

Syntactic Information Hiding in Plain Text

Brian Murphy

M.Phil in Speech and Language Processing

2001

Declaration

I declare that this dissertation has not been submitted as an exercise for a degree at this or any other university and that is entirely my own work.

I agree that the Library may lend or copy this dissertation on request.

Signed:

Date: 8th October 2001

Abstract

Syntactic Information Hiding in Plain Text

Brian Murphy

Information hiding has mostly been concerned with robustly embedding data in rich media types such as audio, image and video files. Some attempts have been made to hide data in the layout and formatting details of text documents, but these can easily be defeated with OCR software. In this paper I suggest some syntactic alternations that preserve both meaning and style. As part of an automated system, they could be used to invisibly store data in the fabric of a text, without affecting the integrity of the document.

In this paper I investigate some candidate structures that I believe could be used in such a system, and then test them as prototype algorithms on a small hand-parsed corpus of English sentences.

The results show that all suggested structures, bar one, are good candidates for such a system, and warrant further testing with larger corpora and more subjects to judge their acceptability.

Acknowledgements

Huge thanks to my supervisor, Carl 'Walking Corpus' Vogel, who was always able to come up with weird and wonderful, but also critically useful sentences to challenge my assumptions. His enthusiasm, dedication and willingness to help were unstinting, and this paper would not be what it is without him.

Many thanks too to Martin Emms, who helped me in the choice of a suitable corpus and, more importantly, with what to do with it afterwards.

Table of Contents

Introduction.....	7
Information Hiding	10
What is Information Hiding?	10
Applications of Information Hiding	10
Information Hiding in Digital Media.....	12
Information Hiding in Text.....	13
Formatted Text.....	13
Plain Text	14
Syntactic Marking.....	17
Design Considerations.....	19
Considerations for Individual Marking Structures	20
Considerations for Specific Applications.....	21
Analysis of Marking Transforms.....	24
Introduction	24
Grammatical Theory	24
General Constraints.....	27
Relative Clauses.....	28
Switching <i>which/who/whom</i> with <i>that</i>	30
Dropping or Inserting <i>which/who/whom/that</i>	32
Reducing and Expanding BE Relative Clauses.....	33
Adjunct Movement	39
<i>It</i> Insertion	42
Extraposition	42
Raising	44
<i>There</i> Insertion.....	48
Conjunction Argument Switching.....	54
Phrasal Verb Particle Movement	56
Verification of the Transforms	59
Introduction	59
The <i>SUSANNE</i> Corpus	60
Methodology	61
The <i>SUSANNE</i> Scheme.....	62
Relative Clauses.....	65
Relative Clause <i>That/Wh...</i> Switch Rule	65
Relative Clause <i>That/Wh...</i> Drop and Insertion Rule	74
Relative Clause Reduction Rule	76
Relative Clause Expansion Rule	80
Adjunct Movement Rule	83
<i>It</i> Insertion	86
Extraposition Rule.....	86
Subject Raising Rule.....	90
Object Raising Rule	94
<i>There</i> Insertion and Deletion Rule.....	96
Conjunction Movement Rule	99
Verb Particle Movement Rule.....	102
Conclusion.....	107
Results.....	107
Suitability for Information Hiding Applications	109
Would it All Work?	111
What Next?	113
Appendices.....	116
Appendix A: Notational Conventions	116

Appendix B: Notational Conventions of the SUSANNE Corpus	117
Appendix C: Sources of the SUSANNE Corpus	119
Appendix D: Verification Examples from the SUSANNE Corpus	123
Appendix D1: Relative Clause <i>That/Wh...</i> Switch Rule.....	123
Appendix D2: Relative Clause <i>That/Wh...</i> Drop and Insertion Rule	128
Appendix D3: Relative Clause Reduction Rule	133
Appendix D4: Relative Clause Expansion Rule.....	137
Appendix D5: Adjunct Movement Rule	139
Appendix D6: Extraposition Rule	143
Appendix D7: Subject Raising Rule	147
Appendix D8: Object Raising Rule	151
Appendix D9: <i>There</i> Insertion and Deletion Rule	151
Appendix D10: Conjunction Movement Rule	154
Appendix D11: Verb Particle Movement Rule	160
References	163

Introduction

Information hiding encompasses two main applications. One is watermarking, where small amounts of attribution data are hidden in a piece of intellectual property to legally establish ownership. The data should be impossible to remove, and invisible to a normal user.

Steganography is the application of similar methods to achieve secret communication. A cover file of some kind is transmitted from sender to recipient, with a hidden message inside. Anyone intercepting the file should be oblivious to the presence of the hidden data.

Information hiding has mostly been concerned with robustly embedding data in rich media types such as audio, image and video files. Some attempts have been made to hide data in the layout and formatting details of text documents, but these can easily be defeated with OCR software.

The objective in this paper is to investigate a new technique to hide data in the fabric of plain text. The data should remain intact whatever the form of the media – spoken, written, electronic or printed.

To do this I will be looking at syntactic alternations that preserve both meaning and style. As part of an automated system, they could be used to invisibly store bits of data in a text, without affecting the integrity of the document. Both literal meaning and the intention of the author should not be compromised.

An example of a candidate alternation is the swapping of co-ordinate conjuncts. For example in the following sentence (1a) we can see three co-ordinate constructions that could have their conjuncts swapped. A typical scheme could encode 1's by changing the constructions, and 0's by leaving them as they are. Thus the processed version (1b) of the sentence below would encode the sequence of bits '101'.

- (1) (a) We looked [under the beds and behind the couches]₀, [in the shed and up in the attic]₀, and through all the [wardrobes and cupboards]₀.

(b) We looked [behind the couches and under the beds]₁, [in the shed and up in the attic]₀, and through all the [cupboards and wardrobes]₁.

The only previous work published on this approach is Atallah, Raskin et al (2001), targeted specifically at watermarking. Their approach is to identify syntactic alternations, such as passivisation or clefting, that change meaning ‘slightly’. They reason that a truly meaning preserving operation is vulnerable to hostile parties, because it could be indiscriminately applied to a file, thus scrambling the watermark. Their solution is this: applying a small number of ‘slightly’ meaning transforming operations will not compromise the document, but applying them globally would damage it unacceptably.

My approach is somewhat different. Firstly, one of my target applications is steganography, where the primary concern is not the robustness of the embedded data (ie how difficult the data is to extract or remove), but that the hidden data remains undetected. Secondly, as regards watermarking, I believe that other constructions can be found, that are totally meaning-preserving, but are irreversible. Thus a scrambled version of a file would still contain the irreversible part of the mark, which on comparison with the original unmarked file retained by the genuine owner, would establish that ownership.

The first part of this paper gives an introduction to information hiding in general, and its application to text in particular. I then describe four different information hiding applications and their special requirements, together with my criteria for individual syntactic constructions.

The next part of the paper introduces nine syntactic phenomena that I think would be suited to this job. Drawing on the literature I review the phenomena, and try formulate rules that would allow them to be performed automatically.

In the third part of the paper, I take these rules and implement them as prototype algorithms. To verify my hypothesised rules, I run these algorithms over a small hand-parsed corpus of

English sentences. By taking a sample of the results and checking them manually, I can evaluate how successful each algorithm is, and correct any flaws exposed.

The results are encouraging. All the structures investigated bar one prove to be good candidates for such a system. But as will be seen, the scale of the verification process limits the strength of my conclusions. Broader tests with a larger corpus, an automatic parser and multiple human subjects are warranted.

Information Hiding

What is Information Hiding?

Information hiding is a relatively new field of study in computer science. The first international workshop on the subject was held in Cambridge in 1996. It grew out of two separate branches of study: the protection of intellectual property; and cryptography, that is the science of codes.

Cryptography has been concerned with the development of increasingly complex code systems, that is methods of encryption. The objective is to guarantee privacy with a code that cannot be broken. An information hiding technique, steganography, though not new¹, presented a completely different approach to secret communication in the computational context. Here the message to be kept secret is hidden in a mass of apparently innocuous information, so that any eavesdropper is completely oblivious to its presence.

In intellectual property protection the conventional approach was to develop permission control systems, that would prevent unauthorised copying or viewing of a document. With the rapid growth of the internet, selling intellectual property over this medium became very attractive, but no effective permission systems were in place. Watermarking, that is the hiding of small amounts of attribution data in a file, for example to attest ownership, was suggested as a method to prevent abuses.

Applications of Information Hiding

Both watermarking and steganography demand that relatively small amounts of data can be embedded in larger files without being perceptible to the ordinary user. Typically the carrier file would be a media file: thus an audio file should not be audibly distorted by the embedding process; an image or video file should remain unchanged to the eye. Another common

¹ For example Anderson & Petitcolas (1998, p.474) mention a method used in medieval Europe, where a wooden grille template was laid over a long dispatch to reveal only the words that made up the hidden message

requirement is that of robustness. Given that a hostile party knows or suspects that a file contains hidden data, they should be prevented from recovering, removing, corrupting or replacing the embedded data.

In steganography (Gr. *hidden writing*), an apparently innocent 'cover' document is used to conceal a confidential message. In contrast to encryption, a hostile party intercepting the processed document (referred to as a 'stego-document') should be oblivious to the presence of hidden information. The requirement of steganography for imperceptibility is particularly exacting. Even if a hostile party cannot isolate and decode a hidden message, once they know steganographic methods are being used, they may be able to prevent communication by intercepting and destroying carrier documents.

Watermarking applications typically insert information describing the originator of a media document, so that intellectual ownership can be legally established. Here the primary requirements are that the mark is imperceptible to the legitimate end user, and that it is robust to attempts to remove or replace it.

A particular application of watermarking, termed fingerprinting, aims to embed data that uniquely identifies each copy of a document distributed. Then end-user display applications can be built to restrict use of media files to authorised users. Fingerprints can also be used for tracing consumers responsible for unauthorised duplication (ie pirating) of copyrighted material.

Another suggested application for fingerprinting is traitor tracing, where copies of confidential documents are individually marked. In the case of a breach of confidentiality, the source of the leak can be determined.

While steganography and watermarking are sometimes seen as alternatives to earlier coding and permission systems, they are often used in a complementary role. For example in

steganographic systems, the message to be hidden is often encrypted first, to provide an extra level of protection against detection and decoding.

Information Hiding in Digital Media

To date most research work in information hiding has concentrated on image, audio and video files (Anderson & Petitcolas, 2001). On the one hand this is because music and film are potentially the most lucrative source of online income for media companies. But they also prove to be particularly amenable to information hiding. Image, audio and video material usually include a fair measure of noise – that is essentially superfluous random data originating in the digitisation process that the user is oblivious to – forming the perfect ‘space’ for information hiding. They also typically contain more information than is perceptible to the human senses – for example, the human eye cannot distinguish all 16 million colour tones possible in a 24-bit RGB image, and the ear, while sensitive to frequency content, is relatively oblivious to the phase information present in audio files.

Some early methods of information hiding in media files simply embedded data directly into the least significant bit (LSB) of each image pixel or sound sample (Shaw 1999, sec 4.2). Such methods were easily detected, since the statistical profile of the LSB data stream is dramatically altered. Consider that, in a typical image, tones of colour generally change rather gradually over a sequence of pixels, while embedded textual data would have an unnaturally random or jagged profile.

Subsequent methods use various techniques to store the information more discreetly and securely. The data is often encrypted before insertion, masking the statistical properties of text, and at the same time making the message more difficult to decode. Some methods simply ‘stretched out’ the hidden data sequence, to give a more gradual and natural profile, while others distribute the data pseudo-randomly through the file. In order to extract the data,

a 'key'² is used that points to the various apparently random locations in the file where hidden data is stored (Bender et al 1996, p.320, 326). Spread spectrum information hiding, which only applies to media with a time domain, such as audio and video, simply overlays the hidden data, but placed in rapidly and pseudo-randomly changing frequency bands.

Information Hiding in Text

The approaches just described are difficult to graft onto the problem of data hiding in text, since, ideally, text does not possess the 'redundancy', as Bender et al put it (1996, p.332), of other media types. Katzenbeisser (2000, p.37) considers it an open question whether plain text steganography can be achieved, while Su et al (1998, p.12) believe it impossible:

"Raw text, such as an ASCII text file or computer source code, cannot be watermarked because there is no 'perceptual headroom' in which to embed hidden information."

Formatted Text

This has not stopped researchers from trying to find the 'perceptual headroom' in the layout of formatted file types such as PDF, PostScript or TeX. Various techniques have been suggested that make minute adjustments to aspects of visual formatting, such as choice of font, shape of individual letters, or the space between letters, words and lines. A notable example is Brassil et al (1994, p.11) who describe a technique where they selectively varied interline spacing by 1/300th of an inch to encode 1's and 0's. The method proved quite robust, even when faxed or photocopied repeatedly.

² A key is usually a long arbitrary numerical sequence that causes an algorithm to perform in a certain way, akin to the seed of random number generation routines. For example in encryption, the algorithm is often in the public domain, but a key held in private by both communicating parties will allow the sender to encrypt a message in a way that only the intended recipient can decode. Similarly, in information hiding, a key is sometimes used to point to a pseudo-random sequence of locations in the carrier file, where the information is stored. As with the grille mentioned in the last footnote, without the key, you do not know where to start looking.

While successful, all these methods are limited in their applicability by their dependence on particular file formats. And the fine details of formatting exploited can be inadvertently lost (or intentionally purged) during file format conversion. Furthermore if the value of a text justifies it and the length allows it, a text can be rescanned using OCR (Optical Character Recognition) software, or even simply retyped.

Plain Text

When steganography is the sole objective, one solution is to generate a cover text from scratch, since its surface semantic content is irrelevant – the so-called mimic method (Wayner 1996). The *NICETEXT* application (Chapman & Davida, 2001) allows grammatical but nonsensical cover text to be randomly generated on demand to encode a secret message. Thus, the short hidden message ‘Hi’ is rendered rather verbosely (in this case in Shakespearean style, Wizard of Oz and Aesop’s Fables also being available) as:

NUT My heart in drums. Audit DAVY Where are you, Ridge John? Audit. Whatsoever, are you buried? Your folly. Say not watery. Believe me, on mine honour, My countermands express my purpose. And so was I; I bear you balcony. Deceive which fear you may, The night is long that never finds the today. Is she without Posthumus? A satchel above such win were felicity.

Chapman & Davida’s method uses context-free grammars and generation by example to produce the cover text. The original message is then recovered with a simple word-by-word code lookup function. While a text like this is unlikely to fool a human interceptor, it could be made to reflect the statistical properties of genuine English and so pass examination by the automatic interception routines that are reportedly used by intelligence agencies. In this way it is a very attractive alternative to encrypted communication, which ‘stands out like a sore thumb’ on the internet (BBC 2001).

However, some researchers have managed to find redundancy in plain text. The first two methods concern presentation, while the second two work with the content of a document:

- Open space formatting: adding redundant spaces and line returns at the end of paragraphs, or distributing inter-word spaces for line justification
- Orthographic formatting: using variations in spelling, punctuation, hyphenation, contractions, and abbreviations
- Synonym substitution: replacing words with alternatives that preserve meaning
- Syntactic substitution: replacing phrase structures with alternatives that preserve meaning

WbStego (Bender 2001) and *TextHide* (Compris 2001) are two commercial applications that use open-space techniques. *TextHide* uses multiple spaces between words and at the end of lines selectively to encode 1's and 0's. Besides the possibility of intentional or accidental reformatting mentioned above, the main weakness of this technique is that varying inter-word spaces are both noticeable and visually disturbing.

WbStego does not alter the number of spaces, but rather substitutes them with alternative encodings – spaces encoded with the standard hex value 0x20 represent 0's, while spaces encoded with 0x00 represent 1's (Bender, personal communication). While *WbStego* achieves a high message capacity of average 1 bit per eight bytes of cover text, it depends on document display applications treating 0x00 as a regular space. This does not apply across the board to all applications or platforms. For example on Microsoft platforms, the standard encoding for spaces is 0x20. The *Notepad* basic text editor displays 0x00 spaces correctly, while for the word processing applications *Word* and *Wordpad* it appears as " (the error character). On Linux (SuSE version 7.0) 0x00 appears as a regular space in the text editor, but is not displayed at all by the `more` command, and appears as '^@' in the `vi` command line editor.

Examples of orthographic variations are given below – (a) versus (b). I am not aware of any commercial application that uses this method. While the examples below successfully preserve meaning and are fully grammatical, they still present problems. Even small variations in style or punctuation usage will be noticeable to the reader.

(2) (a) Tom, Dick and Harry

- (b) Tom, Dick, and Harry
- (3) (a) Computer-based learning
- (b) Computer based learning
- (4) (a) Dr. Zhivago of DCU hasn't arrived
- (b) Doctor Zhivago of Dublin City University has not arrived

Content based methods are much more ambitious, in that they should survive attempts to remove marks through file reformatting, OCR or retyping. A mark could even remain intact if a text was read aloud and re-transcribed.

Winstein (1998) describes a proof-of-concept application *Tyrannosaurus Lex* that makes synonym replacements based on the *WordNet* semantic network, but this is not available for testing online. *TextHide* from the German firm Compris GmbH also uses synonym replacement, but only supports German language texts.

Although initially attractive, the method proves disappointing, owing to the relative lack of truly interchangeable synonyms in language. Most synonyms exhibit some variation in nuance or usage that is intimately bound to context. For example the (a) and (b) versions of the sentences below, while preserving 'literal' meaning, clearly do not say the same thing.³

- (5) (a) She gave me a cordial handshake
- (b) # She gave me a hearty handshake
- (6) (a) I went down the cream cookies (Cockney rhyming slang: *bookies*)
- (b) # I went down the cream biscuits

A quick test of the synonym replacement method used by *TextHide* (using the German article "It's down, it's down!" from the September 11th 2001 edition of the Berlin newspaper *taz*) clearly demonstrates the shortcomings of this approach. The first replacement below, of

³ Here the '#' symbol means that the variation is not semantically or pragmatically equivalent. Refer to Appendix A for the conventions I use for marking grammaticality judgements.

abgeriegelt 'cordoned off' with *abgesperrt* 'closed off' is perfectly acceptable. The next proves clumsy, the full form *Automobil* being out of place in a non-technical context. In the third example the substitution has changed the meaning of the phrase entirely.

- (7) (a) *Das FBI hat Gebäude abgeriegelt, ...*
The FBI cordoned off buildings ...
- (b) *Das FBI hat Gebäude abgesperrt, ...*
The FBI closed off buildings ...
- (8) (a) *... als die Autos auf der Straße, ...*
... as the cars on the street ...
- (b) *... als die Automobils auf der Straße, ...*
... as the automobiles on the street ...
- (9) (a) *... weder Schlaf noch Pause ...*
... neither sleep nor breaks ...
- (b) *... weder Schlaf noch Brotzeit ...*
... neither sleep nor snacks ...

One could argue that the patchy success of this synonym replacement scheme stems from the quality of its synonym list, and that an better list would lead to better results. However the success of this method also demands a degree of sensitivity to context and nuance that is not available to today's parsers.

Syntactic Marking

The final method we will look at is that of syntactic marking, the subject of this paper. In a language such as English there are many arbitrary structural variations that preserve meaning and speaker intention:

- (10) (a) The chairperson of the committee stood to speak
- (b) The committee chairperson stood to speak
- (11) (a) I vaguely remember giving him three Sambucas and a Jameson
- (b) I vaguely remember that I gave him three Sambucas and a Jameson
- (12) (a) The applicant countries include Hungary, Malta and Estonia

- (b) The applicant countries include Malta, Hungary and Estonia
- (c) The applicant countries include Estonia, Malta and Hungary

As far as I am aware, there are no working systems using this approach, and the only previous relevant research is Atallah, Raskin et al (2001). They have developed a prototype system that watermarks text using a small number of syntactic transformations that they class as 'slightly' meaning-transforming. The algorithm works with the output of an off-the-shelf syntactic parser, and uses a keying scheme to distribute the marks through the file. Their rationale is that any truly meaning-preserving transformation could be globally applied to a document to scramble the watermark. A transformation that changes the meaning subtly can be applied very selectively at the watermarking stage without degrading the document significantly, but would damage the document unacceptably if used indiscriminately in an attempted attack. Presumably their method could be equally well applied to steganography, where the ultimate meaning conveyed by the cover document is irrelevant.

I do see some problems with this method. The four syntactic transforms employed (adverb movement, clefting, passivisation and the insertion of 'empty' transitional phrases such as 'in general' or 'basically' at the start of sentences) all exhibit some degree of asymmetry. For example active sentences are much more common in English than passives – and most passives are there for a good reason, such as the indeterminacy of the semantic subject, or emphasis by the writer on the semantic object of a verb. So a possible attack would be to remove all clefts, passives and transitional phrases, while moving all adjuncts to the most typical position. Another possible attack would be for a pirate to leave the text untouched, but backward engineer a key that points to sections of the text making a watermark to support their claim.

Atallah and Raskin's system is a work in progress. They see the syntactic method as an interim solution to be replaced by semantically based methods that would actually change the meaning of some sentences (Atallah, 2001, personal communication) so that a text watermark could also survive translation into another language.

In any case the emphasis of this paper is somewhat different. Here I am trying to identify syntactic structures in English that can be automatically transformed while preserving both semantics and pragmatics. As a result they should be broadly applicable across the range of information hiding applications.

Design Considerations

Before examining individual syntactic structures for their suitability to information hiding, I have to draw up a set of considerations and criteria that can be used to evaluate them. I do not have a single target application, so the requirements will remain rather general. In the conclusion I will discuss how well-suited individual transformations are to various applications.

Broadly the aim is to make changes to text documents that preserve concrete meaning as well as the intention of the author. Furthermore the techniques I employ should be applicable to as wide a range of texts and genres as possible.

In their introduction to watermarking, Kutter & Hartung (2000, p.102) list three general requirements for watermarking methods. I add one final requirement, that of robustness:

- Imperceptible – the mark should not be noticeable on casual or hostile analysis
- Redundant – the mark should be repeated in the carrier file, so that it will survive editing
- Keyed – it should be impossible to recover the hidden data without the key
- Robust – it should be impossible to remove the hidden data without damaging the carrier file

The requirement for imperceptibility holds from three points of view. The casual reader should be oblivious to the presence of the mark, and it should ideally be invisible to expert examination. Finally the author should feel that their text has not been compromised.

Any information-hiding algorithm should use keys, following Kerckhoff's assertion (referred to in Katzenbeisser 2000, p.8-9) that secure communication should always assume that the

algorithm is known. With keyed encryption of the message, and keyed distribution of the mark among the carrier's data, it should be possible to achieve secrecy even when the algorithm has been compromised.

Redundancy guards against attacks where the text is shortened, broken up, or otherwise edited to corrupt the hidden data. Then there is still some chance of retrieving the information from intact fragments of the original.

Other common attacks are collusion (where multiple copies of a marked document are compared to reveal the marking method) and statistical methods (such as looking for unusual frequencies of words or constructions).

Considerations for Individual Marking Structures

In this paper my objective is to identify syntactic structures that could be used as marking methods in information hiding applications. I do not aim to develop a complete system design that tackles questions such as keying methods, parsing techniques and defence strategies. So I will distil the general requirements above into a set that applies to individual marking techniques:

- Semantically neutral – to preserve meaning
- Pragmatically neutral – to preserve writer's intention
- Stylistically neutral – so that they are found across a range of styles, dialects and registers
- Ubiquitous – so that they are found often in carrier texts
- Purely syntactic – so that they can operate with off-the-shelf parsers, not needing any semantic analysis
- Sentence internal – so that they can operate with off-the-shelf parsers, not needing any discourse analysis

To satisfy the requirement of imperceptibility for both users and the author, the variations must be both semantically and pragmatically neutral.

The constructions chosen should of course also be relatively common across a range of texts, hence the requirements to be ubiquitous and stylistically neutral (for example a poetic turn of phrase would be very out of place in a technical document).

Finally the variations should rely on purely syntactic information, and only on the basis of information within the sentence. Thus they should be compatible with off-the-shelf parsers.

There are also two criteria that will govern whether a particular marking structure is appropriate for a particular application:

- Reversibility – whether a transformation can be undone
- Reliability – whether a transformation can be automatically implemented without human verification

Considerations for Specific Applications

For high volume applications such as fingerprinting, where large amounts of files must be automatically and uniquely marked online for every consumer, reliability is obviously a top priority. Also, since the consumer is paying for the document, the quality of the text cannot be compromised.

However, in a traitor tracing scenario, there would typically be a small number of copies of a high value document. Here it might be acceptable to have an application that used human intervention to verify risky transformations, meaning that less reliable transformations can be considered.

For steganography, reliability is not so important, since the cover text is actually a red herring. A certain level of errors is acceptable.

However reversibility is a very important requirement for steganography. Ideally the recipient in a secret communication should not have to have access to the original unprocessed text in

order to extract the secret message. This is due to practical constraints, since for sustained communications, both parties would have to have unlimited banks of identical texts for comparison. The alternatives to this are using the same cover text repeatedly, or sending the processed text together with the original, both of which are bound to arouse suspicion.

It is possible to build applications that can extract a message from a cover document without ever having the original for comparison, if wholly reversible variations are used. Taking restrictive relative clauses as an example, when decoding (13a), all clauses carrying the relativiser *that* could represent 1's, while *wh...* relativisers could represent 0's. The form of the relativisers in the original copy of the document (13b) would be irrelevant.

- (13) (a) The person who (0) you were talking to that (1) arrived yesterday with that girl who's (0) from Istanbul ...
- (b) The person ??? you were talking to ??? arrived yesterday with that girl ??? from Istanbul ...

Reversibility is also important for maintaining the statistical profile of a document. Suppose, for example, that 30% of English restrictive relative clauses use a *wh...* relativiser, 40% *that* and 40% none at all. This balance should be broadly preserved if a document is not to be statistically conspicuous. Steganography is particularly sensitive to this requirement, since once an eavesdropper is aware that it is being employed, they may be able to prevent communication by simply blocking the delivery of cover documents. The strength of steganography is in its inconspicuousness.

For the watermarking techniques reversibility is a mixed blessing. Ideally processed documents should be statistically unmarked, since statistical blips provide a useful starting point for hostile parties who are trying to crack a watermarking scheme. However, assuming that an algorithm is known, reversibility might allow the hostile party to remove a mark completely, and recover the original, before possibly inserting a watermark claiming ownership by them. Furthermore in watermarking scenarios there is no need to extract marks using the marked document alone. One can assume that the owner is able to keep copies of

the original unmarked document. So I can conclude that the marks used in watermarking should be not be reversible, as long as they are not over-applied to the extent that they significantly alter the statistical profile of a document. This is summed up in the table below.

	Semantics	Pragmatics	Stylistic	Ubiquitous	Syntactic	Sentential	Reversible	Reliable
Steganography	✓	✓	✓	✓	✓	✓	✓	○
Watermarking	✓	✓	✓	✓	✓	✓	✗	✓
Fingerprinting	✓	✓	✓	✓	✓	✓	✗	✓
Traitor Tracing	✓	✓	✓	✓	✓	✓	✗	○

Table 1: Requirements for Information Hiding Applications

Analysis of Marking Transforms

Introduction

In this section I identify and discuss syntactic variations that I believe would be suitable marking structures for plain text information hiding. I refer to these as transformations in the everyday sense. They are variations in surface form that preserve the meaning of sentences, as opposed to the controversial deep transformations of GB theory.

