Technical Report

UCAM-CL-TR-223
ISSN 1476-2986

Number 223

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Shallow processing and automatic summarising: a first study

Philip Gladwin, Stephen Pulman,
Karen Spärck Jones

May 1991

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

https://www.cl.cam.ac.uk/

Technical reports published by the University of Cambridge
Computer Laboratory are freely available via the Internet:

https://www.cl.cam.ac.uk/techreports/

ISSN 1476-2986

Abstract

This report describes a study of ten simple texts, investigating various discourse phenomena to see how they might be exploited, in shallow text processing, for summarising purposes. The processing involved was a simulation of automatic analysis which is in principle within reach of the state of the art. Each text was treated by a version of Sidner's focusing algorithm. The products of this were fed into subsidiary stages of analysis to provide an assessment of the activity of the various discourse entities within each text. A concurrent process examined the occurrence of orthographically identical noun phrase forms. Appendices give the ten texts, a complete specification of the version of the focusing algorithm in use, and the full experimental results. These suggest, especially when the brevity of the test texts is taken into account, that the type of information given by focusing has potential but limited value for summarising.

Background

Stephen Pulman and Karen Spärck Jones

This paper reports an experiment, in several senses, in computational linguistics. Here we give some background to the experiment and say what it was trying to achieve.

Our long term concern is the automatic summarisation of texts. This is clearly a very ambitious goal and it can be argued to presuppose solutions to all of the outstanding problems of Natural Language Processing (NLP), knowledge representation and reasoning, at least if anything approaching human performance is sought.

The basis for such an argument can be derived from many psycholinguistic findings.[1] These findings suggest that when people understand texts, they seem to integrate the content of the text into one coherent structure. This structure may include elements not actually present in the text but inferrable on the basis of general or domain specific knowledge. The level of integration is such that people, in subsequent memory tests, are unable to distinguish between propositions actually present in the text and those that were inferred. In the stored representation the propositions have equal status. Literal meanings, and deductions from them, are integrated into one homogeneous representation.

If when summarising a text, one first understands it to this level of integration, it therefore seems likely that inferred elements, which may be just as important a part of the 'content' of the text, will influence the form of the summary. In an informal experiment, we did indeed find several clear instances of this. Six people were asked to produce as short a summary as possible of the 'John and the rabbit' text given in Appendix 1. One of these summaries, for example, stated that "John was sad because the rabbit had disappeared", when the source text in fact only says "The rabbit had disappeared. He (John) went sadly home." That being sad is causally dependent on losing the animal one is watching is of course a plausible inference, which a reader might draw as a way of justifying the placing together of these two events, causally linking them. It is not, however, an explicit part of the input text.

If an automatic system is going to be able to make summaries which are influenced by this type of commonsense reasoning, it follows that the system must indeed be able to carry out processing and inference to this deep level.

However, the fact that to approach human performance in this task might require all of these very difficult problems to have been solved does not entail that no useful level of summary can be obtained at the current state of the art. This is one broad hypothesis that we are currently exploring: that the current or near future state of the art in NLP can provide a level of processing of texts which will enable summaries to be produced that are accurate enough to be useful, and, in contrast to approaches like that exemplified by Cullingford's SAM and DeJong's FRUMP, [2] are not dependent on prespecified and extensive knowledge about the world (although of course nothing would prevent them

---

[1] Bransford, J. D., Barclay, J. R., and Franks, J. J. (1972) *Sentence Memory: a constructive versus interpretive approach*, Cognitive Psychology, 3, 193-209.

[2] Cullingford, R. E. (1981), SAM, in *Readings in Natural Language Processing*, B. Grosz,K. Sparck Jones, and B. Webber, eds., pp627-49, Los Altos: Morgan Kaufmann, 1986 ; De Jong, G. M. (1979), *Skimming newspaper stories by computer*, Dept. of Computer Science, Yale University, Research Report 104

from also using such knowledge where possible.)

The shallowest approaches to summarising depend on word or phrase spotting and statistical techniques. The particular kinds of approach we are currently studying can be called shallow in relying on linguistic as opposed to world knowledge; but they are nevertheless based on automated analysis which uses a rich set of grammatical descriptions exploiting syntactic and some semantic and pragmatic information. It is now possible to process text both to capture more subtle and complex characteristics of individual sentences and to take more account of relationships between sentences, using a purely linguistic apparatus, than many earlier systems did. Linguistic structures of various types give a great deal of information about how an author saw the relative importance of the entities and relations described in the sentences of their text. This information, given a sufficiently refined linguistic description, can in principle be computed without extensive contextual reasoning depending on discourse meaning and relevant world knowledge (although there will always be a residue of 'hard cases' that may need such reasoning).

To give a simple illustration of the kind of essentially linguistic information processing that is being referred to here, consider the two sequences of sentences:

> Presenting his first budget next Tuesday will be the new chancellor, Norman Lamont. He is widely expected to cut interest rates by 1%.
>
> The new chancellor, Norman Lamont, will be presenting his first budget next Tuesday. It is widely expected to cut interest rates by 1%.

Both sequences give approximately the same information overall. But the pair of sentences in the first sequence is pretty clearly primarily 'about' Norman Lamont, and only secondarily about the budget. This effect is achieved by the syntactic structure of the first sentence, reinforced by use of the pronoun in the second. In the second pair, this perspective is less obviously the case. Here the first sentence gives both Chancellor and budget about equal weighting, while the second continues about the budget.

Factors like choice of main clause structure, and pronoun 'focusing' (in the sense of Sidner [3]), therefore give some clue as to what the author thought the important discourse entities to which their text referred were. These are factors which, while still difficult to compute exhaustively, can be characterised as requiring shallow processing (see e.g. Carter [4]) as opposed to the 'deep' processing needed to fully determine lexical senses, identify the substantive entities to which descriptions refer, delimit and relate events, extract implicit causal relations, and so forth, which full-blown NLP demands. To the extent to which these shallow processing factors are good indicators not only of what individual sentences are about, but are therefore also better indicators than exclusively statistical ones of what a text as a whole is about, it should be possible to use them to produce summaries that are at least better than random, without waiting for the solution of all the difficult NLP problems just mentioned.

Thus we arrive at a narrower instance of the broad hypothesis that purely linguistic processing can be a fair base for summarising. The particular instance is that a notion of

[3] Sidner, C., (1983) *Focusing in the comprehension of definite anaphora*, in Grosz et al (eds), Readings in natural language processing, Los Altos CA: Morgan Kaufmann, 1986.

[4] Carter, D. *Interpreting anaphors in natural language texts*, Chichester: Ellis Horwood, 1987.

focus based on structural properties of sentences of the type that are, in small scale systems at least, within the current state of the art, can be used to determine those discourse entities, and specifically those entities represented by the noun phrase (NP) constituents of text, that are in focus, and hence to identify those discourse entities that the text is 'about'. The important point is that the focus mechanism, working with the syntactic information for NPs provided by the use of articles like "the" or of pronouns, will serve to locate the discourse entities the text is 'about' although the specific text expressions used as NPs to refer to them can vary ("A tall man ... the man ... he..."). Of course, when the text is more than a few sentences long, there may be many discourse entities in focus at different times, so some further criteria for selecting the most salient are required. The text will, moreover, predicate various things of the discourse entities that it is 'about' and these predications also need to be taken account of for the production of a summary. But knowing what discourse entities are salient is a good first step towards discovering what the salient relationships between these entities are.

This narrow hypothesis is one of those investigated in the work described in this report. For a series of texts, the entities represented by NPs that a Sidner-like algorithm would characterise as being focused were identified. The hypothesis was that intuitive judgements of what the texts were 'about', would correlate more highly with focused discourse entities than with non-focused ones. Further, counting the number of times different entities were in focus would identify the most salient entities with the best claim to be the entities the texts were most strongly 'about'. The test results presented in the report show that there is something to the hypothesis. The evaluation of the results was, however, done only in an informal and limited way, and the variations among human summaries of the same text, as found with those, mentioned earlier, produced for some of the test texts, emphasise the fact that the idea of evaluating automatic summaries by comparisons with assumed proper human ones is itself problematic.

Of course, texts are 'about' more than just some of the entities to which NPs in them refer. At the very least, the entities are related in various ways by different predicates. This gives rise to a second shallow processing hypothesis: that there will be more predications involving the entities that the texts are 'about', than others. Thus ranking entities for a text according to the number of different predications they are involved in may give a measure of 'aboutness' that is accurate enough to be useful. Again, not all of these predications may be of relevance to the construction of a summary, but knowing which entities may be important because they figure in more predications than others is at least a start in this direction. This hypothesis that entities figuring in the most predications are the ones that the text is 'about' is also tested in what follows.

As a 'neutral' condition, one can compare either of these measures, which deal with discourse entities and hence text content, against a count based explicitly on the surface text data represented by the actual NPs themselves. It is difficult to know on a priori grounds what to expect from such a comparison. One might expect these surface counts to differ from the focus- based entity counts since they cannot take account of the variations in the form of references to the same entity, especially that represented by the use of pronouns, that a focus mechanism is designed to allow for. With longer texts, however, the net effect might be much the same.

The shallow linguistic techniques examined in the work described here are clearly of very limited value when used individually, and appear to be especially problematic for longer

text when the information they do not pick up may have more cumulative effect. It is evident that to take a linguistically based approach further it is necessary to combine different types of information and also, for longer text, to ask whether large scale structure as determined by a rhetorical or discourse grammar has any role to play. It is at the same time necessary to evaluate purely linguistic strategies by comparison not only with statistical ones, but also with strategies, for instance those exploiting scripts, which involve world knowledge.

The report appendices give full details of the test text processing: they are not essential reading, but are included as the data they present on focus behaviour, etc., may be of more general interest than in the summarising context alone.

A Dissection of 10 Short Texts

Philip Gladwin

This section describes investigations of some simple or obvious discourse properties undertaken in an initial study of discourse phenomena potentially relevant to summarising. Specifically, the study examined ten texts from several different points of view in an attempt to establish a relationship between the discourse entities present in a text and an intuitive view of what the text was in some way 'about'. Three of these texts were artificial and seven were taken from the children's weekly newspaper, 'The Indy'. The newspaper stories were occasionally simplified, but never dramatically; on the occasions that texts were altered it was usually to break up particularly complicated clausal noun phrases. All the texts used can be found in Appendix A, with a categorisation of their type. Six different methods of analysis were employed, some related to one another, some independent. They were all applied by hand in a "mechanical simulation", assuming some sort of conventional logical term analysis leading to a predicate argument structure. The different approaches followed were:

- Apply Sidner's Focus Algorithm (Appendix B and C)
- Tabulate the resulting statistics (Appendix D)
- Tabulate the frequency of every NP within each text (Appendix E)
- Tabulate the verb entity predicate/argument list (Appendix F)
- Produce summaries for each text according to the statistics (Appendix G)
- Draw an entity geography chart for each text (Appendix H)

Methods

The algorithm used was a simplified version of Sidner's focusing algorithm, which here is taken be based on the twin, mutually supportive, processes of identifying local discourse focus and resolving all anaphors. [5] The algorithm deals in Discourse Entities (DEs), of which Sidner recognises two kinds; those based on NPs, and those based on VPs. The version used here is described fully in Appendix B; but briefly, each sentence in the text was broken down, as far as possible, into single verb clauses, and for each the following three data structures were identified:

[5] Sidner 1983 Focusing in the comprehension of definite anaphor, in "Readings In Natural Language Processing", eds B.J. Grosz *et al*

- The Discourse Entities in focus after that clause, (the Current Focus)

- All the other Discourse Entities from that clause (The Alternate Focus List). The contents of this list, along with the Current Focus, compete for Current Focus in the next clause.

- The Discourse Entities that have already been Current Focus in this text (the Focus Stack)

The algorithm itself has two halves, both relying on a hypothetical parser to deliver an identification of the various case fillers of the clause, which are here taken (following Sidner) to be: the Verb Complement Theme, the Theme, the Instrument (aka the Goal or the Locative), and the Agent. Thus the kind of sentence analysis that is simulated is one which would deliver interpretations as follows:

Ron played the piano vigorously.

VERB:	played vigorously
THEME:	piano
AGENT:	Ron

His landlady hobbled with her stick to fetch her revolver

VERB:	hobbled
VERB COMPLEMENT THEME:	revolver
INSTRUMENT:	stick
AGENT:	landlady

It was the original intention to use SRI's Core Language Engine as a parser, and to extract the information Sidner's algorithm requires. This was found to be impractical. Sidner's method requires information that is not generated by the CLE, and many of the sentences in the experiment were beyond its syntactical coverage. No automated parsing took place - the DEs were picked from the text by hand. Section A of the algorithm builds up a Default List of candidate foci expected from the components of the first clause in the text, and suggests a focus. This focus is checked against the second clause of the discourse to see if it can be sustained. If so it becomes the Current Focus. Backtracking can take place when a pronoun which does not co-specify with the expected focus is used in the second clause of the discourse, and no anaphor is used to co-specify with the expected focus. Sidner recommends that this checking should be carried out by an independent inference component. Backtracking should select from the alternatives in the Default Focus list. The remaining clauses in the text are then processed by Section B of the algorithm, and the three data structures maintained by simple rules.

There are several complications to the process of choosing the Current Focus of a particular sentence, not least of which is the question of exactly what it means to declare that a DE is "in focus". Intuitively a DE can be seen to be in focus if it is the subject matter of the current region of the discourse: it is what, in a broad sense, is being currently talked about.

Of course, short digressions are permissable from every subject - this does not mean that the focus has to shift onto a new DE in every new clause. Indeed, Sidner stresses that the Current Focus has inertia, usually only changing when our attention is forced onto a new DE by some syntactic mechanism. Her method identifies anaphoric reference as the main means of forcing this shift in attention. The mechanism she describes means that her algorithm only responds to anaphors, and in an order of precedence which tends to favour the direct object of the verb. The Current Focus of the first clause of any discourse is confirmed by anaphor matching or inference and backtracking if necessary, the Current Focus of any other clause in the text is not subject to the same backtracking procedure - once picked it is fixed. In this experiment I assumed that pronominal anaphors were always resolved correctly and I gave the mechanism help if it looked like it needed it. Nominal or implicit anaphors were rather more tricky and it was not always possible to be sure that any assumption was valid. There is further discussion of this below.

Consider the following simple text, and its treatment by the algorithm:

Ron walked along the road. He carried some red flowers. They were for Cath.

The algorithm would produce the following data structures filled with Discourse Entities:

1. Ron walked along the road
   Current Focus: Ron
   Alternate Focus List: [road, walk-verb]
   Focus Stack: []
2. He carried some red flowers
   Current Focus: Ron
   Alternate Focus List: [flowers, carry-verb]
   Focus Stack: []
3. They were for Cath.
   Current Focus: flowers
   Alternate Focus List: [Cath, be-verb]
   Focus Stack: [Ron]

There is no formal definition of a DE, but it seems reasonable to claim that sentence 2 could be composed of three DEs. Two of these coincide roughly with the the two NPs "he", and "some red flowers". The third, tagging the verb "carried", is slightly more complex, and involves an idea of the verb and the DEs that fill its case roles taken together as a whole.

Seven of the texts used in the experiment were based on actual newspaper stories, and were very much more complicated than this example. This effectively prohibited any machine-based implementation of the algorithm. The problems (other than that of automating the parsing of syntactically complex clauses) lay mainly in the area of identifying complex noun phrases as discrete discourse entities. So "John", "car chase", and "their attempt to be the first men to walk unaided to the North Pole" are all Discourse Entities, regardless of the fact the latter two can be decomposed into further DEs. This problem is not properly addressed in this study. For each complex NP it would be necessary to have some sort of method to decide whether the NP or VP could give rise to more free-standing DEs. One, perhaps questionable, approach, which was followed in the NP counting stage

only, was to allow this decision to be influenced by the later occurrence of, or anaphoric reference to, a free-standing fragment of the NP/VP. For example, "car chase", for the benefit of NP counting, is treated as containing three NPs (car, car chase, and chase). Conversely, "cling film" and "sharp boxes" were treated as forming only one NP each. I have attempted to distinguish between true compound nouns and noun+noun combinations. This is obviously a tricky issue, but it seemed plausible that, when "car chase" is referenced elsewhere as "the chase", and in other places there is implicit anaphoric reference to the car involved, there are two discrete nouns being primed by the text. See Page 45 for a limited discussion of the NP "Captain Stanley Lord".

Sidner discusses the possibility of incorporating forms of anaphoric reference other than simple pronominal co-reference, but does not go very far in her treatment. As soon as the real texts were looked at this became a problem. This is a real problem. For example, the text 'Cracked Up Kids' contains the clauses:

Young teenage boys are now being used as runners by inner-city crack dealers.

The crack is manufactured in abandoned flats...

The reference to 'crack' in the second clause is anaphoric, and its comprehension would depend on a sophisticated inference component, which, of course, was not in place. It gets even more complicated. Consider Text 4, which starts as follows:

Two bank robbers were shot dead last Friday

when a high-speed car chase ended in a gun battle in the city centre.

IRA involvement has not been ruled out by detectives.

The drama began shortly after 10am...

The anaphor 'the drama' in clause 4 refers back to a combination of the shooting of two bank robbers, a car-chase, a gun battle, the possibility of IRA involvement, and the fact that it all happened in Dublin city centre. Sidner's algorithm has some ideas about this type of implicit specification, but I did not find them sufficiently developed to be of use in this kind of situation.

Other problems included the fact that there is no mention of the possibility of total novelty within a text. In the case of a clause containing only completely new material there is nothing for the algorithm to work on, and the focus remains - possibly incorrectly - with the DE in focus before the new clause was encountered.[6] It is difficult to suggest a remedy for this, as it would be necessary to have advanced and well-defined ideas regarding implicit co-specification of earlier NPs by the new, often only superficially or loosely connected new NPs. Although some provision for allowing a return to the Expected Focus Algorithm [7] would surely generate results that followed the focus movement, in this experiment I have simply allowed the algorithm to run as it is specified.

