily attract the form of words: "He's on leave: abroad in the Andes and the Straits of Magellan..."

However, if we devise a method for regarding *absence* as a form of *presence*, it turns into *paradox* (not far from the main hypothesis in 1993 of *irony* versus *insincerity*). Holidays are, of course, a logical explanation for a temporary absence of a 'person, personality ...' in Firth's sense (Firth, 1957: 182).

Nonetheless, computational methods, including subtext allow us to test what is missing or has been added to a context of situation (Louw, 2000); and this practice or ability sums up the central meaning of Contextual Prosodic Theory [CPT]. It overflows normal contexts of situation and begins to define worlds, both possible and actual, without recourse to the *putative*. The unreliability of the putative in discovering – *that 'p'* with any accuracy, relies, as we have seen, in this research, upon the certainty that reference corpora, instance by instance, will ‘show it the door’. This has been the case in all instances of political spin that are examined later in this chapter.

The two parameters of *under-provision* and *over-provision* were, in Louw (2000), still insufficiently supported by *method* at the time of writing, but robust methods are now in place. We abandon searches based upon intuition and replace them with those based upon *intuitive opacity* alone.

This may be done category by category within the taxonomy set out by Firth and Malinowski in the table in Firth (1957: 182). The ‘interrogation’ is at last available to philosophy in the form of Dummettian *warrants*, e.g.:

Question: Is Yeats referring to a *ladder*?

Answer: No. The subtext says the ladder is a logical proxy for a *person*.

So, the *ladder* is *under-provision* in general terms and in terms of pathos in *this* instance in the poem ‘The circus animals’ desertion’. It is a tragic and abiding personal loss. Maud Gonne subsists but does not do so where she *belongs*. In Wittgenstein’s terms, the device is simply stated:

What was stated as a symbolic thing, has the status of a cancelled fact.

In terms of CPT the form *ladder* is doubly under-provided. Maud is *absent* as an *object*, *ladder*, because of authorial misrepresentation as a device; she is absent as a ‘person, personality...’ because the logic of the reference corpus says so undeniably and now, at last, forensically. All that remains to be considered is the result of the poem as verbal action: it is made predictably circular by his *heart* to the irrational dictates of which he must return.

Now, if we apply CPT to the poem 'Raptor', we find the dilemma of St Paul: belief with unbelief needing help. After processing large numbers of *delexical* and *rational* absences, we are left with a key 'person or personality' being absent in specific terms but present in faith, word and deed. The aura of the semantic prosody is to be found in the fact that the 'space' of apparent absence is shared by both good and evil. We see something of that in the concordance, in which the *denizens* referred to in the last line of the poem appear, at least numerically, more prominent. Their being *at large* and of no fixed abode seals their classification.

As we study the concordance lines for *abroad in the* from The Bank of English, we realise that the semantic prosody involved in *abroad in the shadows* definitely subsists within the corpus. It is easy to extricate it from the total of 92 lines that refer to destinations with proper names. It is not necessary to refer to them, but we can simply keep the line numbers of those lines that qualify for being mentioned as part of the semantic prosody. Note how collocates such as *evil, power, scourge, Angel of Death*, etc. begin to proliferate at the very moment that we notice that no descript *destination* is visible to the right of the node. Surprise and suspense characterise them as they breach the Firth-Malinowski taxonomy for context of situation and demonstrate through their *absence* that Contextual Prosodic Theory (Louw, 2000) operates by means of *fractured* contexts of situation.

MicroConcord search SW: **abroad in the ***

80 characters per entry

Sort: 1R/SW shifted 3 characters.

```
7 the Serbs are not the only killers abroad in the Balkans. <p> A United
19 s lamp to me from this out, and I abroad in the darkness, spearing salmons
27 evening. I am not only abroad, I am abroad in the <f> evening! <f> I have
42 Victoria that `a kind of madness is abroad in the House of Windsor." <p> A
45 the butterfly fairies, who were abroad in the land at this time of year,
47 Bright, evoking the Angel of Death abroad in the land during the Crimean war,
48 Americans that a deadly scourge was abroad in the land - rule by a financial
50 today?" captured the bitter animus abroad in the land. Dwight felt this same
51 be enough. But something else is abroad in the land: fear. For months we
59 <f> talked about strange things abroad in the modern world. It transpired
61 in his way, and when he goeth abroad in the night on freebooting # i.e.,
62 Here were the deadly creatures, abroad in the night, hunting for her. So
63 bit down on it, hard. Only silence abroad in the night, only silence and <84>
```

72 around, but there was no one else abroad in the place. No person, dog, cat,
83 f a surprise to those not generally abroad in the unkind hours, or who venture
88 are convinced that evil is abroad in the world, they do not believe
89 So much occult knowledge is abroad in the world, so much of the kind
90 power, which I now realized was abroad in the world, I cannot tell what a

Source: The Bank of English

One certainly feels, as one studies these contexts, that Firth must have felt something of its power when he made the first ever collocational selection between *dark* and *night*.

5.4.1 Co-selection as a Contextualising Tool

One of the best examples of the surprise nature of this phenomenon is to be found in the poetry of Yeats. It is easy to accept the symbolic value of the swans in the poem, ‘The Wild Swans at Coole Park’.

The persona establishes in the mind of the reader that he last visited this beautiful sanctuary for swans when he was a much younger man. But the fact that both then and on the current visit that the poem commemorates, the persona not only sees himself *as* a swan but as the intimate lover of one of the swans, which induces in the reader a sense of shock.

He leads up this revelation in fairly subtle ways and, even after it has been stated, he does not labour the matter. In fact, many readers gain their insights into the sexual connection only after many readings and often only because the commentaries on the poem refer to it half-jokingly. But for Yeats, the idea is a serious one. Swans mate for life and are prodigiously active sexually for their entire lives. The fact that he counts the swans and finds 59 of them is our best indication that there must be among their number a vacancy for a mate and lover. It is a point that is made conclusively by the collocates as a contextualising tool.

“... *The persona makes the risky assumption that this swan will be female...*

*...In Chaucer’s terms, we are told that lovers “bathen in gladnesse”. The relexicalisation we are watching out for can only occur in the context of **birds**, as this fragment of the entry for **tread** (meaning number 5 below) from the Collins English dictionary proves. But Yeats adds meaning 6 below in order to retain his detachment as a voyeur.*

tread (tr, d) vb. treads, treading, trod, trodden or trod. 1. to walk or trample in, on, over, or across (something). 2. (when intr., foll. by on) to crush or squash by or as if by treading: to tread grapes; to tread on a spider. 3. (intr.; sometimes foll. by on) to subdue or repress, as by doing injury (to): to tread on one's inferiors. 4. (tr.) to do by walking or dancing: to tread a measure. **(5) (tr.) (of a male bird) to copulate with (a female bird).** **6. tread lightly, to proceed with delicacy or tact.** 7. tread on (someone) toes, to offend or insult (someone), esp. by infringing on his sphere of action, etc. 8. tread water, to stay afloat in an upright position by moving the legs in a walking motion. (Collins English Dictionary)

*The relexicalisation of **tread** and **trod** is **not** a manner of walking (sense 1) but of copulation that, for swans at least, implies no institutional censure or public stigma: “passion or conquest, wander where they will”. Popular culture claims that swans mate for life, but this particular persona appears to want to keep his options open. The swans are as free as the air they **mount** (also relexicalised as sexuality). Their sexual freedom refreshes them, leaving them unlike the persona, “unwearied still” after an absence of 19 years. The persona wants the convenience and luxury of the likes of Chauntecleer:*

Real he was, he was namoore aferd.
 He fethered Pertelote twenty tyme,
 And trad hir eke as ofte, er it was pryeme.

(The Canterbury Tales, The Nun's Priest's Tale, lines 3176-3178)

(Louw, 2010b: 650)

5.5 Corpus-Derived Subtext and the Use of its Products in Literary Appreciation

One condition for the use of subtext in practical terms must always be that the two lines that share logical form ‘converse’ with one another. This is a Firthian requirement: that we shall only understand language once we understand conversation. The fact that the *static* corpse in a war poem can become the *dynamic cause célèbre* to end all wars, will always be dialogic. My first examples bear this out (Louw, 2010b: 642). Once we see that the persona has brushed aside this own advice, chosen paganism and travelled to Byzantium, *conversation* sets it all in motion. If the meaning of a proposition is the *method* of its verification, then the dialogic function of the most frequent variables (*reason* and *excuse*) will need to *interrogate*, conversationally, why such irrational

conduct has taken place. Subtext is a huge missing dimension within critical discourse analysis.

5.6 Subtext as a Practical Method

In my conversations about subtext with philosophers, they often express concern as to the extent to which the procedure which splits the Vienna Circle also falsifies it. My answer is that the Circle's members would have applauded the *method* if and only if computers had been available to demonstrate the heuristic potential of that method. Omelettes can only be made by breaking eggs: the empiricism that the convergence of logical form brings about can only come from that source. The philosophers were all unhappy with the solution I proposed (as a test) as follows:

1. Split logic from metaphysics using wild-cards and a reference corpus;
2. Interrogate the target line using the empiricism obtained in step one;
3. Then, purport to re-unite logic and metaphysics and claim ignorance of the source of your own questions.