The reader may find the following descriptions of syntactic phenomena somewhat limited, uncritical and vague. This is intentional. Firstly I am not attempting to give complete descriptions of each structure addressed. For example, in the section on relative clauses I ignore free relatives completely. Rather I try to identify parts of the phenomena that are predictable, and amenable to automatic processing.

Secondly, I am keeping my hypothesised rules relatively simple in this chapter. I have not gone out of my way to identify all the exceptions and problems associated with each construction. Finally, I have not yet formalised the rules, since they should be applicable to a range of syntactic formalisms.

The subtleties and caveats will be exposed when I formalise the rules and try them out on data from the *SUSANNE* corpus in the next chapter.

Grammatical Theory

I have kept my discussion as general and uncontroversial as possible with regard to competing theories of grammar and constituent structure. For example I do not address whether a certain structure is best represented with the binary branching trees of GB, the multiple branching of phrase structure grammars, or the shallow trees most often used by automatic parsers and in corpora.

In the next chapter I will illustrate how the general rules derived here can be developed to suit a specific scheme of syntactic parsing – Sampson’s *English for the Computer* scheme (1995) as implemented in the *SUSANNE* corpus.

This paper assumes the existence of a reliable off-the-shelf syntactic parser, such as *LINGO* (Copestake 1999). The operations described should not require any semantic analysis. However, where lexical resources exist or can be created, I will consider how they may be of use. Generally I will assume that the basic parts-of-speech and phrasal constituents (noun, verb, prepositional, determiner, adjectival, adverbial) can be isolated.

I will also assume that the parser is able to identify the basic grammatical relations in a sentence, and distinguish adjuncts from complements, whether phrasal or clausal. Obviously this requires that a parser have a comprehensive lexicon of argument structure.

However, I will address how to identify phrasal boundaries, and higher level clausal constituents. For example, adjuncts often present problems for parsers, in determining which constituent they modify. Here it is unclear whether it is the proposition *read the report* that happened *on Tuesday* (14b) or *the report* that concerns *Tuesday* (14c).

- (14) (a) He read the report on Tuesday
 (b) He [[read the report]_{VP} [on Tuesday]_{PP}]_{VP}
 (c) He [read [the report [on Tuesday]_{PP}]_{NP}]_{VP}

An example of difficult clauses is *that* complements (15a), which can sometimes be mistaken for relative clauses (15b), when they appear together with a noun.

- (15) (a) He read [the report [that three people were missing]_{THAT}]_{NP}
 (b) He read [the report [that he had been given]_{RELATIVE}]_{NP}

In this case the lack of a direct object for the verb *give* tells us that the clause in (15b) is gapped, and so must be a relative. In the discussion ahead I will address this kind of question when it comes up.

As Schütze (1996, p.52f) comments in his work on empirical studies in linguistics, the methodology of linguistic judgements is often not clearly set out, and does not reflect the subtleties of introspective reactions to sentences. So the system used should ideally be unambiguous, consistent and differentiate between syntactic and semantic acceptability.

In this work I have adopted a simple three-way scale for judging grammaticality. Sentences that are totally grammatical are unmarked while those I judge to be totally ungrammatical are marked '*'. Anything between these two absolutes is marked '?'. Sentences that I judge questionable or ungrammatical, but that are acceptable in some dialects are marked '%'.

For semantic judgements I used '@' to signal that a sentence is semantically or pragmatically anomalous (ie it doesn't make sense, despite being grammatical from a syntactic point of view). When comparing transformed sentences with originals, sentences are flagged '#' if the semantic or pragmatic sense has not been preserved by the process.

Unfortunately time has not allowed me to test my judgements systematically among a group of speakers. Ideally this would be done on a more graduated scale of acceptability, such as the magnitude estimation methodology suggested by Sorace et al (1996), so that the judgements of different subjects could be quantifiably compared.

The grammatical facts of English discussed below are amalgamated from many sources, so individual references are rarely given. The primary sources used are Quirk & Crystal (1985) (a classical grammar), Hewings (1999) (a pedagogical grammar), Liles (1979) (a transformational grammar) and McCawley (1988) (a modern grammar incorporating phrase-structure approaches). Other sources used extensively were Sag & Wasow (1999), Baker (1989), Borsley (1991) and Trask (1993).

General Constraints

Before addressing specific constructions I will look at some general issues that one has to be wary of when performing any syntactic transformation.

An obvious issue is that of idioms and collocations. Taking co-ordinate structure switching as an example, that is swapping around the arguments of *and* or *or*, it is obvious that idioms (16) and collocations (17) do not permit this. The results are grammatical, but have not preserved meaning.

- (16) (a) I'll have a word with your trouble and strife. (Cockney rhyming slang: *wife*)
(b) # I'll have a word with your strife and trouble.
- (17) (a) Who's big, bad and ugly now?
(b) # Who's bad, ugly and big now?

Anaphora can also present problems when transforming sentences. To take a co-ordinate structure again, we can see that in the processed version (b), it is no longer clear who *her* refers to.

- (18) (a) Sally_i and her_i husband have not arrived.
(b) # Her_j husband and Sally_i have not arrived.

The next example illustrates two potential sources of problems – case and person agreement. Looking at *easy* constructions, where the object of an embedded clause is raised to the subject position of the main clause, we see that the case of the moved constituent must be changed from *them* to *they* and the form of the verb from *is* to *are*. Otherwise we are left with the ungrammatical sentence (b).

- (19) (a) It is easy to take them for a ride.
(b) * Them is easy to take for a ride.
(c) They are easy to take for a ride.

The indefinite article *a* also must agree with the noun that follows it, as we see in this example of adjective reordering:

- (20) (a) It is an arid loveless place.
 (b) * It is an loveless arid place.
 (c) It is a loveless arid place.

Contractions are another phenomena that I have to be aware of, when words are substituted. In this case the relativiser replacement of *that* with *which* produces the non-existent contraction *which's*.

- (21) (a) It's Kerry that's made a fool of themselves.
 (b) * It's Kerry which's made a fool of themselves.

When moving constituents, the presence of long or complex elements, while not affecting grammaticality, can produce unwieldy sentences. In this example of phrasal verb particle movement, compare the perfectly acceptable (22b) with (23b). It may also be advisable to avoid processing special text units such as quotations, titles, headings and captions.

- (22) (a) I couldn't make out what Pauline meant.
 (b) I couldn't make what Pauline meant out.
- (23) (a) I couldn't make out what Pauline, having arrived late for work, meant when she excused herself with some story about leaves on the line between Tallaght and Rathmines.
 (b) ? I couldn't make what Pauline, having arrived late for work, meant when she excused herself with some story about leaves on the line between Tallaght and Rathmines out.

A final issue is that of orthography. Care must be taken that capitalisation and punctuation are appropriate after transformation. However, this paper is not specifically concerned with the written form of English, and so I will not address this question further.

Relative Clauses

Relative clauses are finite but gapped clauses that follow a nominal phrase as a modifier. They can be semantically classified as restrictive or non-restrictive (sometimes also referred

to as appositive). A restrictive relative clause (24) clarifies the reference of the noun phrase of which it forms a part, while a non-restrictive clause (25) simply adds further information for elaboration. Non-restrictive relative clauses are often flagged in speech with bounding pauses, reflected by the use of commas in written English:

- (24) [[The person_i]_{N'} [that GAP_i looks like Mr. T]_{RESTRICTIVE}]_{NP} is an embroidery teacher
- (25) [[Joan_j]_{N'}, [who GAP_j is a dancer]_{NON-RESTRICTIVE}]_{NP}, loves to watch *Neighbours*

The gapped element may be the subject, direct object or prepositional object of the relative clause. The noun phrase head then semantically stands in the same grammatical relation as the gapped constituent:

- (26) The people_{SEMANTIC SUBJECT} [that GAP_{SUBJECT} did that]_{RELATIVE} deserve a medal
- (27) The constituencies_{SEMANTIC DIR. OBJECT} [that the party represents GAP_{DIR. OBJECT}]_{RELATIVE} are unrepresentable
- (28) The theory_{SEMANTIC PREP. OBJECT} [that you believe in GAP_{PREP. OBJECT}]_{RELATIVE} is discredited

Parsers can usually identify relative clauses by the fact that they are headed by a relativiser and directly follow a noun phrase. Sometimes, however, the relativiser may be absent, as in (29) below (to be discussed in more detail later in this section).

- (29) The theory you believe in is discredited

One variety of relative clause is the dislocated relative clause. Here the relative clause has been de-coupled from the noun phrase to which it belongs, and is moved to the end of the sentence (see example below). It is most often used when the relative clause is long and complex, and so might disrupt the flow of the sentence. Since dislocated relatives exhibit a different behaviour than regular relatives (for example they cannot drop the relativiser) I will exclude them from my study and the algorithm.

- (30) (a) I saw the man who you mentioned having been at school with Elvis yesterday
- (b) I saw the man yesterday who you mentioned having been at school with Elvis

In fact what crucially distinguishes relative clauses is that they modify a noun *and* that they

are gapped clauses – that is that they have a missing argument in their predicate structure. Here we see two *that* clauses following a noun. The *that* clause in the first example is clearly a *that* complement, since the sentence it contains (*you made it*) is complete. The second example contains an incomplete sentence, lacking its object, and so must be a relative clause. To differentiate between these two cases, the parser must have a list of verbs' argument structure – in this case to tell it that the verb *made* requires a direct object.

(31) I resent [the suggestion [that you made it]_{THAT}]_{NP}

(32) I resent [the suggestion [that you made GAP_{OBJECT}]_{RELATIVE}]_{NP}

Switching *which/who/whom* with *that*

One common variation found in relative clauses is the switching of the generic relativiser *that* with *which*, *who* or *whom* depending on the context. So the restrictive sentence above (24) can just as well be phrased:

(33) The person who looks like Mr. T is an embroidery teacher

However, the equivalent switch cannot be made with non-restrictive relatives:

(34) * Joan, that is a dancer, loves to watch *Neighbours*

Restrictive relatives can switch the relativiser regardless of what grammatical relation the gapped element occupies. Compare (35)-(37) with (26)-(28):

(35) The people [who GAP_{SUBJECT} did that]_{RELATIVE} deserve a medal

(36) The constituencies [which the party represents GAP_{DIRECT OBJECT}]_{RELATIVE} are unrepresentable

(37) The theory [which you believe in GAP_{PREP. OBJECT}]_{RELATIVE} is discredited

For restrictive clauses the choice of *who* or *which* is determined by whether the nominal head that it modifies refers to a human or not⁴.

⁴ In fact, this rule is not always followed by speakers of English. The *COBUILD Bank of English* contains many examples of *which* being used with human head nouns. For example ... *there are many people which display these*

- (38) (a) The person_{HUMAN} that looks like Mr. T is an embroidery teacher
 (b) The person_{HUMAN} who looks like Mr. T is an embroidery teacher
 (c) * The person_{HUMAN} which looks like Mr. T is an embroidery enthusiast
- (39) (a) The book_{NON-HUMAN} that you mentioned is sold-out
 (b) * The book_{NON-HUMAN} who you mentioned is sold-out
 (c) The book_{NON-HUMAN} which you mentioned is sold-out
- (40) (a) The alien_{NON-HUMAN} that you made love to is playing hard to get
 (b) ? The alien_{NON-HUMAN} who you made love to is playing hard to get
 (c) The alien_{NON-HUMAN} which you made love to is playing hard to get

Although it is dying out in many dialects of English, the *who/whom* distinction may be important in some contexts. It can be shown that *who* is used when the relativised nominal acts as the semantic subject of the embedded verb (Liles 1979, p.126), and *whom* if it acts as the semantic object. However, note that the last example would still be acceptable to many speakers of English.

- (41) (a) That man [who GAP_{SUBJECT} mentioned you] supports Shamrock rovers
 (b) * That man [whom GAP_{SUBJECT} mentioned you] supports Shamrock rovers
- (42) (a) That man [whom you mentioned GAP_{OBJECT}] supports Shamrock rovers
 (b) % That man [who you mentioned GAP_{OBJECT}] supports Shamrock rovers

Generally this process is reversible – that is *that* can be substituted for *wh...* relativisers. However, when the gapped element of the relative clause is a prepositional object, the preposition can be fronted as in (44a). This fronted configuration only allows *who/which/whom* as relativiser, and not *that*.

- (43) (a) The theory [which you believe in]_{RELATIVE} is discredited
 (b) The theory [that you believe in]_{RELATIVE} is discredited
- (44) (a) The theory [in which you believe]_{RELATIVE} is discredited

manifestations ... (BOE ref. S1000910704). . However this is a non-standard alternative, and so will not be included in my analysis.

- (b) * The theory [in that you believe]_{RELATIVE} is discredited

Based on this evidence I can formulate the following variation rule:

(45) **Relative Clause *That/Wh...* Switch Rule**

- In the case of restrictive relative clauses beginning with a relativiser, and modifying a nominal head denoting a non-human entity: *that* can be reversibly switched with *which*.
- In the case of restrictive relative clauses beginning with a relativiser, and modifying a nominal head denoting a human entity: *that* can be reversibly switched with *who*. For registers that use *whom*, it is used when the nominal head functions as a semantic object of the relative clause verb.

The head of a noun clause can usually be identified as the rightmost direct daughter of its deepest N' constituent, in this case below *dog*. By deepest I mean the N' shown below – in some theories of grammar there would be several other N' constituents (such as ADJP N' or N' PP) before reaching NP level.

- (46) [That]_{DETP} [stupid]_{ADJP} [hooligan dog]_{N'} [of yours]_{PP}

My application must also somehow be capable of determining the 'human-ness' of any noun. How this might be done is addressed in the next chapter.

Dropping or Inserting *which/who/whom/that*

In some cases the relativiser can be dropped altogether. The first restriction found is that, again, it cannot be applied to non-restrictive relatives, regardless of whether the relative is object or subject gapped:

- (47) (a) Mr. Hyde [who you told me about GAP_{OBJECT}]_{NONRESTRICTIVE} likes his steak rare
 (b) * Mr. Hyde [you told me about GAP_{OBJECT}]_{NONRESTRICTIVE} likes his steak rare
- (48) (a) Mr. Hyde [who GAP_{SUBJECT} told me about you]_{NONRESTRICTIVE} likes his steak rare
 (b) * Mr. Hyde [GAP_{SUBJECT} told me about you]_{NONRESTRICTIVE} likes his steak rare

Turning to restrictive relatives, it appears that the relativiser can be dropped, but only if the subject is explicit (ie not gapped) in the relative clause. Compare these examples below with the full form (41a) and (42a).

- (49) (a) * That man [GAP_{SUBJECT} mentioned you] supports Shamrock rovers
 (b) That man [you mentioned GAP_{OBJECT}] supports Shamrock rovers

This makes perfect sense, since in the absence of a relativiser or a subject, the relative clause verb could be mistaken for the main verb (a garden path sentence of sorts).

Care must also be taken not to mistake passive sentences, such as that below, for relativiser free relatives. Passives can be easily recognised by the *by* phrase (a), or the gap (b) in the verb's argument structure.

- (50) (a) [The man]_{SUBJECT} mentioned [by you]_{SEMANTIC SUBJECT} supports Shamrock rovers
 (b) [The man]_{SUBJECT} mentioned GAP_{SEMANTIC SUBJECT} supports Shamrock rovers

So to identify relative clauses with a dropped relativiser, I must look for nouns, followed directly by a clause that has an explicit subject and lacks an explicit object.

This yields the next rule:

- (51) **Relative Clause *That/Wh...* Drop and Insertion Rule**
- In the case of restrictive relative clauses which have an explicit subject and lack an explicit object: the relativiser can be reversibly dropped.

Reducing and Expanding BE Relative Clauses

When the main verb of a relative clause is BE the relativiser can often be dropped together with the verb to yield a reduced (or 'whiz-deleted') relative clause:

- (52) (a) The woman [who was afflicted with a fit of hiccups] is recovering in hospital now
 (b) The woman [afflicted with a fit of the hiccups] is recovering in hospital now
 (53) (a) That man [who is leaving by the back door] is a crook
 (b) That man [leaving by the back door] is a crook

- (54) (a) The crowd [that are on their way to the concert] are a lively bunch
- (b) The crowd [on their way to the concert] are a lively bunch
- (55) (a) That person [that is with you] could do with some manners
- (b) That person [with you] could do with some manners
- (56) (a) The car [that is outside] could do with a paint job
- (b) The car [outside] could do with a paint job

The sentences seen all involve subject gapped relatives. Direct and prepositional object relatives do not allow clause reduction. Note that the second example can be reduced by just dropping *that* (58c), as was shown in the last section.

- (57) (a) The predicament [that Simon is in] could have been avoided
- (b) * The predicament [Simon in] could have been avoided
- (58) (a) The kind of person [that Bob is] scares the bejaysus out of me
- (b) * The kind of person [Bob] scares the bejaysus out of me
- (c) The kind of person [Bob is] scares the bejaysus out of me

All the above examples are of restrictive relative clauses. Here again, non-restrictive relatives behave differently. They can undergo the same operation, but the results are semantically anomalous. For example (60b) seems to say that Jimmy is a crook while leaving the room, but not at other times. Note that while (59b) remains grammatical, its default reading is that the affliction of hiccups and the hospital stay are contemporaneous, which is not consistent with the full form (59a):⁵

- (59) (a) Gertrude, who was afflicted with a fit of the hiccups, is recovering in hospital now
- (b) # Gertrude, afflicted with a fit of the hiccups, is recovering in hospital now
- (60) (a) Jimmy, who is leaving by the back door, is a crook
- (b) @ Jimmy, leaving by the back door, is a crook
- (61) (a) Mary's mates, who are on their way to the concert, are a lively bunch

⁵ These (b) sentences are acceptable to some speakers, if there are clear intonational breaks around the embedded participle clause as in *Jimmy – leaving by the back door – is a crook*

- (b) @ Mary's mates, on their way to the concert, are a lively bunch
- (62) (a) Tom's car, which is outside, could do with a paint job
- (b) @ Tom's car, outside, could do with a paint job

Above I have shown that several kinds of BE complements allow this transformation: present-participle clauses, past-participle clauses, prepositional phrases and adverbs of time and place. Noun complements (63) do not, while adjective phrase complements only allow it if they themselves carry a complement (65):

- (63) (a) The person that is secretary of the meeting should distribute the memos
- (b) * The person secretary of the meeting should distribute the memos⁶
- (64) (a) Any patients that are feverous must be isolated
- (b) * Any patients feverous must be isolated
- (65) (a) Any patients that are feverous with the DTs must be isolated
- (b) Any patients feverous with the DTs must be isolated

Additionally, care must be taken to distinguish true participles from deverbal adjectives.

Deverbal adjectives do not allow clause reduction as can be seen below:

- (66) (a) The contortionist that is scared of enclosed spaces hates his job
- (b) ? The contortionist scared of enclosed spaces hates his job
- (67) (a) A girl who is ravishing never fails to turn heads
- (b) * A girl ravishing never fails to turn heads

It turns out that the tense of BE also plays a role in the transformation. Looking first at adverbial reduced relatives, it appears that default assumptions can be made about the tense of the reduced clause on the basis of the tense of the main clause. From the examples below it appears that main clauses in the past simple (68) and past perfect tense (69) lead us to assume a past simple form of BE (ie *was/were*) in the corresponding full relative.

- (68) (a) The man at the end of his tether **lashed** out

⁶ One American speaker judged this sentence to be grammatical. This may be an issue of dialect.

- (b) # The man that **is** at the end of his tether **lashed** out
 - (c) The man that **was** at the end of his tether **lashed** out
 - (d) # The man that **will be** at the end of his tether **lashed** out
- (69)
- (a) The man at the end of his tether **had lashed** out
 - (b) # The man that **is** at the end of his tether **had lashed** out
 - (c) The man that **was** at the end of his tether **had lashed** out
 - (d) # The man that **will be** at the end of his tether **had lashed** out

Main clauses in a present and future tense imply that the relative is in the present simple (ie *am/is/are*). Here the examples are in the present simple (70), present perfect (71), present continuous (72) and future simple (73):

- (70)
- (a) The man at the end of his tether **lashes** out
 - (b) The man that **is** at the end of his tether **lashes** out
 - (c) # The man that **was** at the end of his tether **lashes** out
 - (d) # The man that **will be** at the end of his tether **lashes** out
- (71)
- (a) The man at the end of his tether **has lashed** out
 - (b) The man that **is** at the end of his tether **has lashed** out
 - (c) # The man that **was** at the end of his tether **has lashed** out
 - (d) # The man that **will be** at the end of his tether **has lashed** out
- (72)
- (a) The man at the end of his tether **is lashing** out
 - (b) The man that **is** at the end of his tether **is lashing** out
 - (c) # The man that **was** at the end of his tether **is lashing** out
 - (d) # The man that **will be** at the end of his tether **is lashing** out
- (73)
- (a) The man at the end of his tether **will lash** out
 - (b) The man that **is** at the end of his tether **will lash** out
 - (c) # The man that **was** at the end of his tether **will lash** out
 - (d) # The man that **will be** at the end of his tether **will lash** out

Present participle and complementised adjectival reduced relatives turn out to exhibit the same behaviour as the adverbials seen above. Past tense main verbs imply a past simple be

in the relative: (74), (75), (80), (81). Present and future tense main verbs imply a present simple BE in the relative: (76)-(77), (82)-(85).

- (74) (a) The woman running the show **called** all the shots
- (b) The woman who **was** running the show **called** all the shots
- (75) (a) The woman running the show **had called** all the shots
- (b) The woman who **was** running the show **had called** all the shots
- (76) (a) The woman running the show **calls** all the shots
- (b) The woman who **is** running the show **calls** all the shots
- (77) (a) The woman running the show **has called** all the shots
- (b) The woman who **is** running the show **has called** all the shots
- (78) (a) The woman running the show **is calling** all the shots
- (b) The woman who **is** running the show **is calling** all the shots
- (79) (a) The woman running the show **will call** all the shots
- (b) The woman who **is** running the show **will call** all the shots
- (80) (a) Any patients feverous with the DTs **were** isolated
- (b) Any patients that **were** feverous with the DTs **were** isolated
- (81) (a) Any patients feverous with the DTs **had been** isolated
- (b) Any patients that **were** feverous with the DTs **had been** isolated
- (82) (a) Any patients feverous with the DTs **are** isolated
- (b) Any patients that **are** feverous with the DTs **are** isolated
- (83) (a) Any patients feverous with the DTs **have been** isolated
- (b) Any patients that **are** feverous with the DTs **have been** isolated
- (84) (a) Any patients feverous with the DTs **are being** isolated
- (b) Any patients that **are** feverous with the DTs **are being** isolated
- (85) (a) Any patients feverous with the DTs **will be** isolated
- (b) Any patients that **are** feverous with the DTs **will be** isolated

On the other hand, past participle reduced relatives always imply a main clause in the past simple:

- (86) (a) The man arrested **protested** his innocence
- (b) The man that **was** arrested **protested** his innocence
- (87) (a) The man arrested **had protested** his innocence
- (b) The man that **was** arrested **had protested** his innocence
- (88) (a) The man arrested **has protested** his innocence
- (b) The man that **was** arrested **has protested** his innocence
- (89) (a) The man arrested **protests** his innocence
- (b) The man that **was** arrested **protests** his innocence
- (90) (a) The man arrested **is protesting** his innocence
- (b) The man that **was** arrested **is protesting** his innocence
- (91) (a) The man arrested **will protest** his innocence
- (b) The man that **was** arrested **will protest** his innocence

The judgements above apply to default readings of sentences in isolation. But it is easy to think of situations where another reading is more natural. For example, two people are in a room and one says to the other:

- (92) (a) “The person outside was at the party”
- (b) # “The person that was outside was at the party”
- (c) “The person that is outside was at the party”

Here, the expansion with a past simple verb, as suggested by the evidence above, is clearly not the most obvious or natural reading. This is a matter of context, a supra-sentential phenomenon that would require robust pragmatic analysis to predict, and so falls outside the scope of my requirements. However, for the moment I will run with the assumptions made above. How much of a problem context poses will be seen when I test the algorithm on corpus data.

Recognising reduced relatives in text is not straightforward, as there is no relativiser or finite verb to indicate this. I can look for noun phrases that are directly followed by participle

clauses, complementised adjectives, or time/place adverbs and prepositional phrases. As will emerge in the next chapter, this is not always the case.

Using all this evidence I can formulate a new rule:

(93) **Relative Clause Reduction Rule**

- In the case of subject gapped restrictive relative clauses, headed by the verb BE and directly followed by a past participle or present participle clause or by a complementised adjectival, time/place adverbial or prepositional phrase: then the relative clause can be reduced by removing the relativiser together with the verb BE.

(94) **Relative Clause Expansion Rule**

- Generally in the case of reduced relative clauses: the clause can be expanded by inserting the relativiser *that* together with the appropriate form of BE. If the main clause is in a present or future tense, BE must be in the present simple tense. If the main clause is in a past tense, BE must be in the past simple tense. BE must also agree in person and number with the subject of the main clause.
- However, in the case of reduced past participle clauses: the clause can be expanded by inserting the relativiser *that* together with the appropriate form of BE. BE must be in the past simple tense and agree in number with the subject of the main clause.

Adjunct Movement

Several types of adjunct prove to be quite mobile. Adjuncts of place, time, manner and contingency can all be positioned at the start or end of the sentence they attach to. They usually take the form of a prepositional phrase or an adverbial clause introduced by a subordinating conjunction (such as *since, because, while, as*, etc):⁷

(95) (a) [In the morning]_{TIME} I went to work

(b) I went to work [in the morning]_{TIME}

(96) (a) Life was better [in the home country]_{PLACE}

⁷ For McCawley (1988, pp.195-196) all adjuncts, including those introduced by subordinating conjunctions, are prepositional phrases

- (b) [In the home country]_{PLACE} life was better
- (97) (a) I'll call him [if I can find my phone]_{CONTINGENCY}
- (b) [If I can find my phone]_{CONTINGENCY} I'll call him
- (98) (a) He sealed victory [with a winning try]_{MANNER}
- (b) [With a winning try]_{MANNER} he sealed victory

Some other adjunct types can also be moved. While grammatical, this results in rather clumsy or marked phrasing. I will exclude these for not conforming to the requirement that transformations be pragmatically and stylistically neutral:

- (99) (a) I can assure you of it [with certainty]_{MODALITY}
- (b) [With certainty]_{MODALITY} I can assure you of it
- (100) (a) I made the cake [for your mother]_{BENEFACTIVE}
- (b) [For your mother]_{BENEFACTIVE} I made the cake
- (101) (a) I skipped along [towards the gate]_{DIRECTION}
- (b) [Towards the gate]_{DIRECTION} I skipped along
- (102) (a) She will be on the diet [for six weeks]_{DURATION}
- (b) [For six weeks]_{DURATION} she will be on the diet

One case to be wary of is sentences with clausal constituents as their final arguments. After moving an adjunct to the sentence final position, it may be interpreted as modifying the embedded clause, changing the meaning of the sentence. For instance, in (103a) the telling clearly happened *on Tuesday*, not the engagement. In the transformed version (103b) it no longer clear which event occurred *on Tuesday*.⁸

- (103) (a) [On Tuesday]_{ADJUNCT} he told me [that Francis got engaged]_{THAT COMPLEMENT}
- (b) # He told me that Francis got engaged on Tuesday
- (c) He told me on Tuesday that Francis got engaged
- (104) (a) [At lunch]_{ADJUNCT} I saw that man [you mentioned]_{RELATIVE CLAUSE}

⁸ Here again this sentence could be interpreted otherwise by some speakers, if given exceptional intonational marking – something like *On Tuesday – he told me – that Francis got engaged*

- (b) # I saw that man you mentioned at lunch⁹
- (c) ? I saw at lunch that man you mentioned

Conversely, a sentence containing an embedded clause that is post-modified by an adjunct (105a), can be mistaken for a post-modified sentence (105b). This incorrect analysis would lead to the transformation (105c), rather than the correct (105d).