Focus Sets were a feature of Sidner's original algorithm that I didn't use. There were relatively few of them in the texts.

[6] See, for example, Text 5 Clause 5.

[7] cf Appendix B

Overview of Appendices

Appendix A gives the ten texts used in this study. Appendix B describes the version of Sidner's algorithm used. Apendix C gives the complete record, sentence by sentence, of the focus states established by the algorithm when processing the test texts.

In Appendix D, we analyse how many times each entity occurred as a Current Focus, and how many times it appeared as a candidate for Current Focus. The count thus included resolved anaphors (so "John" and a co-referential "he" would score 2). The intuition was that the set of focused objects in a text would resemble the set of the things that the text could be said to be 'about'. This intuition proved to have some foundation, but in general apppealing results could only be found in text thickly bound together with anaphors. It is likely that in the real world, and especially with the sort of condensed style commonly found in newspaper stories, that a text will contain a sequence of at least two or three clauses which don't contain tractable real anaphoric reference. The focus of the discourse shifts within these clauses, but there is no way for Sidner's algorithm, based on anaphoric reference, to detect this. See for example 'Cracked Up Kids', and how in, clauses 9 and following, the Current Focus remains boys-1 while the subject of the text changes. A similar effect occurs throughout 'Animal Medicine'.

Appendix E tabulates the frequency of every NP within each text. This is a list of all the noun phrases in the text, and a count of their frequency. This stage was kept deliberately simple, carried out on a purely orthographic basis. All the nouns were counted; each different version of the same noun, for example the plural and singular cases, having its own count. There was no attempt to do any anaphor resolution, this being simply a count of how many times, for example, the word "rabbit", occurred in the text. This tested an extremely simple minded version of the hypothesis that the things that are most frequently mentioned are in the most literal sense what the text is 'about'. Some interesting results were produced.

Appendix F is an attempt at a shallow processing notion of 'semantic density'. Every clause in the analysis has a verb at its centre; this section lists them, complete with their arguments (DEs). This gives an idea of which elements interact with which elements, and where. Intuitively the most interactive, or connected, things may prove to be at the centre of the subject matter. After each text is treated the results are summarised in a table that lists the most highly active entities in order. (This material also contributes to Appendix G.)

In Appendix G, the results from each of the preceding three subsections are laid out in an attempt to compare the three methods as a means of identifying what is important in a text. Specifically, applying the obvious idea that the most frequent items (however "items" is defined) are most important, Appendix G gives the most frequent foci, most frequent NPs and most active entities respectively for each text, as they might figure in a notional summary statement "This text is about X".

Appendix H represents an attempt to graphically illustrate the places in the text that referenced a particular entity, and is included for completeness' sake. Each graph is composed of the following elements:

X-axis : the verb/sentence number
Y-axis : the discourse entity

Points : a mention of a particular discourse entity in a particular clause

Evaluation and Conclusions

The analyses produced some interesting results, and the comparison between the three methods of identifying important text information, namely Focus frequency, NP frequency, or entity activity, illustrated in action in Appendix G, is instructive. It seems possible to extract, by one route or another, and in a way that could conceivably be automated, the main players from short texts. It therefore seems to be possible to find out who, or what objects, the texts are about. However, the critical question of 'What are these people, or objects, *doing* to each other?', remains unattacked. This is due to the fact that the analysis is entitity-based rather than action-based. We need ways of categorising VPs into types of action; generalising them, classifying them, identifying causalities, etc.

The specific points to be made refer, first, to the operation of the Focus Algorithm itself as a necessary and important contribution to discourse interpretation, and second, to the information derived from its application concerning important text elements in the focus frequency counts of Appendices D and G. There are also comments on the alternative methods of identifying important elements (the active entity counts of Appendices F and G, which rely on the Focus and Anaphor resolution results, and the simple NP counts in Appendices E and G, which do not exploit resolution but simply use the text directly), and on the comparison between the results for "summarising" given for the three methods in Appendix G.

The biggest problem with the Focus Algorithm lies in properly discriminating and identifying the DEs. There has been a discussion of this in Section 4; for now it will do to note that as long as this area remains ill-defined the algorithm will only give approximate results. An associate, but perhaps more slightly more soluble problem lies in the algorithm's inability to recognise novelty of subject matter, (intuitively coincident with the introduction of novel DEs). This issue is avoidable in very short texts, such as are used in Sidner's paper, but in texts of any length the issue of novelty must be addressed. If a mechanism had existed for deciding that a DE was in fact new to the text, and the Expected Focus Algorithm had been employed at such points, then it is certain that the Focus Algorithm sections of Appendix G would have shown a higher success rate.

More loosely, there seems to be a fading of the algorithm towards the end of texts. In several cases the method produces less and less appealing results as it progresses - an extreme version of which can be seen in the Animal Medicine text, where the only Current Focus the algorithm manages to produce is the subject of the very first clause of the text. This would imply that, given the anaphor-seeking basis of the algorithm, anaphors become less common towards the end of texts. In fact, and this can be observed on the charts in Appendix H, the process of backwards reference often becomes both more frequent and deeper. That is to say that the reference is not to DEs mentioned in the last clause, but to DEs from the early parts of the text. Often therefore the links are too stretched for the normal methods of anaphoric reference to be sufficient, and the referenced DE must be reintroduced as a new DE. This involves a degree of renaming, and there lies the confusion. The algorithm is not nearly good enough to handle this degree of implicit reference.

All three of the methods suffer from the critical problem of recognising the same concept however it is mentioned; but they fail in different ways. For example, in text 9 neither of the summaries produced according to the list of active entities or the list of most frequent NPs catch the fact that the separate mentions of different, abstract, 'dog' entities activate the concept of 'dog' several times. There seem to be two equally problematic alternatives at the moment:

- categorise by kind, giving every abstract mention of a dog the same token, dog-57, (which leads rapidly to intuitively useless unions, such as "little dogs-57 are scared of big dogs-57.") This is what happens in the NP counting method.

- or the method followed here of giving every abstract mention of dogs a new DE label, in which case the entity activation model fails to reflect what is really going on. This is what happens in both the active entity count and the focus resolution method.

What is needed is a mapping from the *entity* activation table to a *concept* activation table. This would of course require rather considerable inferencing ability.

It is dangerous to draw conclusions from this small sample, and in particular without any sort of well defined idea of what constitues a good summary. In general however all the methods have considerable success in drawing out the main DEs in each text. They do all also frequently highlight non-central DEs, or otherwise fail by omission, or by lack of discrimination. The text type could be an important factor - the Focusing Algorithm for example seems as though it is best at delivering the main players in the texts with some Narrative element to it. Of course, it is the case that the central DEs in Narrative text would tend to persist over more than a few sentences - and all three of these counting methods are designed to respond to different types of this persistence. While any of these three methods could have some use in crude text-indexing operations, work needs to be carried out, as mentioned earlier, on identifying what these DEs are actually doing to each other before any real summary could be approached. What is required, therefore, are ways of identifying and comparing complex concepts with internal structural relations at different levels of complexity. This could include, for instance, a consideration of the real character of events.

In general the results produced by the investigations described here were interesting, but hardly conclusive. Each of the three methods had their comparative successes, and failures. It would be useful to carry out this survey on a larger body of texts, and to try and pinpoint some reasons for the differences in success rates. The new CLARE release seems to have some valuable tools for automating the NP counting phase, but a more rigorous idea of anaphor resolution should be implemented before Sidner's method could be said to be properly applied. An alternative approach to tracking focus is detailed in "A Centering Approach to Pronouns" (Brennan, Friedman, Pollard, in ACL 1987). This seems more elegant, and if these experiments were to be repeated, to be worth attention.

Appendix A: The Ten Texts

The test texts are all simple enough to be subject at least in principle to current NLP technology. The first three are artificial texts constructed for another purpose, with very simple sentence forms and a plain style. The last seven are real, but from a source with a straightforward style. The categories are obvious, intuitive ones.

The following three texts are artificial.

0.1 John and the Rabbit

TEXT TYPE: Narrative.

There was a little boy. His name was John. John woke early once. The sun was shining. John dressed quickly. He went into the garden. The birds were singing. John saw a rabbit. The rabbit hopped across the lawn. John watched it. It went into the field. John went carefully after it. It ate some grass. It ate a dandelion. John crept towards it. The rabbit saw John. It hopped away. John followed quietly. He walked slowly. The rabbit stopped. It ate another dandelion. John crept towards the rabbit. The rabbit was frightened. It fled. John ran after it. The rabbit came to the hedge. It ran into a burrow. John was sorry. The rabbit had disappeared. John was sad. He returned to the house.

0.2 Destroyers

TEXT TYPE: Description

Destroyers are the important ships in the fleet. They are armoured ships. They have guns and torpedoes. The American fleet has many destroyers. Some are in the Pacific. Some are in the Atlantic. Their guns fire big shells. Their range is ten miles. Their torpedoes range over five miles. The Russians have many destroyers. Most are in the Atlantic. Some are in the Pacific. Their guns are very accurate. They fire huge shells. Their range is ten miles. The Russians have very powerful torpedoes. They range over ten miles. The Russians have eighty destroyers. The Americans have a hundred destroyers. The Russian ships are better. They are faster than the American ships. They have the best weapons.

0.3 Biographies

TEXT TYPE: Argument

Biographies are the best books. They are about real things. They tell a true story. They

are about particular people. They have lessons for their readers. Smiles wrote biographies. They are about successful workers. These biographies inspire their readers. Their readers imitate these workers. History books are about real things. They tell true stories. They are about states. They are not about particular people. They do not inspire their readers. Novels tell a story. They are about particular characters. They are not about real things. They are dangerous. Biographies are valuable books. Histories are useful. Novels are deceitful books. Give children biographies. Give them histories. Do not give them novels.

The next seven texts are taken from the children's weekly newspaper "The Indy". The original dates of publication appear after the title. As noted in the body of the report, the texts have sometimes been simplified by, for example, breaking up complex noun phrases.

0.4 Dublin 12.7.90

TEXT TYPE: Report

Two bank robbers were shot dead last Friday when a high-speed car chase ended in a gun battle in the city centre. IRA involvement has not been ruled out by detectives. The drama began shortly after 10am when the men held up a bank 5 miles west of Dublin. They made off with 2,500 pounds but their car was spotted by armed detectives on the edge of the city and a high-speed car chase followed. The Garda car came under repeated fire from the raiders. The chase ended when the robbers tried to smash through a second Garda car that was blocking the road. A prolonged shoot-out followed as more armed Gardai arrived. The two men inside were reported dead on arrival in hospital. Gardai could not say who fired the first shot.

0.5 The Beautiful Game 12.7.90

TEXT TYPE: Argument

To become fully-developed human beings we should cultivate our minds and our bodies. Sport provides a unique opportunity to do this and it therefore ranks as one of the most important of all human activies. Competition is essential. It encourages people to develop their abilities, and to appreciate those of others. People naturally admire individual sportsmen and women and develop loyalties to them. The feeling of being represented in a competition by an individual or group develops our own sense of regional or national identity. Competitions allow us to give vent to natural feelings of competition which, without the safety-valve of sport, could lead to large-scale violent confrontations.

0.6 Titanic 19.7.90

Report

The Captain of the Californian, who was accused of failing to aid the stricken liner Titanic, is to have his conduct re-examined 78 years after the tragedy. The Titanic sank with the loss of 1,503 lives after hitting an ice-berg during her maiden voyage in 1912. The new inquiry follows a 30-year old campaign to clear the name of Captain Stanley Lord, Master of the Californian. The 1912 inquiry accused Captain Lord of not acting, although his was the nearest ship. The inquiry found that the Californian was less than 10 miles from the Titanic, but she ignored distress rockets and had her radio switched off. Captain Lord insisted he was 19 miles away, and new evidence on the location of the Titanic wreck could confirm that claim.

0.7 North Pole 3.5.90

TEXT TYPE: Report

The British Explorers Sir Ranulph Fiennes and Dr Michael Stroud were forced to abandon their attempt to be the first men to walk unassisted to the North Pole - with only 90 miles to go. Their North Pole expedition was halted by stretches of open water 47 days and 439 miles after setting out. Fiennes and Stroud have still scored a remarkable achievement. Their record-breaking trip has so far raised over 400,000 pounds for Multiple Sclerosis Research. Several times they nearly lost their lives. Both once fell through the ice and had to be saved from the water. In pain, and fatigued, eventually the shortage of food supplies defeated them. This and the melting ice forced them to summon help.

0.8 Cracked Up Kids 5.4.90

TEXT TYPE: Description

Young teenage boys are now being used as runners by inner-city crack dealers. The crack is manufactured in abandoned flats, often heavily fortified, and it is then given to the boys who sell it on the walkways and the street. The drug is wrapped in cling film and they carry it in their mouths, so they can swallow it if police approach them. More and more young people are being exposed to the drug world. A survey of ten Liverpool primary schools revealed that pupils as young as four were being approached by dealers. Liverpool schools have also been warned of the dangers of pupils contracting the AIDS virus and hepatitis from discarded hypodermic syringes and needles. Warning posters and 'sharp boxes' need to be provided for disposing of such equipment.

0.9 Animal Medicine 5.4.90

TEXT TYPE: Description

The ancient treatment of acupuncture is coming to the veterinary surgery for pet dogs and cats. Acupuncture is used to treat chronic pain. Dogs' slipped discs, severe arthritic pain and muscle and tendon strains are now helped with this ancient art. For veterinary medicine is entering a new stage, and acupuncture together with successful hip replacements for crippled dogs and dental fillings is becoming common. The dentist's drill can keep dogs' and cats' teeth in healthy and sparkling condition. Or regular brushing with liver flavoured toothpaste can prevent decay altogether.

0.10 Sunbed Salesman 5.4.90

TEXT TYPE: Report

A father and his son were jailed last week after they kidnapped a sun-bed salesman when he refused to give them a refund. They squeezed the salesman into the boot of a car and drove him to a lorry park, where he was bound and gagged and forced into a trailer. The bemused salesman was then told that the trailer was going to Poland the next day and he was going with it. The salesman's luck changed when he discovered the trailer sides were made of canvas and he was able to escape by cutting a hole with his broken watch strap.

Appendix B: The Focus Algorithm (After Sidner)

The full discussion of this algorithm can be found in Sidner's 1983 paper Focusing in the comprehension of definite anaphor, in "Readings In Natural Language Processing", eds B.J. Grosz *et al.*

Part A. Expected Focus Algorithm (Suggests a focus for the first sentence).

0a. IF the sentence is an IS-A or THERE-insertion sentence
THEN the EF is the subject of the sentence
0b. ELSE the EF is the first instantiated member of the following Default List:
THEME from verb complement (if exists)
THEME
(INSTRUMENT or GOAL or LOCATIVE)
AGENT
WHOLE VERB PHRASE
1. Check the EF against the second clause using the (unspecified) inference component. Retry A1 until this instruction succeeds.
2. The Current Focus (CF) = the EF.
END of Section A

Part B. Focusing Algorithm (Repeat for each new clause in the rest of the text)

Extract the discourse entities to instantiate the members of the Default List given in Step A1. This list constitutes the Alternate Focus List (the ALFL). Then proceed with the following test. When a HALT is reached, start Section B again with the next clause in the text.

0. IF there is a THIS-anaphor
THEN move the focus to the co-specifier for the THIS anaphor
1. IF clause contains DO-anaphora then take the last member of the ALFL as Discourse Focus. Stack the CF in the Focus Stack.
HALT
2a. IF a Current Focus anaphor in clause
AND an ALFL anaphor
AND CF-anaphor not in Agent position
THEN retain CF as focus.
b. IF a CF anaphor in clause
AND an ALFL anaphor
AND CF-anaphor in the Agent position
THEN move focus to ALFL member
c. IF a CF anaphor in clause
AND an ALFL anaphor
AND neither are in Agent position
AND only ALFL anaphor is mentioned by a pronoun

THEN MOVE focus to ALFL member

d. IF a CF anaphor in clause
AND an ALFL anaphor
AND neither in Agent position
THEN retain CF
HALT.

3. IF there are only anaphors co-specifying the CF
THEN retain the CF.
HALT.

4a. IF the anaphors only co-specify a member of the ALFL
THEN move the focus to it.

b. IF several members of the ALFL are co-specified
THEN choose the focus using the Section A algorithm.
HALT.

5. IF anaphors only co-specify a member of the Focus Stack
THEN move the focus to the stack member by popping the stack.
HALT.

6. IF NO Anaphors co-specifying any of CF, ALFL or Focus Stack
BUT the CF can fill a non-obligatory case of the verb
OR if the VP is related to the CF by nominalisation
THEN retain the CF
HALT.

7. IF there are no members of the Focus Stack mentioned
THEN retain the CF as focus.

8. If there are no clause-external anaphors
THEN retain the CF at it stands.

NB: Sidner's original algorithm (Sidner, *op cit*) includes a concept of Focus Sets, and of Implicit Specification. They have both been omitted for the experiment here (because few, if any Focus Sets occur in these texts, and the idea of Implicit Specification was underspecified in her original paper).

The paper also fails to make clear what happens when a clause contains no anaphor. Following her discussion of the algorithm, and looking at, for example Case D29 step 5, I have assumed that where there is no clause external anaphor the CF persists. Rule 8 is therefore my own.

Sidner has no mention of the situation where an anaphor specifies the CF, whilst another anaphor specifies a member of the Focus Stack. In all cases where this situation arises (eg. Text 1 Clause 22) Rule B3 is taken to be the best matching rule.

Appendix C: Focus Tracking

This appendix details the results of applying Sidner's Focus Algorithm, in simulated mechanical processing, to each text. Every data structure is composed of a discourse entities tagged with a number indicating its clause of origin. The Rule slot indicates which rule in the algorithm was used to produce the Current Focus for that particular clause.