It was plain from these discussions with philosophers that they regard the methodological gains involved in splitting the atom of the Vienna Circle as too good to pass up. They felt (1) that the opportunity costs involved in wandering in the wilderness of metaphysics for 4,000 years had been too high; and (2) that the procedure always involved the identification of *non-p* and not only asserted but proved the dominance of *logic* over *metaphysics* that Wittgenstein refers to in his early notes for the production of *Philosophical Investigations*. Philosophical method needs to adjust to the fact that the empiricism of natural language data is highly informative in methodological terms. There was some consensus to the effect that the methodology for the extraction of subtext needs to be made available to students of philosophy even at a time when active measures are being taken to deny students in linguistics such access. Damage-limitation

in science will always occur and the next generation accepts as scientific fact that which was censored.

Max Planck (1858-1947) wrote the following words. They may well show us something of the future for our efforts to extract subtext:

A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.

(Planck, 1949, translated by Gaynor)

5.7 Collocation finds the *a priori*

It is often the case when researching collocation that one feels that there must be some benefit that flows from linguistics towards philosophy. It was Emmanuel Kant who foresaw that proof of the existence of the *a priori* would come about if and only if a paradigm shift in science took place in order to demonstrate it. He wanted to

*‘show that our intuitions about the implausibility of synthetic *a priori* propositions are mistaken, something he claims will take a Copernican revolution in the way we think about metaphysics and the nature of the mind.’*

(Garvey and Stangroom, 2012 :258; emphasis added)

Linguistic computing and the research of the linguists J. R. Firth (1957) and its verification using computers by J. M. Sinclair (1970; 1991) brought this very paradigm shift to fruition in the area of lexicography by 1987 with the publication by Sinclair of the first ever computer-concordanced dictionary, *The Collins Cobuild English Language Dictionary*. Collocation, or the tendency of linguistic forms to co-occur, was the instrumentation that allowed that early dictionary to draft itself (Clear, 1993: 271).

This paradigm shift within our empirical access to all and any linguistic forms now needs to be applied to philosophy. It is a tribute to Kant that we find the *a priori* located on the very rim of the dividing line between logic and metaphysics. All that was needed in order to reveal it was the ability to use computers to perform the one form of sacrilege that the Vienna Circle dreaded: the splitting scientifically of logic and metaphysics, but, as a crumb of comfort, within natural language rather than philosophy itself. A philosopher was required for the task: a person with *locus* to provide ‘the

method of its verification' and in so doing, assuage the anger of the *Wiener Kreis*. Who better than Bertrand Russell from whose formula we not only accomplish the surgery, but open the entire field of *subtextual* meaning through the recovery of the quasi-propositional variables?

Russell asserts:

'A perfectly logical natural language will have a grammar but no vocabulary at all.'

(Russell, in Pears, 1985: 24; emphasis added)

The next step was to proceed to search reference corpora for the grammar strings alone, using the asterisk, or wild-card, to find all possible 'vocabulary' as quasi-propositional variables.

Because the method proceeds from the highest point of *intuitive opacity*, the cognitive now stands falsified in Popperian terms. As the revolution proceeds it becomes plain, as Whitehead once quipped, that 'all of our enemies have the same neck!' The *a priori* resides in the most frequent vocabulary collocates of the grammar strings. Such vocabulary is 'granted', by those very grammar strings, *immunity* from semantic change over time. This phenomenon falsifies Saussure's synchronic/diachronic distinction. The more frequent collocates are in the nature of 'logicised' metaphysics, but always with a pronounced Wittgensteinian 'family resemblance'. The entire exercise is proof of Sinclair's (2004) assertion that collocation reveals the *function* of lexis. Wittgenstein adopted the same position.

It is appropriate that the most frequent collocates of a grammar string show their affiliation with the *logic* of natural language by refusing to age.

5.8 The Logical Fallacy Involved in Text Mining

Having found that metaphysics has a role that is subservient to the logic of natural language, we need to ask whether any new developments have become casualties to the age-old practice of allowing all activities within corpora and empiricism to be dominated by the issue of vocabulary.

Large sums of money have been invested in text-mining within NLP studies. The texts operate within fairly facile criteria: large corpora of, say, texts on the subject of pharmaceuticals are searched with a view to finding ‘new ontologies’ within such texts. Aspects of the technique were perhaps copied uncritically from the tradition of work that was evolved by John Sinclair for use within various parts of the reference corpora that were set up in order to deal with the writing of dictionaries. There was a time when monitor corpora were felt to need human, intuitive monitoring. Subtext will, as it becomes established as instrumentation, gradually take over much of the function of monitoring.

At the time, all sampling for such reference corpora was argued from and based upon subject-areas and their vocabulary, much of it specialised. It was assumed that once very large corpora relating to pharmaceuticals had been captured, their interactions would be captured comprehensively also.

However, this often turned out *not* to be the case. Large specialised corpora often take rudimentary pharmacology and pharmacognosy for granted. Corpus linguistics fought back with local grammar (Barnbrook, 2002) on the basis that if every grammatical use and function that characterises the development and use of drugs could be stated, all of their uses would potentially be captured within one large corpus.

The same forms of reasoning began to prevail within all subject areas; and many of these studies, especially those that took the trouble to create local grammars, bore fruit. However, the same fallacy that limited the logic of Aristotle, confining its potential to a ‘small corner’ (Magee, 1978: 98), has now begun to hobble corpus linguistics: the call for subject-specific corpora has resulted in the provision of huge corpora in the mistaken belief that they will contain all aspects of a discipline. Furthermore, the size of the corpora is often a great selling point.

The method reached its *reductio ad absurdum* in mid 2010. In that year the Poetics and Linguistics Association (PALA) Conference was held in Genoa, Italy. Professor Ron Carter was to have been the keynote speaker for the area now termed Corpus Stylistics. The conference theme was ‘Landscapes’. Ron was unable to attend, but sent an email to Lesley Jeffries, Chair of PALA, recommending the creation of a ‘landscapes corpus’. His talk was successful, and his desire for a subject-area corpus has caught on. Specialised subject-corpora are on the rise.

However, with discovery of computer-derived subtext the need for ‘horses for courses’ corpora has effectively vanished and now that method needs to be tested to determine whether it is falsifiable.

I recall asking John Sinclair to summarise what he felt Firth meant by the expression ‘abstracted at the level of syntax’. He explained how the OSTI team had spent hours debating it and that he believed they had examined all possibilities.

Well, in a moment of idleness between searches for subtext, I searched the BNC, NOT for “... its collocability with dark” but for *is its * with **. The result is astounding and set out below. I already have colleagues in philosophy demanding to know how to use it. One of them claims that it will save hours of fruitless discussion in metaphysics. I believe that this will result in the dropping of the ‘horses for courses’ fallacy. It is doubtful that NLP will either use it or drop it because, in the case of NLP, *funding for corpus development* is a major financial consideration. Procedure is likely to ‘follow the money’.

But, note how the brief concordance below is a major finding and singles out the brilliance of Firth. All of the instances below, point to the *right* of the node, *outside* of disciplines. They are all *exophoric*. Only Firth, alone, says of collocation that the solutions provided by collocation are effectively, *endophoric*. They point *inside* language to one *collocate* making a ‘reading’ of another collocate and asserting necessary meaning between them! The empiricism of the logic of natural language (Dummett, 1976) may well be capable of distilling the hours spent on vocabulary as metaphysics long ago in an OSTI Seminar, in Edinburgh. Such a development would be highly non-trivial.

MicroConcord search SW: *is its * with **

80 characters per entry

Sort : 1R/SW unshifted.

```
1 er in section 3.6. </p> <p> One is its association with the rather questionable
2 er feature of the rural economy is its relationship with urban labour markets. T
3 everal attractive features. One is its con&rehysonance with what is generally ac
4 has come through to the public is its preoccupation with sex. The idea that dre
5 inguishing feature of the model is its concern with the secular development of t
6 e developed from these sketches is its interaction with social forces – literall
7 ion). At the deepest level this is its symmetry with the social order within whi
```

8 eat strength of 123 For Windows is its compatibility with all the earlier versio
9 mathematically, but that, such is its concern with the quantitative aspects of
10 along a bay. What sets it apart is its situation, with the massive range of the
11 set the TGI100 apart. The first is its conformity with the General MIDI Specific
12 p> A common image of later life is its association with residential care and ins
13 integral part of Eo's strategy is its partnerships with the companies, it says,
14 st strength, on the other hand, is its compatibility with Windows. It's a streng
15 n important feature of research is its concern with the nature of the event unde

Source: BNC

“The doctrine of forced acquaintance is the foundation of Russell’s logical atomism.”

(Pears, 1987: 63)

The concordance above is proof of this doctrine and is a major finding.

The spirit of the movie, *The Dead Poets Society*, is in the process of being made a reality by collocation and the empiricism of natural language and logic alone! The variables of the first wild-card alone and their consistency demonstrate the need for the total re-valuation of Wittgenstein’s assertion (one reason for his refusal to join the *Wiener Kreis*) that natural language logic must be placed above metaphysics.

5.9 Political Spin and Collocation: Intriguing but *non-p*

The possibility that there may be a Third Way in dealing with political problems successfully and without the paucity of logic behind spin operations being detected, has, in the past twenty or so years, been a constant theme within politics, especially in the UK. Prime Minister Tony Blair set up a spin department at 12 Downing Street under Mandelson and Campbell; but its roots went back as far as Kim Philby’s school for persuasive writing in the 1940s at Brickendonbury House.

Modern versions of spin were extremely sophisticated and some were successful, but *all* suffered from the problem that although the computer and collocation could draft the persuasive aspect of spin, the same computer could be used to falsify the writing, in ways that were often as dramatic as they were conclusive.