- (105) (a) We applauded as [he sealed victory [with a winning try]_{ADJUNCT}]_S
- (b) [We applauded as he sealed victory [with a winning try]_{ADJUNCT}]_S
- (c) # With a winning try we applauded as he sealed victory
- (d) We applauded as, with a winning try, he sealed victory

This presents serious problems for parsing. It is difficult to see how this distinction can be determined automatically. It may be that sentences that end in an embedded sentence cannot be post-modified at all, precisely to prevent this kind of confusion. The pre-modified sentence (106a) is perfectly acceptable, while (106b) though grammatical requires some mental acrobatics to come up with a context in which it is plausible. In any case, it does not mean the same thing.

- (106) (a) On the stands we applauded as he sealed victory
- (b) # We applauded as he sealed victory on the stands

Accordingly one approach would be to exclude all sentences containing embedded sentences from this transformation. However, in this paper I am using *SUSANNE* for verification, a hand-checked corpus that should have parsed these cases correctly.

So now I can formulate the adjunct movement rule:

(107) **Adjunct Movement Rule**

- In the case adjuncts of time, place, manner or contingency, that take the form of a

⁹ While the (b) version does have a reading equivalent to the original (a), it is flagged ‘#’ because a non-equivalent reading is also possible, and so would make the transformation unsafe.

prepositional phrase, or an adverbial clause introduced by a subordinating conjunction, and that attach to a sentence whose last complement is not clausal: the adjunct may be freely moved to the start or end of the sentence that it attaches to.

It Insertion

Extraposition

Extraposition is the phenomena whereby a clausal subject may be replaced with *it*, and moved to the end of the sentence, becoming a further object complement:

- (108) (a) [That he leaves the top off the toothpaste]_{SUBJECT} annoys me intensely
 (b) It annoys me intensely [that he leaves the top off the toothpaste]

Several types of clausal subject allow this kind of movement. Besides *that* complements as seen above, infinitival and *for-to* clauses (or complements, as McCawley would term them, 1988, pp.116-138) can undergo the same operation:

- (109) (a) [That James knew]_{THAT} shocked Mary
 (b) It shocked Mary [that James knew]_{THAT}
 (110) (a) [To err]_{INFINITIVAL} is human
 (b) It is human [to err]_{INFINITIVAL}
 (111) (a) [For Captain Kirk to lose his cool]_{FOR-TO} would be a disaster
 (b) It would be a disaster [for Captain Kirk to lose his cool]_{FOR-TO}

That and infinitival complements can also be extraposed in their interrogative forms:

- (112) (a) [Whether he will lead or follow]_{INTERROGATIVE THAT} is irrelevant
 (b) It is irrelevant [whether he will lead or follow]_{INTERROGATIVE THAT}
 (113) (a) [Which way to go]_{INTERROGATIVE INFINITIVAL} remains unclear
 (b) It remains unclear [which way to go]_{INTERROGATIVE INFINITIVAL}

All the examples above use transitive verbs. For intransitive verbs both variations are possible, but the non-extraposed version is somewhat clumsy:

- (114) (a) It grates that he is here
 (b) ? That he is here grates

It is also possible, but unwieldy, to extrapose a clausal subject over a clausal object:

- (115) (a) [That she came at all]_{SUBJECT} implies [that she agrees with our plans]_{OBJECT}
 (b) ? It implies that she agrees with our plans that she came at all

Finally, when reversing extraposition, one must also be wary of raising verbs, which have a slightly different behaviour. Raising verbs can undergo extraposition, but demand that they retain an object:

- (116) (a) It seems unlikely that your milkman is a Soviet mole
 (b) That your milkman is a Soviet mole seems unlikely
 (117) (a) It seems that your milkman is a Soviet mole
 (b) * That your milkman is a Soviet mole seems

I do not see any particular parsing problems associated with this transformation. The constituents I am looking for always appear as the first or last element of a sentence, and the clauses are always flagged with some identifying word such as *to*, *that* or *for*.

The Extraposition Rule then looks like this:

(118) **Extraposition Rule**

- In the case of sentences with a *that* complement, interrogative *that* complement, infinitival clause, interrogative infinitival clause or *for-to* clause as subject: then the subject can be replaced by *it*, and moved to the end of the sentence. However this should be avoided if the sentence already has a similar clausal object.
- Conversely, in the case of sentences with *it* as subject, and ending with a *that* complement, interrogative *that* complement, infinitival clause, interrogative infinitival clause or *for-to* clause as object: then the final clausal object can be moved to the subject position, replacing *it*. However this should be avoided if the clausal object is the only object (or *must* be avoided in the case of raising verbs).

Raising

Raising is the phenomena in English where a verbal argument is removed from the clause where it logically belongs, and is moved up into a higher clause where it has no place in the semantic argument structure:

(119) Eddie Murphy seems to realise that the good days are gone

(120) He is easy to get a rise out of

In the first example, termed subject raising (Trask 1993, p.267) *Eddie Murphy* is semantically the subject argument of *realise*, but appears as the syntactic subject of the higher clause verb *seems*. Similarly in the second example *He* though semantically the object argument of the embedded verb *get*, appears as the subject of the main clause (object raising).

Unsurprisingly, these kinds of sentences can be transformed into a shape that reflects their semantic structure more directly. It can be seen that the result resembles extraposed sentences, in that *it* is inserted as a dummy subject in the main clause, followed by a verbal complement:

(121) It seems that Eddie Murphy realises that the good days are gone

(122) It is easy to get a rise out of him

Subject-raising verbs include *seem*, and *appear*. As the name suggests they are followed by a verb phrase whose subject can be 'raised' to become the subject of the main verb. The embedded finite *that* complement becomes a subject-gapped infinitival clause:

(123) (a) It appears [that Mr. Ben_{SUBJECT} likes cross-dressing]_{THAT}

(b) [Mr. Ben]_{SUBJECT} appears [GAP_{SUBJECT} to like cross-dressing]_{INFINITIVAL}

In combination with the verb BE, some adjectives of modality (eg *likely*, *certain*, *rumoured*, *thought*) function identically:

(124) (a) It is certain [that Mr. Ben_{SUBJECT} likes cross-dressing]_{THAT}

(b) [Mr. Ben]_{SUBJECT} is certain [GAP_{SUBJECT} to like cross-dressing]_{INFINITIVAL}

As just shown, the *it* insertion process is reversible, but only when the raising predicate has a single *that* complement. Raising verb sentences with multiple arguments cannot be raised:

- (125) (a) It seems [clear] [that he is content]_{THAT}
 (b) * He seems [clear] [to be content]_{INFINITIVAL}

However adjuncts can intervene, and they retain their position through the raising process:

- (126) (a) Apparently, Mr. Ben really appears to like cross-dressing on the weekends
 (b) Apparently, it really appears that Mr. Ben likes cross-dressing on the weekends

It is also important not to confuse subject control verbs such as *tend* or *promise* with subject raising verbs. Here the subject only appears in the higher clause, but is a semantic argument of both main and embedded clauses. Although control verb sentences can have the same surface structure, *it* insertion is not allowed:

- (127) (a) [Sundays]_{SUBJECT} tend [GAP_{SUBJECT} to be lazy days]_{INFINITIVAL}
 (b) * It tends [that Sundays_{SUBJECT} are lazy days]_{THAT}

It should also be noted that not all adjectives of modality are subject raisers:

- (128) (a) It is probable [that Mr. Ben_{SUBJECT} likes cross-dressing]_{THAT}
 (b) * [Mr. Ben]_{SUBJECT} is probable [GAP_{SUBJECT} to like cross-dressing]_{INFINITIVAL}

Thus I have to conclude that parsed syntactic structure is not a guide to a verb's membership of the class of raising (or for that matter, control) verbs, and so my algorithm will not be able to determine that membership. A list of raising verbs is needed. *COMLEX* (Grisham et al, 1994), an automatically generated lexicon of English argument structure, would be a prime candidate.

Finally I have to address the tense of the finite *that* complement created by this operation. It turns out that the expanded infinitival clause takes present simple tense if the main predicate is in a present or future tense (129-132), and takes past simple if the main predicate is in a

past tense (133, 134). This pattern is identical to that already found when expanding relative clauses from reduced clauses:

- (129) (a) Mohammed Ali **seems**_{PRESENT SIMPLE} to have a bee in his bonnet
 (b) It **seems** that Mohammed Ali **has**_{PRESENT SIMPLE} a bee in his bonnet
- (130) (a) ? Mohammed Ali **is seeming**_{PRESENT PROGRESSIVE} to have a bee in his bonnet
 (b) ? It **is seeming** that Mohammed Ali **has**_{PRESENT SIMPLE} a bee in his bonnet
- (131) (a) ? Mohammed Ali **has seemed**_{PRESENT PERFECT} to have a bee in his bonnet
 (b) ? It **has seemed** that Mohammed Ali **has**_{PRESENT SIMPLE} a bee in his bonnet
- (132) (a) Mohammed Ali **will seem** to have a bee in his bonnet
 (b) It **will seem** that Mohammed Ali **has**_{PRESENT SIMPLE} a bee in his bonnet
- (133) (a) Mohammed Ali **seemed** to have a bee in his bonnet
 (b) It **seemed** that Mohammed Ali **had**_{PAST SIMPLE} a bee in his bonnet
- (134) (a) Mohammed Ali **had seemed** to have a bee in his bonnet
 (b) It **had seemed** that Mohammed Ali **had**_{PAST SIMPLE} a bee in his bonnet

Again an important caveat is that the tenses inserted above do not reflect the *only* possible interpretation. Rather I judge them to be the default interpretation of the sentence taken in isolation.

Similarly to the extraposition case, the critical constituents for this transformation are clearly defined and appear in predictable positions in the sentence. I do not foresee any particular difficulties in parsing.

I am now ready to formulate the subject-raising rule:

(135) **Subject Raising Rule**

- In the case of sentences headed by a subject raising predicate, with an embedded infinitival clause following the predicate, and that infinitival clause's semantic subject NP preceding the predicate: then the embedded clause can be reconstituted as a finite *that* complement, with the semantic subject NP moving to the subject gap position, and *it* replacing it as subject of the raising predicate. If the main clause is in a present or

future tense, the embedded clause must be in the present simple tense. If the main clause is in a past tense, the embedded clause must be in the past simple tense. The finite embedded verb must agree in person and number with the moved subject.

- Conversely, in the case of sentences headed by a subject raising predicate, with *it* as syntactic subject, and a full *that* complement as object: then the NP subject of the *that* complement can be moved to the raising predicate subject position, replacing *it*, while the complement becomes an infinitival clause. The raising predicate verb must agree in person and number with the moved subject

Object raising constructions, also known as *easy* or *tough* constructions, are the combination of the verb BE with an adjective of difficulty. In this case it is the object of the embedded clause that is raised to subject of the main clause:

- (136) (a) It is impossible [to learn [the fiddle]_{OBJECT}]_{INFINITIVAL}
 (b) [The fiddle]_{SUBJECT} is impossible [to learn GAP_{OBJECT}]_{INFINITIVAL}

As in subject raising constructions, adjuncts do not interfere with the transformation:

- (137) (a) Honestly, the fiddle is impossible to teach yourself in 21 days
 (b) Honestly, it is impossible to teach yourself the fiddle in 21 days

And again like subject raising constructions, membership of the class of object raising adjectives is lexically determined, not syntactically or semantically. *Possible* is an example of an adjective of difficulty that does not work well in raising constructions:

- (138) (a) With Concorde it is possible to cross the Atlantic in three hours
 (b) ? With Concorde the Atlantic is possible to cross in three hours

Object raising constructions can raise both bare (ie direct object) noun phrases, and noun phrases embedded in prepositional phrase objects:

- (139) (a) It is hard to pick a present for her
 (b) A present is hard to pick for her
 (c) She is hard to pick a present for

While tense is not an issue here, as there is no change to the clause or verb form, case must be changed when a pronoun moves from subject to object position, and the verb BE must agree with the subject:

- (140) (a) It is difficult to convince them
 (b) * Them is difficult to convince
 (c) They are difficult to convince

The object-raising rule looks somewhat like its subject-raising cousin:

(141) **Object Raising Rule**

- In the case of sentences headed by an object raising predicate, with an embedded infinitival clause following the predicate, and that infinitival clause's semantic object NP preceding the predicate: then the semantic object can be moved to the object gap position in the infinitival clause, with *it* replacing it as subject NP of the raising predicate. The moved object must have objective case, and the main verb BE must agree in person and number with *it*.
- Conversely, in the case of sentences headed by an object raising predicate, with *it* as subject, and an infinitival clause as object: then a bare or PP-embedded NP object of the embedded clause can be moved to the raising predicate subject position, replacing *it*. The moved object must have nominative case, and the verb BE must agree with it in number and person.

There Insertion

In the GB sense *there* insertion is a deep transformation that results in sentences where the semantic subject has been replaced by *there*, and moved to object position:

- (142) There_{SUBJECT} is a horse_{SEMANTIC SUBJECT} galloping across the field
 (143) There_{SUBJECT} was_{SEMANTIC SUBJECT} someone electrocuted while shaving
 (144) There_{SUBJECT} is a cordon_{SEMANTIC SUBJECT} around the scene
 (145) There_{SUBJECT} is something_{SEMANTIC SUBJECT} wrong with Mary

What makes this interesting for us is that many of these sentences have a surface counterpart without *there*:

- (146) A horse is galloping across the field
- (147) Someone was electrocuted while shaving
- (148) A cordon is around the scene
- (149) Something is wrong with Mary

Many intransitive verbs allow this transformation – in fact Levin (1993, pp.88-91) lists more than two hundred. But many of the resulting sentences are either stylistically highly marked or simply clumsy:

- (150) (a) There arose an awful racket
- (b) ? An awful racket arose
- (151) (a) High up on the hill there stood an imposing ruin
- (b) ? High up on the hill an imposing ruin stood
- (c) ? There stood an imposing ruin high up on the hill
- (d) An imposing ruin stood high up on the hill

Most verbs do not allow *there* insertion under any circumstances:

- (152) (a) A police officer transfixed him with her stare
- (b) * There transfixed a police officer him with her stare
- (153) (a) Some branches floated down the river
- (b) * There floated some branches down the river

So here I will restrict the operation to the verb BE, because its behaviour proves to be more predictable, and because of its abundance in the language¹⁰. From the examples above it can be seen that sentences with clauses or prepositional phrases as complements are candidates

¹⁰ BE is the most common verb in the British National Corpus. The nearest competitor is HAVE with 30% of the incidence of BE, followed by DO, SAY and GO. The incidence of all other verbs is less than 5% compared to BE.

for this transformation. Sentences with noun phrase complements cannot be transformed, while not all adjective complements can:

- (154) (a) Many students are party members
- (b) * There are many students party members
- (155) (a) Many students are stingy
- (b) * There are many students stingy
- (156) (a) Many students are hungover
- (b) There are many students hungover

An important issue that must be considered is that of reference. In all the successful examples above the moved subject carries an indefinite article. Levin (1993, p.91) states that the subject must be of indefinite reference for *there* insertion to take place, or 'existential' as McCawley (1988, p.97) terms it. Subjects of definite reference do not allow *there* insertion, as shown in (157) and (158). Note that these sentences do have a grammatical reading where *there* is a demonstrative pronoun – hence they are marked '#' to indicate that the meaning has been changed.

- (157) (a) Our trifold is in the back-garden
- (b) # There is our trifold in the back garden
- (158) (a) The boys are making their way over
- (b) # There are the boys making their way over

Unfortunately the absence of a definite determiner is not enough to guarantee an indefinite reference. A generic reference is sometimes inferred, as shown by the paraphrases of (159a) and (160a), which disallows *there* insertion:

- (159) (a) An Irishman is likely to spend two-thirds of his income on drink and horses
- [a typical Irishman]
- (b) # There is an Irishman likely to spend two-thirds of his income on drink and horses
- [a particular Irishman]
- (160) (a) Programmers are having trouble finding jobs these days
- [programmers in general]

- (b) # There are programmers having trouble finding jobs these days
[some programmers]

Since the syntax of the subject alone is no reliable guide to its reference being existential or not, I will see if the verbal complement has any effect. Here I label in brackets what I judge to be the semantic nature of the subject reference:

- (161) (a) The programmers_{DEF, PL} are having trouble_{TRANS} [given]
(b) Programmers_{S, PL} are having trouble_{TRANS} [generic]
(c) Some programmers_{INDEF, PL} are having trouble_{TRANS} [existential]
(d) The programmer_{DEF, SG} is having trouble_{TRANS} [given]
(e) A programmer_{INDEF, SG} is having trouble_{TRANS} [existential]
- (162) (a) The programmers_{DEF, PL} are deprived of love_{ENDUR} [given]
(b) Programmers_{S, PL} are deprived of love_{ENDUR} [generic]
(c) Some programmers_{INDEF, PL} are deprived of love_{ENDUR} [generic]
(d) The programmer_{DEF, SG} is deprived of love_{ENDUR} [given]
(e) A programmer_{INDEF, SG} is deprived of love_{ENDUR} [generic]
- (163) (a) The programmers_{DEF, PL} were arrested for speeding_{TRANS} [given]
(b) Programmers_{S, PL} were arrested for speeding_{TRANS} [existential]
(c) Some programmers_{INDEF, PL} were arrested for speeding_{TRANS} [existential]
(d) The programmer_{DEF, SG} was arrested for speeding_{TRANS} [given]
(e) A programmer_{INDEF, SG} was arrested for speeding_{TRANS} [existential]
- (164) (a) The programmers_{DEF, PL} are from Mars_{ENDUR} [given]
(b) Programmers_{S, PL} are from Mars_{ENDUR} [generic]
(c) Some programmers_{INDEF, PL} are from Mars_{ENDUR} [generic]
(d) The programmer_{DEF, SG} is from Mars_{ENDUR} [given]
(e) ? A programmer_{INDEF, SG} is from Mars_{ENDUR} [?]
- (165) (a) The programmers_{DEF, PL} are on the picket line_{TRANS} [given]
(b) Programmers_{S, PL} are on the picket line_{TRANS} [existential]
(c) Some programmers_{INDEF, PL} are on the picket line_{TRANS} [existential]
(d) The programmer_{DEF, SG} is on the picket line_{TRANS} [given]

- (e) A programmer_{INDEF, SG} is on the picket line_{TRANS} [existential]

By examining these sentences it turns out that one feature of the complements is critically important – something akin to aspect. The complements can be classified as denoting a transient (labelled as such_{TRANS}) or enduring (and as such_{ENDUR}) state of affairs. The same distinction is applied to adjectives, which can be classed as ‘inherent’ or ‘non-inherent’ (Quirk & Crystal 1985, p.435)¹¹. It appears that a subject carrying an indefinite article, or no article, together with a transient (ie non-inherent) complement, results in an existential reading, and so, according to McCawley (1988, p.97), allows *there* insertion:

- (166) There are some programmers_{INDEF, PL} having trouble_{TRANS} [existential]
 (167) There is a programmer_{INDEF, SG} having trouble_{TRANS} [existential]
 (168) There were programmers_{PL} arrested for speeding_{TRANS} [existential]
 (169) There were some programmers_{INDEF, PL} arrested for speeding_{TRANS} [existential]
 (170) There was a programmer_{INDEF, SG} arrested for speeding_{TRANS} [existential]
 (171) There are programmers_{PL} on the picket line_{TRANS} [existential]
 (172) There are some programmers_{INDEF, PL} on the picket line_{TRANS} [existential]
 (173) There is a programmer_{INDEF, SG} on the picket line_{TRANS} [existential]

This interpretation fits in well with earlier evidence that noun phrase complements prevented *there* insertion, since they are by their nature enduring. This is demonstrated by the following sentence pair, where adding the noun *person* changes the interpretation of the endurance of the state of being miserable.

- (174) (a) He is miserable (currently, ie transient, non-inherent)
 (b) He is a miserable person (by nature, ie enduring, inherent)

¹¹ The two Spanish forms of the verb BE are another example of this distinction. Generally *ser* is used with inherent qualities, and *estar* with non-inherent. Compare *Estoy cansado* ‘I am tired’ (non-inherent), with *Soy cansado* ‘I am tiresome’ (inherent).

The enduring/transient distinction can also explain the ambiguous evidence I had on adjective complements (175)-(176), as well as predicting that proper noun subjects disallow *there* insertion:

- (175) (a) Many students are stingy_{ENDUR}
- (b) * There are many students stingy_{ENDUR}
- (176) (a) Many students are hungover_{TRANS}
- (b) There are many students hungover_{TRANS}
- (177) (a) Joan is hungover [given]
- (b) * There is Joan hungover

Unfortunately this distinction is semantic, and one of my requirements is to use syntactic considerations only. So I will make some assumptions about the transience of some complements, depending on their syntax. I have shown that PP and adjectival complements can be transient (165), (176) or enduring (164), (175) depending on their content, so I cannot safely include them in my algorithm. What I will assume is that past perfect and present progressive tensed clauses are by their nature transient, regardless of their other semantic content. Whether this is a safe assumption will be tested against corpus data in the next chapter.

When reversing the transformation, things are much simpler – if a sentence allows *there* insertion in the first place, it should therefore be existential. Generally the insertion process is fully reversible. One limitation is that a *there* sentence must have multiple complements, so that after transformation BE still has an object:

- (178) (a) There is a god
- (b) * A god is
- (179) (a) There are plenty of good reasons for that
- (b) * Plenty of good reasons for that are

Another restriction is that the transformation only applies to existential *there* sentences, not those with the demonstrative pronoun *there* as subject:

- (180) (a) There is John running down the street
 (b) # John is running down the street

While I will have no trouble implementing this in the corpus, because *SUSANNE* differentiates explicitly between existential and demonstrative *there*, this would present problems for an automatic parser. Examining the following sentence in isolation, it is impossible to know which of the two paraphrases, (b) or (c), represents the right interpretation.

- (181) (a) There is a person who can do it.
 (b) Over there is a person who can do it
 (c) There exists a person who can do it

This yields two separate rules:

(182) ***There* Insertion and Deletion Rule**

- In the case of BE sentences with a common noun phrase as subject, carrying either an indefinite article or no article, and with a past perfect or present progressive clause as complement: then the subject can be moved between the verb and the complement, being replaced by *there*. The verb continues to agree in person and number with the semantic subject.
- Conversely, in the case of BE sentences with existential *there* as syntactic subject, and with an object noun phrase as semantic subject, followed by a past perfect or present progressive clause: then the semantic subject can be moved to the syntactic subject position, replacing *there*. The verb continues to agree in person and number with the semantic subject.

Conjunction Argument Switching

The English logical conjunctions *and* and *or* can take a wide range of constituents as their arguments, which can be rearranged at will:

- (183) (a) [Layabouts]_{NP}, [gurriers]_{NP} and [good-for-nothings]_{NP}, the lot of them!
 (b) [Gurriers]_{NP}, [good-for-nothings]_{NP}, and [layabouts]_{NP}, the lot of them!
 (c) [Good-for-nothings]_{NP}, [layabouts]_{NP}, and [gurriers]_{NP}, the lot of them!

- (184) (a) I'm thinking of [learning Esperanto]_{V'} or [taking up yoga]_{V'}
 (b) I'm thinking of [taking up yoga]_{V'} or [learning Esperanto]_{V'}
- (185) (a) She throttled him [on the back]_{PP} and [over the head]_{PP}
 (b) She throttled him [over the head]_{PP} and [on the back]_{PP}
- (186) (a) They offered her [a reduction in hours]_{NP} or a [pay rise]_{NP}
 (b) They offered her [pay rise]_{NP} or a [a reduction in hours]_{NP}

The words *and* and *or* have another distinct function, that could be argued to constitute separate lexical entries. Here *and* is a causal operator that roughly means *if ... then ...*, while *or* means *if not ... then* Obviously the conjuncts cannot be switched:

- (187) (a) Play that *Titanic* theme tune again, and I'll strangle you.
 (188) (b) # I'll strangle you, and play that *Titanic* theme tune again.
- (189) (a) Get up to bed or I'll take my hand off your face!
 (190) (b) # I'll take my hand off your face or get up to bed!

These causal uses of the conjunctions are easy to recognise syntactically, since the first clause is always imperative, followed by a declarative sentence.

While the conjunct constituents above are balanced, ellipsis can lead to incomplete conjuncts:

- (191) (a) The suspect was said to be wearing a [black]_{AP} or [[navy blue]_{AP} jacket]_{NP}
 (b) The suspect was said to be wearing a [navy blue]_{AP} or [[black]_{AP} jacket]_{NP}
- (192) (a) You [can [spend the money]_{V'}]_{VP} or [save it]_{V'}
 (b) You [can [save the money]_{V'}]_{VP} or [spend it]_{V'}
- (193) (a) [[Eamonn]_{NP} knows about it]_S and [Clodah]_{NP} does too
 (b) [[Clodah]_{NP} knows about it]_S and [Eamonn]_{NP} does too

Here one has to identify which elements of the conjuncts are matched, as indicated above, before making the switch.

This could present problems for parsers. While parsers usually can recognise gapped constituents and ellipsis, they may have trouble in establishing the equivalence between these and their co-referential counterparts in other parts of the sentence.

Another problem faced is with co-ordinated conjuncts that have some implicit order.

Reordering these results in grammatical but semantically anomalous sentences:

- (194) (a) The deficit rose from 1.8 to 2.1 million
(b) @ The deficit rose from 2.1 to 1.8 million
- (195) (a) See you on Monday or Tuesday
(b) @ See you on Tuesday or Monday
- (196) (a) Francisco came home and put on the dinner
(b) @ Francisco put on the dinner and came home

To avoid these problems all conjuncts that include numeral or temporal elements must be excluded. Verbal phrases of all kinds will also be left out.

This yields the following rule:

(197) **Conjunction Movement Rule**

- In the case of co-ordinate structures linked by the conjunctions *and* and *or*, where the conjuncts do not contain verbal, numeral or temporal constituents of any kind, and where equivalent and matching constituents can be identified within each and every of the conjuncts: then the matching constituents can be exchanged in position at will.

Phrasal Verb Particle Movement

Many English verbs can carry associated particles such as prepositions and adverbs. Often the combination of the verb and particle has a completely different meaning from that the verb has alone.

- (198) [Look]_{VERB} [up]_{PARTICLE} (ie search for in a reference) the number in the directory
- (199) Private Ryan [took]_{VERB} [part]_{PARTICLE} (ie participated) in the landing
- (200) The U2 tickets [sold]_{VERB} [out]_{PARTICLE} (ie were all sold) in minutes

Some verb-particle combinations also allow movement of the particle to a position after the direct object NP:

- (201) (a) Give [the hostages]_{DIRECT OBJECT} [up]_{PARTICLE} !
(b) Give [up]_{PARTICLE} [the hostages]_{DIRECT OBJECT} !