0.11 John and the Rabbit

```
1. There was a little boy.

                    Rule:  A0a

           Current Focus:  boy-1

  Alternate Focus List:  [be-vb-1]

            Focus Stack:  []


2. His name was John.

                    Rule:  B3

           Current Focus:  boy-1

  Alternate Focus List:  [name-2,
                          be-vb-2]

            Focus Stack:  []

NOTE: After its initial appearance I have
taken the string "John" to be as anaphoric
in nature as the phrases "the rabbit" or "the boy"


3. John woke early once.

                    Rule:  B3

           Current Focus:  boy-1

  Alternate Focus List:  [wake-vb-3]

            Focus Stack:  []


4. The sun was shining.

                    Rule:  B8

           Current Focus:  boy-1

  Alternate Focus List:  [sun-4,
                          shine-vb-4]

            Focus Stack:  []


5. John dressed quickly.

                    Rule:  B3

           Current Focus:  boy-1

  Alternate Focus List:  [dress-vb-5]

            Focus Stack:  []


6. He went into the garden.

                    Rule:  B3

           Current Focus: boy-1

  Alternate Focus List: [garden-6,
                         go-vb-6]

            Focus Stack: []


7. The birds were singing.

                    Rule:  B8

           Current Focus: boy-1

  Alternate Focus List: [birds-7,
                         sing-vb-7]

            Focus Stack: []


8. John saw a rabbit.

                    Rule:  B3

           Current Focus: boy-1

  Alternate Focus List: [rabbit-8,
                         see-vb-8]

            Focus Stack: []


9. The rabbit hopped across the lawn.

                    Rule:  B4a

           Current Focus: rabbit-8

  Alternate Focus List: [lawn-9,
                         hop-vb-9]
```

Focus Stack: [boy-1]

10. John watched it.

Rule: B4a

Current Focus: rabbit-8

Alternate Focus List: [boy-1, watch-vb-10]

Focus Stack: [boy-1]

11. It went into the field.

Rule: B3

Current Focus: rabbit-8

Alternate Focus List: [field-11, go-vb-11]

Focus Stack: [boy-1]

12. John went carefully after it.

Rule: B3

Current Focus: rabbit-8

Alternate Focus List: [boy-1, go-vb-12]

Focus Stack: [boy-1]

13. It ate some grass.

Rule: B3

Current Focus: rabbit-8

Alternate Focus List: [grass-13, eat-vb-13]

Focus Stack: [boy-1]

14. It ate a dandelion.

Rule: B3

Current Focus: rabbit-8

Alternate Focus List: [dandelion-14, eat-vb-14]

Focus Stack: [boy-1]

15. John crept towards it.

Rule: B2a

Current Focus: rabbit-8

Alternate Focus List: [boy-1, creep-vb-15]

Focus Stack: [boy-1]

16. The rabbit saw John.

Rule: B2b

Current Focus: boy-1

Alternate Focus List: [rabbit-8, see-vb-16]

Focus Stack: [rabbit-8, boy-1]

17. It hopped away.

Rule: B4a

Current Focus: rabbit-8

Alternate Focus List: [hop-vb-17]

Focus Stack: [boy-1, rabbit-8, boy-1]

18. John followed quietly.

Rule: B5

Current Focus: boy-1

Alternate Focus List: [follow-vb-18]

Focus Stack: [rabbit-8 boy-1, rabbit-8, boy-1]

19. He walked slowly.

Rule: B3

Current Focus: boy-1

Alternate Focus List: [walk-vb-19]

Focus Stack: [rabbit-8 boy-1, rabbit-8, boy-1]

20. The rabbit stopped.

Rule: B5

Current Focus: rabbit-8

```
Alternate Focus List: [stop-vb-20]

        Focus Stack: [boy-1,
                      rabbit-8
                      boy-1,
                      rabbit-8,
                      boy-1]
```

21. It ate another dandelion.

```
                Rule: B3

      Current Focus: rabbit-8

Alternate Focus List: [dandelion-21,
                       eat-vb-21]

        Focus Stack: [boy-1,
                      rabbit-8
                      boy-1,
                      rabbit-8,
                      boy-1]
```

22. John crept towards the rabbit.

```
                Rule: B3

      Current Focus: rabbit-8

Alternate Focus List: [boy-1,
                       creep-vb-22]

        Focus Stack: [boy-1,
                      rabbit-8
                      boy-1,
                      rabbit-8,
                      boy-1]
```

23. The rabbit was frightened.

```
                Rule: B3

      Current Focus: rabbit-8

Alternate Focus List: [be-vb-23]

        Focus Stack: [boy-1,
                      rabbit-8
                      boy-1,
                      rabbit-8,
                      boy-1]
```

24. It fled.

```
                Rule: B3

      Current Focus: rabbit-8

Alternate Focus List: [flee-vb-24]

        Focus Stack: [boy-1,
                      rabbit-8
                      boy-1,
                      rabbit-8,
                      boy-1]
```

25. John ran after it.

```
                Rule: B3

      Current Focus: rabbit-8

Alternate Focus List: [boy-1,
                       run-vb-25]

        Focus Stack: [boy-1,
                      rabbit-8
                      boy-1,
                      rabbit-8,
                      boy-1]
```

26. The rabbit came to the hedge.

```
                Rule: B3

      Current Focus: rabbit-8

Alternate Focus List: [hedge-26,
                       arrive-vb-26]

        Focus Stack: [boy-1,
                      rabbit-8
                      boy-1,
                      rabbit-8,
                      boy-1]
```

27. It ran into a burrow.

```
                Rule: B3

      Current Focus: rabbit-8

Alternate Focus List: [burrow-27,
                       run-vb-27]

        Focus Stack: [boy-1,
                      rabbit-8
                      boy-1,
                      rabbit-8,
                      boy-1]
```

28. John was sorry.

```
                Rule: B5

      Current Focus: boy-1

Alternate Focus List: [be-vb-28]

        Focus Stack: [rabbit-8,
                      boy-1,
                      rabbit-8
                      boy-1,
                      rabbit-8,
                      boy-1]
```

29. The rabbit had disappeared.

Rule: B5

Current Focus: rabbit-8

Alternate Focus List: [disappear-vb-29]

Focus Stack: [boy-1,
rabbit-8,
boy-1,
rabbit-8
boy-1,
rabbit-8,
boy-1]

30. John was sad.

Rule: B5

Current Focus: boy-1

Alternate Focus List: [be-vb-30]

Focus Stack: [rabbit-8,
boy-1,
rabbit-8,
boy-1,
rabbit-8
boy-1,
rabbit-8,
boy-1]

31. He returned to the house.

Rule: B3

Current Focus: boy-1

Alternate Focus List: [house-31,
return-vb-31]

Focus Stack: [rabbit-8,
boy-1,
rabbit-8,
boy-1,
rabbit-8
boy-1,
rabbit-8,
boy-1]

0.12 Destroyers

1. Destroyers are the
important ships in the fleet.

Rule: AOa

Current Focus: destroyers-1

Alternate Focus List: [ships-1,
fleet-1,
be-vb-1]

Focus Stack: []

2. They are armoured ships.

Rule: B3

Current Focus: destroyers-1

Alternate Focus List: [ships-2,
be-vb-2]

Focus Stack: []

3. They have guns and torpedoes.

Rule: B3

Current Focus: destroyers-1

Alternate Focus List: [guns-3,
torpedoes-3,
have-vb-3]

Focus Stack: []

4. The American fleet has many destroyers.

Rule: B8

Current Focus: destroyers-1

Alternate Focus List: [fleet-4,
have-vb-4,
am-dests-4]

Focus Stack: []

5. Some are in the Pacific.

Rule: B4a

Current Focus: am-dests-4

Alternate Focus List: [pacific-5,
be-vb-5]

Focus Stack: [destroyers-1]

6. Some are in the Atlantic.

Rule: B3

Current Focus: am-dests-4

Alternate Focus List: [atlantic-6,
be-vb-6]

Focus Stack: [destroyers-1]

7. Their guns fire big shells.

Rule: B3

Current Focus: am-dests-4

Alternate Focus List: [guns-7, shells-7, fire-vb-7]

Focus Stack: [destroyers-1]

8. Their range is ten miles.

Rule: B4a

Current Focus: guns-7

Alternate Focus List: [range-8, be-vb-8]

Focus Stack: [am-dests-4, destroyers-1]

9. Their torpedoes range over five miles.

Rule: B5

Current Focus: am-dests-4,

Alternate Focus List: [torpedoes-9, range-vb-9]

Focus Stack: [guns-7, am-dests-4, destroyers-1]

10. The Russians have many destroyers.

Rule: B8

Current Focus: am-dests-4,

Alternate Focus List: [russians-10, russ-dests-10, have-vb-10]

Focus Stack: [guns-7, am-dests-4, destroyers-1]

11. Most are in the Atlantic.

Rule: B4a

Current Focus: russ-dests-10

Alternate Focus List: [atlantic-6, be-vb-11]

Focus Stack: [am-dests-4, guns-7, am-dests-4, destroyers-1]

12. Some are in the Pacific.

Rule: B3

Current Focus: russ-dests-10

Alternate Focus List: [pacific-5, be-vb-12]

Focus Stack: [am-dests-4, guns-7, am-dests-4, destroyers-1]

13. Their guns are very accurate.

Rule: B3

Current Focus: russ-dests-10

Alternate Focus List: [guns-13, be-vb-13]

Focus Stack: [am-dests-4, guns-7, am-dests-4, destroyers-1]

14. They fire huge shells.

Rule: B4a

Current Focus: guns-13

Alternate Focus List: [shells-14, fire-vb-14]

Focus Stack: [russ-dests-10, am-dests-4, guns-7, am-dests-4, destroyers-1]

15. Their range is ten miles.

Rule: B3

Current Focus: guns-13

Alternate Focus List: [range-15, be-vb-15]

Focus Stack: [russ-dests-10, am-dests-4, guns-7, am-dests-4, destroyers-1]

16. The Russians have very powerful torpedoes.

Rule: B8

Current Focus: guns-13

Alternate Focus List: [torpedoes-16, russians-10, have-vb-16]

Focus Stack: [russ-dests-10, am-dests-4,

guns-7,
am-dests-4,
destroyers-1]

17. They range over ten miles.

Rule: B4a

Current Focus: torpedoes-16

Alternate Focus List: [range-vb-17]

Focus Stack: [russ-dests-10,
am-dests-4,
guns-7,
am-dests-4,
destroyers-1]

18. The Russians have eighty destroyers.

Rule: B5

Current Focus: russ-dests-10

Alternate Focus List: [russians-10,
have-vb-18]

Focus Stack: [torpedoes-16,
russ-dests-10,
am-dests-4,
guns-7,
am-dests-4,
destroyers-1]

19. The Americans have a hundred destroyers.

Rule: B5

Current Focus: am-dests-4

Alternate Focus List: [americans-19,
have-vb-19]

Focus Stack: [russ-dests-10,
torpedoes-16,
russ-dests-10,
am-dests-4,
guns-7,
am-dests-4,
destroyers-1]

20. The Russian ships are better.

Rule: B5

Current Focus: russ-dests-10

Alternate Focus List: [be-vb-20]

Focus Stack: [am-dests-4,
russ-dests-10,
torpedoes-16,
russ-dests-10,
am-dests-4,
guns-7,
am-dests-4,
destroyers-1]

21. They are faster than the American ships.

Rule: B3

Current Focus: russ-dests-10

Alternate Focus List: [am-dests-4,
be-vb-21]

Focus Stack: [am-dests-4,
russ-dests-10,
torpedoes-16,
russ-dests-10,
am-dests-4,
guns-7,
am-dests-4,
destroyers-1]

22. They have the best weapons.

Rule: B3

Current Focus: russ-dests-10

Alternate Focus List: [weapons-22,
have-vb-22]

Focus Stack: [am-dests-4,
russ-dests-10,
torpedoes-16,
russ-dests-10,
am-dests-4,
guns-7,
am-dests-4,
destroyers-1]

0.13 Biographies

1. Biographies are the best books.

Rule: A0a

Current Focus: biographies-1

Alternate Focus List: [books-1,
be-vb-1]

Focus Stack: []

2. They are about real things.

Rule: B3

Current Focus: biographies-1

Alternate Focus List: [things-2,
be-vb-2]

Focus Stack: []

3. They tell a true story.

Rule: B3

Current Focus: biographies-1

Alternate Focus List: [story-3, tell-vb-3]

Focus Stack: []

4. They are about particular people.

Rule: B3

Current Focus: biographies-1

Alternate Focus List: [people-4, be-vb-4]

Focus Stack: []

5. They have lessons for their readers.

Rule: B3

Current Focus: biographies-1

Alternate Focus List: [readers-5, lessons-5, have-vb-5]

Focus Stack: []

6. Smiles wrote biographies.

Rule: B8

Current Focus: smiles-biogs-6

Alternate Focus List: [smiles-6, write-vb-6]

Focus Stack: [biographies-1]

7. They are about successful workers.

Rule: B3

Current Focus: smiles-biogs-6

Alternate Focus List: [workers-7 be-vb-7]

Focus Stack: [biographies-1]

8. These biographies inspire their readers.

Rule: B3

Current Focus: smiles-biogs-6

Alternate Focus List: [readers-8, inspire-vb-8]

Focus Stack: [biographies-1]

9. Their readers imitate these workers.

Rule: B3

Current Focus: smiles-biogs-6

Alternate Focus List: [workers-7, readers-8, imitate-vb-9]

Focus Stack: [biographies-1]

10. History books are about real things.

Rule: B8

Current Focus: smiles-biogs-6

Alternate Focus List: [things-10, history-bks-10, be-vb-10]

Focus Stack: [biographies-1]

11. They tell true stories.

Rule: B4a

Current Focus: history-bks-10

Alternate Focus List: [stories-11, tell-vb-11]

Focus Stack: [smiles-biogs-6, biographies-1]

12. They are about states.

Rule: B3

Current Focus: history-bks-10

Alternate Focus List: [states-12, be-vb-12]

Focus Stack: [smiles-biogs-6, biographies-1]

13. They are not about particular people.

Rule: B3

Current Focus: history-bks-10

Alternate Focus List: [people-13, be-vb-13]

Focus Stack: [smiles-biogs-6, biographies-1]

14. They do not inspire their readers.

Rule: B3

Current Focus: history-bks-10

Alternate Focus List: [readers-14, inspire-vb-14]

Focus Stack: [smiles-biogs-6, biographies-1]

15. Novels tell a story.

Rule: B8

Current Focus: history-bks-10

Alternate Focus List: [novels-15, story-15, tell-vb-15]

Focus Stack: [smiles-biogs-6, biographies-1]

16. They are about particular characters.

Rule: B4a

Current Focus: novels-15

Alternate Focus List: [characters-16, be-vb-16]

Focus Stack: [history-bks-10, smiles-biogs-6, biographies-1]

17. They are not about real things.

Rule: B3

Current Focus: novels-15

Alternate Focus List: [things-17, be-vb-17]

Focus Stack: [history-bks-10, smiles-biogs-6, biographies-1]

18. They are dangerous.

Rule: B3

Current Focus: novels-15

Alternate Focus List: [be-vb-18]

Focus Stack: [history-bks-10, smiles-biogs-6, biographies-1]

19. Biographies are valuable books.

Rule: B5

Current Focus: biographies-1

Alternate Focus List: [books-19, be-vb-19]

Focus Stack: [novels-15, history-bks-10, smiles-biogs-6, biographies-1]

20. Histories are useful.

Rule: B5

Current Focus: history-bks-10

Alternate Focus List: [be-vb-20]

Focus Stack: [biographies-1, novels-15, history-bks-10, smiles-biogs-6, biographies-1]

21. Novels are deceitful books.

Rule: B5

Current Focus: novels-15

Alternate Focus List: [books-21, be-vb-21]

Focus Stack: [biographies-1, novels-15, history-bks-10, smiles-biogs-6, biographies-1]

22. Give children biographies.

Rule: B5

Current Focus: biographies-1

Alternate Focus List: [children-22, give-vb-22]

Focus Stack: [novels-15, biographies-1, novels-15, history-bks-10, smiles-biogs-6, biographies-1]

23. Give them histories.

Rule: B4a

Current Focus: children-22

Alternate Focus List: [history-bks-10, give-vb-23]

Focus Stack: [biographies-1, novels-15, biographies-1, novels-15, history-bks-10, smiles-biogs-6, biographies-1]

24. Do not give them novels.

Rule: B3

Current Focus: children-22

Alternate Focus List: [novels-15, give-vb-24]

Focus Stack: [biographies-1, novels-15, biographies-1, novels-15, history-bks-10, smiles-biogs-6, biographies-1]

0.14 Bank Robbers Die in Dublin Shoot-out

1. Two bank robbers were shot dead last Friday

Rule: A0a

Current Focus: robbers-1

Alternate Focus List: [shoot-vb-1]

Focus Stack: []

2. when a high-speed car chase ended in a gun battle in the city centre.

Rule: B8

Current Focus: robbers-1

Alternate Focus List: [chase-2, battle-2, centre-2, end-vb-2]

Focus Stack: []

3. IRA involvement has not been ruled out by detectives.

Rule: B8

Current Focus: robbers-1

Alternate Focus List: [ira-involvement-3, detectives-3, rule-vb-3]

Focus Stack: []

4. The drama began shortly after 10am

Rule: B8

Current Focus: robbers-1

Alternate Focus List: [drama-4, begin-vb-4]

Focus Stack: []

5. when the men held up a bank 5 miles west of Dublin.

Rule: B3

Current Focus: robbers-1

Alternate Focus List: [bank-5, Dublin-5, hold-up-vb-5]

Focus Stack: []

6. They made off with 2,500 pounds

Rule: B3

Current Focus: robbers-1

Alternate Focus List: [2500-6, make-off-vb-6]

Focus Stack: []

7. but their car was spotted by armed detectives on the edge of the city

Rule: B3

Current Focus: robbers-1

Alternate Focus List: [car-7, detectives-7, city-edge-7, spot-vb-7]

Focus Stack: []

8. and a high-speed car chase followed.