Two early examples come to mind. No sooner had John Sinclair demonstrated that the contents of a concordance was indicative of a term's *function*, than the newspaper version of the expression *natural justice* could be shown to suffer huge amounts of *denial* and *breach* in newspaper reports of trials. There was even the suggestion that the BNC had been homogenised in order to offer a more attractive take on the matter. Publications that attempted to draw attention to the problem tended to appear in print with key aspects of the matter removed. Louw (2007a) finally did justice to the issue, but only this month has the heat gone out of it, with the publication of the second edition of Paul Simpson's book on Stylistics (Simpson, 2014: 101) where the analysis is given its first independent reference.

Arguably the most damaging example was the huge campaign that was mounted (with assistance from Europe) in South Africa in order to lend credence to the idea that organised mass-waiver might be termed a Truth and Reconciliation Commission (Louw, 2003). Such commissions sprang up all over the world in the wake of the South African prototype. However, their compellingly dishonest aspect is *easily* stated in collocational terms. If *truth* and *reconciliation* are to have the same status as a legal institution, these terms, like *truth* and *justice*, must, at the very least *have been associated with one another in the past*. Sadly they have no such association and are accordingly and notwithstanding the presence of Archbishop Tutu, a product of spin and not of precedent. Unlike the case of natural justice, truth commissions continue to be big, if not fraudulent, business.

5.10 Negotiating

All negotiating is effectively calculated to use language in ways that are to one's own advantage. This means that because the use of language is tactical, its outcomes will be easy to trace within a corpus of negotiating (Louw, 2006).

Of course, once we have shown that certain language choices bring about particular outcomes, there will be a descript sense that some forms that allow for open choice in every day uses of language will, within negotiating, show the consistency of the device to which they have become attached as part of the ideology of negotiation.

Good examples, some with very high frequency, abound within certain ideologically charged situations. For example, the use of the term *learned friend*, *prima facie* diminishes any offence being taken between opposing counsel in a court of law. However, the expression is used so routinely that it sometimes masks any illocutionary force that is in use. It is routinely understood that the adversarial context that obtains between counsel propels the illocutionary force of the term to mean something uncomplimentary, such as ‘complete fool’ in the context of an only dubiously successful line of argument.

Greater solidarity between counsel is achieved once one of the parties drop the use of the term *learned* in such exchanges as: “... as I am sure *my friend* will readily concede...”. This form of usage may be shown, in say the Sizewell Corpus, to show moments of agreement on issues of common cause, such as the undesirability of nuclear contamination or even proliferation. The Sizewell investigation was an informal investigation. The term *my friend* is hugely prevalent within the Sizewell Corpus and although issues relating to blameworthiness may be more hotly contested, it is plain that counsel on both sides are swayed by issues that relate to those issues within the nuclear debate that relate to the ‘public interest’.

On the other hand, the form in a newspaper might find itself evenly divided between legal references and the name of a racehorse.

MicroConcord search SW: **learned friend**

80 characters per entry

Sort : 1R/SW unshifted.

```
1 Hide said: "This course owed My Learned Friend a win. He cracked his near-fore
2 ere for the Tony Hide-trained My Learned Friend. A winner of a small race at Lin
3 iately, if unoriginally named My Learned Friend is owned jointly by Jeremy Rober
4 it must be true. Just because my learned friend Mr Leveson (Brian Leveson, QC, p
5 t possible to snuggle up to your learned friend secure in the lonely and guilty
6 Paper> <Story> <Group> MY LEARNED FRIEND, who broke a leg at Haydock last
7 s we lawyers say, except another learned friend with a right of audience. This i
```

Source: The Times Corpus, 1995

Within philosophy, Michael Dummett (Craig, 2005: 205) takes the view that where a philosopher recognises both (1) the *locus* for speaking out when evidence for *non-p* is detected and (2) that the moment of recognition *warrants* an intervention, such interventions ought to satisfy the warrant and be made. It may take some time before the ideology that prevents the act of speaking out is overcome. The example, *natural justice* falls into this category.

The two concordances below demonstrate how the term *happy to* collocates with situations of cover-up followed by compliance and *sorry to* with insincere apology in situations where interruption is important for obtaining a particular outcome in a court setting.

MicroConcord search SW: **happy to** and **sorry to**

80 characters per entry

Sort : 1R/SW shifted 4 characters.

Source Sizewell Corpus

21 t struck the right balance, we will be happy to comply with what you regard as
22 some other way we shall, of course, be happy to comply. 166. Sir,
37 gories? A. Not without notice, but I am happy to have the information supplied
49 blem in the US. Again, I would be very happy to make that available to the Boa
50 gain, this is something I would be very happy to make available to you on Ameri
52 y document, sir, but would be quite happy to obtain a copy of this for Rev.
53 e by giving those references - I am happy to omit them - the reason I do is
55 that is why I said initially we are happy to provide the factual informatio
56 o not have them to hand. I shall be happy to provide them. 478. Q. Than
57 Thank you, Mr. Baker. I would be quite happy to provide you with my calculatio
58 Table 11. A. Certainly, we will be happy to provide it. 66. Q. The oth
59 inding the Department would be very happy to provide it. The decision to e
60 y. We have told Mr. Sweet that we are happy to provide that material to him
61 eady in evidence, we would be happy to put it into evidence.
62 is not already in evidence, we would be happy to put it into evidence. It
63 the total was 24%. Sir, I shall be happy to put that in evidence. 493.
69 hing for Mr. Priddle, we have, I am happy to say up to a point, exceede
70 e safety case. So, while I am very happy to schedule these changes so that
71 - A. But I think the Board would be happy to see a completed station before
72 Department of Energy would be quite happy to see stations ordered in advanc
17 AM _____ 410. MR. POPHAM: I am sorry to interject, sir. I do not think the
18 aps I could go on... 320. Q. I am sorry to interrupt you. I am asking you for
19 s, indeed. 247. THE INSPECTOR: I am sorry to interrupt you, Mr. Taylor, but the
20 97. THE INSPECTOR: Mr. Howell, I am sorry to interrupt you, but the shorthand wr
21 _____ 660. Q. Professor, I am sorry to interrupt you. Could you go with m
22 02. THE INSPECTOR: Mr. Howell, I am sorry to interrupt you, but you have identif
23 . 944. A MEMBER OF THE PUBLIC: I am sorry to interrupt this, but you have just b
24 7. MR. FITZGERALD: Mr. Jenkin, I am sorry to interrupt you. I wonder if I could

25 upply... 137. Q. Mr. Parker, I am sorry to interrupt you, but can we trace th
 26 which is just above... 46. Q. I am sorry to interrupt you, Le Baylais 3, when y
 27 st l mtce 768. Q. Professor, sorry to interrupt you. A. Where is the oil
 28 _____ 660. Q. Professor, I am sorry to interrupt you. Could you go with m
 29 upply... 137. Q. Mr. Parker, I am sorry to interrupt you, but can we trace th
 30 THE INSPECTOR: Mr. Fitzgerald, I am sorry to interrupt you, but we promised we w
 31 262. Q. THE INSPECTOR: I am sorry to interrupt you, but you keep on usi
 32 ou not? A. That is correct. 605. Q. Sorry to interrupt you. A. To some extent, t
 33 THE INSPECTOR: Mr. Fitzgerald, I am sorry to interrupt, but it has happened on a
 34 CPRE/S/155. 39. THE INSPECTOR: I am sorry to interrupt, Mr. Taylor. What is this
 35 PROF. FOSTER: Thank you very much. Sorry to interrupt. 730. Q. MR. BARTLETT: Mr
 36 like, of the supply curve. 331. Q. Sorry to interrupt. Can you continue please?
 37 0. MR. TAYLOR: Mr. Fitzgerald, I am sorry to intervene, but you will recall in r
 38 ly accept it, yes. 269. MR. TAYLOR: Sorry to intervene, Lord Silsoe, but you do
 42 are... 237. Q. THE INSPECTOR: I am sorry to keep interrupting you, but characte
 43 oil prices into UK terms. 181. I am sorry to keep on coming back to it, but as f

5.11 Branding and Advertising

Intuitive opacity applies as much to advertising agencies as it does to members of the public. One might have expected the information age to have corrected this by now, but it has signally failed to do so. Advertising campaigns are still routinely sucked out of the thumbs of fairly untalented account executives and copy-writers and this is likely to remain the case unless and until some corpus skills are provided to members of their client base. Once that occurs, there will be a fairly immediate reaction.

For example, MBA students believe that for the money they are paying to advertising agencies there ought to be some form of proof that the language they are sold actually works. The fact of the matter is that no ordinary agency ever provides subtext or knows how to extract it. And it gets worse, because some of the invented slogans, although catchy to the ear, have very little consistent empiricism behind them in the real world. So, all of the apparent inspiration behind the cellular giant's slogan

Inspired to change your world

is a mere puff and produces some very hit and miss material, as we see in the concordance below. The numbers for the concordance have been altered because of the removal of the word *need* which was unwanted. The search line is **ed to *your **.

MicroConcord search SW: *ed to * your *

80 characters per entry

Sort: 1R/SW unshifted.