These are commonly called phrasal verbs, as opposed to prepositional verbs, such as *take part* and *give in*, that do not allow this operation. Trask (1993, p.215) lists particle movement as one of the tests used to distinguish phrasal from prepositional verbs:

- (202) (a) Never give in to intimidation
(b) * Never give to intimidation in

There are also a number of phrasal verbs that only allow the particle in the post-object position:

- (203) (a) I have to get that fax off before five
(b) # I have to get off that fax before five

Given a phrasal verb that does allow movement, there are two situations in which this is blocked. If the object is a pronoun, the particle must follow it. In the case of other single word objects, putting the particle after the verb can be clumsy:

- (204) (a) The police set him up
(b) * The police set up him
(205) (a) I'm not sure if I can take time off
(b) ? I'm not sure if I can take off time

Secondly, if the object phrase is very long, placing the particle after it can be awkward:

- (206) (a) The police set up the man they had detained two weeks earlier for loitering
(b) ? The police set the man they had detained two weeks earlier for loitering up

Automatic recognition of phrasal verbs can be difficult. Verb particles are often orthographically identical to prepositions or adverbs. One strategy is to look for 'loose' prepositions (ie prepositions with no complement, see (207)) or adverbs in unusual places. However other types of sentences can also have loose prepositions, for example (208).

(207) The rocket took [off]_{VERB ARTICLE}

(208) What planet is he [from]_{LOOSE PREPOSITION}?

Another issue is one related to the lexicon used by parsing systems. To parse particle verbs correctly, a system must have a comprehensive lexicon of verb argument structure. It must be particularly careful of verbs that can be both phrasal or prepositional (usually with different senses).

(209) (a) Three people made up_{PREPOSITIONAL} the team (ie comprised)

(b) * Three people made the team up

(210) (a) Three people made up_{PHRASAL} the story (ie fabricated)

(b) Three people made the story up

Now I can formulate the phrasal verb rule. For the moment I will not try to decide on what constitutes a 'short' phrase. This should become clearer at the verification stage.

(211) **Verb Particle Movement Rule**

- In the case of phrasal verbs that are known to support particle movement, followed by a short direct object of at least two words: the particle can be freely moved between the pre-object and post-object position.

Verification of the Transforms

The large print giveth, and the small print taketh away

Lyric from *Step Right Up* by Tom Waits

Introduction

This chapter will provide a reality check for the informal hypotheses developed in the last. As the reader will have noticed, the English examples used so far are of the relatively simple and highly idealised kind frequently used in linguistic analysis.¹² Such analyses often break down when confronted with large amounts of genuine material.

Thus the objectives of this chapter are several. Testing my rules on corpus data will hopefully expose some errors, while turning up new patterns that had not been considered in the last chapter. Additionally it should give us an idea of the practicality of the transforms in an automatic system, and give an indication of the abundance (or lack) of candidate sentences in typical text.

The corpus chosen was selected for the depth and accuracy of its parsing scheme. Because it is manually annotated, I can be reasonably confident that any problems that appear are due to flaws in the algorithm, rather than shortcomings of the parsing. The corpus also includes a wide spread of texts, with material drawn equally from journalism, technical literature, fiction and *belles lettres*.

¹² ‘And so the “modern” linguist spends his or her time starring or unstaring terse unlikely sentences like “John, Bill and Tom killed each other” (to pick one at random from a recent journal), which seethe with repressed frustration and are difficult to work into a conversation. These example sentences bear no discernible resemblance to the sentences which compose the text that purportedly explains them – yet the linguist’s own sentences are also alleged (implicitly) to be drawn from the same English Language!’ – Joseph Becker quoted in Schütze (1996, p.3)

The *SUSANNE* Corpus

The manually annotated *SUSANNE* corpus was created under the direction of Geoffrey Sampson at the University of Sussex, as part of a project to develop a comprehensive scheme for the grammatical annotation of English. Sampson's *English for the Computer* (1995) describes the scheme in detail. The work developed out of an unpublished project to manually annotate parts of the Lancaster/Oslo-Bergen (LOB) corpus. The part-of-speech and phrasal/clausal annotation scheme was based on the Lancaster set, but with finer distinctions and the addition of 'logical' structure – that is semantic grammatical relations (eg the syntactic or 'surface' subject in a passive construction is its 'logical' or semantic object).

The corpus comprised 64 texts of over 2000 words (giving a total of 125,000+ words in 6000+ sentences) drawn from the Brown University corpus – a collection of over one million words of written American English collect in the 1960's. The texts were equally drawn (16 each) from one of four genres defined in the Brown corpus:

- A press reportage
- G *belles lettres*, biography, memoirs
- J learned (mainly scientific and technical) writing
- N adventure and Western fiction

Where I use sentences from the corpus in this paper they are accompanied by their full reference, for example J03:0390.06. The first three letters of the reference are the name of the file from which the text is taken, and the remainder points to the first record (ie word) of the sentence. The full source information for all 64 files is listed in Appendix C.

Sampson (2001) claims that, being manually annotated, the *SUSANNE* corpus and scheme can draw finer and more precise distinctions than comparable automatically parsed corpora, such as the Pennsylvania Treebank.

Methodology

To test my hypotheses I aimed to extract an representative sample of sentences from the corpus. Because the corpus is broken up into genres and files, I used stratified random sampling (Oakes 1998, p.10). In cases where at least one matching candidate sentence was found in every file, the first match was extracted from each file to give 64 sentences – sixteen from each of the four genres. This being the case allows me to generalise my results the over the genres and the whole corpus.¹³

In many cases however, matching candidates for a particular transformation were not found in every file, and an incomplete sample was extracted. In extreme cases of very low incidence, all matching candidates were extracted, often more than one from each file.

Scripts written in *perl* were used to search for and extract the candidate sentences. The *perl* language was chosen because it is a scripting language (ie not a compiled language), making it efficient for prototyping applications. It is also powerful in terms of string searching and manipulation. A real application would probably use a combination of declarative and procedural compiled languages, such as *Prolog* and *C*.

After extraction, the candidate sentences were examined and transformed manually. Following transformation I classified the sentences as successful, doubtful or failed, and calculated a rate of success as a simple percentage. Reasons for rejection were be ungrammaticality, or failure to preserve semantic and pragmatic content.

The scripts also recorded the total number of candidate sentences in the corpus, which was used to calculate a rate of incidence (ie how common the construction is).

¹³ In fact this is not strictly random, only arbitrary, since it was always the first available sentence in each file that was extracted.

Often the first run of a script exposed weak points in the hypothesis or the algorithm, and I ran a modified algorithm again to see if better results could be obtained. All the sentences analysed are listed in Appendix D. Insertions made by the transformation are underlined, while deletions are ~~struck-through~~.

The *SUSANNE* Scheme

The defining characteristic of the *SUSANNE* corpus is that it has been fully parsed by hand, with annotation of part-of-speech, constituent structure, including phrasal and clausal elements, grammatical relations, gapped elements and ‘logical’ structure. Many larger corpora (eg the British National Corpus) are only tagged for part-of-speech.

The structure employed in *SUSANNE* is rather flat. For example, in a noun phrase, all constituent determiners and modifiers branch directly from one node (212). Similarly, verbs, their arguments and their modifiers all branch directly from the sentence level (213).¹⁴

```
(212) ...  
      [NP  
        [Det the]  
        [Adj long]  
        [N fight]  
        [PP  
          [P for]  
          [NP  
            [Det the]  
            [N measure]  
          ]  
        ]  
      ]  
    ]  
    ...
```

¹⁴ See Appendix A for the grammatical annotation conventions I use in this paper.

```

(213) ...
      [S
        [NP
          [N Daniel]
        ]
        [AdvP
          [Adv personally]
        ]
        [VP
          [V led]
        ]
        [NP
          [Det the]
          [N fight]
        ]
      ]
    ...

```

The examples just shown use a conventional annotation reduced from the much richer *SUSANNE* scheme. These extracts actually appear as such in the corpus:

```

(214) ...
      [S
        [Nns:s
          [NP1s Daniel]
        ]
        [Nns:s]
        [R:h
          [RR personally]
        ]
        [R:h]
        [Vd
          [VVDv led]
        ]
        [Vd]
        [Ns:o
          [AT the]
          [NN1n fight]
        ]
        [P
          [IF for]
        ]
        [Ns
          [AT the]
          [NN1n measure]
        ]
        [Ns]
        [P]
        [Ns:o]
      ]
    S]

```

Part-of-speech tags (referred to as wordtags) always begin with two capital letters, that denote their class. To give an idea of the richness of the taxonomy, we can look at a few examples. The word *Daniel* is tagged as class **NP**, that is a proper noun, while the appended **1** denotes singular, and **s** surname. *Measure* is tagged **NN1n**, meaning a singular common noun that can be mass or countable. *Led* is marked as 'base form of verb having transitive

and intransitive uses', or $\forall\text{DV}$. The scheme also allots tags to special function words, such as existential *there* EX , or the conjunction *that* CST .

Phrasal and clausal constituents (phrasetags and clausetags) all begin with a single capital letter, optionally followed by lowercase letters to denote subclass. For example the phrase *the fight for the measure* is tagged class N (noun phrase) and s for singular, while Vd stands for a verb phrase beginning with a past tense verb. The reader will notice that some elements are followed by a colon and further lowercase letters. These are the functiontags that are primarily used to indicate roles. The $:\text{s}$ tags surrounding *Daniel* indicate that this phrase is the subject of the verb. Similarly $:\text{o}$ indicates that *the fight for the measure* is the object. The $\text{R}:\text{h}$ tags surrounding *personally* indicate that this is an adverbial phrase of manner.

Of particular interest is the representation of 'logical' structure. Looking at the following example of a passive sentence, we see that the syntactic subject *His contention* is tagged with the uppercase functiontag $:\text{S}$, meaning that it is the surface subject without being the semantic or 'logical' subject of the verb phrase *was denied*. The *by* phrase is tagged $:\text{a}$ agentive.

```
(215) ...
      [Ns:S
        [APPGm His]
        [NN1n contention]
      Ns:S]
      [Vsp
        [VBDZ was]
        [VVNt denied]
      Vsp]
      [Pb:a
        [IIb by]
        [Np
          [DA2q several]
          [NN2 bankers]
        Np]
      Pb:a]
      ...
```

In the last example we see how gapped or 'ghost' elements are represented and co-referenced. The uppercase functiontag $:\text{O}$, indicates that the phrase *the Texas Education Agency*, while appearing as the syntactic object of *authorize*, is not the 'logical' object. In the

infinitival clause (Ti) that follows (all finite clausetags begin with F, non-finite with T) the first element shown is the gapped subject of *establish*, and as can be seen from the numerical index 129, it is co-referential with the syntactic object of the main clause.

```
(216) ...
      [Ni:s
        [PPH1 It]
      Ni:s]
      [Vdc
        [VMd would]
        [VV0t authorize]
      Vdc]
      [Nns:0129
        [AT the]
        [NP1p Texas]
        [NN1u Education]
        [NNJ1n Agency]
      Nns:0129]
      [Ti:o
        [s129-
          s129]
        [Vi
          [TO to]
          [VV0t establish]
        Vi]
      ...
```

Space would not allow me to give a complete description of this scheme here, and in any case, many of the details of the scheme are not relevant to this study. However a partial listing of wordtags, phrasetags, clausetags and functiontags is given in Appendix B. Sampson (1995) provides the whole story. I will draw further attention to particular features of the scheme where they are pertinent.

Relative Clauses

Relative Clause *That/Wh...* Switch Rule

Let us remind ourselves of the rule:

- In the case of restrictive relative clauses beginning with a relativiser, and modifying a nominal head denoting a non-human entity: *that* can be reversibly switched with *which*.
- In the case of restrictive relative clauses beginning with a relativiser, and modifying a nominal head denoting a human entity: *that* can be reversibly switched with *who*. For

registers that use *whom*, it is used when the nominal head functions as a semantic object of the relative clause verb.

The first problem I need to solve is to determine the ‘human-ness’ of nominal heads. This is a semantic, not a syntactic, quality. One of my basic requirements is that I rely on purely syntactic information. However computational resources do exist that contain this sort of data. *WordNet* (Miller et al 1990) is an example of a freely available semantic lexicon of English that could be used to determine if common nouns denote humans or not.

For example in *WordNet* the word *contortionist* is linked to *human* through four hypernym links (... is a kind of ...): *contortionist* > *acrobat* > *athlete* > *contestant* > *human*. For the purposes of verification I will assume I have access to this semantic information.

Another challenge is that the *SUSANNE* schema, being purely syntactic, does not distinguish between restrictive and non-restrictive clauses. Luckily non-restrictive relatives are usually bounded by commas, while restrictive clauses are not, so I will try to use this as the criteria.

As regards the tree structure of relative clauses, I will first look at conventional theory. McCawley (1988, p.432) represents a relative clause as a gapped complement (COMP') that attaches to an N'. The COMP' in turn consists of an NP containing the relativiser, and a gapped sentence S'.¹⁵

¹⁵ See Appendix A for the notation I use when illustrating conventional constituent structure

```

(217) ...
      [NP
        [Det The]
        [N'
          [N' two fluids]
          [Comp'
            [NP
              [N which]
            ]
            [S'
              [V'
                [V gave]
                [NP
                  [Det the]
                  [N'
                    [AdjP
                      [Adj small]
                      [AdjP
                        [Adj negative]
                      ]
                    ]
                  ]
                  [N'
                    [N pressures]
                  ]
                ]
              ]
            ]
          ]
        ]
      ]
    ]
  ]
  ...
  (J03:0390.06: The two fluids which gave the small negative pressures were
  polybutenes with molecular weights which were stated to be 520 and 300.)

```

The structure used by *SUSANNE* is much flatter. Here the head noun, determiners and modifiers all branch directly from the NP level. Inside the embedded clause a flat structure can also be seen, with the object NP being a direct daughter of the clause, rather than being subsumed by the VP. Relative clauses are seen as a kind of finite clause (*FinC*), and the relativiser is classed a determiner.

```

(218) ...
      [NP
        [Det The]
        [Num two]
        [N fluids]
        [FinC
          [DetP
            [Det which]
          ]
          [VP
            [V gave]
          ]
          [NP
            [Det the]
            [Adj small]
            [Adj negative]
            [N pressures]
          ]
        ]
      ]
    ]
  ]
  ...

```

The tree above has been simplified with a reduced and more conventional set of notation. The next tree (219) shows the full detail of the *SUSANNE* scheme. Relative clauses have a specific clausetag of their own *Fr*. The relativisers have individual wordtags, in this case *DDQr* for *which*. Here the relativiser is functioning as the subject (signalled by the :*s* in the determiner phrase) of the embedded phrase, and is linked to its logical subject, the head noun (signalled by the matching cross-references, in this case 143).¹⁶

```

(219) [Np:s143
      [AT The]
      [MC two]
      [NN2 fluids]
      [Fr
        [Dq:s143
          [DDQr which]
        ]
        Dq:s143]
      [Vd
        [VVDv gave]
      ]
      Vd]
      [Np:o
        [AT the]
        [JJ small]
        [JJ negative]
        [NN2 pressures]
      ]
      Np:o]
      Fr]
      Np:s143]

```

¹⁶ See Appendix B for the notation scheme used by *SUSANNE*

This flatter structure makes determination of the head noun trivial – it is simply the last noun before the relative clause starts. With the representation of relatives in *SUSANNE* clarified, I can construct the search and transformation algorithm in pseudocode thus (here I have chosen not to support *whom*):

- search for a relative clause
- the nearest noun before the relative clause is the head noun
- if (there is a comma between the relative clause and the head noun) { abort }
- if (the first word in the relative clause is not a relativiser) { abort }
- determine if the head noun is semantically human or not
- if (unsure) { abort }
- if (human) {
 - if (relativiser is *who*) {replace with *that*}
 - else {replace with *who*}
- else {
 - if (relativiser is *which*) {replace with *that*}
 - else {replace with *which*}
- }
- the end

I implemented part of this code as a *perl* script to identify all candidate sentences in the *SUSANNE* corpus. Of these one sentence was extracted from each of the 64 texts in the corpus, to be manually verified (listed in Appendix D1). The relativiser replacement was done by hand, since it relies on the human/non-human distinction, which in a real system would be implemented with a semantic network such as *WordNet*.

Altogether 586 matching relative constructions were identified in the corpus (in 550 sentences, as some sentences contained multiple instances), giving a per sentence incidence of about 1 in 11 (9.4%). Of the 64 sentences examined manually, 55 were correctly identified

by the algorithm. Six of the sentences that were wrongly identified turned out to be non-restrictive relatives that were not introduced by a comma.¹⁷

- (220) (A10:0030.15) * Ierulli will replace Desmond D. Connall that ~~who~~ has been called to active military service but is expected back on the job by March 31.

One failed sentence used the relativiser *whose* that my method did not allow for:

- (221) (J22:0010.39) * As the historic processes of modernization gradually gain momentum, their cohesion will be threatened by divisive forces, the gaps between rulers and subjects, town and country, will widen; new aspirants for power will emerge that ~~whose~~ ambitions far exceed their competence; old rulers may lose their nerve and their sense of direction.

With a further two sentences it was questionable whether the relative was restrictive or not.

This made the transformation unsafe:

- (222) (A14:0120.27) ? Van Brocklin, the quarterback that ~~who~~ led the Eagles to the title, was signed by the Vikings last Wednesday.
- (223) (J01:0030.24) ? The results of present observations of the thermal radio emission of the moon are consistent with the very low thermal conductivity of the surface layer that ~~which~~ was derived from the variation in the infrared emission during eclipses (e.g., Garstung, 1958).

Of the 55 sentences that were successfully recognised, 54 transformed successfully, for example (224). Three of these (A06, G02 and J17) contained two instances of the relative construction I was looking for, for example (225):

- (224) (N11:0130.33) Lewis was a man that ~~who~~ had made a full-time job of cow stealing.
- (225) (J17:0710.27) Although no drugs act exclusively on the hypothalamus or a part of it,

¹⁷ Here and in the appendices ~~struckthrough~~ text denotes deletions, and underlined text denotes insertions made during the transformation. The codes in brackets that precede examples are the *SUSANNE* reference, the first three characters being the filename, and the portion after the colon denoting the specific record at which the sentence starts in the file

there is sufficient specificity to distinguish drugs that ~~which~~ shift the hypothalamic balance to the sympathetic side from those that ~~which~~ produce a parasympathetic dominance.

As for the one sentence that failed, I am unclear as to why it is not wholly acceptable. It may be because a metaphor is employed, making it ambiguous whether the *sweetest flower* is semantically human or not. It may also be that *flower that grows* is a collocation, or that *which* is inappropriate in this rather poetic style:

(226) (N09:0060.27) ? To me you'll always be the girl o' my dreams, an' the sweetest flower which ~~that~~ grows”.

On the basis of these results I can revise my algorithm. Since punctuation is not a reliable guide to the restrictive or non-restrictive quality of relative clauses, I will have to limit my search method to *that* relatives, that are always restrictive:

- search for a relative clause
- the nearest noun before the relative clause is the head noun
- if (the first word in the relative clause is not the relativiser *that*) { abort }
- determine if the head noun is semantically human or not
- if (unsure) { abort }
- if (human) { replace *that* with *who* }
- else { replace *that* with *which* }
- the end

With the new version of the algorithm, 231 constructions were identified in the corpus, yielding an incidence of 3.7%, or one in every 27 sentences. Only 50 examples were extracted for manual examination, since fourteen of the corpus files contained no appropriate candidates.

Of the 50 examined, all were correctly identified, and 47 were transformed successfully, for example:

(227) (N18:0370.48) There was something about the contour of her face, her smile which ~~that~~ was like New Orleans sunshine, the way she held her head, the way she walked — there

was scarcely anything she did which did not fascinate me.

One sentence that was doubtful was that which failed in the last run:

(228) (N09 : 0060 . 27) ? To me you'll always be the girl o' my dreams, an' the sweetest flower
which ~~that~~ grows”.

Another doubtful case was the following sentence. The second variation of *which* with *that* is unsafe. *The stitch in time that saved nine* seems to be a collocation, and the quotes surrounding it reinforce this view:

(229) (G02 : 0450 . 12) ? For lawyers, reflecting perhaps their parochial preferences, there has been a special fascination since then in the role played by the Supreme Court in that transformation — the manner in which its decisions altered in “the switch in time which ~~that~~ saved nine”, President Roosevelt's ill-starred but in effect victorious “Court-packing plan”, the imprimatur of judicial approval time which ~~that~~ was finally placed upon social legislation.

The only to fail completely was for two reasons. Firstly, the algorithm would have identified the head noun as *one*, not *you*, and so conclude that it is non-human. Secondly, as a result of the original relativiser *that* being part of a contraction the transformation yielded the non-existent *which*'s:

(230) (N03 : 0790 . 27) * You're the one which ~~that~~'s taking the big chance.

So how does this method shape up? On the first run the rate of incidence was very high at one in twelve sentences. But the final failure rate was an unacceptable 17%. This was primarily due to non-restrictive relative clauses not being tagged with a comma. Interestingly most of these examples were in pieces of sports journalism.

By restricting my search to *that* relatives, I guarantee that they are all restrictive. The substitution can then proceed with a very high level of reliability (94%), at the cost of lower

incidence (about one in thirty). These results depend on having a comprehensive and accurate semantic lexicon.

Despite my initial requirements demanding purely syntactic analysis, this kind of semantically based decision-making is perfectly viable, because given the appropriate lexical resource it is a simple look-up operation, requiring no semantic analysis as such. *WordNet* would be the most obvious choice for this element of a system.

A drawback of restricting ourselves to *that* relatives is that it makes the process non-reversible. Restrictive *that* relatives can be transformed into *who/which* relatives, but not vice-versa. This will affect what applications this transformation can be used in.

The problem of disambiguating restrictive from non-restrictive relatives is not an easy one to resolve. In speech non-restrictives are flagged with clear pauses and intonation, but it appears that in written contexts authors' use of commas for the same purpose is not consistent enough to be useful to us. One alternative approach would be to exclude relatives that are attached to proper nouns, which are usually already given in the discourse, for example:

(231) (A10:0030.15) Ierulli will replace [Desmond D. Connall]_{PROPER NOUN} [who has been called to active military service but is expected back on the job by March 31.]_{NON-RESTRICTIVE}

However the algorithm would still have to recognise examples like that below, where the relative clause, though attached to a common noun (*righthander*), ultimately refers to a proper noun:

(232) (A11:0080.18) Over the first five frames, [Jack Fisher]_{PROPER NOUN}, the big [righthandler]_{COMMON NOUN} [who figures to be in the middle of Oriole plans for a drive on the 1961 American League pennant,]_{NON-RESTRICTIVE} held the A's scoreless while yielding three scattered hits.

This would need some kind of deep semantic and pragmatic analysis, requiring considerable real-world knowledge, and so falls well outside my terms of reference. Whether it is at all achievable is debatable.

Relative Clause *That/Wh...* Drop and Insertion Rule

- In the case of restrictive relative clauses which have an explicit subject and lack an explicit object: the relativiser can be reversibly dropped.

This is what a relative clause with explicit subject looks like in *SUSANNE* (233). Since there is no relativiser, it must be restrictive. The first constituent in the clause is the clause subject. Here again there is a fairly flat structure with the clausal complement (an infinitival clause – *Ti*) a direct daughter of the relative clause. The gapped object argument follows this (*o*₁₅₅). The *o* indicates that this is the object, and the number 155 shows that it is co-referential with its semantic object, the head noun (*s*₁₅₅).

```
(233) [Ns:s155
      [AT The]
      [JJT largest]
      [NN1c hurdle]
      [Fr
        [Np:s
          [AT the]
          [NN2 Republicans]
        Np:s]
        [Vdc
          [VMd would]
          [VH0 have]
        Vdc]
        [Ti:z
          [Vi
            [TO to]
            [VV0v face]
          Vi]
          [o155-
            o155]
        Ti:z]
      Fr]
    Ns:s155]
    ...
    (A01:1000.06: The largest hurdle the Republicans would have to face is a state law
    which says that before making a first race...)
```

This time I am still looking for restrictive clauses, so I will look for relatives with *that* as the relativiser, or none at all. After this the next constituent should be the subject. In pseudocode:

- search for a relative clause
- if (the first word in the relative clause is the relativiser *that* && the next constituent is a subject noun phrase) { remove relativiser }
- else if (the first word in the relative clause is a subject noun phrase && it is not a relativiser) { insert *that* at start of relative clause }
- the end

Again I implemented this as a prototype with *perl* to identify and extract candidate sentences.

I then manually verified if the transformations are possible (see Appendix D2).

The results are encouraging. Incidence is reasonable at 3.4% (about one in 30) and of the 52 sentences extracted for examination, only three failed. One sentence failed because it was incorrectly parsed in the corpus:

- (234) (a) (G05:0910.12) It was to provide a safe and spacious crossing for these caravans, and also to make a pleasance for the city, [that Shah Abbas II in about 1657 built, of sun-baked brick, tile, and stone, the present bridge.]_{THAT COMPLEMENT}
- (b) * It was to provide a safe and spacious crossing for these caravans, and also to make a pleasance for the city, Shah Abbas II in about 1657 built, of sun-baked brick, tile, and stone, the present bridge.

This (234a) is an extraposed sentence, of which the phrase headed by *that* is not a relative clause as indicated in the corpus code, but a *that* complement. As a result the transformation fails (234b).

Two other sentences failed because of a flaw in the algorithm. I failed to provide for relative clauses that start with a noun phrase but still carry a relativiser elsewhere, such as:

- (235) (a) (G02:1100.39) For better or for worse, we all now live in welfare states, [[the organizing principle of [which]_{RELATIVISER}]_{SUBJECT NP} is collective responsibility for individual well-being.]_{RELATIVE}
- (b) * For better or for worse, we all now live in welfare states, that the organizing

principle of which is collective responsibility for individual well-being.

So I will have to amend the code to check for relativisers in any relative clauses that start with a subject noun phrase:

- search for a relative clause
- if (the first word in the relative clause is the relativiser *that* && the next constituent is a subject noun phrase) { remove relativiser }
- else if (the first word in the relative clause is a subject noun phrase && the clause does not contain a relativiser) { insert *that* at start of relative clause }
- the end

This time the results improved. The incidence fell somewhat to 3.2% of sentences, but the reliability increased to 98%. The only sentence that failed was the incorrectly parsed *that* complement mentioned above (234). So overall the method proves successful.

One thing to notice from the results is that *that* insertions are much more common than *that* deletions. Of the 208 constructions identified, only 33 were candidates for insertion. This means that some applications using this method would have been designed to ensure that *that* deletions were roughly balanced by insertions, so as not to disturb the statistical profile of the document.

Relative Clause Reduction Rule

- In the case of subject gapped restrictive relative clauses, headed by the verb BE and directly followed by a past participle or present participle clause or by a complementised adjectival, time/place adverbial or prepositional phrase: then the relative clause can be reduced by removing the relativiser together with the verb BE.