Rule: B8

Current Focus: robbers-1

```
Alternate Focus List:  [chase-8,
                        follow-vb-8]

Focus Stack:  []
```

9. The Garda car came under repeated fire from the raiders.

```
Rule:  B8

Current Focus:  robbers-1

Alternate Focus List:  [car-9,
                        fire-9,
                        come-under-vb-9]

Focus Stack:  []
```

10. The chase ended

```
Rule:  B4a

Current Focus:  chase-8

Alternate Focus List:  [end-vb-10]

Focus Stack:  [robbers-1]
```

NOTE: This assumes it is possible to reference chase-8 in some way, in order to use Rule B4a. The assumption I have made throughout is to say that if the referenced AFL member is no more than three clauses behind, then the co-specification will generally be permitted.

11. When the robbers tried to smash through a second Garda car

```
Rule:  B5

Current Focus:  robbers-1
Alternate Focus List:  [car-11,
                        try-vb-11]

Focus Stack:  [chase-8,
               robbers-1]
```

12. that was blocking the road.

```
Rule:  B4a

Current Focus:  car-11

Alternate Focus List:  [road-12,
                        block-vb-12]

Focus Stack:  [robbers-1,
               chase-8,
               robbers-1]
```

13. A prolonged shoot-out followed

```
Rule:  B8

Current Focus:  car-11

Alternate Focus List:  [shoot-out-13,
                        follow-vb-13]

Focus Stack:  [robbers-1,
               chase-8,
               robbers-1]
```

14. as more armed Gardai arrived.

```
Rule:  B8

Current Focus:  car-11

Alternate Focus List:  [gardai-14,
                        arrive-vb-14]

Focus Stack:  [robbers-1,
               chase-8,
               robbers-1]
```

15. The two men inside were reported dead on arrival in hospital.

```
Rule:  B5

Current Focus:  robbers-1

Alternate Focus List:  [hospital-15,
                        report-vb-15]

Focus Stack:  [car-11,
               robbers-1,
               chase-8,
               robbers-1]
```

16. Gardai could not say who fired the first shot.

```
Rule:  B8

Current Focus:  robbers-1

Alternate Focus List:  [shot-16,
                        gardai-16,
                        say-vb-16]

Focus Stack:  [car-11,
               robbers-1,
               chase-8,
               robbers-1]
```

0.15 The Beautiful Game

1. To become fully-developed human beings

Rule: AOb

Current Focus: humans-1

Alternate Focus List: [become-vb-1]

Focus Stack: []

2. we should cultivate our minds and our bodies.

Rule: B8

Current Focus: humans-1

Alternate Focus List: [minds-2,
bodies-2,
readers-2,
cultivate-vb-2]

Focus Stack: []

3. Sport provides a unique opportunity to do this

Rule: BO

Current Focus: cultivate-vb-2

Alternate Focus List: [opportunity-3,
sport-3,
provide-vb-3]

Focus Stack: [humans-1]

4. and it therefore ranks as one of the most important of all human activies.

Rule: B4a

Current Focus: sport-3

Alternate Focus List: [activities-4,
rank-vb-4]

Focus Stack: [cultivate-vb-2,
humans-1]

5. Competition is essential.

Rule: B8

Current Focus: sport-3

Alternate Focus List: [competition-5,
be-vb-5]

Focus Stack: [cultivate-vb-2,
humans-1]

6. It encourages people to develop their abilities,

Rule: B3

Current Focus: competition-5

Alternate Focus List: [abilities-6,
people-6,
develop-vb-6]

Focus Stack: [sport-3,
cultivate-vb-2,
humans-1]

7. and to appreciate those of others.

Rule: BO

Current Focus: abilities-6

Alternate Focus List: [people-7,
appreciate-vb-7]

Focus Stack: [competition-5,
sport-3,
cultivate-vb-2,
humans-1]

8. People naturally admire individual sportsmen and women

Rule: B8

Current Focus: abilities-6

Alternate Focus List: [sports-people-set-8,
people-8,
admire-vb-8]

Focus Stack: [competition-5,
sport-3,
cultivate-vb-2,
humans-1]

9. and develop loyalties to them.

Rule: B4a

Current Focus: sports-people-set-8

Alternate Focus List: [loyalties-9,
develop-vb-9]

Focus Stack: [abilities-6,
competition-5,
sport-3,
cultivate-vb-2,
humans-1]

10. The feeling of being represented in a competition by an individual or group develops our own sense of regional or national identity.

Rule: B8

Current Focus: sports-people-set-8

Alternate Focus List: [identity-10,
competition-10,
individual-10,
group-10,
feeling-10,
readers-10,
represent-vb-10]

Focus Stack: [sports-people-set-8,
abilities-6,
competition-5,
sport-3,
cultivate-vb-2,
humans-1]

11. Competitions allow us to give vent to natural feelings of competition

Rule: B4a

Current Focus: readers-10

Alternate Focus List: [feelings-11,
competition-5,
competitions-11,
allow-vb-11]

Focus Stack: [sports-people-set-8,
abilities-6,
competition-5,
sport-3,
cultivate-vb-2,
humans-1]

12. which, without the safety-valve of sport, could lead to large-scale violent confrontations.

Rule: B0

Current Focus: feelings-11

Alternate Focus List: [confrontations-12,
sport-3,
safety-valve-12,
lead-vb-12]

Focus Stack: [readers-10
sports-people-set-8,
abilities-6,
competition-5,
sport-3,
cultivate-vb-2,
humans-1]

0.16 Titanic enquiry to be re-opened

1. The Captain of the Californian, who was accused of failing to aid the stricken liner Titanic,

Rule: A0b

Current Focus: captain-1

Alternate Focus List: [californian-1,
liner-1,
titanic-1,
accuse-vb-1]

Focus Stack: []

2. is to have his conduct re-examined 78 years after the tragedy.

Rule: B3

Current Focus: captain-1

Alternate Focus List: [tragedy-2,
conduct-2,
examine-vb-2]

Focus Stack: []

3. The Titanic sank with the loss of 1,503 lives

Rule: B8

Current Focus: captain-1

Alternate Focus List: [loss-3,
lives-3,
titanic-1,
sink-vb-3]

Focus Stack: []

4. after hitting an ice-berg during her maiden voyage in 1912.

Rule: B4a

Current Focus: titanic-1

Alternate Focus List: [voyage-4,
ice-berg-4,
hit-vb-4]

Focus Stack: [captain-1]

5. The new inquiry follows a 30-year old campaign to clear the name of Captain Stanley Lord, master of the Californian.

Rule: B8

Current Focus: titanic-1

Alternate Focus List: [name-5,
californian-1,
campaign-5,
inquiry-5,
follow-vb-5]

Focus Stack: [captain-1]

NOTE: A similar issue to recognising that ''John'' and ''the boy'' are the same entity, made more complex in this text by the occurrence of ''the master'',''the Captain,'', ''Captain Lord'', and ''Captain Stanley Lord''. This is an area that requires much work

6. The 1912 inquiry accused Captain Lord of not acting,

Rule: B3

Current Focus: captain-1

Alternate Focus List: [not-acting-6,
inquiry-6,
accuse-vb-6]

Focus Stack: [titanic-1,
captain-1]

7. although his was the nearest ship.

Rule: B3

Current Focus: captain-1

Alternate Focus List: [californian-1,
be-vb-7]

Focus Stack: [titanic-1,
captain-1]

8. The inquiry found

Rule: B4a

Current Focus: inquiry-6

NOTE: Dubious use of this Rule.

Alternate Focus List: [find-vb-8]

Focus Stack: [captain-1,
titanic-1,
captain-1]

9. that the Californian was less than 10 miles from the Titanic,

Rule: B3

Current Focus: titanic-1

Alternate Focus List: [californian-1,
be-vb-9]

Focus Stack: [inquiry-6,
captain-1,
titanic-1,
captain-1]

10. but she ignored distress rockets

Rule: B4a

Current Focus: californian-1

Alternate Focus List: [rockets-10,
ignore-vb-10]

Focus Stack: [titanic-1,
inquiry-6,
captain-1,
titanic-1,
captain-1]

11. and had her radio switched off.

Rule: B3

Current Focus: californian-1

Alternate Focus List: [radio-11,
have-vb-11]

Focus Stack: [titanic-1,
inquiry-6,
captain-1,
titanic-1,
captain-1]

12. Captain Lord insisted he was 19 miles away,

Rule: B5

Current Focus: captain-1

Alternate Focus List: [insist-vb-12]

Focus Stack: [californian-1,
titanic-1,
inquiry-6,
captain-1,
titanic-1,
captain-1]

13. and new evidence on the location of the Titanic wreck could confirm that claim.

Rule: B0

Current Focus: insist-vb-12

Alternate Focus List: [location-13,
wreck-13,
claim-13,
evidence-13,

confirm-vb-13]

Focus Stack: [captain-1,
californian-1,
titanic-1,
inquiry-6,
captain-1,
titanic-1,
captain-1]

0.17 North Pole

1. The British Explorers Sir Ranulph Fiennes and Dr Michael Stroud were forced to abandon their attempt to be the first men to walk unassisted to the North Pole - with only 90 miles to go.

Rule: AOa

Current Focus: explorers-1

Alternate Focus List: [attempt-1,
pole-1,
men-1,
miles-1,
abandon-vb-1]

Focus Stack: []

2. Their North Pole expedition was halted by stretches of open water 47 days and 439 miles after setting out.

Rule: B3

Current Focus: explorers-1

Alternate Focus List: [attempt-1,
water-2,
days-2,
miles-2,
halt-vb-2]

Focus Stack: []

3. Fiennes and Stroud have still scored a remarkable achievement.

Rule: B3

Current Focus: explorers-1

Alternate Focus List: [achievement-3,
score-vb-3]

Focus Stack: []

4. Their record-breaking trip has so far raised over 400,000 pounds for Multiple Sclerosis research.

Rule: B3

Current Focus: explorers-1

Alternate Focus List: [trip-4,
400k-4,
research-4,
raise-vb-4]

Focus Stack: []

5. Several times they nearly lost their lives.

Rule: B3

Current Focus: explorers-1

Alternate Focus List: [lives-5,
times-5,
lose-vb-5]

Focus Stack: []

6. Both once fell through the ice

Rule: B3

Current Focus: explorers-1

Alternate Focus List: [ice-6,
fall-vb-6]

Focus Stack: []

7. and had to be saved from the water.

Rule: B8

Current Focus: explorers-1

Alternate Focus List: [water-7,
save-vb-7]

Focus Stack: []

8. In pain, and fatigued, eventually the shortage of food supplies defeated them.

Rule: B3

Current Focus: explorers-1

Alternate Focus List: [pain-8,
shortage-8,
supplies-8,

defeat-vb-8]

Focus Stack: []

9. This and the melting ice forced them to summon help.

Rule: B0

Current Focus: shortage-8

Alternate Focus List: [ice-6, help-9, force-vb-9]

Focus Stack: [explorers-1]

0.18 Cracked Up Kids

1. Young teenage boys are now being used as runners by inner-city crack dealers.

Rule: A0a

Current Focus: boys-1

Alternate Focus List: [runners-1, dealers-1, use-vb-1]

Focus Stack: []

2. The crack is manufactured in abandoned flats, often heavily fortified,

Rule: B4a

Current Focus: crack-2

NOTE: By implicit use of Rule B4a

Alternate Focus List: [flats-2, make-vb-2]

Focus Stack: [boys-1]

3. and it is then given to the boys

Rule: B3

NOTE: Again, no specific Rule to deal with this situation of a Cf anaphor and an FS anaphor - B3 seems the best match

Current Focus: crack-2

Alternate Focus List: [boys-1, give-vb-3]

Focus Stack: [boys-1]

4. who sell it on the walkways and the street.

Rule: B2b

Current Focus: crack-2

Alternate Focus List: [walkways-4, street-4, boys-1, sell-vb-4]

Focus Stack: [boys-1]

5. The drug is wrapped in cling film

Rule: B3

Current Focus: crack-2

Alternate Focus List: [film-5, wrap-vb-5]

Focus Stack: [boys-1]

6. and they carry it in their mouths,

Rule: B2b

Current Focus: crack-2

Alternate Focus List: [boys-1, mouths-6, carry-vb-6]

Focus Stack: [boys-1]

7. so they can swallow it

Rule: B2b

Current Focus: crack-2

Alternate Focus List: [boys-1, swallow-vb-7]

Focus Stack: [boys-1]

8. if police approach them.

Rule: B4a

Current Focus: boys-1

Alternate Focus List: [police-8,
approach-vb-9]

Focus Stack: [crack-2,
boys-1]

9. More and more young people are
being exposed to the drug world.

Rule: B8

Current Focus: boys-1

Alternate Focus List: [people-9,
world-9,
exposed-vb-9]

Focus Stack: [crack-2,
boys-1]

10. A survey of ten liverpool primary
schools revealed

Rule: B8

Current Focus: boys-1

Alternate Focus List: [school-10,
survey-10,
reveal-vb-10]

Focus Stack: [crack-2,
boys-1]

11. that pupils as young as four were
being approached by dealers.

Rule: B8

Current Focus: boys-1

Alternate Focus List: [pupils-11
dealers-11,
approach-vb-11]

Focus Stack: [crack-2,
boys-1]

12. Liverpool schools have also been warned
of the dangers of pupils contracting
the AIDS virus and hepatitis from discarded
hypodermic syringes and needles.

Rule: B8
Current Focus: boys-1

Alternate Focus List: [schools-12,
dangers-12,
pupils-12,
virus-12,
hepatitis-12,
syringes-12,
needles-12,
warn-vb-12]

Focus Stack: [crack-2,
boys-1]

13. Warning posters and 'sharp boxes'
need to be provided

Rule: B8

Current Focus: boys-1

Alternate Focus List: [posters-13,
boxes-13,
need-vb-13,]

Focus Stack: [crack-2,
boys-1]

14. for disposing of such equipment.

Rule: B8

Current Focus: equipment-14

Alternate Focus List: [dispose-vb-14]

Focus Stack: [boys-1,
crack-2,
boys-1]

0.19 Animal Medicine

1. The ancient treatment of acupuncture
is coming to the veterinary surgery
for pet dogs and cats.

Rule: A0a

Current Focus: acupuncture-1

Alternate Focus List: [surgery-1,
dogs-1,
cats-1,
come-vb-1]

Focus Stack: []

2. Acupuncture is used to treat chronic pain.

Rule: B3

Current Focus: acupuncture-1

Alternate Focus List: [pain-2,
treat-vb-2]

Focus Stack: []

3. Dogs' slipped discs, severe arthritic pain
and muscle and tendon strains
are now helped with this ancient art.

Rule: B0

Current Focus: acupuncture-1

Alternate Focus List: [discs-3,
pain-3,
strains-3,
help-vb-3]

Focus Stack: []

4. For veterinary medicine is entering a new stage,

Rule: B8

Current Focus: acupuncture-1

Alternate Focus List: [stage-4,
medicine-4,
enter-vb-4]

Focus Stack: []

5. and acupuncture together with successful
hip replacements for crippled
dogs and dental fillings is becoming common.

Rule: B3

Current Focus: acupuncture-1

Alternate Focus List: [dogs-5,
fillings-5,
replacements-5,
become-vb-5]

Focus Stack: []

6. The dentist's drill can keep dogs'
and cats' teeth in healthy and
sparkling condition.

Rule: B8

Current Focus: acupuncture-1

Alternate Focus List: [condition-6,
teeth-6,
drill-6,
keep-vb-6]

Focus Stack: []

7. or regular brushing with liver flavoured
toothpaste can prevent decay altogether.

Rule: B8

Current Focus: acupuncture-1

Alternate Focus List: [decay-7,
toothpaste-7,
brushing-7,
prevent-vb-7]

Focus Stack: []

0.20 Sunbed Salesman

1. A father and his son were jailed last week

Rule: A0a

Current Focus: father-son-1

Alternate Focus List: [week-1,
jail-vb-1]

Focus Stack: []

2. after they kidnapped a sun-bed salesman

Rule: B3

Current Focus: father-son-1

Alternate Focus List: [salesman-2,
kidnap-vb-2]

Focus Stack: []

3. when he refused to give them a refund.

Rule: B2a

Current Focus: father-son-1

Alternate Focus List: [refund-3,

salesman-2,
refuse-vb-3]

Focus Stack: []

4. They squeezed the salesman into the boot of a car

Rule: B2b

Current Focus: salesman-2

Alternate Focus List: [car-4,
boot-4,
father-son-1,
squeeze-vb-4]

Focus Stack: [father-son-1]

5. and drove him to a lorry park,

Rule: B3

Current Focus: salesman-2

Alternate Focus List: [park-5,
drive-vb-5]

Focus Stack: [father-son-1]

6. where he was bound

Rule: B3

Current Focus: salesman-2

Alternate Focus List: [bind-vb-6]

Focus Stack: [father-son-1]

7. and gagged

Rule: B8

Current Focus: salesman-2
Alternate Focus List: [gag-vb-7]
Focus Stack: [father-son-1]

8. and forced into a trailer.

Rule: B8

Current Focus: salesman-2

Alternate Focus List: [trailer-8,
force-vb-8]

Focus Stack: [father-son-1]

9. The bemused salesman was then told

Rule: B3
Current Focus: salesman-2
Alternate Focus List: [tell-vb-9]
Focus Stack: [father-son-1]

10. that the trailer was going to Poland the next day

Rule: B4a

Current Focus: trailer-8

Alternate Focus List: [day-10,
poland-10,
go-vb-10]

Focus Stack: [salesman-2,
father-son-1]

11. and he was going with it.