(Source: BNC)

3 <Group> I was both dismayed and alarmed to read your article "Heritage police g
4 From Mrs M. Symonds Sir, I was distressed to read your article on Auckland, Ne
5 m Mrs Eileen Craine Sir, I was interested to read your article about British G
6 rict mountain rescue team, I was stunned to read your article on the wearing of
7 he beauties of Homer and Virgil, referred to in your article, reside not solely
8 nable to find out much, so I was pleased to read your article. Yours faithfull
9 rdings of "womb sounds" that you played to get your baby to sleep. Then somebody
10 ck in business. Next week, jeans designed to give your backside a lift will hit
11 bership. In addition, you may be entitled to have your benefits backdated. Hur
12 icle, February 21). We were also pleased to note your brief report of February 7
13 desk one, so that others are not forced to take your calls. If you are confronte
14 umes, to leather and wood aromas designed to make your car smell like new. Thi
16 m. Thwack! But of course, I'd be delighted to hold your cardigan. Splat! No real
17 t easy-to-use computer programs, guaranteed to set your CD-Rom spinning. My fi
18 OT so many years ago, if you had wanted to take your children on a ski holiday y
19 le to cope. And NPI has at least promised to pay your conveyancing fees, togethe
20 using plastic overseas. You are advised to keep your credit card in sight at al
21 , with the swipe machine that is supposed to read your credit card. These are su
22 ction. There is nothing better guaranteed to ruin your day than a night of tor
23 of large sums of money which you used to pay your debts and finance a number of
25 Enya, if you like. Aren't you supposed to pay your dues before getting a reco
26 s in the nursery, moments of fun designed to open your eyes to the unexpected de
27 rescription lenses or especially moulded to fit your face. Cutler A Gross's st
28 rom Mr D.P. Hosking Sir, I was interested to read your feature on the use of c
31 93 Leasehold Reform Act, you are entitled to buy your freehold provided: You h
32 terests me is why Havel Merchant agreed to sell your fur coats in the first plac
33 side a neighbour's house but you continued to make your getaway. "As a result
34 well in La Baule you have to be prepared to put your hand in your pocket. A set
40 can look so horrible if you are forced to take your jacket off because you are
41 to drive them. If you are too tired to lift your left arm to switch on the wi
42 The conventional phone line that used to be your link with the outside world has
43 <Group> In boxing, you are not allowed to save your man from injury; in Americ
44 of investor regulation, that is designed to keep your money safe and to keep ro
45 now good value, without feeling emboldened to put your money where their mouths
46 like a T-shirt, but once you get used to it, your new ribcage will give you adde
47 our FSAVC, you might not even be allowed to take your occupational pension until
48 State for Transport Sir, I was disappointed to see your one-sided attack on th
49 etired formally in 1990, but had continued to edit Your Own Business, The Times'
50 till a moneyed sport and you are expected to make your own way, which means comp
51 ey decisions for you. If you are minded to pick your own trust for each differen
53 ly you can't do that, you're not supposed to take your own life with someone els
54 tificial way. If you cannot be bothered to make your own, then I doubt you wil
55 em works, and how best it can be managed to meet your particular needs. The temp
56 Investment freedom: you are not obliged to turn your pension fund into cash on
57 erving places that so far have failed to make your personal "worth a detour" o

59 d mine in the Orange Free State, omitted to tell your readers that, in a radio i
60 Mr Peter Marsden Sir, We were pleased to see your recent coverage of the situ
61 m Sir Peter Lazarus Sir, I was astounded to read your report of December 30, h
62 om Mr Charles Sussex Sir, I am delighted to read your report today that in Ame
63 om Mr Gregory Segal Sir, I was interested to read your report (June 9) stating
64 er> <Story> <Group> I WAS interested to read your report (News, June 4) t
65 tick-up collars, before you were allowed to buy your set. Volume V by Iverach
66 > <Story> <Group> WE WERE pleased to see your short extract from a letter
67 reathalyser, although you may be allowed to call your solicitor first. If the ma
69 it after the weekend was awful, used to turn your stomach on a Sunday night. I
70 at what altitude, but do not be tempted to grab your summer beach-holiday lefto
71 a perennial family feud. I was overjoyed to read your warm "Welcome to the Pre
72 s investment managers, or may be tempted to give your wealth to spendthrift chil
73 tion. She tells the composer: "I wanted to meet your wife and now I have seen he
74 the legal system that it is less wicked to kill your wife than to kill a strange
75 r. Exam technique: You will be expected to show your working and reasoning whe
76 he "soft fact-find". You will be asked to describe your dreams of a new home, ca
77 their own line manager." Being asked to leave your job may be the worst-case s
78 ave spotted in good time had you bothered to check your tyres regularly. Take yo
79 tory outlet. And if you can't be bothered to leave your armchair, there's the re
80 t." As far as music especially commissioned to heighten your emotions in a fil
81 during your lifetime, and can be converted to store your body upon your demise.
82 ure that exercise areas are well cushioned to protect your child in the event of
83 treatment you received when you decided to cancel your cover. You were clobbere
84 culous letter? You have wisely decided to invest your fill a month into a build
85 my intervention, the RAC has now decided to refund your membership fee for this
86 t your meagre income. If you are determined to spend your year in a particular t
87 city's cathedral. Should you be disposed to reduce your journeying through 30,0
88 he Law Society says: "We are not disposed to accept your suggestion." The appoin
89 ou may disagree, but you are not entitled to support your standpoint by ridiculi
90 authentically Thai and is almost guaranteed to anaesthetise your taste buds for
91 ossibility of your own demise is guaranteed to alter your perspective on the imp
92 those annoying cartons the ones guaranteed to splash your shirt before you leav
93 d figures, the more you might be inclined to leave your cash on deposit, if not
94 also a "rite of purification ... intended to cleanse your life". The state jud
116 ing my intervention, the company offered to collect your computer and get a seni
117 redress, and now Gold Medal has offered to refund your advance payment for the
118 pposed to taste good for you, as opposed to delicious. Your parents force you to
119 other, says Legat, unless you're prepared to commit your money for at least five
120 s in this industry unless you're prepared to invest your leisure time as well. B
121 p me and contradict me"; and "Be prepared to drink your cocktails standing for a
122 u in touch with an agent that is prepared to handle your timeshare and as a spec
123 now tells me it always has been prepared to return your cash (less 10% for the
124 satisfy most appetites. Just be prepared to adjust your real-life meal- times,
125 ing do not bother unless you are prepared to commit your money for at least five
126 le jester makes you wish you had remembered to bring your crossbow. But the glos
127 hotels and flights. You are not required to vacate your room in Goa. Brochures
128 ich people rather than music are selected to share your tropical solitude. I end
129 ng too often, as I was once when telephoned to confirm your acceptance of an ear
130 frightening, times." You may be tempted to stick your hand up and ask: what the
131 ter and the landscape. You'll be tempted to think your cab driver is taking you

132 d... oh anyway, if you were ever tempted to remove your critical blinkers and re
133 ciety, but you do admit that you tended to close your eyes to your mortgage arra
134 At one point, the Woolwich even threatened to repossess your daughter's home. S
135 n (ret'd) Sir, The photograph used to illustrate your report today on the cha
136 ecognise the video graphics card used to drive your monitor and will let you cha
137 y and root strength before being used to repair your lawn. All the above lawnc
138 eally means melo-drama: music is used to heighten your feelings, comment on the
139 ly, elements of my plan could be used to complement your own efforts. Start wi
140 ster was a would-be murderer who wanted to sabotage your aircraft?" Witness firm
141 how would you like it if someone wanted to murder your family pet? Frankly, I
142 <Group> If you have always wanted to paddle your own canoe, or go punting

Some of the most successful brands were established intuitively as far back as the Victorian period, as we see in the case of an organisation called the Old Mutual. As Wittgenstein might have noted inherent in the title and common to its two words is the absent collocate *friend*. *Old friend* has fairly huge, if common place, empiricism; but the term *mutual* straddles the boundary between friendship and the insider language of the investment and insurance industry.

MicroConcord search SW: mutual CW: friend

80 characters per entry

Sort : 1L/SW unshifted.

1 ce> <Date>25 August 1995</Date> <Headline>Mutual friend</headline> <byline
2 the lovely and wonderful Beverley. A mutual friend introduced us, we went out fo
3 ury period. She met Waitz through "a mutual friend", Gerard Hartmann, McColgan's
4 owed to bring a companion, ideally a mutual friend or brother, to whom she can c
5 im the highest political office? A mutual friend, Ken Hamer, doubts she would
6 g photographed with the prince. As a mutual friend of hers and Charles put it: "
7 enables's press briefing came from a mutual friend, so we will preserve his anon
8 blic with Camilla Parker Bowles at a mutual friend's party. He seems determined
9 London, where he was introduced by a mutual friend. In their opening conversatio
10 e over a private lunch arranged by a mutual friend. The subsequent conversation
11 tactable with Dixon at the home of a mutual friend, yesterday issued a statement
12 gh his business connections. Using a mutual friend, he approached Nicholson offe
13 . Seconds after Tony, myself and a mutual friend Mark, a photographer from Den
14 dy left his wife. White also cites a mutual friend whom he regards as an excepti
15 was introduced to a businessman by a mutual friend and he was going to do a sale
16 n 1976 Dylan needed a girl singer. A mutual friend recommended Harris and the re
17 ost attractive man I had ever met. A mutual friend had brought him along to a Ch
18 few years ago I sent a message via a mutual friend to Meryle Secrest, the Englis
19 th of a Los Angeles earthquake and a mutual friend had suggested we dine togethe
20 She told Christopher Fairweather, a mutual friend, about how she wanted to get

21 papers." He tells of the death of a mutual friend and thanks God for the relief
 22 ment of Rena Costello. The trio, and mutual friend Isa McNeill, had moved to Glo
 23 he would never marry her. A close mutual friend spoke last week of the affect
 24 ght change completely," said a close mutual friend of the prince and Parker Bowl
 25 >20 July 1995</Date> <Headline>Our mutual friend</headline> <byline></byline>
 26 October 1995</Date> <Headline>Our mutual friend at the Treasury</headline>
 27 Gad's Hill Place and to complete Our Mutual Friend. Mr Tripp says: "The exact
 28 uctions of Martin Chuzzlewit and Our Mutual Friend. More work on television fo
 29 r is best described in Dickens's Our Mutual Friend, a masterpiece of the urban i
 30 <Headline>The building societies' mutual friend</headline> <byline>Sara McC
 31 milieu. Her frank portrait of their mutual friend Camille Paglia caused the now

Source: The Times, 1995

Advertising agencies will need to address the intuitive deficit. Even the ancient Greek
 that is at the heart of the term *stochastic* no longer means 'capable of guessing' as we
 see in this definition from the Collins Dictionary.