Here again I am looking for restrictive clauses, so in the algorithm I will limit the search to clauses with *that* as their relativiser. This should be followed by a form of the verb BE, and since I am looking for object gapped relatives, no subject noun phrase should intervene:

- search for a relative clause
- if(the first constituent is the relativiser *that* && the next verb phrase contains a form of BE && that no subject noun phrase intervenes between the relativiser and the verb && if the BE verb phrase is followed by a present participle, or past participle clause, or by an complex adjectival, time/place adverbial, or time/place prepositional phrase) { remove *that* and the verb BE }
- the end

The *SUSANNE* scheme has specific notation for present and past participle clauses. It also includes tags for prepositional phrases and adverbs of both time and place/direction. It even has a tag Jh for what it calls ‘post-modified’ adjectival phrases, for example (236), which seems to correspond to the complementised adjectives I have been looking at.

(236) The number of nuclei is [Jh dependent [P upon the physical and chemical properties of ...
P] Jh] (Sampson 1995, p.214)

The algorithm succeeded in finding a few examples (thirteen), but had suspiciously few past and present participle relatives (in fact only one). In the *SUSANNE* documentation (Sampson 1995, pp.264-268) the tags Tn and Tg are introduced to denote past and present participle clauses respectively. One might expect structures such as the following:

(237) ...
[Fr
 [CST that]
 [s133-
 s133]
 [Vzu
 [VBZ 's]
 Vzu]
 [Tg
 [Vzu
 [VVGv taking]
 Vzu]
 [Ns:o
 [AT the]
 [JJ big]
 [NN1n chance]
 Ns:o]
 Tg]
Fr]

However after some manual searching through the texts I found that structures like this are represented differently. The present or past participle is generally considered part of the BE verb phrase as seen below:

```
(238) [S
      [Ny:s
      [PPY You]
      Ny:s]
      [Vab
      [VBR 're]
      Vab]
      [Ms:e133
      [AT the]
      [MC1 one]
      [Fr
      [CST that]
      [s133-
      s133]
      [Vzu
      [VBZ 's]
      [VVGv taking]
      Vzu]
      [Ns:o
      [AT the]
      [JJ big]
      [NN1n chance]
      Ns:o]
      Fr]
      Ms:e133]
      S]
      (N03:0790.27: You're the one that's taking the big chance.)
```

So I modified the algorithm to also look for present and past participles in the relative verb phrase itself, yielding much more results (see Appendix D3). The total was 42, giving an incidence of 0.67%, or about one in 150 sentences. Because there were so few, all 42 were extracted to be examined manually. Of these 37 were correctly selected (88%), while three of these failed to transform, giving a final success rate of 83%.

Three of the wrongly selected sentences contained modals such as *could*, *would* or *may*:

```
(239) (N08:1400.18) * There was a peculiar density about it, a thick substance that could be
      sensed but never identified, never actually perceived.
```

However the tense modals (*have*, *will*) do allow this:

```
(240) (A07:1070.18) In the $40,000,000 budget that has been submitted for Congressional
```

approval, \$26,000,000 would be spent through universities and private voluntary agencies.

- (241) (A12 : 0070 . 03) Moritz doesn't even have a nose guard or hip pads but he's one of the most valuable members of the Longhorn team ~~that will be~~ heavily favored Saturday over Oklahoma in the Cotton Bowl.

One sentence to be wrongly identified was due to incorrect parsing. From the first clause it is clear that the second clause is a *that* complement, not a relative clause as indicated in the corpus:

- (242) (G13 : 1050 . 33) * It is sex that obsesses them, sex ~~that is~~ at the basis of their aesthetic creed.

In the next failed sentence a collocation leads to incorrect transformation. Here *at the same time* is treated as a regular prepositional phrase, while in fact it is functionally a connecting adverb:

- (243) (J23 : 2070 . 33) * Religion has its own supernatural prescriptions ~~that are~~ at the same time codes of behavior for the here and now.

It is difficult to say why sentence (244) is not wholly acceptable. It appears that the relative clause is non-restrictive – *that was forming a pool of blood on the floor* does not help us decide which *ugly gash* is being talked about. Rather it provides some incidental information.

- (244) (N03 : 1690 . 03) ? The bullet had torn through the flesh just above the knee, inflicting an ugly gash ~~that was~~ forming a pool of blood on the floor.

So I could claim that the sentence was originally flawed, and should have used the relativiser *which*. But there were other sentences in the sample that transformed successfully, and appeared to be non-restrictive while using *that*. For example in (245) it is clear in this context that *society* refers to general society. Thus the relative clause cannot have a reference clarifying role, and must be non-restrictive.

- (245) (N11 : 1690 . 09) The hands and their bosses saw him as a lone knight of the range, waging a dedicated crusade against a lawless new society ~~that was~~ threatening a beloved

way of life.

Another possibility is that the sentence is logically flawed, in that it cannot be the *ugly gash* itself that is *forming a pool*, but rather blood from the *ugly gash*.

The last sentence to fail is somewhat clumsy. This is because the head noun is separated from its relative clause by a prepositional phrase modifier.

(246) (A07:0060.24) ? Some anti-organization Democrats saw in the program an opportunity to end [the bitter internal [fight]_{HEAD} [within the Democratic party]_{PP MODIFIER} [~~that has been~~ going on for the last three years]_{RELATIVE}]_{NP}.

So I reran the algorithm after modifying it to exclude modals other than *have*, *shall* and *will*. No new sentences were found, but all the failed modal sentences were excluded. This increased the success rate to a respectable 90%. But as we will see at the end of the next section, it is not reliably reversible.

Relative Clause Expansion Rule

- Generally in the case of reduced relative clauses: the clause can be expanded by inserting the relativiser *that* together with the appropriate form of BE. If the main clause is in a present or future tense, BE must be in the present simple tense. If the main clause is in a past tense, BE must be in the past simple tense. BE must also agree in person and number with the subject of the main clause.
- However, in the case of reduced past participle clauses: the clause can be expanded by inserting the relativiser *that* together with the appropriate form of BE. BE must be in the past simple tense and agree in number with the subject of the main clause.

It turns out that the *SUSANNE* corpus has a dedicated phrase tag (*z*) for reduced relatives. In the example below we see a reduced relative of a prepositional phrase of time or place. In this scheme *II* stands for preposition, *P* for prepositional phrase, and *:p/ :t* for place or time respectively.


```

(247) ...
      [Ns:108
        [AT the]
        [NN1n power]
        [NN1n system]
        [Z
          [s108-
            s108]
          [P:p
            [II in]
            [NN1n existence]
          P:p]
          [P:t
            [II during]
            [Np
              [AT the]
              [MD&
                [MD eighteenth]
                [MD+
                  [CC and]
                  [MD nineteenth]
                  MD+]
              MD&]
              [NNT2 centuries]
            Np]
          P:t]
        Z]
      Ns:108]
...
(G02:0110.30: That corpus of law was a reflection of the power system in existence during
the eighteenth and nineteenth centuries.)

```

However to test the strength of my hypotheses meaningfully I will try to identify reduced relatives from their internal structure, in a reversal of the last algorithm. This means that I will be looking for noun phrases directly followed by past or present participle clauses, by adverbs or prepositional phrases of time or place, or by complex (ie complementised) adjectival phrases. In pseudocode:

- search for a noun phrase
- if(the next constituent is a present participle clause, past participle clause, complex adjectival phrase, time/place adverbial phrase, or time/place prepositional phrase) { insert *that* and the verb BE }
- if(the constituent was a past participle clause) { BE must be in the past simple tense, and agree in number and person with the noun phrase }
- else if(the main clause is in a present or future tense) { BE must be in the present simple tense, and agree in number and person with the noun phrase }

- else if(the main clause is in a past tense) { BE must be in the past simple tense, and agree in number and person with the noun phrase }
- the end

The results of this as seen in Appendix D4 are the least successful yet. For a start, 35% of the sentences extracted were incorrectly identified. The majority of these were noun phrases that take a present participle as a modifier, but are not derived from *that BE* relatives. If (248) is derived from a relative, it would be something like (249).

(248) (J12:0390.51) * Some of these were obviously filling from interlobular branches of the bronchial arteries while others were filling from direct hilar branches that were following along the pleural surface.

(249) Some of these were obviously filling from interlobular branches of the bronchial arteries while others were filling from direct hilar branches that follow along the pleural surface.

More worrying is the failure of so many of the correctly identified sentences to transform successfully. As mentioned in the last chapter, the rules for determining the form and tense of the inserted BE may be affected by supra-sentential factors. For example in (250a) I cannot be sure if energy transfers only in these specific conditions, in which case (250b) would be an appropriate paraphrase; or if this happens generally, making (250c) the best option.

(250) (a) (J02:0020.42) Thus, the energy transferred from the arc to the anode was partly fed back into the arc.

(b) Thus, the energy that was transferred from the arc to the anode was partly fed back into the arc.

(c) Thus, the energy that is transferred from the arc to the anode was partly fed back into the arc.

Similarly, by looking at the sentence (251a) in isolation, it is impossible to know which, if any, of the sentences in (251b-e) mean the same thing.

(251) (a) (N03:0200.33) No doubt there would be men guarding the horses.

(b) No doubt there would be men that were guarding the horses.

- (c) No doubt there would be men that would be guarding the horses.
- (d) No doubt there would be men that are guarding the horses.
- (e) No doubt there would be men that guarding the horses.

Since the constituent structure of the incorrectly identified sentences is identical to that of my target structures, I cannot see a syntactically based solution. Of course I could search for *SUSANNE*'s reduced relative tag *z* but this would be a rather trivial, since I cannot expect automatic parsers to have the same capability in identifying them. Interestingly all the *z* tagged examples I have found are of place/time adverbial or prepositional phrases, which do not appear to have the problems of the participle reduced relatives.

Nevertheless the problem of verb tense and form would remain. To resolve this would require some sort of intelligent model of context, which I do not consider viable. So I am forced to abandon this method.

Adjunct Movement Rule

- In the case adjuncts of time, place, manner or contingency, that take the form of a prepositional phrase, or an adverbial clause introduced by a subordinating conjunction, and that attach to a sentence whose last complement is not clausal: the adjunct may be freely moved to the start or end of the sentence that it attaches to.

In pseudocode the first attempt at the Adjunct Movement Rule is:

- search for an adverbial clause or prepositional phrase, whose function is of time, place, manner or contingency
- if(the sentence's last complement is clausal) { abort }
- if(the adjunct is at the start of the sentence) { move it to the end }
- else if(the adjunct is at the end of the sentence) { move it to the start }
- the end

Here I get 72 target constructions – an incidence of just under one in a hundred. But the success rate is dismal, at only 50 percent. Looking at the failed sentences, most contain co-ordinate sentence structures, which the algorithm was not sensitive to. For example (252a) is transformed into (252b), because it does not recognise that the adjunct in question *in 1900* attaches to the co-ordinated sentence *the South was only 15% urban*. The correct result would be (252c).

- (252) (a) (G08 : 0290 . 18) In 1900 the South was only 15% urban; in 1950 it had become 47.1% urban.
- (b) * The South was only 15% urban; in 1950 it had become 47.1% urban in 1900.
- (c) The South was only 15% urban in 1900; in 1950 it had become 47.1% urban.

Some sentences failed because the algorithm was also not sensitive to multiple adjuncts in series – it only transformed the first one, resulting in incorrect transformations such as (253b). The correct form would be (253c)

- (253) (a) (A08 : 0090 . 12) From June, 1942, until December, 1945, Pfaff served in the Army Air Corps.
- (b) @ Until December, 1945, Pfaff served in the Army Air Corps from June, 1942.
- (c) Pfaff served in the Army Air Corps from June, 1942, until December, 1945.

A couple of sentences failed because, after transformation, there was no obvious subject to the sentence:

- (254) (G18 : 0010 . 06) * ~~AMONG THE RECIPIENTS of the Nobel Prize for Literature~~ More than half are practically unknown to readers of English among the recipients of the Nobel Prize for Literature.

The result of one sentence was clumsy, since there were adjuncts at both ends of the original. In (255) it is unclear whether *in 1* modifies the sentence, or the modifier *of complex numbers* that it follows.

- (255) (J21 : 0010 . 15) ~~In 1~~ We investigate a new series of line involutions in a projective space of three dimensions over the field of complex numbers in 1.

So I modified the pseudocode to exclude these problem cases:

- identify a bottom level sentential constituent (including co-ordinate sentences)
- search for a single or multiple adverbial clauses or prepositional phrases, whose function is of time, place, manner or contingency
- if(the sentential constituent has adjuncts at both ends) { abort }
- if(the sentential constituent's last complement is clausal) { abort }
- if(the adjunct is at the start of the sentential constituent) { move it to the end }
- else if(the adjunct is at the end of the sentential constituent) { move it to the start }
- if(the sentential constituent now has no subject) { abort }
- the end

Now the results are much more encouraging. Only one sentence was incorrectly identified, and one further sentence failed to transform, giving a final success rate of 94%, and incidence of 1.1%.

The incorrectly identified sentence was (256a). Here *In Laos* appears to modify the whole sentence, but in fact only refers to the first instance of *the administration*. The algorithm produced the sentence (256b), which does not preserve meaning.

- (256) (a) (A04:1410.27) In Laos, the administration looked at the Eisenhower administration efforts to show determination by sailing a naval fleet into Southeast Asian waters as a useless gesture.
- (b) # The administration looked at the Eisenhower administration efforts to show determination by sailing a naval fleet into Southeast Asian waters as a useless gesture, in Laos.

The last sentence to fail was due to a functional issue. Despite having the syntactic appearance of a sentential modifier, *At the same time*, functions as a discourse marker acting as a connective between this sentence and the text that precedes it. Obviously it cannot be moved to the end of the sentence.

(257) (a) (A07:0040.18) At the same time reaction among anti-organization Democratic leaders and in the Liberal party to the Mayor's reported plan was generally favorable.

(b) # Reaction among anti-organization Democratic leaders and in the Liberal party to the Mayor's reported plan was generally favorable at the same time.

Some improvement could be made in the performance of this method by providing a list of discourse markers that should be ignored by the algorithm. However I cannot see a syntactic method for picking out cases like (256).

It Insertion

Extraposition Rule

- In the case of sentences with a *that* complement, interrogative *that* complement, infinitival clause, interrogative infinitival clause or *for-to* clause as subject: then the subject can be replaced by *it*, and moved to the end of the sentence. However this should be avoided if the sentence already has a similar clausal object.
- Conversely, in the case of sentences with *it* as subject, and ending with a *that* complement, interrogative *that* complement, infinitival clause, interrogative infinitival clause or *for-to* clause as object: then the final clausal object can be moved to the subject position, replacing *it*. However this should be avoided if the clausal object is the only object (or *must* be avoided in the case of raising verbs).

Here my objective is to find certain subject complement clauses, or certain object complement clauses with *it* as subject. In the *SUSANNE* scheme, subjects are usually annotated with the lowercase letter *s*, which they refer to as logical subject. However, when an element appears in the subject position (ie is the syntactic subject), but is not the logical subject (ie semantic

subject) it is annotated with a capital S. This applies to the expletive *it* I am looking for.

Semantic subjects in positions other than that of the syntactic subject, continue to be annotated as logical subject.

- (258) (a) [That he won the prize]_{LOGICAL SUBJECT} is disgraceful
(b) [It]_{SURFACE SUBJECT} is disgraceful [that he won the prize]_{LOGICAL SUBJECT}

This means that the complement clauses I am looking for are tagged as subjects, whether or not they are in the extraposed position (259). This will simplify my code a little. As seen before, *SUSANNE* also has tags for the various types of complement I am looking for.

```
(259) [S
      [Ni:S
        [PPH1 It]
      Ni:S]
      [Vzb
        [VBZ is]
      Vzb]
      [J:e
        [JJ impossible]
      J:e]
      [Ti:s
        [Vi
          [TO to]
          [VV0v get]
        Vi]
        [Ns:o
          [AT1 a]
          [JJ fair]
          [NN1n trial]
        Ns:o]
      ...
      (A03:0120.30 It is impossible to get a fair trial when some of the defendants made
      statements involving themselves and others.)
```

One problem is that the scheme does not explicitly mark raising verbs, so I will have to identify them manually for the time being. However computational resources are available that could perform this task automatically, for example the COMLEX lexicon (Grisham et al, 1994).

- search for sentences that have a *that* complement, *for-to* complement or infinitival complement, whether interrogative or declarative, as their logical subject
- if(the subject complement starts the sentence and the verb does not already have a complement of the same type as an object) { move the complement to the end of the sentence, and insert *it* as the surface subject }
- if(the subject complement ends the sentence and *it* starts

```

the sentence) { unless(the main verb is a raising verb and
there is no non-clausal object) { move the complement to
the start of the sentence, replacing it as the surface
subject } }

```

- the end

The script found 68 candidate constructions in all (incidence 1.1%). Of these 37 were extracted and examined. I found all to have been correctly identified, but only 30 transformed correctly (81%).

Some of the sentences exposed some holes in the rule. I already had a condition that raising verbs must have non-clausal object to allow reversal of extraposition. The examples examined revealed that the object must be direct (260), and that this condition applies to all copula verbs (261).

- (260) (a) (J22:0230.30) It may seem [to some of them]_{INDIRECT OBJECT} that success can be purchased much less dearly by fishing in the murky waters of international politics than by facing up to the intractable tasks at home.
- (b) * That success can be purchased much less dearly by fishing in the murky waters of international politics than by facing up to the intractable tasks at home may seem [to some of them]_{INDIRECT OBJECT}.
- (261) (a) (N07:0470.30) It [was]_{COPULA VERB} to him that Barton had sent Carl Dill on Dill's release from the prison.
- (b) * That Barton had sent Carl Dill on Dill's release from the prison [was]_{COPULA VERB} to him.

As seen in some earlier sections, contractions were a problem for one sentence (262) and collocations refused to transform elegantly, in this case *it is well* (G07:0050.06), and *it is because ... that ...* (G12:1060.15):

- (262) (a) (N14:1010.06) It'll be a pleasure for you to return this money to Colcord and tell him about it.
- (b) * For you to return this money to Colcord and tell him about it'll be a pleasure.

While many sentences ended up rather unwieldy because of the length of the extraposed constituent, I only judged one to be unacceptable:

- (263) (a) (G08:0180.09) It is interesting, however, that despite this strong upsurge in Southern writing, almost none of the writers has forsaken the firmly entrenched concept of the white-suited big-daddy colonel sipping a mint julep as he silently recounts the revenue from the season's cotton and tobacco crops; of the stereotyped Negro servants chanting hymns as they plow the fields; of these and a host of other antiquated legends that deny the South its progressive leaps of the past century.
- (b) ? That despite this strong upsurge in Southern writing, almost none of the writers has forsaken the firmly entrenched concept of the white-suited big-daddy colonel sipping a mint julep as he silently recounts the revenue from the season's cotton and tobacco crops; of the stereotyped Negro servants chanting hymns as they plow the fields; of these and a host of other antiquated legends that deny the South its progressive leaps of the past century is interesting, however.

Finally there was one sentence that I judged not to be fully acceptable, but where I could not determine the reason. There are many other examples with longer or more complex extraposed constituents that transform successfully (eg G10:1540.15).

- (264) (a) (J05:1830.33) It is evident that the requirements imposed by these effects upon any one detergent constituent acting alone are severe.
- (b) ? That the requirements imposed by these effects upon any one detergent constituent acting alone are severe is evident.

Based on these results I reran the analysis with some changes to the algorithm:

- search for sentences where the main verb is not contracted, and that have a *that* complement, *for-to* complement or infinitival complement, whether interrogative or declarative, as their logical subject
- if(the subject complement starts the sentence and the verb does not already have a complement of the same type as an

- ```

object) { move the complement to the end of the sentence,
and insert it as the surface subject }

```
- if(the subject complement ends the sentence and *it* starts the sentence) { unless(the main verb is a copula verb and there is no direct object) { move the complement to the start of the sentence, replacing *it* as the surface subject } }
  - the end

The results improved considerably, achieving 88% success. No new unsuccessful sentences emerged, and the sentences that remained were those that failed because of constituent length or collocations.

## Subject Raising Rule

- In the case of sentences headed by a subject raising predicate, with an embedded infinitival clause following the predicate, and that infinitival clause's semantic subject NP preceding the predicate: then the embedded clause can be reconstituted as a finite *that* complement, with the semantic subject NP moving to the subject gap position, and *it* replacing it as subject of the raising predicate. If the main clause is in a present or future tense, the embedded clause must be in the present simple tense. If the main clause is in a past tense, the embedded clause must be in the past simple tense. The finite embedded verb must agree in person and number with the moved subject.
- Conversely, in the case of sentences headed by a subject raising predicate, with *it* as syntactic subject, and a full *that* complement as object: then the NP subject of the *that* complement can be moved to the raising predicate subject position, replacing *it*, while the complement becomes an infinitival clause. The raising predicate verb must agree in person and number with the moved subject

We have already met all the constituents that we will be dealing with in this section.

Identification of subject raising predicates is a problem though. As mentioned in the last chapter, I cannot identify them from the syntactic structure that surrounds them, so they must be listed in the algorithm. Neither of the two works on verb argument structure that I consulted

– Levin (1993) or COBUILD (1996) – supply a list of raising verbs. For the purpose of this test I am limiting the search to *seem*, *appear*, *BE certain* and *BE likely*, so the incidence figures will be of limited value.

In terms of the *SUSANNE* scheme, the *that* complements will always be marked as the logical subject, while where it appears *it* will be the surface subject. The first attempt at the algorithm in pseudocode is:

- search for sentences where the main predicate is known to support subject raising
- if(*it* is the surface subject, and a logical subject *that* complement follows the verb) { replace *it* with the subject of the *that* complement and change the remains of the *that* complement into an infinitival clause }
- if(*it* is not the surface subject, and an infinitival complement follows the verb) { replace the surface subject with *it*, and move it into the infinitival clause, forming a *that* complement. if(the raising predicate is in a past tense, the *that* complement must be in the past simple) else(the *that* complement must be in the present simple ) }
- the end

Here the scripts identified 74 sentences of which 40 were extracted for examination. Of those 40, fifteen failed to transform properly. Eight of these were incorrect identifications. They were cases of ellipsis in co-ordinated sentences, such as (265), sentences with a *wh...* word as their subject (266), or embedded prepositional phrases, such as (267). Since I had not foreseen sentences of this type, some could not be transformed.

(265) (N05:0800.21) “He works for my father”, the girl said, and [then [GAP]<sub>ELLIPSIS</sub> seemed to change her mind]<sub>COORDINATED SENTENCE</sub>.

(266) (a) (J12:0210.39) In this instance, because of incomplete septation, the secondary lobule does not constitute in itself [what]<sub>SUBJECT</sub> appears to be a small individual lung as in type I.

(b) \* In this instance, because of incomplete septation, the secondary lobule does not constitute in itself it appears that what is small individual lung as in type I.

- (267) (a) (G12:0330.30) He seems, [by some unconscious division of labor]<sub>pp</sub>, to have given them that one function and no other, leaving communication to the rest of the face.
- (b) ? It seems, by some unconscious division of labor, that he has given them that one function and no other, leaving communication to the rest of the face.

Cases of ellipsis cannot be transformed this way, and so should be excluded from the search. Prepositional phrases like that above belong before the infinitival in raised structures, and after *that* in unraised versions (268).

- (268) It seems that, [by some unconscious division of labor]<sub>pp</sub>, he has given them that one function and no other, leaving communication to the rest of the face.

Two sentences failed because I had not incorporated the limitation on multiple arguments into the algorithm. As shown in the last chapter (125), these cannot be raised.

- (269) (a) (J17:0130.12) It seems to follow that by and large an antagonism exists between the paleo- and the neocortex as far as emotional reactivity is concerned, and that the balance between the two systems determines the emotional responsiveness of the organism.
- (b) \* An antagonism seems to follow by and large to exist between the paleo- and the neocortex as far as emotional reactivity is concerned, and that the balance between the two systems determines the emotional responsiveness of the organism.

I also found that modals together with the raising verb or embedded verb prevented transformation, either because the modal attaches to the subject (270), or because no infinitival form of the modal exists (271)

- (270) (a) (A12:0420.12) I've tried to teach some of the other boys to kick and some of them [can]<sub>MODAL</sub> 't seem to get the feel.
- (b) \* I've tried to teach some of the other boys to kick and it can't seem that some of them get the feel.
- (c) I've tried to teach some of the other boys to kick and it seems that some of them can't get the feel.
- (271) (a) (A13:1010.09) Just when it seems baseball might be losing its grip on the masses

up pops heroics to start millions of tongues to wagging.

- (b) \* Just when baseball seems to might be losing its grip on the masses up pops heroics to start millions of tongues to wagging.

Finally one sentence failed because of the issues of tense discussed in the last chapter.

Following the rules I have developed results in (272b), which does not correctly communicate that the act of covering has already been completed. (272c) would be more appropriate.

- (272) (a) (J08:1040.30) It is quite likely that an even greater area was covered, particularly downwind.
- (b) # An even greater area is quite likely to be covered, particularly downwind.
- (c) An even greater area is quite likely to have been covered, particularly downwind.

The algorithm was modified and run again:

- search for sentences where the main predicate is known to support subject raising
- if(a modal verb is in combination with either the main predicate or the embedded predicate) { abort }
- if(the complement contains co-ordinate structures) { abort }
- if(subject is a *wh...* word) { abort }
- if(non-PP arguments intervene between the main predicate and the complement) { abort }
- if(*it* is the surface subject, and a logical subject *that* complement follows the verb) { replace *it* with the subject of the *that* complement and change the remains of the *that* complement into an infinitival clause }
- if(*it* is not the surface subject, and an infinitival complement follows the verb) { replace the surface subject with *it*, and move it into the infinitival clause, forming a *that* complement. if(the raising predicate is in a past tense, the *that* complement must be in the past simple) else(the *that* complement must be in the present simple }
- the end

The number of candidate structures found dropped to 45, but since I do not have a comprehensive list of raising predicates this does not tell us much. Reliability was much improved (97%) but still one questionable sentence remained from the last round, which seems grammatical but very awkward.

(273) (a) (G09:0160.27) He began with masks to make the dancer identify himself with the creature he appeared to be.

(b) ? He began with masks to make the dancer identify himself with the creature it appeared that he was.

## Object Raising Rule

- In the case of sentences headed by an object raising predicate, with an embedded infinitival clause following the predicate, and that infinitival clause's semantic object NP preceding the predicate: then the semantic object can be moved to the object gap position in the infinitival clause, with *it* replacing it as subject NP of the raising predicate. The moved object must have objective case, and the main verb BE must agree in person and number with *it*.
- Conversely, in the case of sentences headed by an object raising predicate, with *it* as subject, and an infinitival clause as object: then a bare or PP-embedded NP object of the embedded clause can be moved to the raising predicate subject position, replacing *it*. The moved object must have nominative case, and the verb BE must agree with it in number and person.

Here the *SUSANNE* scheme makes the job of identifying target sentences much easier, because it marks subjects and objects explicitly as syntactic, semantic or gapped. I am looking for sentences based around object raising predicates, and since I do not have a lexicon with a comprehensive list of these, I will restrict the search to *BE easy, difficult, tough, impossible or hard*.

Unraised sentences can be recognised by the fact that the subject *it* is tagged as a surface subject, and that the infinitival complement contains a logical object NP. Raised sentences

have a logical subject, which is co-referential with the corresponding object gap in the infinitival complement.