Rule: B2b

Current Focus: trailer-8

Alternate Focus List: [salesman-2,
go-vb-11]

Focus Stack: [salesman-2,
father-son-1]

12. The salesman's luck changed

Rule: B4a

Current Focus: salesman-2

Alternate Focus List: [luck-12,
change-vb-12]

Focus Stack: [trailer-8,
salesman-2,
father-son-1]

13. when he discovered the trailer sides were made of canvas

Rule: B3

Current Focus: salesman-2

Alternate Focus List: [canvas-13,
sides-13,
discover-vb-13]

Focus Stack: [trailer-8,
salesman-2,
father-son-1]

14. and he was able to escape

Rule: B3

Current Focus: salesman-2

```
Alternate Focus List:  [escape-vb-14]

        Focus Stack:  [trailer-8,
                       salesman-2,
                       father-son-1]
```

15. by cutting a hole with his broken watch strap.

```
                Rule:  B3


     Current Focus:  salesman-2

Alternate Focus List:  [strap-15,
                        hole-15,
                        cut-vb-15]

      Focus Stack:  [trailer-8,
                     salesman-2,
                     father-son-1]
```

Appendix D: Tabulate the Focus Statistics

This appendix summarises of the results of the focus tracking given in Appendix C, tabulating each discourse entity, the number of it entered the Alternate Focus List, and the it was actually Current Focus. It is impossible for an entity to be in the AFL and the Focus List in the same clause.

Summary information showing the items most in Current Focus in each text is given in Appendix G.

0.21 John and the Rabbit

ENTITY	IN AFL	IN FOCUS
1. boy-1	5	14
2. be-vb-1	1	-
3. name-2	1	-
4. be-vb-2	1	-
5. wake-vb-3	1	-
6. sun-4	1	-
7. shine-vb-4	1	-
8. dress-vb-5	1	-
9. garden-6	1	-
10. go-vb-6	1	-
11. birds-7	1	-
12. sing-vb-7	1	-
13. rabbit-8	2	17
14. see-vb-8	1	-
15. lawn-9	1	-
16. hop-vb-9	1	-
17. watch-vb-10	1	-
18. field-11	1	-
19. go-vb-11	1	-
20. go-vb-12	1	-
21. grass-13	1	-
22. eat-vb-13	1	-
23. dandelion-14	1	-
24. eat-vb-14	1	-
25. creep-vb-15	1	-
26. see-vb-16	1	-
27. hop-vb-17	1	-
28. follow-vb-18	1	-
29. walk-vb-19	1	-
30. stop-vb-20	1	-
31. dandelion-21	1	-
32. eat-vb-21	1	-
33. creep-vb-22	1	-
34. be-vb-23	1	-
35. flee-vb-24	1	-
36. run-vb-25	1	-
37. hedge-26	1	-
38. arrive-vb-26	1	-
39. burrow-27	1	-
40. run-vb-27	1	-
41. be-vb-28	1	-
42. disappear-vb-29	1	-
43. be-vb-30	1	-
44. house-31	1	-
45. return-vb-31	1	-

0.22 Destroyers

ENTITY	IN AFL	IN FOCUS
1. destroyers-1	-	4
2. ships-1	1	-
3. fleet-1	1	-
4. be-vb-1	1	-
5. ships-2	1	-
6. be-vb-2	1	-
7. guns-3	1	-
8. torpedoes-3	1	-
9. have-vb-3	1	-
10. fleet-4	1	-
11. have-vb-4	1	-
12. pacific-5	2	-
13. american-destroyers-4	2	6
14. be-vb-5	1	-
15. atlantic-6	2	-
16. be-vb-6	1	-
17. guns-7	1	1
18. shells-7	1	-
19. fire-vb-7	1	-
20. range-8	1	-
21. be-vb-8	1	-
22. torpedoes-9	1	-
23. range-vb-9	1	-
24. russians-10	3	-
25. russian-destroyers-10	1	7
26. have-vb-10	1	-
26. be-vb-11	1	-
27. be-vb-12	1	-
28. guns-13	1	3
29. be-vb-13	1	-
30. shells-14	1	-
31. fire-vb-14	1	-
32. range-15	1	-
33. be-vb-15	1	-
34. torpedoes-16	1	1
35. have-vb-16	1	-
36. range-vb-17	1	-
37. have-vb-18	1	-
38. americans-19	1	-
39. have-vb-19	1	-
40. be-vb-20	1	-
41. be-vb-21	1	-
42. weapons-22	1	-
43. have-vb-22	1	-

0.23 Biographies

ENTITY	IN AFL	IN FOCUS
1. biographies-1	-	7
2. books-1	1	-
3. be-vb-1	1	-
4. things-2	1	-
5. be-vb-2	1	-
6. story-3	1	-
7. tell-vb-3	1	-
8. people-4	1	-
9. be-vb-4	1	-
10. lessons-5	1	-
11. readers-5	1	-
12. have-vb-5	1	-
13. smiles-6	1	-
14. write-vb-6	1	-
15. smiles-biogs-6	-	5
16. workers-7	2	-
17. be-vb-7	1	-
18. readers-8	2	-
19. inspire-vb-8	1	-
20. imitate-vb-9	1	-
21. things-10	1	-
22. history-bks-10	2	6
23. be-vb-10	1	-
24. stories-11	1	-
25. tell-vb-11	1	-
26. states-12	1	-
27. people-13	1	-
28. be-vb-12	1	-
29. be-vb-13	1	-
29. readers-14	1	-
30. inspire-vb-14	1	-
31. novels-15	2	4
32. story-15	1	-
33. tell-vb-15	1	-
34. characters-16	1	-
35. be-vb-16	1	-
36. things-17	1	-
37. be-vb-17	1	-
38. be-vb-18	1	-
39. books-19	1	-
30. be-vb-19	1	-
41. be-vb-20	1	-
42. books-21	1	-
43. be-vb-21	1	-
44. children-22	1	2
45. give-vb-22	1	-
46. give-vb-23	1	-
47. give-vb-24	1	-

0.24 Dublin

ENTITY	IN AFL	IN FOCUS
1. robbers-1	-	12
2. shoot-vb-1	1	-
3. chase-2	1	-

4. battle-2	1	-
5. centre-2	1	-
6. end-vb-2	1	-
7. ira-involvement-3	1	-
8. detectives-3	1	-
9. rule-vb-3	1	-
10. drama-4	1	-
11. begin-vb-4	1	-
12. bank-5	1	-
13. Dublin-5	1	-
14. hold-up-vb-5	1	-
15. 2500-6	1	-
16. make-off-vb-6	1	-
17. car-7	1	-
18. detectives-7	1	-
19. city-edge-7	1	-
20. spot-vb-7	1	-
21. chase-8	1	1
22. follow-vb-8	1	-
23. car-9	1	-
24. come-under-vb-9	1	-
25. fire-9	1	-
26. end-vb-10	1	-
27. car-11	1	3
28. try-vb-11	1	-
29. road-12	1	-
30. block-vb-12	1	-
31. shoot-out-13	1	-
32. follow-vb-13	1	-
33. gardia-14	1	-
34. arrive-vb-14	1	-
35. hospital-15	1	-
36. report-vb-15	1	-
37. shot-16	1	-
38. say-vb-16	1	-
39. gardai-16	1	-

0.25 The Beautiful Game

ENTITY	IN AFL ENTERED AFL	IN FOCUS
1. humans-1	-	2
2. become-vb-1	1	-
3. readers-2	1	-
4. minds-2	1	-
5. bodies-2	1	-
6. cultivate-vb-2	1	1
7. opportunity-3	1	-
8. sport-3	2	2
9. provide-vb-3	1	-
10. activities-4	1	-
11. rank-vb-4	1	-
12. competition-5	2	1
13. be-vb-5	1	-
14. abilities-6	1	2
15. people-6	1	-
16. develop-vb-6	1	-
17. people-7	1	-
18. appreciate-vb-7	1	-
19. sports-people-set-8	1	2
20. people-8	1	-
21. admire-vb-8	1	-
22. loyalties-9	1	-
23. develop-vb-9	1	-
24. identity-10	1	-
25. competition-10	1	-
26. individual-10	1	-
27. group-10	1	-
28. feeling-10	1	-
29. represent-vb-10	1	-
30. readers-10	1	1
31. feelings-11	1	1
31. competitions-11	1	-
32. allow-vb-12	1	-
33. confrontations-12	1	-
34. safety-valve-12	1	-
35. lead-vb-12	1	-

0.26 Titanic

ENTITY	IN AFL ENTERED AFL	IN FOCUS
1. captain-1	-	6
2. californian-1	4	2
3. liner-1	1	-
4. titanic-1	1	3
5. accuse-vb-1	1	-
6. tragedy-2	1	-
7. conduct-2	1	-
8. examine-vb-2	1	-
9. loss-3	1	-
10 lives-3	1	-
11. sink-vb-3	1	-
12. voyage-4	1	-
13. ice-berg-4	1	-
14. hit-vb-4	1	-
15. name-5	1	-
16. campaign-5	1	-
17. inquiry-5	1	-
18. follow-vb-5	1	-
19. not-acting-6	1	-
20. inquiry-6	1	1
21. accuse-vb-6	1	-
22. be-vb-7	1	-
23. find-vb-8	1	-
24. be-vb-9	1	-
25. rockets-10	1	-
26. ignore-vb-10	1	-
27. radio-11	1	-
28. have-vb-11	1	-
29. insist-vb-12	1	1
30. location-13	1	-
31. wreck-13	1	-
32. claim-13	1	-
33. evidence-13	1	-
34. confirm-vb-13	1	-

0.27 North Pole

ENTITY	IN AFL ENTERED AFL	IN FOCUS
1. attempt-1	2	-
2. explorers-1	-	8
3. pole-1	1	-
4. men-1	1	-
5. miles-1	1	-
6. abandon-vb-1	1	-
7. water-2	1	-
8. days-2	1	-
9. miles-2	1	-
10. halt-vb-2	1	-
11. achievement-3	1	-
12. score-vb-3	1	-
13. trip-4	1	-
14. 400k-4	1	-
15. research-4	1	-
16. raise-vb-4	1	-
17. lives-5	1	-
18. times-5	1	-
19. lose-vb-5	1	-
20. ice-6	2	-
21. fall-vb-6	1	-
22. water-7	1	-
23. save-vb-7	1	-
24. pain-8	1	-
25. shortage-8	1	1
26. supplies-8	1	-
27. defeat-vb-8	1	-
28. help-9	1	-
29. force-vb-9	1	-

0.28 Cracked Up Kids

ENTITY	IN AFL ENTERED AFL	IN FOCUS
1. boys-1	4	7
2. runners-1	1	-
3. dealers-1	1	-
4. use-vb-1	1	-
5. crack-2	-	6
6. flats-2	1	-
7. make-vb-2	1	-
8. give-vb-3	1	-
9. walkways-4	1	-
10. street-4	1	-
11. sell-vb-4	1	-
12. film-5	1	-
13. wrap-vb-5	1	-
14. carry-vb-6	1	-
15. mouths-6	1	-

15. swallow-vb-7	1	-
16. police-8	1	-
17. approach-vb-8	1	-
18. people-9	1	-
19. world-9	1	-
20. exposed-vb-9	1	-
21. school-10	1	-
22. survey-10	1	-
23. reveal-vb-10	1	-
24. pupils-11	1	-
25. dealers-11	1	-
26. approach-vb-11	1	-
27. schools-12	1	-
28. dangers-12	1	-
29. pupils-12	1	-
30. virus-12	1	-
31. hepatitis-12	1	-
32. syringes-12	1	-
33. needles-12	1	-
34. warn-vb-12	1	-
35. posters-13	1	-
36. boxes-13	1	-
37. need-vb-13	1	-
38. equipment-14	-	1
39. dispose-vb-14	1	-
26. discover-vb-13	1	-
27. escape-vb-14	1	-
28. strap-15	1	-
29. hole-15	1	-

0.29 Animal Medicine

ENTITY	IN AFL ENTERED AFL	IN FOCUS
1. acupuncture-1	-	7
2. surgery-1	1	-
3. dogs-1	1	-
4. cats-1	1	-
5. come-vb-1	1	-
6. pain-2	1	-
7. treat-vb-2	1	-
8. discs-3	1	-
9. pain-3	1	-
10. strains-3	1	-
11. help-vb-3	1	-
12. stage-4	1	-
13. medicine-4	1	-
14. enter-vb-4	1	-
15. dogs-5	1	-
16. fillings-5	1	-
17. replacements-5	1	-
18. become-vb-5	1	-
19. condition-6	1	-
20. teeth-6	1	-
21. drill-6	1	-
22. keep-vb-6	1	-
23. decay-7	1	-
24. toothpaste-7	1	-
25. brushing-7	1	-
26. prevent-vb-7	1	-

0.30 Sunbed Salesman

ENTITY	IN AFL	IN FOCUS
1. father-son-1	1	3
2. week-1	1	-
3. jail-vb-1	1	-
4. salesman-2	3	10
5. kidnap-vb-2	1	-
6. refund-3	1	-
7. refuse-vb-3	1	-
8. car-4	1	-
9. boot-4	1	-
10. squeeze-vb-4	1	-
11. park-5	1	-
12. drive-vb-5	1	-
13. bind-vb-6	1	-
14. gag-vb-7	1	-
15. force-vb-8	1	-
16 trailer-8	1	2
17. tell-vb-9	1	-
18. day-10	1	-
19. poland-10	1	-
20. go-vb-10	1	-
21. go-vb-11	1	-
22. luck-12	1	-
23. change-vb-12	1	-
24. canvas-13	1	-
25. sides-13	1	-

Appendix E: Tabulate the Frequency of every NP

This section tabulates the result of applying a hypothetical, simple NP counter to each text. Each occurrence of a particular noun spelling was counted. This means that any noun that is inflected will have a separate entry on the table for each inflection; so, for example, "destroyer" and "destroyers" would merit two entries. Of course, anaphor resolution is not assumed to have been applied. Compound nouns, like "car chase", will sometimes contribute to three separate counts,[8] corresponding to: "car chase", "car" and "chase".

This makes some attempt to compensate for the times when an abbreviation for the compound is used in the rest of the text. It has, however, the drawback of occasionally unbalancing the analysis. "Car chase", for example, occurs first in Text 4 in this compound form, but then is later referenced simply as "the chase". If, however, the abbreviation of a compound noun is not used then an unbalancing count is made. This can be seen in the case of "Captain Lord" in text 6. The initial occurrence is "Captain", then "Captain Stanley Lord", and then as "Captain Stanley lord". I have counted "Stanley Lord" as one, but I think that "Captain Lord" must be treated as other compound nouns, and have three entries.

This section gives the full NP count for each text - the summaries derived from this count, found in Appendix G, will treat those NPs with a count of only one as being insignificant.

0.31 John and the Rabbit

NP	COUNT
1. boy	1
2. John	13
3. sun	1
4. garden	1
5. birds	1
6. rabbit	8
7. lawn	1
8. field	1
9. grass	1
10. dandelion	2
11. hedge	1
12. burrow	1
13. house	1

0.32 Destroyers

NP	COUNT
1. destroyers	5
2. ships	4
3. fleet	2
4. guns	3
5. torpedoes	3
6. pacific	2
7. atlantic	2
8. shells	2
9. range	2
10. miles	4
11. russians	3
12. weapons	1

[8] See section 2.1, page 4 for discussion of this point

0.33 Biographies

NP	COUNT
1. biographies	5
2. books	4
3. things	3
4. story	2
5. people	2
6. lessons	1
7. readers	4
8. smiles	1
9. workers	2
11. history books	1
12. stories	1
13. states	1
15. novels	3
16. characters	1
17. children	1
18. history	1

0.34 Dublin

NP	COUNT
1. bank robbers	1
2. friday	1
3. chase	3
4. battle	1
5. city centre	1
6. involvement	1
7. detectives	2
8. drama	1
9. men	2
10. bank	1
11. Dublin	1
12. 2,500 pounds	1
13. car	5
15. city	2
16. fire	1
17. raiders	1
18. road	1
19. shoot-out	1
20. gardai	2
21. hospital	1
22. shot	1
23. robbers	2
24. car chase	2
25. gun battle	1
26. centre	1

0.35 The Beautiful Game

NP	COUNT
1. human beings	1
2. minds	1
3. bodies	1
4. sport	2
5. opportunity	1
6. activities	1
7. competition	3
8. people	2
9. abilities	1
10. sportsmen	1
11. women	1
12. loyalties	1
13. feeling	2
14. individual	1
15. group	1
16. sense	1
17. identity	1
18. confrontations	1
19. beings	1
20. competition	1
21. feelings	1

0.36 Titanic

NP	COUNT
1. captain	4
2. Californian	3
3. liner	1

4. Titanic	4
5. conduct	1
6. years	1
7. tragedy	1
8. loss	1
9. lives	1
10. ice-berg	1
11. voyage	1
12. inquiry	3
13. campaign	1
14. name	1
15. master	1
16. ship	1
17. miles	2
18. rockets	1
19. radio	1
20. evidence	1
21. location	1
22. wreck	1
23. claim	1
24. Captain Lord	2
25. Captain Stanley Lord	1
26. Titanic wreck	1

0.37 North Pole

NP	COUNT
1. explorers	1
2. attempt	1
3. men	1
4. north pole	2
5. miles	2
6. expedition	1
7. stretches	1
8. water	2
9. days	1
10. achievement	1
11. trip	1
12. 400,000	1
13. research	1
14. lives	1
15. ice	2
16. pain	1
17. shortage	1
18. food supplies	1
19. help	1
20. Fiennes	2
21. Stroud	2
22. food	1
23. supplies	1

0.38 Cracked Up Kids

1. boys	2
2. runners	1
3. dealers	2
4. crack	2
5. flats	1
6. walkways	1
7. street	1
8. drug	2
9. cling film	1
10. mouths	1
11. police	1
12. people	1
13. drug world	1
14. survey	1
15. schools	2
16. pupils	2
17. dangers	1
18. aids virus	1
19. hepatitis	1
20. syringes	1
21. needles	1
22. posters	1
23. sharp boxes	1
24. equipment	1
25. crack dealers	1
26. world	1
27. liverpool	1
28. virus	1
29. aids	1

0.39 Animal Medicine

1. treatment	1
2. acupuncture	3
3. surgery	1
4. dogs	4
5. cats	2
6. pain	2
7. slipped discs	1
8. strains	1
9. art	1
10. veterinary medicine	1
11. stage	1
12. hip replacements	1
13. fillings	1
14. drill	1
15. teeth	1
16. condition	1
17. brushing	1
18. toothpaste	1
19. decay	1
20. discs	1
21. muscle	1
22. tendon	1
23. hip	1
24. replacement	1
25. dentist	1

0.40 Sunbed Salesman

1. father	1
2. son	1
3. week	1
4. salesman	4
5. refund	1
6. boot	1
7. car	1
8. lorry park	1
9. trailer	3
10. poland	1
11. day	1
12. luck	1
13. sides	1
14. canvas	1
15. hole	1
16. watch strap	1
17. sun-bed	1
18. sun-bed salesman	1
19. lorry	1
20. park	1
21. trailer sides	1

Appendix F: Tabulate Verb-Entity List

This is an attempt to cross-reference all the discourse entities by the verbs that used them. After anaphor resolution has taken place each verb (roughly corresponding one to one to every clause in the text) is listed with the entities filling its thematic roles.[9] At the end of each subsection there is another table, detailing the entities that occurred with the most verbs, i.e. in the most different clauses.