Stochastic adj. 1. Statistics. a. (of a random variable) having a probability
 distribution usually with finite variants. b. (of a process) involving a random
 variable the successive values of which are not independent. c. (of a matrix)
 square with non-negative elements that add to unity in each row. 2. Rare.
 involving conjecture. [C17: from Greek *stokhastikos* capable of guessing. from
stokhazesthai to aim at, conjecture, from *stokhos* a target] *sto*'chastically adv.

(Emphasis added)

5. 12 Conclusion

The selection of applications that have been demonstrated in this chapter represent only
 a small fraction of verified applications within my research more generally. However,
 one advantage that can be gained by creating a selection is that they can be represented
 by way of a conclusion in tabular form. The tabular form is in a sense Firthian in its
 method, concentrating as it does upon outcomes both in the external world and in the
 confirmation of the hypotheses set out for examination in this thesis.

Table 5.1: Practical Applications of Collocation

	Application	Date	Mechanism in Collocation	Outcomes	Hypotheses borne out
1.	Delexicalisation and Relexicalisation	Louw, 1991	Relexicalisation as device	Intensification of meaning	Numbers 2, 5
2.	Semantic Prosody	Louw, 1993	Irony/Insincerity and prospection in literary texts	Ability to trace and prospect them from first lines	Number 3
3.	Contextual Prosodic Theory [CPT]	Louw, 2000	Fractured Contexts of Situation	The sinister nature of 'absence'	Numbers 1, 2, 3
4.	Corpus-Derived Subtext	Louw, 2010a, 2010b	Carnapian Subtext; Russellian Subtext	One key to authorial insincerity	Numbers 2, 6
5.	Subtext as a practical method	Louw And Milojkovic, 2014a	An extended example of the Cambridge Handbook Of Stylistics	Logical Atomism applied	Numbers 2, 4, 6
6.	Collocation finds the <i>a priori</i>	This volume	Collocation improves method in philosophy and proves itself as science	The beginning of collocation studies in philosophy	Numbers 4, 6
7.	The Logical Fallacy Involved in Text Mining	This volume	The dangers of arguing from metaphysics alone	'Horses for courses' fallacy demonstrates the importance of natural language logic	Numbers 4, 6
8.	Aspects of Political spin	Louw, 2004	Useful for deception but always contain <i>non-p</i>	Its dangers revealed, because philosophy uncovers <i>non-p</i>	Number 6
9.	Negotiating	Louw, 2006	A paradigm shift in negotiating	Produces a new class of negotiator	Numbers 5, 6
10.	Branding and Advertising	This volume	Suasive aspects in Subtext	Persuasion with surgical accuracy	Number 6

CHAPTER 6: CONCLUSIONS

6.0 Introduction

In this concluding chapter I shall deal with the following:

1. Simultaneous paradigm shifts within the research of language itself and within applications
2. The value of intuitive opacity as a Popperian strenuous test

6.1. Simultaneous Paradigm Shifts within the Research of Language itself and within Applications

It is often the case that paradigm shifts show up our own analogue taxonomic efforts in a poor light. A good example would be that the massive contribution to lexicography of John Sinclair was simply accepted by the community of dictionary-users, rather than treated as a major discovery within the *science* of dictionary-writing. But in 1987 there was spirited resistance to Sinclair's new and revolutionary method. However, the novel aspects of the first edition were insufficiently played up, even by the publishers, who classified the new dictionary as an ELT book and it was sold in book shops in their ELT sections! The result of this was that the digital revolution within lexicography was slowed down and even the publishers, whose misclassification of their own dictionary reduced their profits, were initially lukewarm about the fact that it was a landmark dictionary, produced by a visionary linguist.

And yet, Sinclair was aware all along that the landmark status of his dictionary was, in scientific terms, intended to clear the way for a bigger landmark paradigm shift within language itself. The dictionary, for Sinclair, had a dual function: it was to revolutionise lexicography as a dry-run for furnishing instrumentation for meaning based upon J. R. Firth's notion of meaning by collocation.

Open resistance to this idea was never voiced; and it soon became obvious to Sinclair that a Collocation Dictionary within Cobuild was unlikely to materialise. Furthermore, he realised with even greater disappointment that his Project LUCID, that would have made direct claims to the status of a new theory of language, was likely to remain at the level of planning only. He never demonstrated to me any applications that it was likely to have, and the fate of his notes, if any, is unknown.

However, even though any speculation about these two planned ventures appears to be nugatory, they do serve a useful purpose: They allow us to infer the presence of an almost unique scientific occurrence. The *methodology* that was about to change lexicography forever, and did change it, was on the brink of altering linguistic *theory* forever also. In the current jargon of US diplomacy, this fact alone would have been a ‘game-changer’. As such, the term implies that decision-making at a higher level than the analyst may then take place.

But it raises an important issue for the pursuit of science: Can a mere *methodology* ever ‘eat’ a *theory*? Can steps be taken beyond the realm of the analyst and his/her science to prevent these forms of ‘predation’ from occurring spontaneously within disciplines in general and also within the very medium in which all scholarly and general writing takes place, in particular?

It is bewildering, of course, but we do seem to be driven towards the fact that alterations to theory are always contested; and that this contest is sombre. As Mephistopheles puts it in Part 1 of *Faust*:

*Grau, teurer Freund, ist alle Theorie,
Und grün des Lebens goldner Baum.*
All theory, dear friend, is grey,
but the golden tree of actual life sprints ever green.
(Goethe, 1808)

If resistance to change is a major issue that accompanies all nascent theory, we may need to remind ourselves of the simplicity of theory as properly defined:

'... a set of propositions which provides principles of analysis or explanation of a subject matter. Even a single proposition can be called a theory.'

(Mautner, 2000: 563)

This thesis has preformed both analysis and explanation. The newest dimensions of these apply primarily to *subtext*. However, what singles subtext out as a special case is the fact that upon close inspection, we find that specific reference to natural language logic is rather sparse in the works of both Firth and Sinclair. And yet, they are the mainspring of the work of the philosophers, especially Frege, who makes space within mathematical propositions for a form of context of situation. Dummett (1981: 176) sets out ways in which such contexts can be amplified by means of challenging warrants.

But it is collocation that seals this in the form of a linguistic theory whilst simultaneously improving philosophy itself. If philosophy adopts collocation into natural language philosophy and finally hands it over to science, the stage will be set for launching a fully verifiable subtextual theory of language and language study. However, all of this comes with the *caveat* of J.R. Firth himself that scientific facts do not exist until they are *claimed*. I hope that this thesis and the decades of research that underpins it will between them bring closer the possibility of and the *locus* for making such a claim; and out of an abundance of caution (*ex abundante cautela*) I hereby lay claim to collocation as a theory of language, lest my endeavours fall victim to a technicality.

A theory of collocational subtext will not suffer from the malady of being grey. On the contrary, it will be as colourful as Goethe visualised life, not only in theory, but as represented in the paintings that are still *in situ* to this day in the Goethe House in Frankfurt. The reason for this colour-saturation is to be found in the criterion of over-provision contained in my founding article for Contextual Prosodic Theory (Louw, 2000).

In other words, it is CPT that allows us to experience those collocates of states of affairs that happen to be absent from our own lives, *not as fiction*, but as verifiable *truth*. The reference corpus seen in analogue terms, needs to be *read*, as Ron Carter said in his opening remarks for the 35th Annual Conference of ICAME in Nottingham in 2014. However, the timeless and digital dimensions obtained from the situational logic contained in any corpus may, in analogue terms, appear to have unsuitable *subject*

matter. But they will provide more concentrated contexts of situation than we experience. In future, they will paradoxically be furnished by the grammar strings of each sentence of the target text in the form of the full range of lexical collocates that are germane to one line of argument at a time and consolidated cumulatively using what I am calling the ‘tendrils and network’ model for collocation of Williams (2008). If the world is, as Wittgenstein proposes, the ‘totality of facts’, then their provision in a linear form ought not to be too great a challenge for present-day computers to provide.