(274) It<sub>SURFACE SUBJECT</sub> is easy to tire of [writing theses]<sub>LOGICAL OBJECT</sub>

(275) [Writing theses]<sub>LOGICAL SUBJECT</sub> is easy to tire of GAP<sub>OBJECT</sub>

So the pseudocode is:

- search for sentences where the main predicate is known to support object raising and that carry an infinitival complement
- if(the subject is *it* and the infinitival complement contains an object NP) { move the object NP into subject position, replacing *it* }
- if(the subject is not *it* and the infinitival complement contains an object gap) { move the subject NP into the object gap position, replacing it with *it* }
- the new subject must be nominative case
- the verb must agree in person and number with the new subject
- the new object must be objective case
- the end

This algorithm yielded very few sentences, only eight in fact. That is little more than one in a thousand sentences. Among those there were examples of both raising (276), and reversing raising (277).

(276) (A03:0120.30) It A fair trial is impossible to get a ~~fair trial~~ when some of the defendants made statements involving themselves and others.

(277) (G08:1080.51) Writers openly admit that ~~the Negro~~ it is easier to write the Negro than the white man; but they obviously mean by this, not a Negro personality, but a Negro type.

The only sentence that failed was due to ellipsis in co-ordinated sentences, as seen in several of the subject raising sentences:

(278) (a) (J12:0780.24) In addition, it would be difficult to imagine chronic generalized emphysema occurring in a cow, considering its marked lobular development but,

conversely, not difficult to imagine this occurring in the horse or the dog.

(b) \* In addition, chronic generalized emphysema occurring in a cow would be difficult to imagine, considering its marked lobular development but, conversely, not difficult to imagine this occurring in the horse or the dog.

The code could be modified to exclude co-ordinate structures in the infinitival complement, and so increase reliability to 100%, but since the sample is so small, it would not tell us much.

### **There *Insertion and Deletion Rule***

- In the case of BE sentences with a common noun phrase as subject, carrying either an indefinite article or no article, and with a past perfect or present progressive clause as complement: then the subject can be moved between the verb and the complement, being replaced by *there*. The verb continues to agree in person and number with the semantic subject.
- Conversely, in the case of BE sentences with existential *there* as syntactic subject, and with an object noun phrase as semantic subject, followed by a past perfect or present progressive clause: then the semantic subject can be moved to the syntactic subject position, replacing *there*. The verb continues to agree in person and number with the semantic subject.

There is one slight complication when looking for sentences that are candidates for *there* insertion. As in the section on relative clause reduction, *SUSANNE* is somewhat inconsistent in its treatment of present and past participles following BE – compare (279) and (280). The participle is sometimes treated as part of the BE verb phrase, but also sometimes follows the verb phrase as the head of a participle clause.

- (279) (a) She [is taking]<sub>VP</sub> [a break]<sub>NP</sub>  
(b) She [is]<sub>VP</sub> [taking a break]<sub>PRESENT PARTICIPLE CLAUSE</sub>
- (280) (a) He [is tired]<sub>VP</sub> [of life]<sub>NP</sub>  
(b) He [is]<sub>VP</sub> [tired of life]<sub>PAST PARTICIPLE CLAUSE</sub>



So my algorithm must look for both configurations. *SUSANNE* also has specific tags for definite and indefinite determiners and quantifiers, so this should present no problems.

The pseudocode for the algorithm is:

- search for *BE* sentences preceded by an *NP* subject and followed by a past or present participle clause
- if(the subject head is not a common noun) { abort }
- unless(the subject head has no article or the indefinite article) { abort }
- move the subject between the verb and complement, and put *there* in the subject position
- search for *BE* sentences preceded by existential *there* and followed by an *NP* semantic subject and a past or present participle clause
- move the semantic subject to the subject position, replacing *there*
- the end

The script did not find many instances in the corpus. All 32 were extracted, and 8 failed to transform properly, giving a success rate of 75%. The main reason for failures was where subjects with no determiner nevertheless had a definite reference. I had posited that subjects with no determiner, paired with a transient complement would force an existential reading, and so allow *there* insertion. This turned out not to be the case.

- (281) (a) (A13:1010.09) Just when it seems baseball might be losing its grip on the masses  
up pops heroics to start millions of tongues to wagging.
- (b) \* Just when it seems there might be baseball losing its grip on the masses up pops  
heroics to start millions of tongues to wagging.

One sentence failed because the algorithm did not take account of modifiers at the head of the sentence, that can follow *there* but must precede a noun phrase subject when *there* is deleted.

- (282) (a) (A04:1620.42) There certainly was not any more energy displayed after it was  
clear the United States would not back the pro-Western government to the hilt.

(b) \* Not any more energy certainly was displayed after it was clear the United States would not back the pro-Western government to the hilt.

(c) Certainly not any more energy was displayed after it was clear the United States would not back the pro-Western government to the hilt.

The last sentence was clumsy because of the length and complexity of the moved subject:

(283) (a) (A06:1800.15) Fifteen members of the Republican State Committee who are retiring — voluntarily — this year were honored yesterday by their colleagues.

(b) ? There were fifteen members of the Republican State Committee who are retiring — voluntarily — this year honored yesterday by their colleagues.

On the basis of these results I modified the algorithm to handle modifier position properly, and to exclude all sentences without an indefinite determiner.

- search for BE sentences preceded by an NP subject and followed by a past or present participle clause
- if(the subject head is not a common noun) { abort }
- unless(the subject head has an indefinite determiner) { abort }
- move the subject between the verb and complement, and put *there* in the subject position
- search for BE sentences preceded by existential *there* and followed by an NP semantic subject and a past or present participle clause
- move the semantic subject to the subject position, and delete *there*
- the end

Now the results were much improved, with a success rate of 100% over 24 sentences. My hypothesis was that subjects with no determiner, or an indefinite determiner would transform if they were accompanied by a present progressive or past perfect complement. The corpus data showed that this does not hold for subjects without a determiner. However it is interesting to note that sentences with an indefinite determiner transformed successfully despite occurring together with other complement tenses, such as the subjunctive

(G11:1110.33) and the past progressive (N03:1100.30). However the sample obtained from the corpus is probably too small to draw any definitive conclusions from.

## ***Conjunction Movement Rule***

- In the case of co-ordinate structures linked by the conjunctions *and* and *or*, where the conjuncts do not contain verbal, numeral or temporal constituents of any kind, and where equivalent and matching constituents can be identified within each and every of the conjuncts: then the matching constituents can be exchanged in position at will.

In this corpus co-ordinate structures are represented, rather unconventionally, as embedded structures – see (284a). The whole co-ordinated constituent is marked with an ampersand, and each constituent conjunct together with its connecting conjunction is marked with a plus sign. A more traditional representation might be (284b) (McCawley 1988, p.59).

- (284) (a) [ [Tom]<sub>N'</sub> [ [and]<sub>CONJ</sub> [Jerry]<sub>N'</sub> ]<sub>NP+</sub> ]<sub>NP&</sub>  
 (b) [ [Tom]<sub>N'</sub> [and]<sub>CONJ</sub> [Jerry]<sub>N'</sub> ]<sub>NP</sub>

As seen in earlier sections of this chapter, the *SUSANNE* corpus does not mark ellipsis comprehensively. Therefore I will restrict my search to complete balanced strings of nouns, prepositions, adverbs and adjectives. By balanced I mean that each of the arguments contain the same sequence of part-of-speech and phrasal constituents. *SUSANNE* has a very rich part-of-speech scheme, for example *NN<sub>C</sub>* is 'countable plural noun' such as *sheep* and *NNU<sub>C</sub>* is 'countable plural noun of unit' such as *hertz*. So when determining the balance of arguments, I will limit the comparison of wordtags to their first two letters, that denote their class. For example the two wordtags mentioned above start with *NN*, which denotes 'common noun'. Similarly, only the first letter of the phrasetag or clausetag will be considered.

Excluding verbal, numeral and temporal constituents is trivial. Verbal phrasetags all start with *v*, numeral phrases with *M*, while the phrasetags with a temporal function carry the *:t* functiontag. In pseudocode the rule becomes:

- search for sentences with noun, prepositional, adjectival

or adverbial co-ordinate structures

- if(the arguments contain any verbal, numeral or temporal constituents) { abort }
- unless(the number and part-of-speech class of the argument constituents are equivalent) { abort }
- swap the arguments around the conjunction
- the end

This script found a very large number of candidate constructions, being present in almost one in every twelve sentences of the corpus. Of these a sample of 61 sentences was automatically extracted to be examined by hand.

Twelve sentences failed, giving a success rate of 80%. The most common reason for failure (10 sentences) was collocation, eg (285).

- (285) (a) (J08:0030.27) It is the exploitation of the inherent potential of infectious disease agents by scientific research and development, resulting in the production of BW weapons systems.
- (b) ? It is the exploitation of the inherent potential of infectious disease agents by scientific development and research, resulting in the production of BW weapons systems.

Some of these collocations were fixed proper names:

- (286) (a) (A19:0010.06) The Baltimore and Ohio Railroad announced yesterday it would reduce the total amount of its payroll by 10 per cent through salary cuts and lay-offs effective at 12.01 A.M. next Saturday.
- (b) \* The Ohio and Baltimore Railroad announced yesterday it would reduce the total amount of its payroll by 10 per cent through salary cuts and lay-offs effective at 12.01 A.M. next Saturday.

One transformed collocation resulted in what is for many speakers of English, an ungrammatical sentence (287b), the more common form being (287c):

- (287) (a) (N12:0700.51) You and I have a little talking to do.
- (b) % I and you have a little talking to do.

- (c) Me and you have a little talking to do.

One sentence failed because there was some implicit order between the constituents. In the original version (288a) it is implied that the antibodies are first isolated, and only then can they be studied. Although it is a rather subtle difference, (288b) does not preserve this implication.

- (288) (a) (J09:0010.06) A variety of techniques have been directed toward the isolation and study of blood group antibodies.
- (b) # A variety of techniques have been directed toward the study and isolation of blood group antibodies.

The final sentence failed because of incorrect parsing. The prepositional phrase *of points of color on white* was parsed as modifying the whole co-ordinate NP *backdrop and costumes* (289a), while from reading the text it is clear that the prepositional phrase only modifies *costumes* (289b). This led to the incorrectly transformed version (289c). Note that besides changing in meaning, the sentence is now also ungrammatical because of the form of the indefinite article (\* *a matching costumes*).

- (289) (a) (G09:0060.48) He has designed a matching [backdrop and costumes]<sub>NP</sub> [of points of color on white]<sub>PP</sub> for Mr. Cunningham's *Summerspace*, so that background and dancers merge into a shimmering unity.
- (b) He has designed a matching [backdrop]<sub>NP</sub> and [costumes [of points of color on white]<sub>PP</sub>]<sub>NP</sub> for Mr. Cunningham's *Summerspace*, so that background and dancers merge into a shimmering unity.
- (c) \*# He has designed a matching [costumes and backdrop]<sub>NP</sub> [of points of color on white]<sub>PP</sub> for Mr. Cunningham's *Summerspace*, so that background and dancers merge into a shimmering unity.

It is not possible for me to rewrite my script to exclude collocations, but there are some strategies I can use to make them less likely. One is to exclude structures containing personal pronouns, whose sequence is often governed by social convention (eg *his and hers*). Another

is to exclude arguments that include proper nouns and capitalised letters. The algorithm was modified thus:

- search for sentences with noun, prepositional, adjectival or adverbial co-ordinate structures
- if(the arguments contain any verbal, numeral or temporal constituents) { abort }
- if(the arguments contain any personal pronouns, proper nouns or capital letters) { abort }
- unless(the number and part-of-speech class of the argument constituents are equivalent) { abort }
- swap the arguments around the conjunction
- the end

This improved performance considerably. Incidence dropped to one in fifteen sentences, while reliability climbed to a respectable 87%. In the extracted sample, there was one new failed sentence – another case of collocation.

### ***Verb Particle Movement Rule***

- In the case of phrasal verbs that are known to support particle movement, followed by a short direct object of at least two words: the particle can be freely moved between the pre-object and post-object position.

*SUSANNE* classifies phrasal verb particles as a type of adverb, and has a functiontag :n to identify them specifically. So the algorithm has to find instances of adverbs tagged as these particles, and then identify the verb and direct object. It will also check if the object is too short (less than two words) or too long (as yet undefined) for safe transformation.

- search for sentences containing a phrasal verb particle
- identify the verb and direct object
- if(the direct consist of one word) { abort }
- if(the direct is too long) { abort }
- if(the particle follows the verb, move it after the direct object)
- else if(the particle follows the direct object, move it after the verb)

- the end

The sample was not completely arbitrary this time. I was not familiar with the particle verbs in five of the sentences originally extracted (due to dialect difference presumably). These sentences appear under the heading 'excluded' in Appendix D10. Where possible I replaced these sentences with the next sentence available from the same corpus file. This leads me to be somewhat cautious of the conclusions that will follow in this section, since it seems to suggest that particle verb combinations are particularly sensitive to dialectical variation. Remember that the *SUSANNE* corpus is of American English.

Incidence was passable at one in 65 sentences. Six of the sentences extracted were cases of incorrect identification, and the final success rate was 80%.

Two of the incorrectly identified sentences (290), (291) contained prepositional verbs, namely *take on* (ie *assume*) and *make up* (ie *comprise*).

(290) (a) (N08 : 0810 . 21) The forest took on an impersonal aspect.

(b) \* The forest took an impersonal aspect on.

(291) (a) (J05 : 1890 . 09) In the over-all process, it is difficult to assign a "pure" role to each constituent of a built-detergent formulation; and, indeed, there is no more reason to separate the interrelated roles of the active, builder, antiredeposition agent, etc. than there is to assign individual actions to each of the numerous isomers making up a given commercial organic active.

(b) \* In the over-all process, it is difficult to assign a "pure" role to each constituent of a built-detergent formulation; and, indeed, there is no more reason to separate the interrelated roles of the active, builder, antiredeposition agent, etc. than there is to assign individual actions to each of the numerous isomers making a given commercial organic active up.

Interestingly both of these have alternative senses (*undertake* and *fabricate* respectively, which are phrasal verbs, as can be seen in (292-295).

(292) (a) Charles took the mafia on.

- (b) Charles took on the mafia.
- (293) (a) Charles took on a new identity.
- (b) \* Charles took a new identity on.
- (294) (a) The lads made up an alibi.
- (b) The lads made an alibi up.
- (295) (a) The lads made up a great team.
- (b) \* The lads made a great team up.

The remaining four incorrectly identified sentences contained verbs with an adverbial of place or direction (*pick up*, *turn over*, *sputter away* and *take in*), which can resemble a phrasal verb. This can be shown by arbitrarily substituting another adverbial as seen in (296). Although, these sentences do still transform correctly, they constitute a different phenomena, that of adverb movement, which is in turn a special case of adjunct movement.

- (296) (a) “Fred was mighty crude about the way he took in cattle” his own hired man, Andy Ross, mentioned later.
- (b) “Fred was mighty crude about the way he took home cattle” his own hired man, Andy Ross, mentioned later.
- (c) “Fred was mighty crude about the way he took back cattle” his own hired man, Andy Ross, mentioned later.
- (c) “Fred was mighty crude about the way he took out cattle” his own hired man, Andy Ross, mentioned later.

Again some of these verbs have another sense that is phrasal, for example *turn over* in the sense of *hand over*, or *take in* in the sense of *con*.

Altogether eight sentences failed to transform successfully. Four of these were borderline cases, where placing the particle after a long and complex direct object produced unwieldy results.

- (297) (a) (A02:0150.21) It permits the state to take over [bank accounts, stocks and other personal property of persons missing for seven years or more]<sub>DIRECT OBJECT</sub>.



- (b) It permits the state to take [bank accounts, stocks and other personal property of persons missing for seven years or more]<sub>DIRECT OBJECT</sub> over.

In terms of length, the direct objects in failed sentences ranged from eight to fifteen words. The also all include multiple embedded prepositional phrases, that may be a contributing factor. Compare the following sentence, that despite having seven words in the direct object, is much flatter in terms of structure, and transforms perfectly well.

- (298) (a) (G17:0060.03) In town after town my companion pointed out [the Negro school and the White school]<sub>DIRECT OBJECT</sub>, and in every instance the former made a better appearance (it was newer, for one thing).
- (b) In town after town my companion pointed [the Negro school and the White school]<sub>DIRECT OBJECT</sub> out, and in every instance the former made a better appearance (it was newer, for one thing).

The two prepositional verbs discussed above failed (290), (291), and lastly two phrasal verbs came up that only allow the particle in the post-object position (*get ... off* and *look ... over*):

- (299) (a) (N14:0510.21) Looking the setup over, Stevens started coolly for the rear of the place.
- (b) \* Looking over the setup, Stevens started coolly for the rear of the place.
- (300) (a) (A20:0240.39) In their suburban cottage the crown charges, the Krogers received secrets from the mystery man, usually on the first Saturday evening of each month, and spent much of the week-end getting the secrets off to Moscow, either on a powerful transmitter buried under the kitchen floor or as dots posted over period marks in used books.
- (b) \* In their suburban cottage the crown charges, the Krogers received secrets from the mystery man, usually on the first Saturday evening of each month, and spent much of the week-end getting off the secrets to Moscow, either on a powerful transmitter buried under the kitchen floor or as dots posted over period marks in used books.

The problems met here are primarily to do with attributes of specific lexemes. I have found that the *SUSANNE* scheme does not annotate phrasal verbs in a way consistent with my aims. To improve the algorithm a lexical resource (eg *COMLEX*) would be needed to inform the algorithm on which particle verb combinations allow the particle both before and after the direct object, regardless of verb sense.

## Conclusion

### *Results*

It is hard to draw any truly firm conclusions from the results of the last section. A very small corpus was used, and within that, the extracted sample sizes were even smaller. As a result there is a considerable degree of uncertainty about the figures. Despite that, they prove very useful as pointers that can be used to lead on to wider analysis with a larger corpus and more subjects to verify linguistic judgements.

Below is a summary of the results from the last chapter. The reader will notice that one of the requirements has been changed. Previously my aim was that all the analyses be based on syntactic considerations only. Now I have widened that to allow the use any lexical resources (syntactic and semantic) that are amenable to fully automatic processing (ie lists). Accordingly the new criteria is termed 'automatic' in the table below, and the judgements made assume the availability of these lexical resources.

I have marked the transforms as satisfying '✓' a requirement or not '✗'. 'O' indicates an inconclusive result. I have quantified the ubiquity requirement as per-sentence incidence (ie the construction can be found in one out of x typical sentences) and reliability by the percentage of sentences that transform successfully. Care must be taken with these numerical results, since the samples used were relatively small. The margin of error on the reliability figures ranges between  $\pm 12\%$  and  $\pm 35\%$ <sup>18</sup>.

The judgements on semantic, pragmatic and stylistic neutrality are the least objective of the criteria, and so the most open to criticism. The reader will notice that, in my judgement, they are satisfied across the board – as they were among my primary criteria for the selection of transformations, I only chose to analyse constructions that would satisfy them.

---

<sup>18</sup> The margin of error was calculated using the formula  $1/\sqrt{n}$ , where n is the size of the sample (Utts, 1996)

|                        | Semantics | Pragmatics | Stylistic | Sentential | Automatic | Ubiquity            | Reversible | Reliability       |
|------------------------|-----------|------------|-----------|------------|-----------|---------------------|------------|-------------------|
| Switching Relativiser  | ✓         | ✓          | ✓         | ✓          | ✓         | 1/30                | ✗          | 94%               |
| Dropping Relativiser   | ✓         | ✓          | ✓         | ✓          | ✓         | 1/30                | ✓          | 98%               |
| Reducing Relatives     | ✓         | ✓          | ✓         | ✓          | ✓         | 1/160               | ✗          | 90%               |
| Expanding Relatives    | ✓         | ✓          | ✓         | ✗          | ✗         | 1/40                | ✗          | 53%               |
| Adjunct Movement       | ✓         | ✓          | ✓         | ✓          | ✓         | 1/90                | ✓          | 94%               |
| Extrapolation          | ✓         | ✓          | ✓         | ✓          | ✓         | 1/100               | ✓          | 88%               |
| Subject Raising        | ✓         | ✓          | ✓         | ✓          | ✓         | 1/140 <sup>19</sup> | ✓          | 97%               |
| Object Raising         | ✓         | ✓          | ✓         | ✓          | ✓         | 1/800               | ✓          | 87% <sup>20</sup> |
| <i>There</i> Insertion | ✓         | ✓          | ✓         | ✓          | ○         | 1/300               | ✓          | 100%              |
| Conjunct Movement      | ✓         | ✓          | ✓         | ✓          | ✓         | 1/15                | ✓          | 87%               |
| Verb Particle Movement | ✓         | ✓          | ✓         | ✓          | ✓         | 1/65                | ✓          | 80%               |

Table 2: Evaluating the Transformations

Looking at individual methods, most of the transformations appeared promising, except the Reduced Relative Expansion Rule. Here it was impossible to know from the syntactic form whether a clause constituted a reduced relative or not. Furthermore, deciding what tense expanded relatives should have proved to be a problem I could not solve.

The ‘automatic’ criteria is marked inconclusive for *there* insertions. My rule for this method posited that a combination of tense and complement form would determine whether a sentence could be transformed. However the results were mixed, and the sample was small (only 24 sentences), making it difficult to draw any definitive conclusion.

Turning to the reasons for the failure of sentences, I find that out of 35 sentences whose failure could not be resolved, almost a third (11 sentences) were due to collocations. We

---

<sup>19</sup> No comprehensive list of raising predicates was available for this test, so the incidence figures may not be representative. With a proper list, I would expect incidence to increase.

<sup>20</sup> Here the sample extracted was so small ( $n = 8$ ) that this figure is virtually worthless, with a margin of error of  $\pm 35\%$

found many cases where the algorithm tried to alter a collocation, or mistook a collocation for a regular structure during the search phase. Other common causes of failure were performance issues (ie the length and complexity of constituents to be moved), issues of indeterminate tense (discussed above), and parsing difficulties (to be tackled below).

A final and unexpected conclusion is that the statistical profile of texts appears to be quite idiosyncratic. I had assumed that the incidence of particular syntactic structures would show considerable correlation across texts. However my results suggest that this might not be the case. For example, in the relative clause reduction test, almost one third of all target sentences originated in three of the 64 files in the corpus. These three files (A07, J08 and G08) were also from three different genres, suggesting that it is not a stylistic characteristic. Among the extractions for *there* insertion, almost half of the candidate sentences were from three files (A06, G08, J03), again spanning three of the four genres. This leads me to believe that authors are very varied in their choice of constructions, and that my concerns about inadvertently damaging the statistical profile of texts is largely unfounded.

### ***Suitability for Information Hiding Applications***

Below I have listed the four possible target applications, and evaluated the suitability of each transformation to them. Watermarking and fingerprinting have identical demands. Since both may involve high volume automated operations (see the section *Considerations for Specific Applications* in the *Information Hiding* chapter), they cannot make use of a technique with patchy reliability like expanding reduced relatives. *There* insertion is marked inconclusive for all four applications, since the examples of this structure that were found in the corpus did not allow me to draw any concrete conclusions.

Otherwise all the transforms appeared promising for inclusion in watermarking and fingerprinting applications. Remember that in my introduction to the applications I concluded that reversibility is a mixed blessing for these two applications. On the one hand I do not want to dramatically alter the statistical profile of documents. On the other hand, an entirely reversible scheme offers hostile parties the possibility of reverse engineering the algorithm to

recover original copies of documents, which they can then remark at will. Thus the mix of reversible and non-reversible techniques might work very well. Even a small number of non-reversible structures (the Relativiser Switch Rule and the Relative Reduction Rule in this case) would make it impossible to recover the original, while leaving it statistically largely intact.

|                        | Steganography | Watermarking | Fingerprinting | Traitor Tracing |
|------------------------|---------------|--------------|----------------|-----------------|
| Switching Relativiser  | ○             | ✓            | ✓              | ✓               |
| Dropping Relativiser   | ✓             | ✓            | ✓              | ✓               |
| Reducing Relatives     | ○             | ✓            | ✓              | ✓               |
| Expanding Relatives    | ○             | ✗            | ✗              | ○               |
| Adjunct Movement       | ✓             | ✓            | ✓              | ✓               |
| Extraposition          | ✓             | ✓            | ✓              | ✓               |
| Subject Raising        | ✓             | ✓            | ✓              | ✓               |
| Object Raising         | ✓             | ✓            | ✓              | ✓               |
| <i>There</i> Insertion | ○             | ○            | ○              | ○               |
| Conjunct Movement      | ✓             | ✓            | ✓              | ✓               |
| Verb Particle Movement | ✓             | ✓            | ✓              | ✓               |

Table 3: Suitability for Information Hiding Applications

For traitor tracing, pretty much anything goes. While fully automatic operation is clearly desirable, human assisted operation (ie the system would suggest changes to the text, which a human operator would accept or reject) would often be acceptable, given the small amount of data that would be typically stored and the high value and possibly limited length of the document in question. ‘O’ marks those methods that would require human intervention.

For steganography the same two methods are marked inconclusive, since it is possible to imagine highly secretive scenarios where an operator is willing to vet the changes made by the system. Another two methods (switching the relativiser and reducing relative clauses – both non-reversible) are also marked inconclusive, since again this depends on the scenario in which the application is used. Remember that using non-reversible transforms means that the original unmarked document is needed for comparison when extracting the hidden message. However one can imagine situations where this is not a problem – for example I

could use an online article as the cover text for my secret message. On receiving it, the recipient could download the original from the public internet, and extract the message.

Finally I have made some simple calculations to estimate the message carrying capacity achievable by an application using these transforms. Assuming that only one change would be made per sentence (I assume this since I have not tried to investigate any interference there might be between different techniques) and based on the *SUSANNE* data, I calculate the incidence of markable sentences would be 17.5%, or about one in six<sup>21</sup>. Given a simple binary scheme<sup>22</sup> where each transformed sentence carries only one bit of information, this would give an approximate capacity of one bit of hidden data for every 700 bytes of carrier file, or for every 115 words<sup>23</sup>. The aggregate success rate would be 90% (a weighted average), but as will be discussed in the following section, this could improved with the help of lexical resources.

### ***Would it All Work?***

Although the aggregate success rate just calculated (90%) for a system using these transforms is still not acceptable, I believe that great improvements could be made by adding some off-the-shelf lexical resources. As already noted, almost a third of failed sentences (excluding the relative clause expansion rule that has been discarded) were due to the algorithm's inability to recognise idioms and other collocations. Suitable lists of English collocations may already available. They could also be statistically precompiled, by searching a very large corpus, such as the *COBUILD Bank of English* (40 million words) or the *British National Corpus* (100 million words), for common combinations of words. Obviously a search

---

<sup>21</sup> Calculated as follows: eg I have two transformations that apply to 20% and 25% of sentences respectively. The aggregate percentage of applicable sentences would be  $20\% + (100\% - 20\%) \cdot 25\% = 40\%$ .

<sup>22</sup> Attalah, Raskin et al 2001, suggest an intricate system where each node of a sentence's parse tree can carry information, greatly increasing capacity

<sup>23</sup> In plain text *SUSANNE* totals 785Kbytes in 6248 sentences and 125K words – calculations are  $785K / (6248 \cdot 17.5\%)$  bytes and  $125K / (6248 \cdot 17.5\%)$  words

of this kind would turn up many unwanted results such as *is a* or *and the*, but it could be tuned to concentrate on collections of content words rather than function words.