0.41 John and the Rabbit

VERB ENTITY	ARGUMENTS
be-vb-1	boy-1
be-vb-2	name-2
	boy-1
wake-vb-3	boy-1
shine-vb-4	sun-4
dress-vb-5	boy-1
go-vb-6	garden-6
	boy-1
sing-vb-7	birds-7
see-vb-8	rabbit-8
	boy-1
hop-vb-9	lawn-9
	rabbit-8
watch-vb-10	boy-1
	rabbit-8
go-vb-11	field-11
	rabbit-8
go-vb-12	boy-1
	rabbit-8
eat-vb-13	grass-13
	rabbit-8
eat-vb-14	dandelion-14
	rabbit-8
creep-vb-15	boy-1
	rabbit-8
see-vb-16	boy-1
	rabbit-8
hop-vb-17	rabbit-8
follow-vb-18	boy-1
walk-vb-19	boy-1
stop-vb-20	rabbit-8
eat-vb-21	dandelion-21
	rabbit-8
creep-vb-22	boy-1
	rabbit-8
be-vb-23	rabbit-8
flee-vb-24	rabbit-8
run-vb-25	boy-1
	rabbit-8
arrive-vb-26	hedge-26
	rabbit-8
run-vb-27	burrow-27
	rabbit-8
be-vb-28	boy-1
disappear-vb-29	rabbit-8
be-vb-30	boy-1
return-vb-31	house-31
	boy-1

Treating any more than 1 activation as significant gives the following table:

ENTITY	TIMES ACTIVATED
rabbit-8	19
boy-1	17

[9] Here taken to be: the Verb Complement Theme, the Theme, the Instrument (aka the Goal or the Locative), and the Agent.

0.42 Destroyers

VERB ENTITY	ARGUMENTS
be-vb-1	destroyers-1
	ships-1
	fleet-1
be-vb-2	destroyers-1
	ships-2
have-vb-3	destroyers-1
	guns-3
	torpedoes-3
have-vb-4	american-destroyers-4
	fleet-4
be-vb-5	american-destroyers-4
	pacific-5
be-vb-6	american-destroyers-4
	atlantic-6
fire-vb-7	american-destroyers-4
	guns-7
	shells-7
be-vb-8	guns-7
	range-8
range-vb-9	american-destroyers-4
	torpedoes-9
have-vb-10	american-destroyers-4
	russians-10
	destroyers-1
be-vb-11	russian-destroyers-10
	atlantic-6
be-vb-12	russian-destroyers-10
	pacific-5
be-vb-13	russian-destroyers-10
	guns-13
fire-vb-14	guns-13
	shells-14
be-vb-15	guns-13
	range-15
have-vb-16	torpedoes-16
	russians-10
range-vb-17	torpedoes-16
have-vb-18	russians-10
	russian-destroyers-10
have-vb-19	americans-19
	american-destroyers-4
be-vb-20	russian-destroyers-10
be-vb-21	russian-destroyers-10
	american-destroyers-4
have-vb-22	russian-destroyers-10
	weapons-22

Treating any more than 1 activation as significant gives the following table:

ENTITY	TIMES ACTIVATED
american-destroyers-4	8
russian-destroyers-10	7
destroyers-1	4
russians-10	3
guns-13	3
pacific-5	2
atlantic-6	2
torpedoes-16	2
guns-7	2

0.43 Biographies

VERB ENTITY	ARGUMENTS
be-vb-1	books-1
	biographies-1
be-vb-2	things-2
	biographies-1
tell-vb-3	biographies-1
	story-3
be-vb-4	biographies-1
	people-4
have-vb-5	lessons-5
	readers-5
	biographies-1
write-vb-6	smiles-biogs-6
	smiles-6
be-vb-7	smiles-biogs-6
	workers-7
inspire-vb-8	readers-8
	smiles-biogs-6
imitate-vb-9	smiles-biogs-6
	workers-7
	readers-8
be-vb-10	things-10
	history-bks-10
tell-vb-11	stories-11
	history-bks-10
be-vb-12	states-12
	history-bks-10
be-vb-13	people-13
	history-bks-10
inspire-vb-14	readers-14
	history-bks-10
tell-vb-15	novels-15
	story-15
be-vb-16	characters-16
	novels-15
be-vb-17	novels-15
	things-17
be-vb-18	novels-15
be-vb-19	biographies-1
	books-19
be-vb-20	history-bks-10
be-vb-21	books-21
	novels-15
give-vb-22	biographies-1
	children-22
give-vb-23	children-22
	history-bks-10
give-vb-24	children-22
	novels-15

Treating any more than 1 activation as significant gives the following table:

ENTITY	TIMES ACTIVATED
biographies-1	7
history-bks-10	7
novels-15	6
smiles-biogs-6	4
children-22	3
workers-7	2
readers-8	2

0.44 Dublin

VERB ENTITY	ARGUMENTS
shoot-vb-1	robbers-1
end-vb-2	chase-2
	battle-2
	centre-2
rule-vb-3	ira-involvement-3
	detectives-3
begin-vb-4	drama-4
hold-up-vb-5	robbers-1
	bank-5
	Dublin-5
make-off-vb-6	robbers-1
	2500-6
spot-vb-7	car-7
	detectives-7
	city-edge-7
follow-vb-8	chase-8
come-under-vb-9	robbers-1
	car-9
	fire-9
end-vb-10	chase-8
try-vb-11	robbers-1
	car-11
block-vb-12	car-11
	road-12
follow-vb-13	shoot-out-13

arrive-vb-14	gardai-14
**********	
report-vb-15	robbers-1
	hospital-15
**********	
shot-16	shot-16
	gardai-16

Treating any more than 1 activation as significant gives the following table:

ENTITY	TIMES ACTIVATED ACTIVATED/MENTIONED
robbers-1	6
chase-8	2
car-11	2

0.45 The Beautiful Game

VERB ENTITY	ARGUMENTS
**********	
become-vb-1	humans-1
**********	
cultivate-vb-2	readers-2
	minds-2
	bodies-2
**********	
provide-vb-3	cultivate-vb-2
	opportunity-3
	sport-3
**********	
rank-vb-4	sport-3
	activities-4
**********	
be-vb-5	competition-5
**********	
develop-vb-6	competition-5
	abilities-6
	people-6
**********	
appreciate-vb-7	abilities-6
	people-7
**********	
admire-vb-8	sports-people-set-8
	people-8
**********	
develop-vb-9	sports-people-set-8
	loyalties-9
**********	
represent-vb-10	readers-10
	identity-10
	competition-10
	individual-10
	group-10
	feeling-10
**********	
allow-vb-11	readers-10
	feelings-11
	competition-5
	competitions-11
**********	
lead-vb-12	feelings-11
	confrontations-12
	sport-3
	safety-valve-12

Treating any more than 1 activation as significant gives the following table:

ENTITY	TIMES ACTIVATED
competition-5	3
sport-3	3
feelings-11	2
abilities-6	2
sports-people-set-8	2
cultivate-vb-2	2
readers-10	2

0.46 Titanic

VERB ENTITY	ARGUMENTS
**********	
accuse-vb-1	captain-1
	californian-1
	liner-1
	titanic-1
**********	
examine-vb-2	captain-1
	tragedy-2
	conduct-2
**********	
sink-vb-3	loss-3
	lives-3
	titanic-1
**********	
hit-vb-4	titanic-1
	voyage-4
	ice-berg-4
**********	
follow-vb-5	captain-1
	name-5
	californian-1
	campaign-5
	inquiry-5
**********	
accuse-vb-6	captain-1
	not-acting-6
	inquiry-6
**********	
be-vb-7	captain-1
	californian-1
**********	
find-vb-8	inquiry-6
**********	
be-vb-9	titanic-1
	californian-1
**********	
ignore-vb-10	californian-1
	rockets-10
**********	
have-vb-11	californian-1
	radio-11
**********	
insist-vb-12	captain-1

**********	
confirm-vb-13	insist-vb-12
	location-13
	wreck-13
	claim-13
	evidence-13

ENTITY	TIMES ACTIVATED
explorers-1	8
ice-6	2
shortage-8	2
attempt-1	2

Treating any more than 1 activation as significant gives the following table:

ENTITY	TIMES ACTIVATED
captain-1	6
californian-1	6
titanic-1	4
inquiry-6	2
insist-vb-12	2

0.47 North Pole

VERB ENTITY	ARGUMENTS
**********	
abandon-vb-1	attempt-1
	explorers-1
	pole-1
	men-1
	miles-1
**********	
halt-vb-2	explorers-1
	attempt-1
	water-2
	days-2
	miles-2
**********	
score-vb-3	explorers-1
	achievement-3
**********	
raise-vb-4	explorers-1
	trip-4
	400K-4
	research-4
**********	
lose-vb-5	explorers-1
	lives-5
	times-5
**********	
fall-vb-6	explorers-1
	ice-6
**********	
save-vb-7	water-7
	explorers-1
**********	
defeat-vb-8	explorers-1
	pain-8
	shortage-8
	supplies-8
**********	
force-vb-9	shortage-8
	ice-6
	help-9

Treating any more than 1 activation as significant gives the following table:

0.48 Cracked Up Kids

VERB ENTITY	ARGUMENTS
use-vb-1	boys-1
	runners-1
	dealers-1
make-vb-2	crack-2
	flats-2
give-vb-3	crack-2
	boys-1
sell-vb-4	crack-2
	walkways-4
	street-4
	boys-1
wrap-vb-5	crack-2
	film-5
carry-vb-6	crack-2
	boys-1
	mouths-6
swallow-vb-7	crack-2
	boys-1
approach-vb-8	boys-1
	police-8
exposed-vb-9	people-9
	world-9
reveal-vb-10	school-10
	survey-10
approach-vb-11	pupils-11
	dealers-11
warn-vb-12	schools-12
	dangers-12
	pupils-12
	virus-12
	hepatitis-12
	syringes-12
	needles-12
need-vb-13	posters-13
	boxes-13
dispose-vb-14	equipment-14

Treating any more than 1 activation as significant gives the following table:

ENTITY	TIMES ACTIVATED
crack-2	6
boys-1	6

0.49 Animal Medicine

VERB ENTITY	ARGUMENTS
come-vb-1	acupuncture-1
	surgery-1
	dogs-1
	cats-1
treat-vb-2	acupuncture-1
	pain-2
help-vb-3	acupuncture-1
	discs-3
	pain-3
	strains-3
enter-vb-4	stage-4
	medicine-4
become-vb-5	acupuncture-1
	dogs-5
	fillings-5
	replacements-5
keep-vb-6	condition-6
	teeth-6
	drill-6
prevent-vb-7	decay-7
	toothpaste-7
	brushing-7

Treating any more than 1 activation as significant gives the following table:

ENTITY	TIMES ACTIVATED
acupuncture-1	4

0.50 Sunbed Salesman

VERB ENTITY	ARGUMENTS
jail-vb-1	father-son-1
	week-1
kidnap-vb-2	father-son-1
	salesman-2
refuse-vb-3	refund-3
	salesman-2
	father-son-1
squeeze-vb-4	father-son-1
	car-4
	boot-4
	salesman-2
drive-vb-5	salesman-2
	park-5
bind-vb-6	salesman-2

**********	
gag-vb-7	salesman-2
**********	
force-vb-8	salesman-2
	trailer-8
**********	
tell-vb-9	salesman-2
**********	
go-vb-10	trailer-8
	day-10
	poland-10
**********	
go-vb-11	trailer-8
	salesman-2
**********	
change-vb-12	salesman-2
	luck-12
**********	
discover-vb-13	salesman-2
	canvas-13
	sides-13
**********	
escape-vb-14	salesman-2
**********	
cut-vb-15	salesman-2
	strap-15
	hole-15

Treating any more than 1 activation as significant gives the following table:

ENTITY	TIMES ACTIVATED
salesman-2	13
father-son-1	4
trailer-8	3

Appendix G: Summaries for each text

This section pulls together the results for the three methods, and presents the kind of simplistic summary that could feasibly be constructed in each of the three cases. The Focus summaries are drawn from the results contained in Appendix D, and the entities mentioned in each are listed in decreasing order of frequency of the times they were actually in focus. The NP Count results come from Appendix E, the Active Entity results from Appendix F. These too are listed in descending order of frequency. The labels used in the Focus and the Active Entity sections are the labels constructed by the Focus Mechanism; the labels used in the NP Count section are simply the spelling of the NP as it occurs in the text.

(Only items with a frequency greater than one are seen as significant for inclusion in the summaries.)

0.51 John and the Rabbit

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	rabbit-8	17
	boy-1	14
Summary according to NP count:		
This text is about:	John	13
	rabbit	8
	dandelion	2
Summary based on the more active entities within the text:		
This text is about:	rabbit-8	19
	boy-1	17

0.52 Destroyers

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	russian-destroyers-10	7
	american-destroyers-4	6
	destroyers-1	4
	guns-13	3
	torpedoes-16	1
Summary according to NP count:		
This text is about:	destroyers	5
	ships	4
	miles	4
	Russians	3

	guns	3
	torpedoes	3
	fleet	2
	pacific	2
	atlantic	2
	shells	2
	range	2
Summary based on the more active entities within the text:		
This text is about:	american-destroyers-4	8
	russian-destroyers-10	7
	destroyers-1	4
	russians-10	3
	guns-13	3
	pacific-5	2
	atlantic-6	2
	torpedoes-16	2
	guns-7	2

0.53 Biographies

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	biographies-1	7
	history-bks-10	6
	smiles-biogs-6	5
	novels-15	4
	children-22	2
Summary according to NP count:		
This text is about:	biographies	5
	books	4
	readers	4
	things	3
	novels	3
	story	2
	people	2
	workers	2
Summary based on the more active entities within the text:		
This text is about:	biographies-1	7
	history-bks-10	7
	novels-15	6
	smiles-biogs-6	4
	children-22	3
	workers-7	2
	readers-8	2

0.54 Dublin

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	robbers-1	12
	car-11	3
	chase-8	1
Summary according to NP count:		
This text is about:	car	5
	chase	3
	detectives	2
	men	2
	Gardai	2
	city	2
	robbers	2
	car chase	2
Summary based on the more active entities within the text:		
This text is about:	robbers-1	6
	chase-8	2
	car-11	2

0.55 The Beautiful Game

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	sport-3	2
	abilities-6	2
	humans-1	2
	sports-people-set-8	2
	cultivate-vb-2	1
	competition-5	1
	feelings-11	1
	readers-10	1
Summary according to NP count:		
This text is about:	competition	3
	sport	2
	people	2
	feeling	2
Summary based on the more active entities within the text:		
This text is about:	competition-5	3
	sport-3	3
	feelings-11	2
	abilities-6	2
	sports-people-set-8	2

	cultivate-vb-2	2
	readers-10	2

0.56 Titanic

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	captain-1	6
	titanic-1	3
	californian-1	2
	inquiry-6	1
	insist-vb-12	1
Summary according to NP count:		
This text is about:	Captain	4
	Titanic	4
	Californian	3
	inquiry	3
	miles	2
	Captain Lord	2
Summary based on the more active entities within the text:		
This text is about:	captain-1	6
	californian-1	6
	titanic-1	4
	inquiry-6	2
	insist-vb-12	2

0.57 North Pole

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	explorers-1	8
	shortage-8	1
Summary according to NP count:		
This text is about:	North Pole	2
	miles	2
	water	2
	ice	2
	Fiennes	2
	Stroud	2
Summary based on the more active entities within the text:		
This text is about:	explorers-1	8
	ice-6	2
	shortage-8	2

	attempt-1	2

0.58 Cracked Up Kids

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	boys-1	7
	crack-2	6
Summary according to NP count:		
This text is about:	boys	2
	dealers	2
	crack	2
	drug	2
	schools	2
	pupils	2
Summary based on the more active entities within the text:		
This text is about:	boys-1	6
	crack-2	6

0.59 Animal Medicine

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	acupuncture-1	7
Summary according to NP count:		
This text is about:	dogs	4
	acupuncture	3
	cats	2
	pain	2
Summary based on the more active entities within the text:		
This text is about:	acupuncture-1	4

0.60 Sunbed Salesman

		FREQUENCY OF ITEM IN TEXT
Summary according to Focus:		
This text is about:	salesman-2	10
	father-son-2	3
	trailer-8	2
Summary according to NP count:		

This text is about:	salesman	4
	trailer	3

Summary based on the more active entities within the text:

This text is about:	salesman-2	13
	father-son-1	4
	trailer-8	3

Appendix H: Draw Entity Geography Chart for each Text

This was an attempt to graphically illustrate the places in the text that referenced a particular entity, and is included for completeness' sake. Each graph is composed of the following elements:

- X-axis : the verb/clause/sentence number
- Y-axis : the discourse elements, arranged in the order preferred by the Default Expected Algorithm (See Appendix B). So the new DEs for each clause are lined up along the Y-Axis in the following order:

  THEME from verb complement (if exists)

  THEME, (INSTRUMENT or GOAL or LOCATIVE)

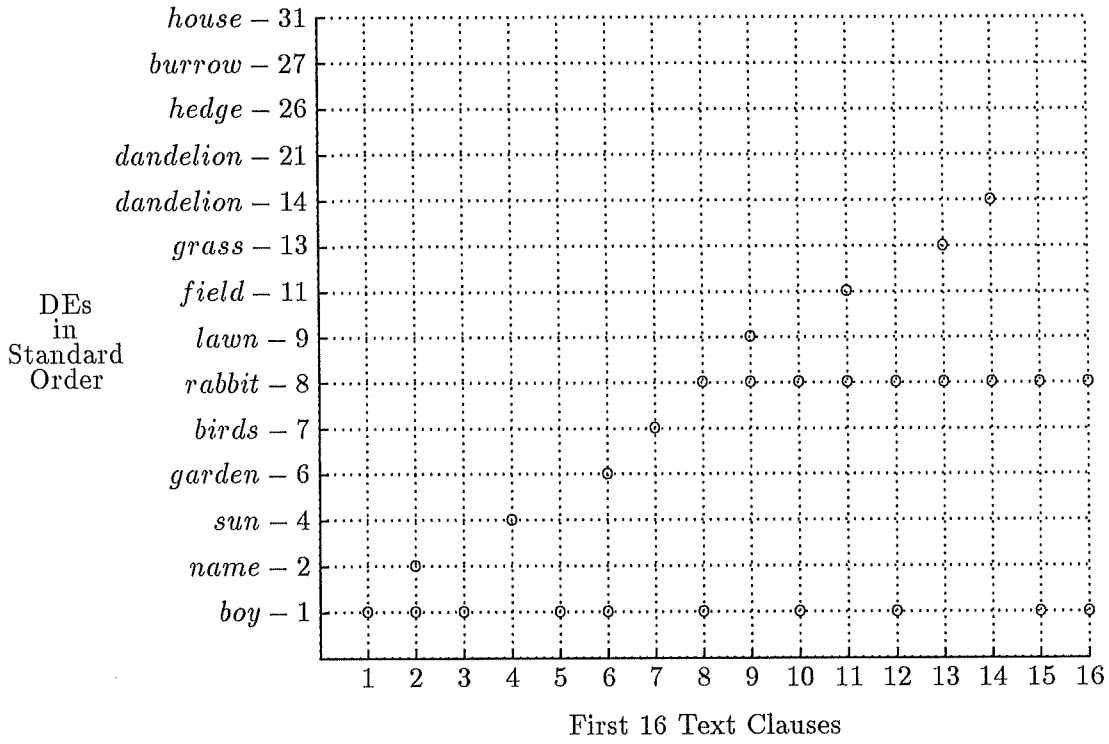
  AGENT, WHOLE VERB PHRASE

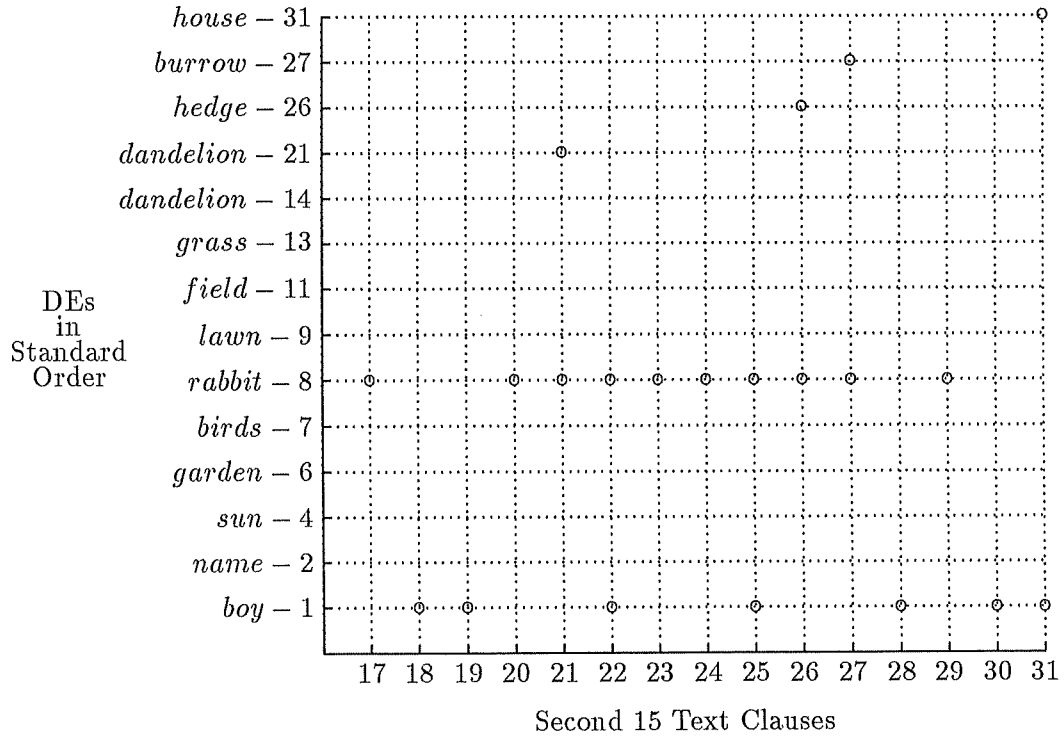
- Points : a mention of a particular discourse entity in a particular clause

I find that the information contained in these graphs is intuitively useful, but of a rather unclassifiable character. I experimented with several different methods of presentation before settling on the current simple plotting. One approach I tried was to connect all the points above the same verb entity with a vertical line, in an attempt to emphasise the spread of the entities in each verb. Roughly , the longer the vertical line was for any verb, the wider its reference. (In terms of referring to entities in far flung parts of the discourse.). A further step was to attempt to find some mid point for each of these vertical lines and to connect them, in an attempt to represent the growth of the text in terms of the new entities introduced. Roughly, the higher the gradient of these secondary lines the more new material is coming in.

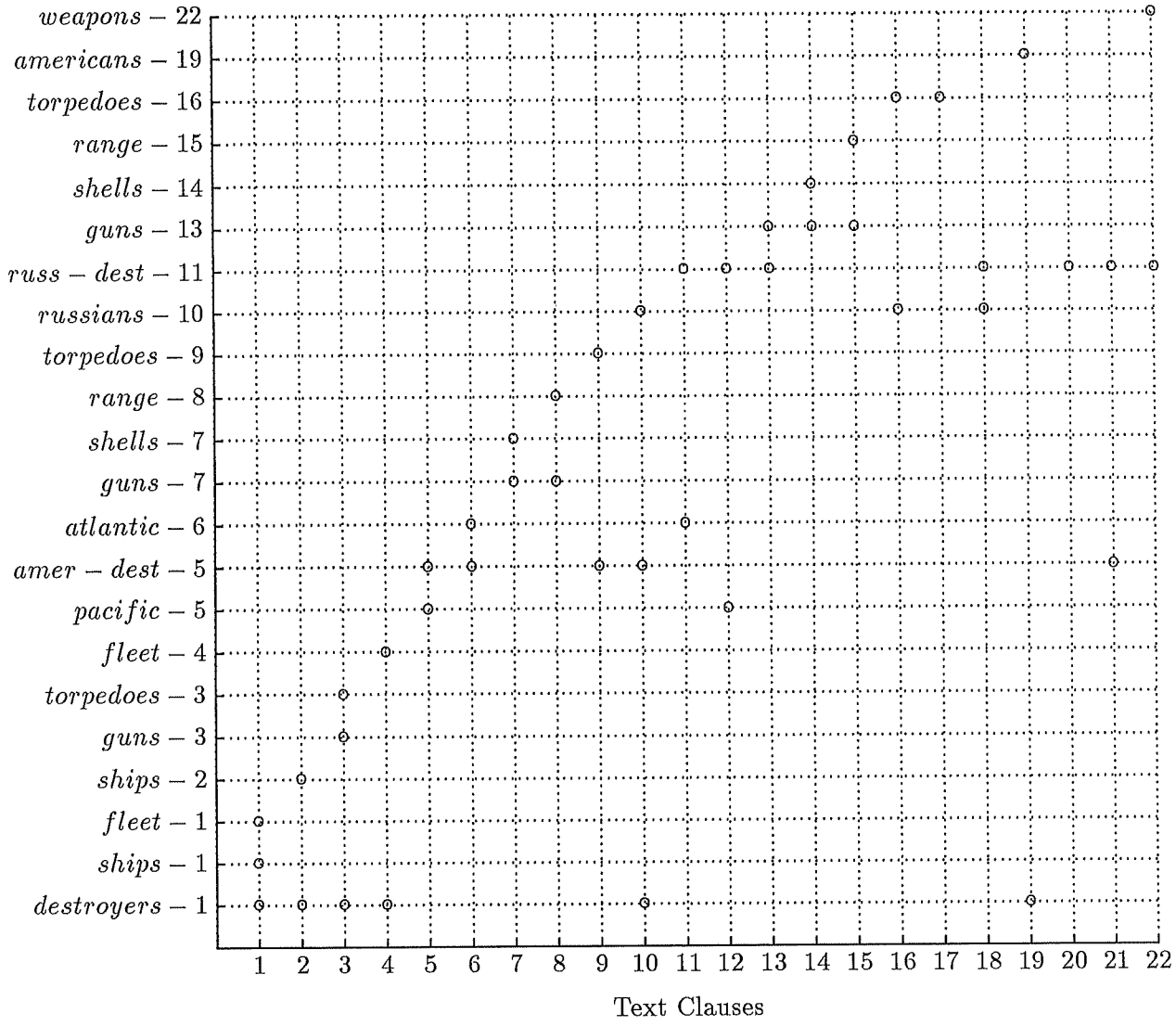
This particular feature of the experiment was not a success; on reflection I decided that it was not particularly meaningful to construct gradients or midpoints in this way. Yet the clustering on these charts does seem, by virtue of the fact that the patterns are more or less similar in each case, to have some significance. This could be a direction for future work.