In Chapter 4 of this thesis, I provided a table to show how the destinations involved in each search for subtext are opaque to our intuition. But they are not opaque to our faculties of reason. Each situational or situated argument stretches over a range of collocational variables that cover what often seems to be an eternity of *time*. And the character of these variables is almost always uniquely part of the subject matter of philosophy. It is singularly gratifying that a *method* derived from a central tenet of the Vienna Circle should, once it has been used to ‘split the atom’ of that esteemed Circle, provide us with *proofs* for the terminology that is common to both philosophy and linguistics. The line ‘Move him into the sun’ as a subtextual line opens a continuum of time that stretches from the *a priori* to the present day and simultaneously demonstrates the continuum of delexicalisation and relexicalisation. The range of permissible corpus-attested terminology (Louw, 2008) will finally be limited, finite and complete if it is based upon the phenomena that subtextual studies uncover. All other terminology will predictably be falsifiable. This task will be part of an equal collaboration between university teachers and students (Louw, 2014d). A major concern of the new method for teaching will involve:

1. the eradication of all intuitively-derived terminology;
2. the evolution of criteria for the attestation by the corpus of new terminology; and
3. methods that will convert the apparent weakness of intuitive opacity into discovery procedures that are both powerfully heuristic and instantly adoptable. This will counteract the trend that has developed (even among keynote presenters at corpus conferences) of enshrining obsolete intuitively-derived software in ever new and increasingly recondite subject-matter, based on NLP-styled types of study.

And so, we begin to discern that although corpus linguistics applies computer technology to disciplines that are still, for historical reasons, largely analogue in their interests and organisation, this variety of linguistics also currently staggers on aimlessly without any apparent theory. It has run out of steam to the point that, unless a solution is found, in the form of a theory, NPL will take it over completely. The boredom is palpable as the well-capped discipline of corpus linguistics thinks only sound thoughts on approved subjects.

The real hope for a new theory of language becomes visible if we turn to collocation, not as a sub-section of corpus linguistics, but as the instrument for moving us from possible worlds into the real world and its contexts of situation. From there, we proceed like Orpheus into the intuitive darkness of the ‘power behind the words’. Within subtext we discover situational logic, its eternal origins of argument and their ageless and invisible collocates that modify the meaning uniquely of whatever lexis actually occurs in every sentence as it builds the world (Carnap, 1928), by means of the logic of natural language.

In the same way that this process allowed Frege to dwarf the logic of Aristotle, we, using corpora (Louw and Milojkovic, 2014c), ought now to begin to read fully any line of any text, rather than only a ‘small corner’ of it.

6.2 The Value of Intuitive Opacity as a Popperian Strenuous Test

If we are to grasp some measure of the distance covered by collocation in the course of this study, it may be worth showing where matters stood, say roughly 70 years ago, and then compare them with the results that we have obtained as a direct result of intuitive opacity.

In a chapter devoted to ‘the process of concept attainment’, Bruner *et al.* (1956: 54) set out the following guidelines in point form. The text is, of course, a cognitive study and pre-dates the arrival of lexical computing in Cobuild by 30 years.

- a. *To insure that the concept will be attained after the minimum number of encounters with relevant instances.*
- b. *To insure that a concept will be attained with certainty, regardless of the number of instances one must test en route to attainment.*
- c. *To minimize the amount of strain on inference and memory capacity while at the same time insuring that a concept will be attained.*
- d. *To minimize the number of wrong categorizations prior to attaining a concept.*

(Bruner *et al.*, 1956: 54)

The objectives are carefully stated and if we consider their relation to the title of Bruner's landmark textbook on the subject, they are as far-reaching as anyone might have hoped to be in 1956. Bruner's objective stated in the title of the main volume was *'to offer an understanding derived as coming from beyond the information given'* (Bruner, 1974). The date of this latter work places it five years after the delivery by John Sinclair of the typed manuscript of the OSTI Report. At the time, lexical computing had already begun at Georgetown University, but was still in its infancy. So where does the problem lie? The problem lies in the fact that from that period until the present day, no substantial progress has apparently been desired or tolerated within the school of thought that Bruner celebrates so ably as early as 1956 and again in a landmark volume in 1974.

As we saw in an earlier chapter, Jakobson (Bazell *et al.*, 1966) contributes a chapter to the volume, *In Memory of J.R. Firth*, eight years after his closing statement made at a stylistics conference at which Firth was present. Jakobson's closing statement famously sets out his 'poetic function' (Sebeok, 1960) and its content not only cries out for collocation but proceeds to 'make do' without it for no better reason than that he never read about it in Firth (1957) or dismissed it because he may have felt that it undermines the syntactic rigour upon which the poetic function was perched. In any event, we see no computational support in the work of Jakobson as late as 1966 (a time at which the OSTI study was underway in Edinburgh, with Andrew Ortony on the staff list!). Whatever was going on on the ground at the time, one thing is clear: The Bruner school of thought must have found itself end-stopped for the lack of computerisation. We cannot even say with any certainty that this was a conscious decision.

However, once we examine developments in philosophy around this period the picture becomes clearer. One analytic philosopher in the USA takes a step that will enshrine mentalism at the highest level of abstractness within philosophy, and so publicly start schools of thought that actively support the schematic structure of the 'given'. The production by Quine of a fresh dogma of empiricism to replace those of Kant (Morick, 1972) led to a raft of other philosophers taking up positions that shore up a philosophical tradition that is, to this day, unshakably mentalist. Quine, who died in his nineties, may have sensed that there was a rising tide of computational empiricism, wrote his own entry for inclusion in *The Penguin Dictionary of Philosophy* and posted it in the mail to the editor Thomas Mautner (2000: 467).

Whatever the effect of the events stated above, the fact that computational work in linguistics and philosophy failed to take off at the same time as it did in Edinburgh is material to the future direction of both linguistic theory and philosophy. We may never be able to unpick fully why this was the case and who the actors were that failed or refused to entertain computing with a view to verifying what scholarship there was and where it was bound.

This situation made progress in both disciplines extremely difficult; but within linguistics there was more confusion of a methodological kind than in philosophy. Philosophers know that science must always be preferred to naïve realism. One philosopher in particular devised strategies for dealing with such problems as sentimentality for the use of and purported strengths of human intuition. The approach of Karl Popper allows us to turn the weakness of intuition into a powerful strenuous test and methodology. Here is an example of his unique form of reasoning. I have felt drawn to his view of science for all of my career.

'I wish to convey here a heroic and romantic idea of science and its workers: men [and women] who humbly devoted themselves to the search for truth, to the growth of our knowledge; men [and women] whose life consisted in an adventure of bold ideas.'

(From: *The Problem of Demarcation*, 1974; in Miller, 1983:118)

Bold ideas are worth fighting for. They are bold because they are prepared to deal with the status quo by facing it down. When scholarship ceases to invite or allow change, Popper maintains that forms of pressure will build up until true scholars will use the

truth of falsification to shift the impasse. Popper says that “It is science in this heroic sense that I wish to study.”

We learn to be bold only gradually. The status quo resists change but falsification strikes unexpectedly and often in Socratic ways. I have made the claim that collocation is instrumentation for meaning. Timing is important. We need to wait until we are attacked and when the moment has been correctly judged to be the right moment, we need to offer proof and to offer it with great simplicity. For example, before I was locked out of teaching, I used to *show* the class the volume of the OSTI Report, *show* them the line in Firth that inspired Sinclair (‘One meaning of night is its collocability with dark’) and produce the concordance for

*is + its + * + with*

and then fall *silent*, to allow the processes of induction and inference free reign. Simplicity, boldness and falsification between them will dislodge the boulder of theoretical stasis and, technically, because this is to do with *our* language and meaning hidden from us, *that* boulder might find a slope to run down and gather both speed and momentum (which a colleague of mine refers to as ‘going viral’).

A further example may better illustrate what Popper regards as desirable: I stated in Chapter 5 that the destination of a subtextual search will never be seen fully by means of intuition. It is deeper than *genre* and all of the boundaries created by our intuitive taxonomic ability. Even if a conjecture is *false*, Popper demands that it be made.

‘But it is not the present acceptance of the theory which I wish to discuss, but its boldness. It was bold because it clashed with all then accepted views, and with the prima facie evidence of the senses. It was bold because it postulated a hitherto unknown hidden reality behind the appearances.’

(Miller, 1983: 119; Emphasis added)

In my respectful view, subtext fulfils this criterion. We stare at the boulder of established theory, as one does at Michelangelo’s David in the piazza in Florence, and, without the prompting of alcohol, we say: “I could have sworn, just a moment ago, that it *moved*.”