Another (computationally expensive) possibility would be to make searches dynamically while transformations are being performed. The incidence of the original version of a structure could be compared to that of the transformed structure. If the original was much more common than the transformed version, that might be good grounds for rejecting the transformation as unsafe. This comparison might be performed on the basis of strings (ie word comparison) or trees (ie constituent structure comparison).

A second common reason for failed transformations proved to be performance limitations. Some sentences produced were grammatical, but very unwieldy because of the length or complexity of the constituent that was moved or jumped over. To resolve this would first require some quantitative measure of the 'size' of constituents, which would probably be some mix of the number of words or syllables and the depth or complexity of the tree structure. We might also learn something from phonetics, where there are well developed theories of the intonational phrase. These are used in speech synthesis systems to determine natural intonational breaks, and the intonational phrase boundaries found often prove to be radically different from the Chomskian variety. See Abney (1992, 1995), for an example of such a scheme.

Other resources that could be added to improve performance would be lexical lists precompiled for particular functions. For example, it would very easy to compile a list of 'human' nouns from the publicly available *WordNet* lexical network (approx. 60,000 noun entries). I could also use *COMLEX* (a computational lexicon of argument structure, containing 6,000 verbs) to make a list of raising predicates and exclusively phrasal verbs. The adjunct movement transformation could do with a list of discourse markers, although I am not sure where this could be obtained.



The biggest open question is that of parsing. I chose to use a hand parsed corpus at this stage to verify my hypotheses – any problems encountered were most likely due to flaws in my rules, rather than incorrect parsing. However, a theoretically sound algorithm is not the same thing as an algorithm that works! As I introduced the different transformations, I briefly touched on parsing difficulties that I could foresee. But I have not investigated parsing strategies in detail, or more importantly, tested my rules against an off-the-shelf parser. So, whether these rules will work as is, with automatically parsed input, is an open question. The broadly positive results I have obtained tell me that, in a syntactically perfect world, my transformations could work, and it is worth trying them out on real parsers.

### ***What Next?***

The scale of this paper meant that I could not pursue fully some other possible candidate constructions. There are many variations in form of verbal complements, from the most mundane, such as (301), through recipient (302) and benefactive (303) arguments swaps, to more exotic structures such as (304). In her huge work Beth Levin (1993) lists 80 such variations that pivot on particular verbal lexemes.

- (301) (a) I don't think that he knows
- (b) I don't think he knows
- (302) (a) Send a bunch of flowers to Mary!
- (b) Send Mary a bunch of flowers!
- (303) (a) He promised to bake a cake for me
- (b) He promised to bake me a cake
- (304) (a) They sprayed slurry onto the field
- (b) They sprayed the field with slurry

Other transformations that might be investigated are adjective swapping (305), adverb movement (306), *wh*-preposition movement (307), and noun-noun compounds (308).

- (305) (a) Augustiner has a clean fresh taste
- (b) Augustiner has a fresh clean taste
- (306) (a) Occasionally we would go down to the seaside to get a 99

- (b) We occasionally would go down to the seaside to get a 99
  - (c) We would occasionally go down to the seaside to get a 99
  - (d) We would go down occasionally to the seaside to get a 99
  - (e) We would go down to the seaside occasionally to get a 99
  - (f) We would go down to the seaside to get a 99 occasionally
- (307) (a) With whom did you come?
- (b) Whom did you come with?
- (308) (a) ... poetry book ...
- (b) ... book of poetry ...

There were some significant limitations in the methodology of this work. As already mentioned, a hand-parsed corpus was chosen so that I could assume any problems encountered were due to flaws in the rules, rather than flaws in parsing. Furthermore, being manually produced, the corpus was very small. To give more weight to my results, it would be necessary to test against a much larger corpus, and with the raw output of an automatic parser. This process would hopefully also expose more shortcomings, and so allow me to refine the algorithms. And I might be surprised by success rates, since there is no guarantee that transformations that work well with hand-parses, also work well with automatic parses.

I have tried not to rely on orthography, so that my rules would be equally applicable to spoken English. So it would be interesting to test the algorithms on a spoken corpus. In fact Geoffrey Sampson, has also produced *CHRISTINE*, a slightly smaller corpus of spoken English, to accompany *SUSANNE*. I might find that some transformations work better with information that is discarded in written texts. For example, I believe intonation would be a much more reliable indicator of a relative clause being restrictive or not than orthographic practise (ie commas).

Another limitation was in the judgement process, where transformed sentences were evaluated for success. I did not perform extensive testing of sentences among different subjects. The judgements used are almost exclusively mine, and included pragmatic judgements, which are notoriously slippery. Ideally a group of naïve subjects should evaluate

the sentences individually, marking them on a magnitude estimation scale (Sorace et al, 1996), so that their results could be meaningfully and quantitatively collated.

## Appendices

### ***Appendix A: Notational Conventions***

#### Grammaticality

|   |                                           |
|---|-------------------------------------------|
| * | totally ungrammatical                     |
| % | grammatical in some dialects              |
| ? | questionably grammatical                  |
| # | semantically/pragmatically non-equivalent |
| @ | semantically/pragmatically anomalous      |

#### Transformations

|                 |               |
|-----------------|---------------|
| <del>word</del> | deleted text  |
| <u>word</u>     | inserted text |

#### Parts of Speech

|       |                 |
|-------|-----------------|
| N     | Noun            |
| V     | Verb            |
| P     | Preposition     |
| Adj   | Adjective       |
| Adv   | Adverb          |
| Num   | Numeral         |
| Gen   | Genitive Marker |
| Det   | Determiner      |
| Conj  | Conjunction     |
| Quant | Quantifier      |

#### Phrasal/Clausal Constituents

|       |                      |
|-------|----------------------|
| NP    | Noun Phrase          |
| VP    | Verb Phrase          |
| PP    | Prepositional Phrase |
| AdjP  | Adjectival Phrase    |
| AdvP  | Adverbial Phrase     |
| NumP  | Numeral Phrase       |
| GenP  | Genitive Phrase      |
| DetP  | Determiner Phrase    |
| FinC  | Finite Clause        |
| NonfC | Non-Finite Clause    |
| QuotC | Quoted Clause        |

## **Appendix B: Notational Conventions of the SUSANNE Corpus**

Note that the contents of this appendix are not comprehensive. I only include classes of constituents that are pertinent to the processes discussed in this document. Sampson (1995) is the authoritative reference for the scheme used in *SUSANNE*.

### Wordtags (Parts of Speech)

|        |                                            |
|--------|--------------------------------------------|
| AT...  | Article                                    |
| CC...  | Co-ordinating Conjunction                  |
| CS...  | Subordinating Conjunction                  |
| CST    | <i>that</i> as a Subordinating Conjunction |
| DD...  | Determiner                                 |
| EX     | Existential <i>there</i>                   |
| Il...  | Preposition                                |
| JJ...  | Adjective                                  |
| MC...  | Cardinal Numeral                           |
| MD...  | Ordinal Numeral                            |
| NN...  | Common Noun                                |
| NP...  | Proper Noun                                |
| PN...  | Pronoun                                    |
| PP...  | Personal Pronoun                           |
| RL...  | Adverb of Place or Direction               |
| RR...  | Adverb                                     |
| RT...  | Adverb of Time                             |
| VB...  | Verb BE                                    |
| VD...  | Verb DO                                    |
| VH...  | Verb HAVE                                  |
| VM...  | Modal Verb                                 |
| VVO... | Base Form of Verb                          |
| VVD... | Past Tense Form of Verb                    |
| VVG... | Present Participle Form of Verb            |
| VVD... | Past Participle Form of Verb               |

### Phrashtags and Claushtags (Phrasal/Clausal Constituents)

|      |                           |
|------|---------------------------|
| D... | Determiner Phrase         |
| Fa   | Adverbial Clause          |
| Fn   | Complement Clause         |
| Fr   | Relative Clause           |
| G... | Genitive Phrase           |
| J... | Adjectival Phrase         |
| M... | Numeral Phrase            |
| N... | Noun Phrase               |
| P... | Prepositional Phrase      |
| R... | Adverbial Phrase          |
| S... | Sentence                  |
| Tg   | Present Participle Clause |
| Tn   | Past Participle Clause    |
| Ti   | Infinitival Clause        |
| Tf   | <i>For-To</i> Clause      |
| V... | Verb Phrase               |
| Z... | Reduced Relative Clause   |

## Functiontags (Roles)

|    |                                 |
|----|---------------------------------|
| :s | Logical Subject                 |
| :o | Logical Direct Object           |
| :i | Indirect Object                 |
| :u | Prepositional Object            |
| :S | Surface but not Logical Subject |
| :O | Surface but not Logical Object  |
|    |                                 |
| :p | Place                           |
| :t | Time                            |
| :q | Direction                       |
| :h | Manner or Degree                |
| :m | Modality                        |
| :c | Contingency                     |
|    |                                 |
| :n | Phrasal Verb Particle           |

## Suffixes

|   |                                        |
|---|----------------------------------------|
| ? | Interrogative                          |
| * | Imperative                             |
| + | Conjunct Introduced by Conjunction     |
| - | Conjunct Not Introduced by Conjunction |

## **Appendix C: Sources of the SUSANNE Corpus**

The *SUSANNE* corpus is made up of a cross-genre sample of 64 texts drawn from the Brown corpus of written American English collected in the 1960's. Below the filename and source of each text is listed. As can be seen, some files contain text from several sources. The first letter of the filename denotes the genre:

- A      press reportage
- G      belles lettres, biography, memoirs
- J      learned (mainly scientific and technical) writing
- N      adventure and Western fiction

- A01    The Atlanta Constitution  
1961.11.04 issue, p. 1, 2 items  
1961.08.17 issue, p. 6, 2 items  
1961.03.06 issue, p. 1, 3 items; p. 18, 1 item
- A02    The Dallas Morning News  
1961.02.17 issue, section 1, p. 5, 5 items; p. 12, 2 items  
Chicago Daily Tribune  
1961.02.10 issue, part 1, p. 4, 1 item
- A03    Chicago Daily Tribune  
1961.07.25 issue, p. 1, 3 items  
1961.02.10 issue, p. 1, 2 items
- A04    The Christian Science Monitor  
1961.05.11 issue, p. 1, 3 items
- A05    The Providence Journal  
1961.07.23 issue, p. 19, 1 item  
1961.07.16 issue, sec. 1, p. 9, 1 item  
1961.07.19 issue, p. 5, 1 item  
1961.07.20 issue, p. 5, 1 item  
1961.07.22 issue, p. 17, 1 item
- A06    Newark Evening News  
1961.03.22 issue, p. 25, 6 items
- A07    The New York Times  
1961.06.19 issue, p. 1, 7 items
- A08    The Times-Picayune [New Orleans]  
1961.01.01 issue, sec. 2, p. 3, 4 items
- A09    The Philadelphia Inquirer  
1961.05.10 issue, p. 49, 4 items  
Chicago Daily Tribune  
1961.02.10 issue, sec. F, p. 9, 1 item  
1961.10.25 issue, sec. I, p. 16, 1 item
- A10    The Oregonian [Portland]  
1961.10.24 issue, p. 8, 5 items  
1961.11.29 issue, p. 12, 3 items  
1961.10.24 issue, p. 8, 1 item
- A11    The Sun [Baltimore]  
1961.03.18 issue, pp. 15 and 18, 6 items

- A12 The Dallas Morning News  
1961.10.10 issue, sec. 2  
p. 1, 1 item  
p. 2, 1 item  
p. 3, 3 items
- A13 Rocky Mountain News [Denver, Colorado]  
1961.05.02 issue  
p. 50, 3 items  
p. 51, 2 items  
The Dallas Morning News  
1961.10.10 issue, sec. 2, p. 1, 1 item
- A14 The New York Times  
1961.01.24 issue, p. 23, 5 items
- A19 The Sun [Baltimore]  
1961.01.08 issue, p. 36, 8 items  
1961.12.10 issue, sec. C, p. 1, 4 items
- A20 Chicago Daily Tribune  
1961.02.10 issue  
pp. 1 and 2, 1 item  
p. 9, 1 item  
p. 2, 2 items  
p. 9, 1 item
- G01 Edward P. Lawton, "Northern Liberals and Southern Bourbons", The Georgia Review, 15 (1961), 254-259
- G02 Arthur S. Miller, "Toward a Concept of National Responsibility", The Yale Review, LI:2 (December 1961), 186-191
- G03 Peter Wyden, "The Chances of Accidental War", The Saturday Evening Post, 1961.06.03, 18-19 and 60-61
- G04 Eugene Burdick, "The Invisible Aborigine", Harper's Magazine, 223:1336 (September 1961), 70-72
- G05 Terence O'Donnell, "Evenings at the Bridge", Horizon, III:5 (May 1961), 26-30
- G06 The American-German Review, October-November 1961  
pp. 26-28: Ruth Berges, "William Steinberg, Pittsburgh's Dynamic Conductor"  
pp. 28-29: Henry W. Koller, "German Youth Looks at the Future"
- G07 Richard B. Morris, "Seven Who Set Our Destiny", The New York Times Magazine, 1961.02.19, 9 and 69-70
- G08 Frank Murphy, "New Southern Fiction: Urban or Agrarian?", The Carolina Quarterly, 13:2 (Spring 1961), 18-25
- G09 Selma Jeanne Cohen, "Avant-Garde Choreography", Criticism A Quarterly for Literature and the Arts, vol. III, no. 1 (Winter 1961), 24-28
- G10 Clarence Streit, "How the Civil War Kept You Sovereign" [chapter 8 of Freedom's Frontier - Atlantic Union Now], Freedom and Union, 16:2 (February 1961), 16-18
- G11 Frank Oppenheimer, "Science and Fear - A Discussion of Some Fruits of Scientific Understanding", The Centennial Review, 5:4 (Fall 1961), 404-409



- G12 Tom F. Driver, "Beckett by the Madeleine", *Columbia University Forum*, 4:3 (Summer 1961), 21-24
- G13 Charles Glicksberg, "Sex in Contemporary Literature", *The Colorado Quarterly*, 9:3 (Winter 1961), 278-82
- G17 Randall Stewart, "A Little History, a Little Honesty: A Southern Viewpoint", *The Georgia Review*, 15:1 (Spring 1961), 10-15
- G18 Charles Wharton Stork, "Verner von Heidenstam", *American-Scandinavian Review*, 49:1 (March 1961), 39-43
- G22 Kenneth Reiner, "Coping with Runaway Technology", *The Ethical Outlook*, XLVII:3 (May-June 1961), 91-95
- J01 Cornell H. Mayer, "Radio Emission of the Moon and Planets", in Gerard P. Kuiper & Barbara M. Middlehurst (eds.), *Planets and Satellites. Vol. 3 of The Solar System*. University of Chicago Press, 1961, pp. 442-446
- J02 Raymond C. Binder et al. (eds.), *Proceedings of the 1961 Heat Transfer and Fluid Mechanics Institute*, Stanford University Press, 1961, pp. 193-196
- J03 Harry H. Hull, "The Normal Forces and Their Thermodynamic Significance", *Transactions of the Society of Rheology*, V (1961), 120-125
- J04 James A. Ibers et al., "Proton magnetic resonance study of polycrystalline  $\text{HCrO}_2$ ", *The Physical Review*, 121:6 (1961.03.15), 1620-1622
- J05 Jay C. Harris & John R. Van Wazer, "Detergent building", in J.R. Van Wazer (ed.), *Phosphorus and its Compounds*, Interscience Publishers, Inc., 1961, pp. 1732-1737
- J06 Francis J. Johnston & John E. Willard, "The exchange reaction between Chlorine and Carbon Tetrachloride", *Journal of Physical Chemistry*, 65 (February 1961), 317-318
- J07 J.F. Vedder, "Micrometeorites", in Francis S. Johnson (ed.), *Satellite Environment Handbook*, Stanford University Press, 1961, pp. 92-97
- J08 LeRoy Fothergill, "Biological Warfare", in Peter Gray (ed.), *The Encyclopedia of the Biological Sciences*, Reinhold Publishing Corporation, 1961, pp. 145-149
- J09 M. Yokoyama et al., "Chemical and serological characteristics of blood group antibodies in the ABO and Rh systems", *The Journal of Immunology*, 87 (1961), 56-60
- J10 B.J.D. Meeuse, *The Story of Pollination*, The Ronald Press Company, 1961, pp. 104-108
- J12 Richard F. McLaughlin et al., "A study of the subgross pulmonary anatomy in various mammals", *The American Journal of Anatomy*, 108 (1961), 154-157
- J17 E. Gellhorn, "Prolegomena to a theory of the emotions", *Perspectives in Biology and Medicine*, 4 (1961), 426-431
- J21 C.R. Wylie, Jr., "Line involutions in  $S^3$  whose singular lines all meet in a twisted curve", *Proceedings of the American Mathematical Society*, 12 (1961), 335-339.
- J22 Max F. Millikan & Donald L.M. Blackmer (eds.), *The Emerging Nations: Their Growth and United States Policy*, Little, Brown and Company, 1961, pp. 136-142

- J23 Joyce O. Hertzler, *American Social Institutions; A Sociological Analysis*, Allyn and Bacon, Inc., 1961, pp. 478-482
- J24 Howard J. Parad, "Preventive casework: problems and implications", *The Social Welfare Forum*, 1961, Columbia University Press for the National Conference on Social Welfare, 1961, pp. 186-191
- N01 Wayne D. Overholser, *The Killer Marshal*, Dell Publishing Co., 1963 [copyright 1961], pp. 53-58
- N02 Clifford Irving, *The Valley*, McGraw-Hill Book Company, Inc., 1961, pp. 262-267
- N03 Cliff Farrell, *Trail of the Tattered Star*, Doubleday & Company, 1961, 168-173
- N04 James D. Horan, *The Shadow Catcher*, Crown Publishers, Inc., 1961, pp. 248-253
- N05 Richard Ferber, *Bitter Valley*, Dell Publishing Company, 1961, pp. 9-17
- N06 Thomas Anderson, *Here Comes Pete Now*, Random House, 1961, pp. 4-12
- N07 Todhunter Ballard, *The Night Riders*, Pocket Books, Inc., 1961, pp. 5-11
- N08 Mary Savage, *Just For Tonight*, Dodd, Mead & Company, 1961, pp. 114-120
- N09 Jim Thompson, *The Transgressors*, The New American Library of World Literature, Inc., 1961, pp. 9-13
- N10 Joseph Chadwick, *No Land Is Free*, Avon Book Division, Hearst Corporation, 1961, pp. 21-26
- N11 Gene Caesar, *Rifle For Rent*, Monarch Books, Inc., 1963 [copyright 1961], pp. 46-51
- N12 Edwin Booth, *Outlaw Town*, Ballantine Books, Inc., 1961, pp. 103-108
- N13 Martha Ferguson McKeown, *Mountains Ahead*, G.P. Putnam's Sons, 1961, pp. 390-395
- N14 Peter Field, *Rattlesnake Ridge*, Jefferson House, Inc., 1961, pp. 164-172
- N15 Donald J. Plantz, *Sweeney Squadron*, Dell Publishing Co., Inc., 1961, pp. 133-138
- N18 Peter Bains, "With Women...Education Pays Off", *Monsieur*, 4:2 (February 1961), 17 and 77-78

## Appendix D: Verification Examples from the SUSANNE Corpus

### Appendix D1: Relative Clause That/Wh... Switch Rule

**First run: searching for *who, whom, whose, which, that* and excluding non-restrictive relatives based on comma punctuation**

Incidence of candidate constructions: 586/6248 (9.4%)

Correctly identified sentences: 55/64 (86%)

Correctly transformed sentences: 54/64 (84%)

#### *Successful*

- (A01:0070.06) The September-October term jury had been charged by Fulton Superior Court Judge Durwood Pye to investigate reports of possible “irregularities” in the hard-fought primary that ~~which~~ was won by Mayor-nominate Ivan Allen Jr.
- (A02:0130.12) He told the committee the measure would merely provide means of enforcing the escheat law that ~~which~~ has been on the books “since Texas was a republic”.
- (A03:0710.18) After reading his statement discharging the 23d ward case, Karns told Wexler that if the seven cases scheduled for trial also involved persons that ~~who~~ had been subpoenaed, he would dismiss them.
- (A04:0040.24) This has been a working session of an organization which ~~that~~, by its very nature, can only proceed along its route step by step and without dramatic changes.
- (A05:0150.39) Mr. Hawksley said he believed there are a number of qualified city residents that ~~who~~ would be willing to take the full-time CD job.
- (A06:0570.21) His reference to ‘discredited carcass’ or ‘tattered remains’ of the president’s leadership is an insult to the man that ~~who~~ led our forces to victory in the greatest war in all history, to the man that ~~who~~ was twice elected overwhelmingly by the American people as president of the United States, and who has been the symbol to the world of the peace-loving intentions of the free nations.
- (A07:0060.24) Some anti-organization Democrats saw in the program an opportunity to end the bitter internal fight within the Democratic party which ~~that~~ has been going on for the last three years.
- (A08:0380.12) Along this avenue that ~~which~~ saw marching soldiers from the War Between the States returning in 1865 is the National Archives building where hundreds of thousands of this country’s most valuable records are kept.
- (A09:0010.06) City Controller Alexander Hemphill charged Tuesday that the bids on the Frankford Elevated repair project were rigged to the advantage of a private contracting company that ~~which~~ had “an inside track” with the city.
- (A19:0340.06) A baby was burned to death and two other children were seriously injured last night in a fire which damaged their one-room Anne Arundel county home.
- (A20:0200.30) A shadowy couple who call themselves Peter Kroger, bookseller, and wife, Joyce.
- (G01:0050.18) The name presumably derives from the French royal house which never learned and never forgot; since Bourbon whiskey, though of Kentucky origin, is at least as much favored by liberals in the North as by conservatives in the South.
- (G02:0080.09) It became the sole “subject” of “international law” (a term that ~~which~~, it is pertinent to remember, was coined by Bentham), a body of legal principle that ~~which~~ by and large was made up of what Western nations could do in the world arena.
- (G03:0050.27) I have just asked these questions in the Pentagon, in the White House, in offices of key scientists across the country and aboard the submarines that prowl for months underwater, with neat rows of green launch tubes that ~~which~~ contain Polaris missiles and which are affectionately known as “Sherwood Forest”.
- (G04:0050.27) It was nothing more than a tiny distant rain squall, a dull gray sheet that ~~which~~ reached from a layer of clouds to the earth.

- (G05:0270.30) The one apparent connection between the two is a score of buildings that ~~which~~ somehow or other have survived and which naturally enough are called “historical monuments”.
- (G06:0670.12) In Frankfurt, too, he directed the Museum and Opera House concerts that ~~which~~, in addition to the standard repertoire, featured novelties like Erdmann's Piano Concerto and Mahler's Sixth Symphony.
- (G07:0050.06) It is well then that in this hour both of “national peril” and of “national opportunity” we can take counsel with the men that ~~who~~ made the nation.
- (G08:0050.33) Thus we are compelled to face the urbanization of the South — an urbanization that ~~which~~, despite its dramatic and overwhelming effects upon the Southern culture, has been utterly ignored by the bulk of Southern writers.
- (G09:0090.21) For Mr. Taylor's *Images and Reflections* he made some diaphanous tents which ~~that~~ alternately hide and reveal the performer, and a girl's cape lined with grass.
- (G10:0130.24) Nothing can show more than this the immensity of the danger to democratic peoples which ~~that~~ lies in even relatively slight deviation from their true concept of sovereignty.
- (G11:0030.33) The useful suggestion of Professor David Hawkins that ~~which~~ considers culture as a third stage in biological evolution fits quite beautifully then with our suggestion that science has provided us with a rather successful technique for building protective artificial environments.
- (G12:0360.21) The voice is light in timbre, with a rough edge which ~~that~~ corresponds to his visage.
- (G13:0440.12) If Wilhelm Reich is the Moses that ~~who~~ has led them out of the Egypt of sexual slavery, Dylan Thomas is the poet who offers them the Dionysian dialectic of justification for their indulgence in liquor, marijuana, sex, and jazz.
- (G17:0520.39) To guard against the tyranny of a numerical majority, Calhoun developed his theory of “concurrent majority”, which, he said, “by giving to each portion of the community that ~~which~~ may be unequally affected by the action of government, a negative on the others, prevents all partial or local legislation”.
- (G18:0020.39) Of these there are surely few that would be more rewarding discoveries than Verner von Heidenstam, the Swedish poet and novelist that ~~who~~ received the award in 1916 and whose centennial was celebrated two years ago.
- (G22:0280.03) So in these pages the term “technology” is used to include any and all means that ~~which~~ could amplify, project, or augment man's control over himself and over other men.
- (J02:0680.27) Thus, direct comparisons can be drawn with free burning arcs that ~~which~~ have been studied in detail during the past years and decades by numerous investigators (Ref. 3).
- (J03:0180.15) The face of one block contained a hole 1/16 in diameter that ~~which~~ led to a manometer for the measurement of the normal pressure.
- (J04:0780.54) However, any oxygen nonequivalence would shorten either the already extremely short [FORMULA] interlayer distance of 2.55 Å or the non-hydrogen-bonded [FORMULA] interlayer interactions that ~~which~~ are already quite short at 2.58 Å.
- (J05:0160.30) A well-publicized entrant that ~~which~~ has achieved success only recently is the built liquid detergent, with which the major problem today is incorporation of builder and active into a small volume using a sufficiently high builder/active ratio.
- (J07:0750.30) One may conclude that most of the detected micrometeoritic material is concentrated in orbital streams that ~~which~~ intersect the Earth's orbit.
- (J09:0260.42) In addition one serum was obtained from a donor (R. E.) that ~~who~~ had been sensitized 6 years previously.
- (J10:0070.24) I must add at once that these animals are what we call “queens”, young females which ~~that~~ have mated in the previous summer or autumn.
- (J12:1331.27) This viewpoint is further supported by Verloop's ('48) demonstration of thickened bronchial artery and arteriolar muscular coats that ~~which~~ are capable of acting as valves.
- (J17:0710.27) Although no drugs act exclusively on the hypothalamus or a part of it, there is sufficient specificity to distinguish drugs that ~~which~~ shift the hypothalamic balance to the sympathetic side from those that ~~which~~ produce a parasympathetic dominance.
- (J21:0050.24) Then in 2 we show that any line involution with the properties that (a) It has no complex of invariant lines, and (b) Its singular lines form a complex consisting exclusively of the lines that ~~which~~ meet a twisted curve, is necessarily of the type discussed in 1.