0.61 John and The Rabbit

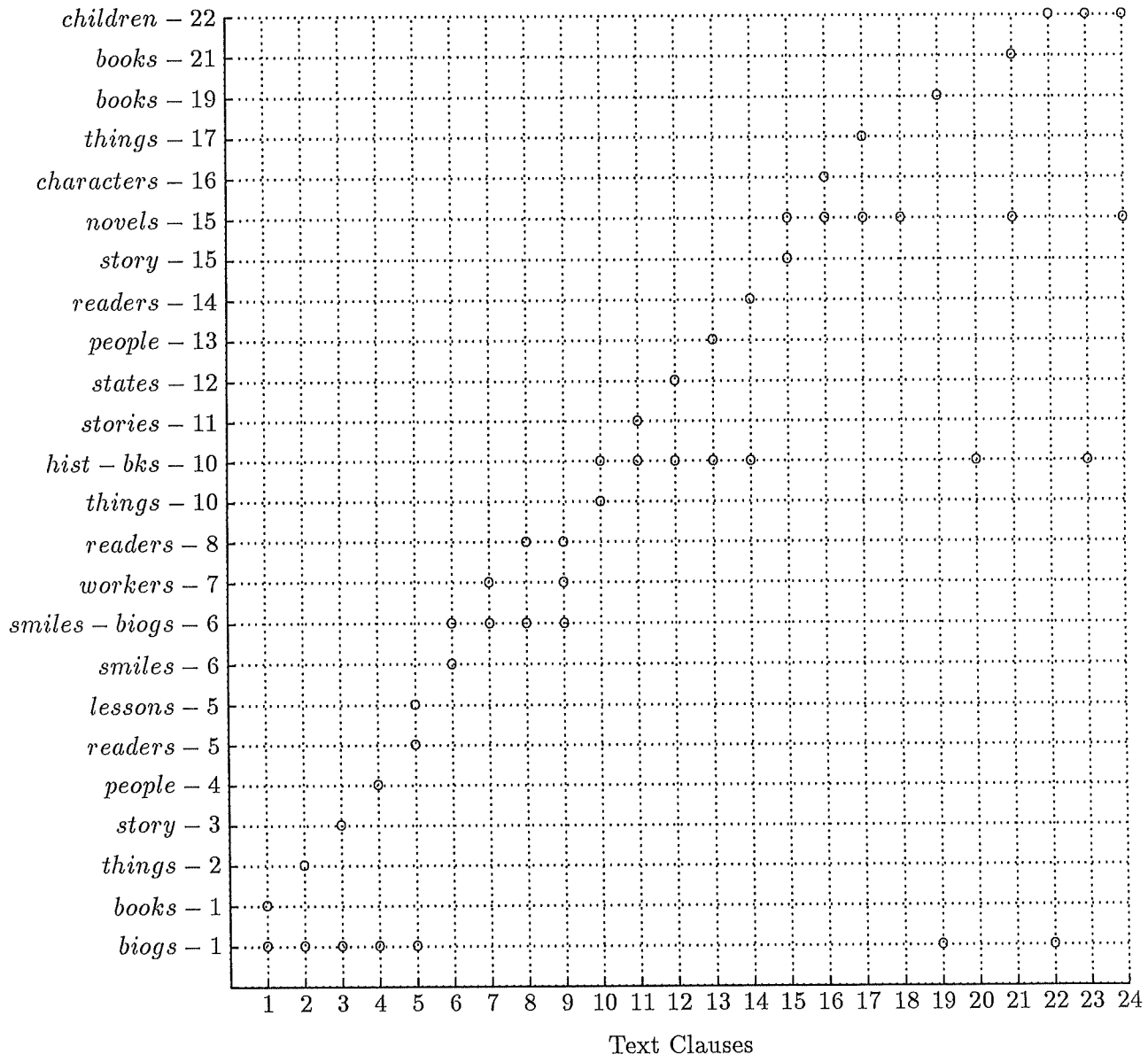




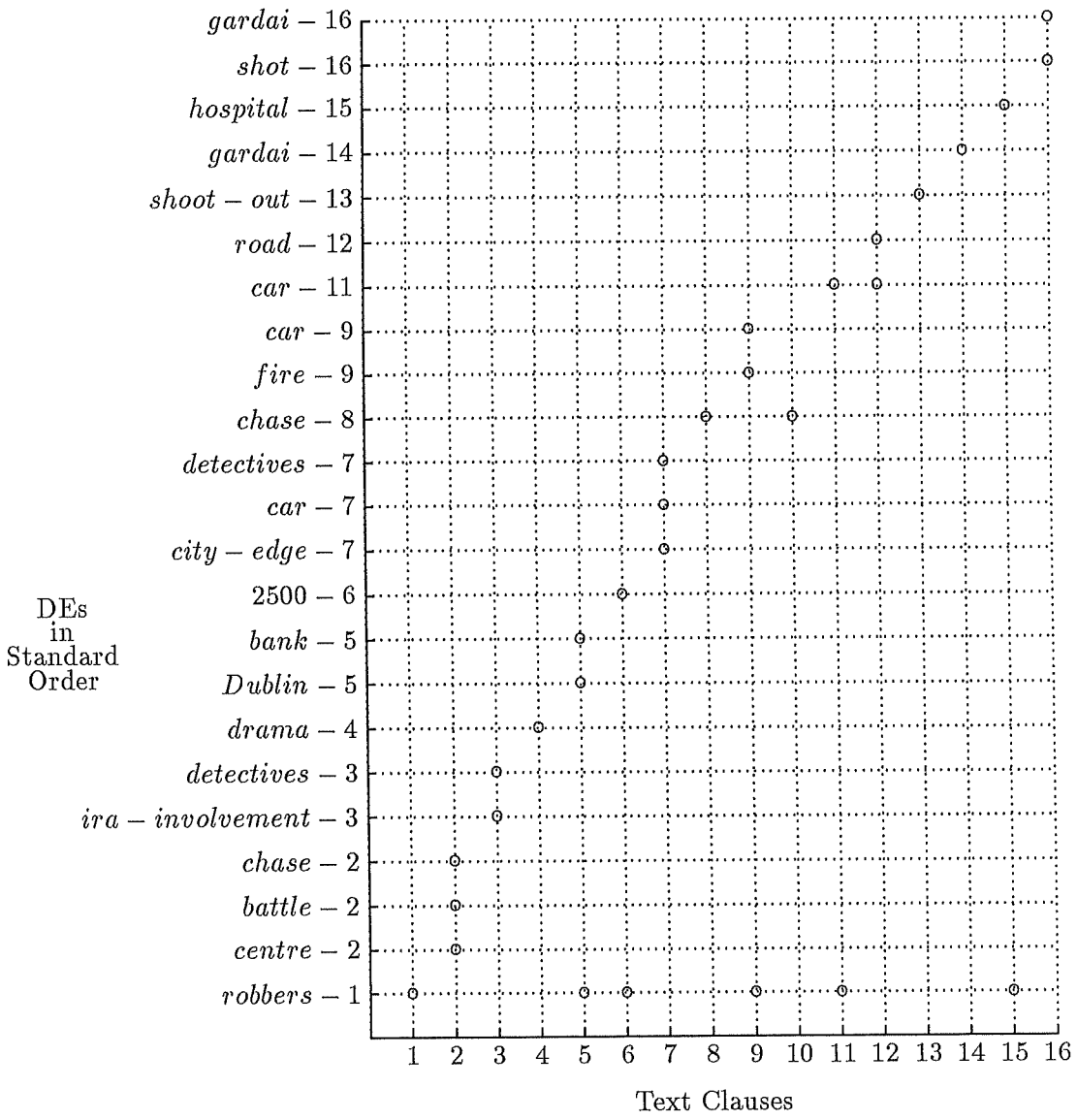
0.62 Destroyers



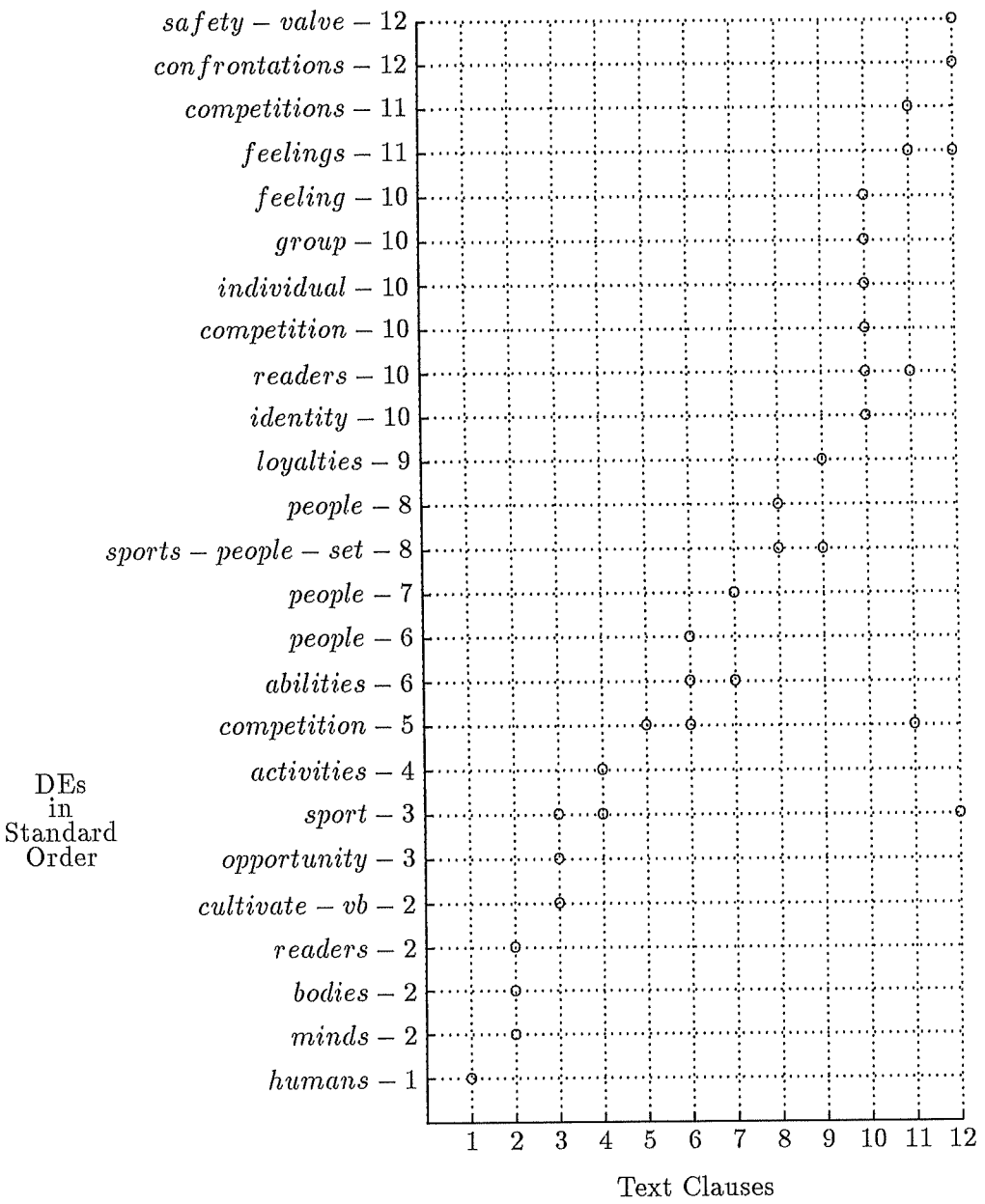
0.63 Biographies



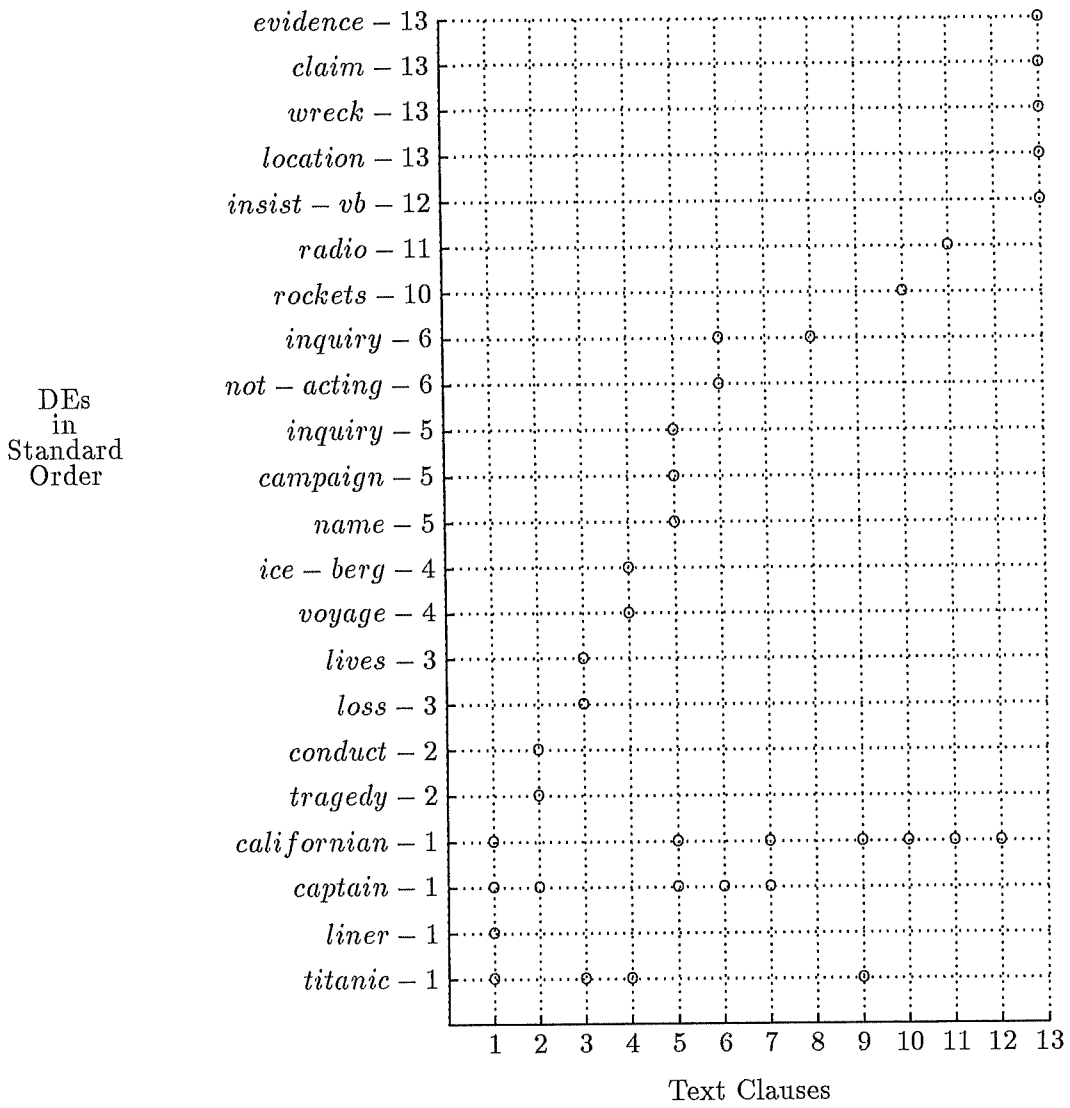
0.64 Dublin


gardai – 16
shot – 16
hospital – 15
gardai – 14
shoot – out – 13
road – 12
car – 11
car – 9
fire – 9
chase – 8
detectives – 7
car – 7
city – edge – 7
2500 – 6
bank – 5
Dublin – 5
drama – 4
detectives – 3
ira – involvement – 3
chase – 2
battle – 2
centre – 2
robbers – 1
DEs in Standard Order
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Text Clauses

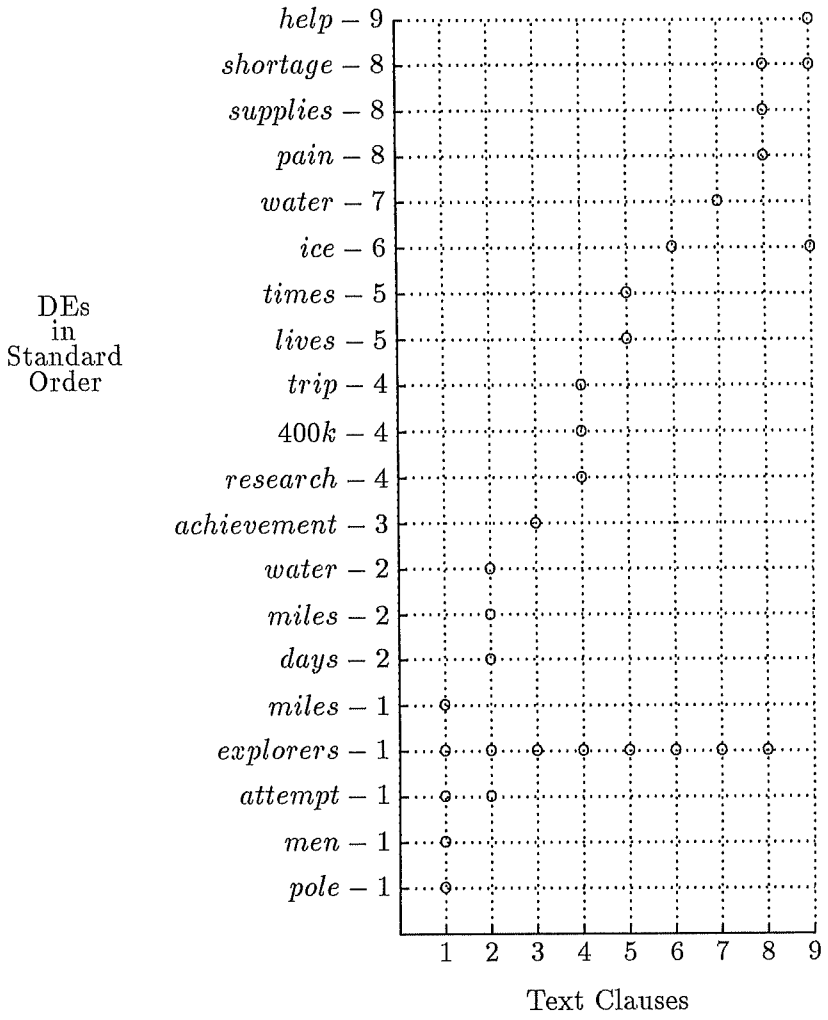
0.65 The Beautiful Game



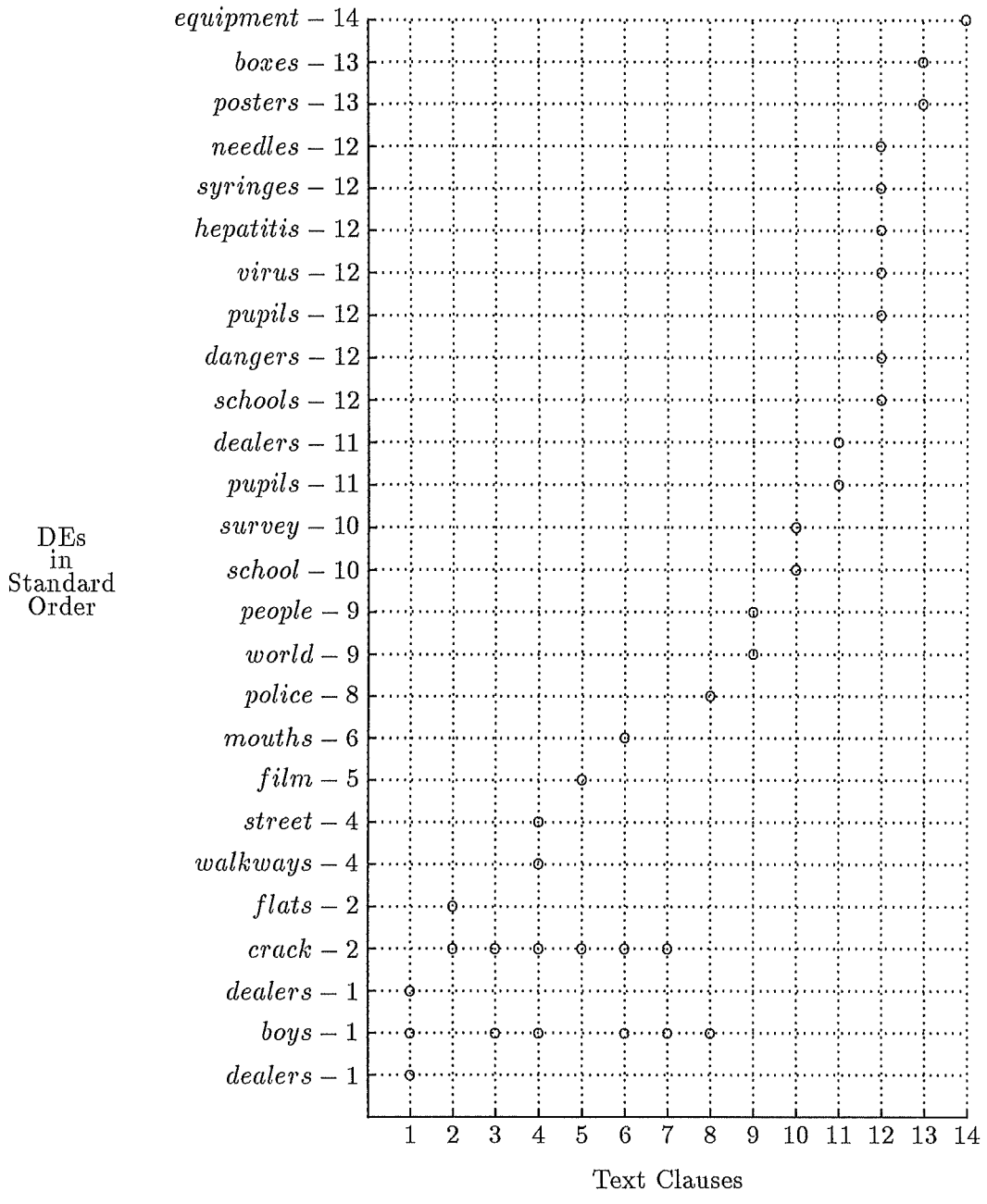
0.66 Titanic


evidence – 13
claim – 13
wreck – 13
location – 13
insist – vb – 12
radio – 11
rockets – 10
inquiry – 6
not – acting – 6
inquiry – 5
campaign – 5
name – 5
ice – berg – 4
voyage – 4
lives – 3
loss – 3
conduct – 2
tragedy – 2
californian – 1
captain – 1
liner – 1
titanic – 1
DEs in Standard Order
1 2 3 4 5 6 7 8 9 10 11 12 13
Text Clauses

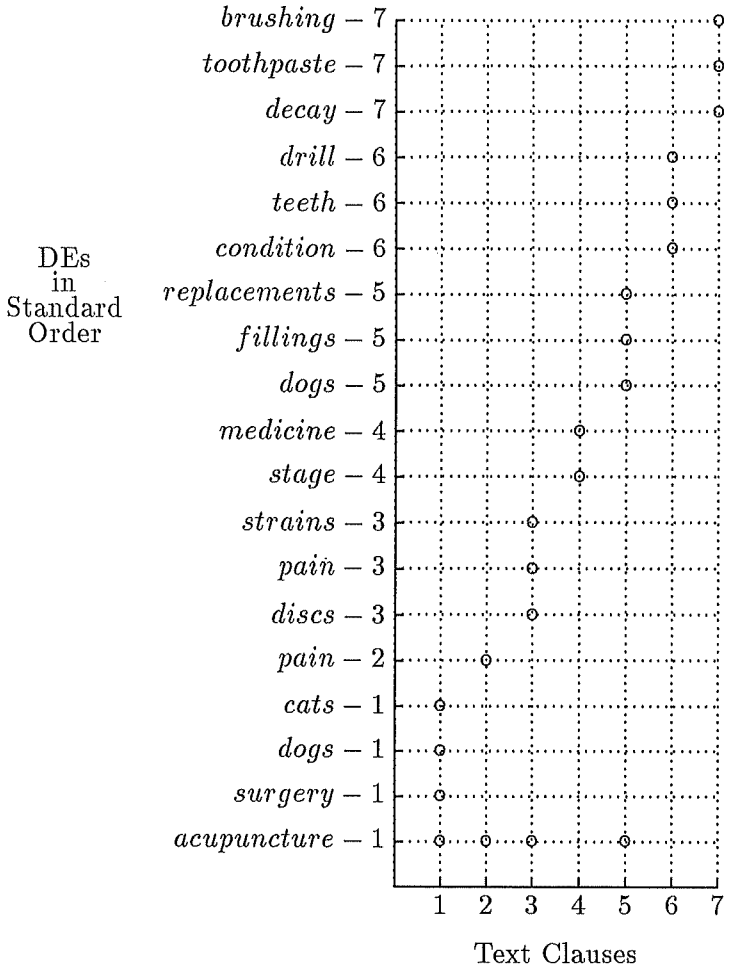
0.67 North Pole



0.68 Cracked Up Kids



0.69 Animal Medicine



0.70 Sunbed Salesman