References

- Alston, W. P. (1996). *A Realist Conception of Truth*. Ithaca, New York: Cornell University Press.
- Ayer, A. J. (1969). *The Foundations of Empirical Knowledge*. London: Palgrave Macmillan.
- Ayer, A. J. (1982). *Philosophy in the Twentieth Century*. London: Weidenfeld and Nicolson.
- Baker, M., Francis, G., & Tognini-Bonelli, E. (eds.) (1993). *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins.
- Barnbrook, G. (2002). *Defining Language: A Local Grammar of Definition Sentences*. Amsterdam: John Benjamins.
- Barnbrook, G., Mason, O., & Krishnamurthy, R. (2013). *Collocation: applications and implications*. London: Palgrave Macmillan.
- Barratt, C. (1965). *Collected Papers on Aesthetics*. Oxford: Blackwell.
- Bazell, C. E., Catford, J. C. Halliday, M. A. K. & Robins, R. H. (eds.), (1966). *In Memory of J. R. Firth*. London: Longman.
- Beaney, M. (ed.) (1997). *The Frege Reader*. Oxford: Blackwell Publishing.
- Benson, M., Benson, E. & Ilson, R. (eds.) (1986). *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. Amsterdam: John Benjamins.
- Berovsky, B. (1971). *Determinism*. Princeton NJ: Princeton University Press.
- Bondi, M. & Scott, M. (eds.) (2010). *Keyness in Texts*. Amsterdam: John Benjamins.
- Broad, C. D. (1963). Two Lectures on the Nature of Philosophy. In Lewis, H. D. (ed.) *Clarity is Not Enough: Essays in Criticism of Linguistic Philosophy*. London: George Allen & Unwin Ltd.
- Bruner, J. S. (1974). *Beyond The Information Given: Studies in the Psychology of Knowing*. London: George Allen & Unwin Ltd.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A Study of Thinking*. New York: John Wiley & Sons, Inc.
- Carnap, R. (1928). *Der Logische Aufbau der Welt*. Berlin: Weltkreis Verlag.
- Carver, N. (1996). Philosophy as Grammar. In *The Cambridge Companion to Wittgenstein*. Sluga H and Stern D G (eds.). Cambridge: Cambridge University Press.
- Chapman, S. & Routledge, C. (eds.) (2005). *Key Thinkers in Linguistics and the Philosophy of Language*. Edinburgh: Edinburgh University Press.
- Clear, J. (1993). From Firth principles: Computational tools for the study of collocations. In *Text and Technology: In Honour of John Sinclair*. Baker, M., Francis, G., & Tognini-Bonelli, E. (eds.). Amsterdam: John Benjamins.
- Craig, E (ed.) (2005). *The Shorter Routledge Encyclopedia of Philosophy*. Oxford: Routledge.

- Cudden, J. A. (1980). *The Penguin Dictionary of Literary Terms*. Harmondsworth: Penguin Books.
- Dancy, J. (1995): 'Arguments from Illusion', *The Philosophical Quarterly*. v 45, 181. 421–438.
- Davidson, D. (1973). On the very idea of a conceptual scheme. In *Proceedings and addresses of the American Philosophical Association*. 5-20. American Philosophical Association.
- Dummett, M. (1976). Is logic empirical? In *Contemporary British Philosophy Personal Statements*. Lewis, H. D. (ed.). London: George Allen and Unwin.
- Dummett, M. (1981). *The Interpretations of Frege's Philosophy*. London: Duckworth.
- Dummett, M. (2004). *Truth and the Past*. New York: Columbia University Press.
- Emmett, D. M. (1945). *The Nature of Metaphysical Thinking*. London: Macmillan & Co.
- Firth, J. R. (1957). *Papers in Linguistics: 1934-1951*. Oxford: Oxford University Press.
- Freeman, D. (ed.) (1970). *Linguistics and Literary Style*. New York: HRW.
- Frege, G. (1884). *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*. Translated by J. L. Austin (1974). Oxford: Blackwell Publishing.
- Garvey, J. & Stangroom, J. (2012). *The Story of Philosophy: A History of Western Thought*. London: Quercus.
- Golinelli, G. (2008). E gli ultimi saranno I primi: una prospettiva indigena sulla civiltà. In: *Il primitivismo e le sue metamorfosi: archeologia di un discorso culturale*. Bologna: Clueb.
- Glock, H. J. (1996). *A Wittgenstein Dictionary*. Oxford: Blackwell Publishing.
- Goodman, N. (1951). *The Structure of Appearance*. Cambridge, Massachusetts: Harvard University Press.
- Grayling, A. C. (1996). *Russell*. Oxford: Oxford University Press.
- Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Jakobson, R. (1960) Closing statement: Linguistics and Poetics. In *Style in Language*. Sebeok, T. A. (ed). Cambridge MA: Massachusetts Institute of Technology Press. 350-377.
- Jakobson, R. (1968). *Poetry of Grammar and Grammar of Poetry*. *Lingua*, 21, 597-609.
- Jeffries, L. & McIntyre, D. (2010). *Stylistics*. Cambridge: Cambridge University Press.
- Johns, T. (2005). *Times 1995 Corpus*. A 44.5-million-word newspaper corpus.
- Kant, I. (1914[1892]). *Kant's Critique of Judgement*. Translated, with Introduction and Notes by J. H. Bernard (2nd ed. revised). London: Macmillan.
- Kenny, A. (2000). *Frege: An Introduction to the Founder of Modern Analytic Philosophy*. Oxford: Blackwell Publishing.
- Kitcher, P. (1993). *The Advancement of Science: Science without Legend, Objectivity without illusions*. Oxford: Oxford University Press.

- Kjellmer, G. (1990). Patterns of collocability. *Theory and practice in corpus linguistics*. 152-174.
- Krishnamurthy, R. (2004). *English Collocation Studies: The OSTI Report*. London: Continuum.
- Kul-want, C. & Klimowski, A. (2011). *Introducing Kant: A Graphic Guide*. Harmondsworth: Penguin Books.
- Lakoff, G. & Johnson, M. (2003[1980]). *Metaphors We Live By*. Chicago: Chicago University Press.
- Landini, G. (1998). *Russell's Hidden Substitutional Theory*. Oxford. Oxford University Press.
- Landini, G. (2011) *Russell*. Oxford: Routledge.
- Losee, J. (1972). *A Historical Introduction to the Philosophy of Science*. Oxford: Oxford University Press.
- Losee, J. (2005). *Theories on the Scrap Heap: Scientists and Philosophers on the Falsification, Rejection and Replacement of Theories*. Pittsburgh: Penn Press.
- Louw, W. E. (1989). Subroutines in the integration of language and literature. In *Teaching English Literature Overseas British Council Documents 131*. Carter, R. (ed.). London: Pergamon Press.
- Louw, W. E. (1991). Classroom concordancing of delexical forms and the case for integrating language and literature. In: *ELR Journal. Vol. 4 Classroom Concordancing*. 151-178. Birmingham University: ELR. (Paper 1 in Appendix).
- Louw, W. E. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In *Text and Technology: In Honour of John Sinclair*. Baker, M., Francis, G., & Tognini-Bonelli, E. (eds.). Amsterdam: John Benjamins. (Paper 2 in Appendix)
- Louw, W. E. (1997). The role of corpora in critical literary appreciation. In *Teaching and Language Corpora*. Wichmann, A., Fligelstone, S., MacEnery, T., & Knowles, G. (eds.). London: Routledge.
- Louw, W. E. (2000). Contextual Prosodic Theory: Bringing semantic prosodies to life. In *Words in Context. A Tribute to John Sinclair on his Retirement*. Heffer, C. & Sauntson, H. (eds.). Birmingham: English Language Research. (CD-ROM) (Paper 4 in Appendix)
- Louw, W. E. (2003). Dressing up waiver: A stochastic-collocational reading of the Truth and Reconciliation Commission (TRC). Harare: mimeo, also available from *Quaderni del CeSLiC: Occasional Papers*
<http://www.lingue.unibo.it/ceslic/e_occ_papers.htm> (Paper 5 in Appendix)
- Louw, W. E. (2004). 'Unravelling the ideological from the authentic'. *Guardian Weekly*. Thursday 20 May 2004. Available on line at:
<http://education.guardian.co.uk/tefl/story/0,,1220377,00html> (accessed August 2006).
- Louw, W. E. (2006). Data-assisted negotiating: Will it produce a new class of negotiator or destroy the ideology of negotiating? *Language Matters: Studies in the Languages of Africa*. 37(2). 183-205. (Paper 6 in Appendix)

- Louw, W. E. (2007a). Truth, literary worlds and devices as collocation. In *Corpora in the foreign language classroom*. Hidalgo, E., Quereda, L., & Santana, J. (eds.), 329-362. Amsterdam: Rodopi. (Paper 7 in Appendix)
- Louw, W. E. (2007b). Collocation as the determinant of verbal art. In *Verbal Art Revisited*. D. Miller & M. Turci (eds.), 149-180. London: Equinox. (Paper 8 in Appendix)
- Louw, W. E. (2008). Consolidating empirical method in data-assisted stylistics: Towards a corpus-attested glossary of literary terms. In *Directions in Empirical Literary Studies. In Honour of Willie van Peer*. [Linguistic Approaches to Literature 5]. Zyngier, S., Bortolussi, M., Chesnokova A., & Auracher, J. (eds.). Amsterdam: John Benjamins. (Paper 9 in Appendix)
- Louw, W. E. (2009). The Analysis and Creation of Humour. In *L'Umore in prospettiva interculturale: Immagini, Aspetti e Linguaggi / Crosscultural Humour : Images, Aspects, and Languages*. (2014). Coloru, O. & Minunno, J. (eds.). Parma, Atelier 65. (Paper 10 in Appendix)
- Louw, W. E. (2010a). Collocation as instrumentation for meaning: a scientific fact. In *Literary Education and Digital Learning: Methods and Technologies for Humanities Studies*. van Peer, W., Zyngier, S. & Viana, V. (eds.). Hershey, PA: IGI Global. (Paper 12 in Appendix)
- Louw, W. E. (2010b). Automating the extraction of literary worlds and their subtexts from the poetry of William Butler Yeats. In *Para por y sobre Luis Quereda*. Falces Sierra M. (ed.). Granada: Granada University Press. (Paper 13 in Appendix)
- Louw, W. E. (2010c) The painting is where the paint is not”: Reflections on the “Bootcamp Debate. *The Bootcamp Discourse and Beyond Special Issue of International Journal of Corpus Linguistics* 15:3. 344-353.
- Louw, W. E. (2011). Interview. In *Perspectives on Corpus Linguistics*. Viana, V., Zyngier, S. & Barnbrook, G. (eds.) Amsterdam: John Benjamins.
- Louw, W. E. (2014). Literary appreciation from the index of first lines: a corpus-based approach. Moscow: MGIMO University Press. (Paper 14 in Appendix)
- Louw, W. E. (2014b). ‘Science without scions: an introduction to the pedagogy of subtextual corpus linguistics’. Paper delivered at AERLINCO, Las Palmas, May 2014. Harare: mimeo.
- Louw, W. E. & Chateau, C. (2010). Semantic prosody for the 21st Century: Are prosodies smoothed in academic contexts? In *Statistical Analysis of Textual Data: Proceedings of the Tenth JADT Conference*. Bolasco S., Chiari I. & Giuliano L. (eds.). 754-764. Rome.
- Louw, W. E. & Milojkovic, M. (2014a). Semantic prosody. In *The Cambridge Handbook of Stylistics*. Stockwell, P. & Whitely, S. (eds.). Cambridge: Cambridge University Press.
- Louw, W. E. & Milojkovic, M. (2014b, in press). *Contextual Prosodic Theory*. Amsterdam: John Benjamins.
- Louw, W. E. & Milojkovic, M. (2014c). ‘Whither the Monitor Corpus?’ Paper presented at the 35th Annual ICAME Conference, Nottingham, May 2014.

- Macdonald, G. F. (ed.) (1979). *Perception and Identity: Essays presented to A. J. Ayer with his Replies to them*. London: Macmillan.
- Magee, B. (1978). *Men of Ideas: Some Creators of Contemporary Philosophy*. Oxford: Oxford University Press.
- Mautner, T. (2000). *The Penguin Dictionary of Philosophy*. Harmondsworth: Penguin Books.
- McEnery, T. & Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- McGinn, M. (1981). The Third Dogma of Empiricism. In *Proceedings of the Aristotelian Society*. 89-101. Aristotelian Society.
- Miller, D. (1980). Science without induction. In *Applications of Inductive Logic. Proceedings of a Conference at The Queen's College, Oxford, 21-24 August, 1978*. Cohen, L. J. & Hesse, M. (eds.). Oxford: Clarendon Press.
- Miller, D. (1983). *A Pocket Popper*. Glasgow: Fontana.
- Milojkovic, M. (2011). Quenched light, or seeing through a glass darkly- A collocation-based view of Larkin's atheism and depression. In *Belgrade English Language and Literature Studies*. Vol. 3. Belgrade: University Press.
- Milojkovic, M. (2013). Is corpus stylistics bent on self-improvement? The role of reference corpora 20 years after the advent of semantic prosody. In *Journal of Literary Semantics*. Vol. 42. Issue 1 June 2013. Amsterdam: De Gruyter.
- Milojkovic, M. (2014). Measuring banality. Paper presented at the Sixth Spanish Conference in Corpus Linguistics. Las Palmas, May 2014.
- Monk, R. (2005). *How to Read Wittgenstein*. London: Granta Press.
- Morick, H. (1972). *Challenges to Empiricism*. Cambridge, MA: Hackett.
- Ong, W. J. (1982|2002). *Orality and Literacy: The Technologizing of the Word*. 2nd edition. New York: Routledge
- Ongley, J. & Carey, R (2013). *Russell: A guide for the perplexed*. London: Bloomsbury.
- Palmer, F. R. (ed.) (1968). *Selected Papers of J. R. Firth 1952-1959*. London: Longmans, Green and Co. Ltd.
- Palmer, F. R. (1981). *Semantics: A New Outline*. 2nd Edition. Cambridge: Cambridge University Press.
- Palmer, H. E. (1933). *Second Interim Report on English Collocations*. Institute for Research in English Teaching. Tokyo: Kaitakusha.
- Pears, D. F. (ed.) (1985). *Bertrand Russell: A Collection of Critical Essays*. New York: Anchor Books.
- Pears, D. F. (1987). *The False Prison: A Study of the Development of Wittgenstein's Philosophy. Volume 1*. Oxford: Oxford University Press.
- Philip, G. (2011). *Colouring Meaning. Collocation and connotation in figurative language*. Amsterdam: John Benjamins.
- Quine, W. V. O. (1981). On the very idea of a third dogma. In *Theories and Things*. Cambridge, Massachusetts: Harvard University Press.

- Renouf, A. (2007). Corpus development 25 years on: from super-corpus to cybercorpus. In *Corpus Linguistics 25 Years On*. Facchinetti R. (ed.) 27-49. Amsterdam: Rodopi.
- Robinson, D. & Groves, J. (2002). *Introducing Bertrand Russell*. New York: Totem Books.
- Ross, J. J. (1970) *The Appeal to the Given : a Study in Epistemology*. London: Allen and Unwin.
- Russell, B. (1912|2001). *The Problems of Philosophy*. Oxford: Oxford University Press.
- Russell, B. (1914|2009). *Our Knowledge of the External World. As a field for scientific method in philosophy*. London: Routledge
- Russell, B. (1918). The Philosophy of Logical Atomism. *The Monist*. Vol. 28. No. 4. 495-527.
- Russell, B. (1946). *History of Western Philosophy*. London: George Allen & Unwin.
- Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. London: Routledge.
- Russell, B. (1960). *Bertrand Russell Speaks his Mind*. New York: World Publishing Company.
- Russell, B. (1967). *The Autobiography of Bertrand Russell: 1872-1914*. Boston: Little, Brown and Company.
- Sampson, G. & McCarthy, D. (2004). *Corpus Linguistics: Readings in a Widening Discipline*. London: Continuum.
- Schilpp, P. A. (1963). *The Philosophy of Rudolf Carnap*. La Salle (Ill): Open Court.
- Schilpp, P. A. (1974). *The Philosophy of Karl Popper*. La Salle (Ill): Open Court.
- Scott, M. & Johns, T. (1993). *MicroConcord*. <<http://lexically.net/software/index.htm>>
- Scott-Kakures, D. Castagnetto, S., Benson, H., & Hurley, P. (1993). *The History of Philosophy*. Houston, Texas: Harper Resource.
- Sebeok, T. (1960). *Style in Language*. Massachusetts: Massachusetts Institute of Technology Press.
- Simpson, P. (2014). *Stylistics: A Resource Book for Students*. Second edition. London: Routledge.
- Sinclair, J. M. (1987). *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. Glasgow: HarperCollins.
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. (2003). *Reading Concordances*. London: Longman.
- Sinclair, J. M. (2004). *Trust the Text*. London: Routledge.
- Sinclair, J., Jones, S. & Daley, R. (1970). In *English Collocation Studies: The OSTI Report*. Krishnamurthy, R. (ed.). London: Continuum.
- Sinclair, J. M., Hanks, P., Fox, G., Moon, R & Stock, P. (1987|1993). *The Collins COBUILD English Language Dictionary*. London: HarperCollins.

- Taylor, B. (2007). In *The Shorter Routledge Encyclopaedia of Philosophy*. Craig, E. (ed.). London: Routledge.
- Teubert, W. (2010). *Meaning, Discourse and Society*. Cambridge: Cambridge University Press.
- Tibballs, G. (ed.) (2011). *The Mammoth Book of Really Silly Jokes*. London: Constable and Robinson.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Urmson, J. O. (ed.) (1960). *The Concise Encyclopaedia of Western Philosophy and Philosophers*. London: Hutchinson & Co. Ltd.
- Weber, J. J. (1996). *The Stylistics Reader*. London: Edward Arnold.
- Whitsitt, S. (2005). A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics*. 10:3. 283-305. Amsterdam: John Benjamins.
- Wichmann, A., Fligelstone, S., MacEnery, T., & Knowles, G. (eds.) (1997). *Teaching and Language Corpora*. London: Longman.
- Williams, G. (2008). The Good Lord and His Works: A corpus-driven study of collocational resonance. In *Phraseology: An Interdisciplinary Perspective*. Granger, S. & Meunier, F. (eds.). 159-173. Amsterdam: John Benjamins.
- Williams, G. & Millon, C. (2009). The General and the Specific: Collocational resonance of scientific language. In *Proceedings of the Corpus Linguistics Conference CL2009*. Mahlberg, M., González-Díaz, V. & Smith, C. (eds.). Liverpool: University of Liverpool.
<http://ucrel.lancs.ac.uk/publications/CL2009/129_FullPaper.doc>
- Williams, H. (2007). *Wrestling in the Mud: New and Selected Poems*. Cinnamon Press.
- Wittgenstein, L. (1913). Notes on Logic. Appendix I. In *Notebooks 1914-1916*. Transl. Anscombe, G. E. M. (1969). Oxford: Blackwell Publishing.
- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. Transl. Pears, D. F. & McGuinness, D. F. (1960). London: Routledge and Kegan Paul.
- Wittgenstein, L. (1929). Some remarks on logical form. *Proceedings of the Aristotelian Society*. 162-171. London.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell Publishing.
- Wittgenstein, L. (1958). *The Blue and Brown Books: Preliminary Studies for the 'Philosophical Investigations'*. Oxford: Blackwell Publishing.
- Wittgenstein, L. (1969). *On certainty*. Transl. Anscombe, G. E. M. & Wright, G. H. (Eds.). Oxford: Blackwell Publishing.
- Wyatt, W. (1960). *Bertrand Russell Speaks his Mind*. New York: World Publishing.