- (J23:0220.33) The significant things in it, at the higher religious levels, are the inner emotional, mental, and spiritual occurrences which ~~that~~ fill the pressing human needs of self-preservation, self-pacification, and self-completion.
- (J24:0700.21) Here there is a specific preventive component that ~~which~~ applies in a more generalized sense to any casework situation.
- (N01:0020.21) He certainly didn't want a wife that ~~who~~ was fickle as Ann.
- (N02:0750.27) The moon had sunk below the black crest of the mountains and the land, seen through eyes which ~~that~~ had grown accustomed to the absence of light, looked primeval, as if no man had ever trespassed before.
- (N03:1040.39) They bawled questions which ~~that~~ were not answered in the uproar.
- (N04:0010.09) "So it wasn't the earthquake which ~~that~~ made him return to his village"!
- (N05:0200.12) She glanced around the clearing, taking in the wagon and the load of supplies and trappings scattered over the ground, the two kids, the whiteface bull which ~~that~~ was chewing its cud just within the far reaches of the firelight.
- (N06:0180.30) He spoke to me in a gruff voice, an affectation that ~~which~~ quite belied his personality.
- (N07:0150.33) There were three other men within this prison that ~~whom~~ Barton would have liked to liberate, but they were in other cell blocks.
- (N08:0050.15) She had offered to walk, but Pamela knew she would not feel comfortable about her child until she had personally confided her to the care of the little pink woman that ~~who~~ chose to be called "Auntie".
- (N10:0020.21) The herd was watered and then thrown onto a broad grass flat that ~~which~~ was to be the first night's bedground.
- (N11:0130.33) Lewis was a man that ~~who~~ had made a full-time job of cow stealing.
- (N12:0390.33) There was no lock on the door, only an iron hook that ~~which~~ he unfastened.
- (N13:0470.30) Then, with a glory which ~~that~~ almost wiped out the deep, downward sags in her careworn face, Matilda leaned over the wheel and shouted to Hez, who was stumbling along in the heat and the dust on the opposite side of the wagon "Pa! Pa! I can see Dan. And he's with the Jacksons".
- (N14:0670.21) Cobb unleashed a single powerful jab which ~~that~~ sent Gyp reeling wildly and crashing down with a whining groan.
- (N15:0030.09) Every plane which ~~that~~ could fly was sent into the air.
- (N18:0040.39) If it were not for an old professor that ~~who~~ made me read the classics I would have been stymied on what to do, and now I understand why they are classics; those who wrote them knew people and what made people tick.

#### *Doubtful*

- (A14:0120.27) Van Brocklin, the quarterback that ~~who~~ led the Eagles to the title, was signed by the Vikings last Wednesday.
- (J01:0030.24) The results of present observations of the thermal radio emission of the moon are consistent with the very low thermal conductivity of the surface layer that ~~which~~ was derived from the variation in the infrared emission during eclipses (e.g., Garstung, 1958).
- (N09:0060.27) To me you'll always be the girl o' my dreams, an' the sweetest flower which ~~that~~ grows".

#### *Failed*

- (A10:0030.15) Ierulli will replace Desmond D. Connall that ~~who~~ has been called to active military service but is expected back on the job by March 31.
- (A11:0080.18) Over the first five frames, Jack Fisher, the big righthandler that ~~who~~ figures to be in the middle of Oriole plans for a drive on the 1961 American League pennant, held the A's scoreless while yielding three scattered hits.
- (A12:0010.21) A Texas halfback that ~~who~~ doesn't even know the team's plays, Eldon Moritz, ranks fourth in Southwest Conference scoring after three games.
- (A13:0370.21) Among the spectators was the noted exotic dancer, Patti Waggin that ~~who~~ is Mrs. Don Rudolph when off the stage.

- (J06:0530.24) The silver chloride was fused under vacuum in the presence of aluminum chips with the resultant product of [FORMULA] that ~~which~~ was sublimed into a flask on the vacuum line.
- (J08:0680.09) This material fluoresces under ultraviolet light which facilitates its sampling and assessment.
- (J22:0010.39) As the historic processes of modernization gradually gain momentum, their cohesion will be threatened by divisive forces, the gaps between rulers and subjects, town and country, will widen; new aspirants for power will emerge that ~~whose~~ ambitions far exceed their competence; old rulers may lose their nerve and their sense of direction.

## Second run: searching for *that*

Incidence of candidate constructions: 231/6248 (3.7%)

Correctly identified sentences: 50/50 (100%)

Correctly transformed sentences: 47/50 (94%)

### *Successful*

- (A01:0510.39) Regarding Atlanta's new multi-million-dollar airport, the jury recommended "that when the new management takes charge Jan. 1 the airport be operated in a manner which ~~that~~ will eliminate political influences".
- (A04:0040.24) This has been a working session of an organization which ~~that~~, by its very nature, can only proceed along its route step by step and without dramatic changes.
- (A05:0210.18) One advantage which ~~that~~ would come to the city in having a full-time director, he said, is that East Providence would become eligible to apply to the federal government for financial aid in purchasing equipment needed for a sound civil defense program.
- (A06:1590.24) Calling the Democrats the "party which ~~that~~ lives, breathes and thinks for the good of the people", Hughes asked, "a representative Democratic vote in the primary for a springboard toward victory in November".
- (A07:0060.24) Some anti-organization Democrats saw in the program an opportunity to end the bitter internal fight within the Democratic party which ~~that~~ has been going on for the last three years.
- (A08:0430.12) Street car tracks run down the center of Pennsylvania, powered with lines which ~~that~~ are underground.
- (A09:0060.03) Hemphill said that the Hughes Steel Erection Co. contracted to do the work at an impossibly low cost with a bid which ~~that~~ was far less than the "legitimate" bids of competing contractors.
- (A10:0560.51) I think everybody is agreed that we need to hear some voice on the national level which ~~that~~ would make some sense and in which we would have some confidence in following.
- (A12:0070.03) Moritz doesn't even have a nose guard or hip pads but he's one of the most valuable members of the Longhorn team which ~~that~~ will be heavily favored Saturday over Oklahoma in the Cotton Bowl.
- (A13:1590.15) It is an irritable rule which ~~that~~ does baseball more harm than good, especially at the minor league level.
- (A14:0570.30) Named by Mayor Wagner three years ago to head a committee which ~~that~~ included James A. Farley, Bernard Gimbel and Clint Blume, Shea worked relentlessly.
- (A19:1730.06) While availability of mortgage money has been a factor in encouraging apartment construction, the generally high level of prosperity in the past few years plus rising consumer income are among the factors which ~~that~~ have encouraged builders to concentrate in the apartment-building field.
- (A20:0710.03) Britain began designing the ship in 1956 but got nowhere until the American government decided to end a ban on sharing military secrets with Britain which ~~that~~ had been imposed after Fuchs blabbed.
- (G01:0310.39) I take this to mean that the intelligent — and therefore necessarily cynical— liberal considers that the need for a national economy with controls which ~~that~~ will assure his conception of social justice is so great that individual and local liberties as well as democratic processes may have to yield before it.

- (G03:0050.27) I have just asked these questions in the Pentagon, in the White House, in offices of key scientists across the country and aboard the submarines which ~~that~~ prowl for months underwater, with neat rows of green launch tubes which contain Polaris missiles and which are affectionately known as “Sherwood Forest”.
- (G04:0150.12) Occasionally, for no reason which ~~that~~ I could see, they would suddenly alter the angle of their trot.
- (G05:0390.15) The promenade, for example, continues to take place on the Chahar Bagh, a mile-long garden of plane and poplar trees which ~~that~~ now serves as the city's principal street.
- (G06:1620.21) I was far from convinced of the truth of my statement, but could not think of anything which ~~that~~ might evoke responses more quickly.
- (G07:0420.03) Author of the Albany Plan of Union, which, had it been adopted, might have avoided the Revolution, he fought the colonists' front-line battles in London, negotiated the treaty of alliance with France and the peace which ~~that~~ ended the war, headed the state government of Pennsylvania, and exercised an important moderating influence at the Federal Convention.
- (G08:0180.09) It is interesting, however, that despite this strong upsurge in Southern writing, almost none of the writers has forsaken the firmly entrenched concept of the white-suited big-daddy colonel sipping a mint julep as he silently recounts the revenue from the season's cotton and tobacco crops; of the stereotyped Negro servants chanting hymns as they plow the fields; of these and a host of other antiquated legends which ~~that~~ deny the South its progressive leaps of the past century.
- (G09:0090.21) For Mr. Taylor's *Images and Reflections* he made some diaphanous tents which ~~that~~ alternately hide and reveal the performer, and a girl's cape lined with grass.
- (G10:0130.24) Nothing can show more than this the immensity of the danger to democratic peoples which ~~that~~ lies in even relatively slight deviation from their true concept of sovereignty.
- (G11:0560.21) We are not now afraid of atomic bombs in the same way which ~~that~~ people once feared comets.
- (G12:0360.21) The voice is light in timbre, with a rough edge which ~~that~~ corresponds to his visage.
- (G13:0470.33) In addition, they have been converted to Zen Buddhism, with its glorification of all which ~~that~~ is “natural” and mysteriously alive, the sense that everything in the world is flowing.
- (G18:0020.39) Of these there are surely few which ~~that~~ would be more rewarding discoveries than Verner von Heidenstam, the Swedish poet and novelist who received the award in 1916 and whose centennial was celebrated two years ago.
- (G22:0940.06) The corporation in America is in reality our form of socialism, vying in a sense with the other socialistic form which ~~that~~ has emerged within governmental bureaucracy.
- (J03:1140.30) There is another means which should show the direction and relative value of the stresses in viscoelastic fluids which ~~that~~ is not mentioned as such in the literature, and that is the shape of the suspended drops of low viscosity fluids in shear fields.
- (J05:0970.27) However, a major factor in binding dirt is the attraction between surfaces which ~~that~~ goes under the name of van der Waal's forces.
- (J06:1750.39) The last column shows the rate of exchange which ~~that~~ would have been observed at a relative intensity of 4 (14.7 cm. distance) calculated on the assumptions that the incident light intensity is inversely proportional to the square of the distance of the lamp from the cell and that the rate is directly proportional to the incident light intensity.
- (J08:1180.39) An enemy would obviously choose an agent which ~~that~~ is believed to be highly infectious.
- (J10:0070.24) I must add at once that these animals are what we call “queens”, young females which ~~that~~ have mated in the previous summer or autumn.
- (J23:0220.33) The significant things in it, at the higher religious levels, are the inner emotional, mental, and spiritual occurrences which ~~that~~ fill the pressing human needs of self-preservation, self-pacification, and self-completion.
- (J24:1930.12) Probably, in the immediate future, we will have to settle for middle-range efforts which ~~that~~ fall short of utopian models.
- (N01:0070.42) The easiest thing would be to sell out to Al Budd and leave the country, but there was a stubborn streak in him which ~~that~~ wouldn't allow it.

- (N02:0750.27) The moon had sunk below the black crest of the mountains and the land, seen through eyes which ~~that~~ had grown accustomed to the absence of light, looked primeval, as if no man had ever trespassed before.
- (N04:0010.09) “So it wasn't the earthquake which ~~that~~ made him return to his village”!
- (N05:0200.12) She glanced around the clearing, taking in the wagon and the load of supplies and trappings scattered over the ground, the two kids, the whiteface bull which ~~that~~ was chewing its cud just within the far reaches of the firelight.
- (N06:0260.03) His hat (the cause of his baldness?) hung on a hook on the wall, and underneath it I could see his tie, knotted, ready to be slipped over his head, a black badge of frayed respectability which ~~that~~ ought never to have left his neck.
- (N07:1600.48) Even Hague was repelled by the machinelike deadliness which ~~that~~ was Kodyke.
- (N08:0170.24) That mistake, she thought, had cost her dearly these past few days, and she wanted to avoid falling into any more of the traps which ~~that~~ the mountain might set for her.
- (N10:1710.48) They brought to it all the odors which ~~that~~ clung to men like themselves, that of their own sweat, of campfire smoke, of horses and cattle.
- (N11:0890.06) A detailed scouring of the entire area revealed nothing beyond a ledge of rocks which ~~that~~ might have been the rifleman's hiding place.
- (N13:0470.30) Then, with a glory which ~~that~~ almost wiped out the deep, downward sags in her careworn face, Matilda leaned over the wheel and shouted to Hez, who was stumbling along in the heat and the dust on the opposite side of the wagon “Pa! Pa! I can see Dan. And he's with the Jacksons”.
- (N14:0670.21) Cobb unleashed a single powerful jab which ~~that~~ sent Gyp reeling wildly and crashing down with a whining groan.
- (N15:0030.09) Every plane which ~~that~~ could fly was sent into the air.
- (N18:0370.48) There was something about the contour of her face, her smile which ~~that~~ was like New Orleans sunshine, the way she held her head, the way she walked — there was scarcely anything she did which did not fascinate me.

#### *Doubtful*

- (N09:0060.27) “To me you'll always be the girl o' my dreams, an' the sweetest flower which ~~that~~ grows”.
- (G02:0450.12) For lawyers, reflecting perhaps their parochial preferences, there has been a special fascination since then in the role played by the Supreme Court in that transformation — the manner in which its decisions altered in “the switch in time which ~~that~~ saved nine”, President Roosevelt's ill-starred but in effect victorious “Court-packing plan”, the imprimatur of judicial approval time which ~~that~~ was finally placed upon social legislation.

#### *Failed*

- (N03:0790.27) You're the one which ~~that~~'s taking the big chance.

## Appendix D2: Relative Clause *That/Wh...* Drop and Insertion Rule

### First run: searching for relative clauses starting with a subject

Incidence of candidate constructions: 212/6248 (3.4%)

Correctly identified sentences: 49/52 (94%)

Correctly transformed sentences: 49/52 (94%)

#### *Successful*

- (A01:1000.06) The largest hurdle ~~that~~ the Republicans would have to face is a state law which says that before making a first race, one of two alternative courses must be taken:
- (A03:0200.24) President Kennedy today pushed aside other White House business to devote all his time and attention to working on the Berlin crisis address ~~that~~ he will deliver tomorrow night to the American people over nationwide television and radio.
- (A04:0880.21) This is another of the modifications of policy on Laos ~~that~~ the Kennedy administration has felt compelled to make.



- (A05:0440.30) One of the first things that he would do, he said, would be to organize classes in first aid.
- (A06:1360.06) He promised nearly 200 Democratic county committee members at the meeting in the Puddingstone Inn: "When I come back here after the November election you'll think, 'You're my man — you're the kind of governor that we're glad we elected'".
- (A07:0190.15) Their view is that last-minute changes that the Mayor is proposing to make in the Democratic ticket only emphasize the weakness of his performance as Mayor.
- (A08:0820.27) Certainly nobody will predict that the next time that the lawmakers come back together Barnett will be able to enjoy a re-enactment of the strange but successful "honeymoon" that he had in the 1960 legislative session.
- (A09:0130.42) The Philadelphia Transportation Co. is investigating the part that its organization played in reviewing the project.
- (A10:0170.06) "We must persuade them to enjoy a way of life which, if not identical, is congenial with ours", he said but adding that if they do not develop the kind of society that they themselves want it will lack rituality and loyalty.
- (A12:0710.39) It was the first time that we've been ahead this season (when John Richey kicked what proved to be the winning field goal)".
- (A13:0930.48) I told him what Liston had said and he said Liston was a double-crosser and said that anything he (Liston) got was through a keyhole.
- (A14:1280.42) From the standpoint of the army of duffers, however, this was easily the most heartening exhibition that they had had since Ben Hogan fell upon evil ways during his heyday and scored an 11 in the Texas open.
- (A19:0890.09) Conceding that several cities to the north were in worse shape than Baltimore after the last storm, Mr. Schaefer listed several improvements that he said should be made in the snow plan here.
- (A20:0820.03) It was there ~~that~~ the two accused civil servants were at work.
- (G01:0580.12) This, for the liberals that I know, would be an understatement.
- (G03:0210.48) "E" stands for "execution" — the moment that a "go order" would unleash an American nuclear strike.
- (G04:0150.12) Occasionally, for no reason ~~that~~ I could see, they would suddenly alter the angle of their trot.
- (G06:1390.03) He was not enthusiastic over the newly acquired Claude Lorrain, but reminisced with pleasure over a Poussin exhibit that he had been able to see in Paris a year ago.
- (G07:0600.27) HIS political opponent and lifetime friend, Thomas Jefferson, achieved immortality through his authorship of the Declaration of Independence, but equally notable were the legal and constitutional reforms that he instituted in his native Virginia, his role as father of our territorial system, and his acquisition of the Louisiana Territory during his first term as President.
- (G08:0550.51) William Styron, while facing the changing economy with a certain uneasy reluctance, insists he is not to be classified as a Southern writer and yet includes traditional Southern concepts in everything that he publishes.
- (G09:0850.18) The process stipulates that the choreographer sense the quality of the initial movement that he has discovered and that he feel the rightness of the quality that is to follow it.
- (G10:0740.24) They recognized that slavery was a moral issue and not merely an economic interest, and that to recognize it explicitly in their Constitution would be in explosive contradiction to the concept of sovereignty that they had set forth in the Declaration of 1776 that "all men are created equal, that they are endowed by their Creator with certain unalienable rights, that among them are life, liberty and the pursuit of happiness ...".
- (G11:0560.21) We are not now afraid of atomic bombs in the same way ~~that~~ people once feared comets.
- (G12:0180.18) My curiosity was sharpened a day or two before the interview by a conversation that I had with a well-informed teacher of literature, a Jesuit father, at a conference on religious drama near Paris.
- (G13:0420.18) This is the Holy Grail that these knights of the orgasm pursue, this is the irresistible cosmic urge to which they respond.
- (G18:0670.09) It consists of fragmentary personal revelations, such as "The Spark": "There is a spark dwells deep within my soul. To get it out into the daylight's glow Is my life's aim both

- first and last, the whole. It slips away, it burns and tortures me. That little spark is all the wealth that I know, That little spark is my life's misery”.
- (G22:0190.48) We are already committed to establishing man's supremacy over nature and everywhere on earth, not merely in the limited social-political-economical context that we are fond of today.
- (J06:0350.33) Contrary to these expectations we have found it impossible to obtain the degree of reproducibility that one would wish, even with extensive efforts to prepare especially pure reagents.
- (J07:1870.03) Most meteoritic material, by the time that it reaches the Earth's surface, has been reduced to dust or to spherules of ablated material in its passage through the atmosphere.
- (J10:0880.18) Not always, though, does the development of a bumblebee colony take place in the smooth fashion that we have just described.
- (J21:0930.33) We now shall show that any involution with these characteristics is necessarily of the type that we have just described.
- (J22:0150.27) The United States can help by communicating a genuine concern with the problems that these countries face and a readiness to provide technical and other appropriate forms of assistance where possible.
- (J23:0450.21) The kind of religious experience ~~that~~ most moderns seek not only provides, clarifies, and relates human yearnings, values, ideals, and purposes; it also provides facilities and incitements for the development of personality, sociality, and creativeness.
- (J24:1660.18) My aim in mentioning this factor obviously is not to give license to “wild therapy” but rather to encourage us to use the time-honored clinical casework skills that we already possess, and to use them with greater confidence, precision, and professional pride.
- (N01:0280.24) It could be some kind of trick that Budd had thought up.
- (N02:0080.51) Clayton tried to call back the face of the man that he had known.
- (N03:0310.51) He was thinking that the way that she had responded to his own kiss hadn't meant what he had believed it had.
- (N04:0300.51) In this country there's a war on every time that the grass turns green.
- (N05:0340.03) It's not the kind of thing ~~that~~ a man would be proud of.
- (N06:0210.09) It was obvious that he wished himself different from the sort of person that he thought he was.
- (N07:0090.60) The only reason that we brought you was to get Miller out.
- (N08:0170.24) That mistake, she thought, had cost her dearly these past few days, and she wanted to avoid falling into any more of the traps ~~that~~ the mountain might set for her.
- (N09:0130.03) He knew that anything that a brainy little lady like her had to say would be plumb important, as well as pleasin' to the ear, and he didn't want to miss a word of it.
- (N11:0300.15) That's all that I ask.
- (N12:0170.21) He turned and looked around at the lobby as though seeing things that he hadn't before noticed.
- (N13:0220.42) From the way that the wound in his head was itching, Dan knew that it would heal.
- (N14:0130.33) It's all that I ask, Stevens”.
- (N15:0050.27) Anything that the enemy flew or floated was his target.
- (N18:0100.48) The arrangement that I had with him was to work four hours a day.

#### *Failed*

- (A02:0830.12) Instead, retailers would sign a certificate of correctness, that violation of which would carry a penalty of one to five years in prison, plus a \$1,000 fine.
- (G02:1100.39) For better or for worse, we all now live in welfare states, that the organizing principle of which is collective responsibility for individual well-being.
- (G05:0910.12) It was to provide a safe and spacious crossing for these caravans, and also to make a pleasance for the city, ~~that~~ Shah Abbas II in about 1657 built, of sun-baked brick, tile, and stone, the present bridge.

**Second run: searching for *that* followed by the clause subject, or the clause subject starting the clause with no relativiser in the sentence**

Incidence of candidate constructions: 208/6248 (3.2%)

Correctly identified sentences: 51/52 (98%)  
Correctly transformed sentences: 51/52 (98%)

*Successful*

- (A01:1000.06) The largest hurdle that the Republicans would have to face is a state law which says that before making a first race, one of two alternative courses must be taken:
- (A02:1230.33) The bill would increase from \$5,000,000 to \$15,000,000 the maximum loan that the state could make to a local water project.
- (A03:0200.24) President Kennedy today pushed aside other White House business to devote all his time and attention to working on the Berlin crisis address that he will deliver tomorrow night to the American people over nationwide television and radio.
- (A04:0880.21) This is another of the modifications of policy on Laos ~~that~~ the Kennedy administration has felt compelled to make.
- (A05:0440.30) One of the first things that he would do, he said, would be to organize classes in first aid.
- (A06:1360.06) He promised nearly 200 Democratic county committee members at the meeting in the Puddingstone Inn: "When I come back here after the November election you'll think, 'You're my man — you're the kind of governor that we're glad we elected'".
- (A07:0190.15) Their view is that last-minute changes that the Mayor is proposing to make in the Democratic ticket only emphasize the weakness of his performance as Mayor.
- (A08:0820.27) Certainly nobody will predict that the next time that the lawmakers come back together Barnett will be able to enjoy a re-enactment of the strange but successful "honeymoon" he had in the 1960 legislative session.
- (A09:0130.42) The Philadelphia Transportation Co. is investigating the part that its organization played in reviewing the project.
- (A10:0170.06) "We must persuade them to enjoy a way of life which, if not identical, is congenial with ours", he said but adding that if they do not develop the kind of society that they themselves want it will lack rituality and loyalty.
- (A12:0710.39) It was the first time that we've been ahead this season (when John Richey kicked what proved to be the winning field goal)".
- (A13:0930.48) I told him what Liston had said and he said Liston was a double-crosser and said anything that he (Liston) got was through a keyhole.
- (A14:1280.42) From the standpoint of the army of duffers, however, this was easily the most heartening exhibition that they had had since Ben Hogan fell upon evil ways during his heyday and scored an 11 in the Texas open.
- (A19:0890.09) Conceding that several cities to the north were in worse shape than Baltimore after the last storm, Mr. Schaefer listed several improvements that he said should be made in the snow plan here.
- (A20:0820.03) It was there ~~that~~ the two accused civil servants were at work.
- (G01:0580.12) This, for the liberals that I know, would be an understatement.
- (G02:1420.18) Almost febrile in intensity, the principle has become worldwide in application — unfortunately at the very time ~~that~~ nationalist fervors can wreak greatest harm.
- (G03:0210.48) "E" stands for "execution" — the moment that a "go order" would unleash an American nuclear strike.
- (G04:0150.12) Occasionally, for no reason ~~that~~ I could see, they would suddenly alter the angle of their trot.
- (G06:1390.03) He was not enthusiastic over the newly acquired Claude Lorrain, but reminisced with pleasure over a Poussin exhibit that he had been able to see in Paris a year ago.
- (G07:0600.27) HIS political opponent and lifetime friend, Thomas Jefferson, achieved immortality through his authorship of the Declaration of Independence, but equally notable were the legal and constitutional reforms that he instituted in his native Virginia, his role as father of our territorial system, and his acquisition of the Louisiana Territory during his first term as President.
- (G08:0550.51) William Styron, while facing the changing economy with a certain uneasy reluctance, insists he is not to be classified as a Southern writer and yet includes traditional Southern concepts in everything that he publishes.

- (G09:0850.18) The process stipulates that the choreographer sense the quality of the initial movement that he has discovered and that he feel the rightness of the quality that is to follow it.
- (G10:1100.30) This would destroy the federal balance between it and the states, and in the end sacrifice to the sovereignty of the states all the liberty that the citizens had gained by their Union.
- (G11:0560.21) We are not now afraid of atomic bombs in the same way that people once feared comets.
- (G12:0180.18) My curiosity was sharpened a day or two before the interview by a conversation that I had with a well-informed teacher of literature, a Jesuit father, at a conference on religious drama near Paris.
- (G13:0420.18) This is the Holy Grail that these knights of the orgasm pursue, this is the irresistible cosmic urge to which they respond.
- (G18:0630.15) It consists of fragmentary personal revelations, such as "The Spark": "There is a spark dwells deep within my soul. To get it out into the daylight's glow Is my life's aim both first and last, the whole. It slips away, it burns and tortures me. That little spark is all the wealth that I know, That little spark is my life's misery".
- (G22:0190.48) We are already committed to establishing man's supremacy over nature and everywhere on earth, not merely in the limited social-political-economical context that we are fond of today.
- (J06:0350.33) Contrary to these expectations we have found it impossible to obtain the degree of reproducibility that one would wish, even with extensive efforts to prepare especially pure reagents.
- (J07:1870.03) Most meteoritic material, by the time that it reaches the Earth's surface, has been reduced to dust or to spherules of ablated material in its passage through the atmosphere.
- (J10:0880.18) Not always, though, does the development of a bumblebee colony take place in the smooth fashion that we have just described.
- (J21:0930.33) We now shall show that any involution with these characteristics is necessarily of the type that we have just described.
- (J22:0150.27) The United States can help by communicating a genuine concern with the problems that these countries face and a readiness to provide technical and other appropriate forms of assistance where possible.
- (J23:0450.21) The kind of religious experience that most moderns seek not only provides, clarifies, and relates human yearnings, values, ideals, and purposes; it also provides facilities and incitements for the development of personality, sociality, and creativeness.
- (J24:1660.18) My aim in mentioning this factor obviously is not to give license to "wild therapy" but rather to encourage us to use the time-honored clinical casework skills that we already possess, and to use them with greater confidence, precision, and professional pride.
- (N01:0280.24) It could be some kind of trick that Budd had thought up.
- (N02:0080.51) Clayton tried to call back the face of the man that he had known.
- (N03:0310.51) He was thinking that the way that she had responded to his own kiss hadn't meant what he had believed it had.
- (N04:0300.51) In this country there's a war on every time that the grass turns green.
- (N05:0340.03) It's not the kind of thing that a man would be proud of.
- (N06:0210.09) It was obvious that he wished himself different from the sort of person that he thought he was.
- (N07:0090.60) The only reason that we brought you was to get Miller out.
- (N08:0170.24) That mistake, she thought, had cost her dearly these past few days, and she wanted to avoid falling into any more of the traps that the mountain might set for her.
- (N09:0130.03) He knew that anything that a brainy little lady like her had to say would be plumb important, as well as pleasin' to the ear, and he didn't want to miss a word of it.
- (N11:0700.18) "Fred was mighty crude about the way that he took in cattle" his own hired man, Andy Ross, mentioned later.
- (N12:0170.21) He turned and looked around at the lobby as though seeing things that he hadn't before noticed.
- (N13:0380.03) Much as they had to look forward to, they didn't begrudge a moment of the time that they spent seeing them go.

- (N14:0130.33) It's all that I ask, Stevens".  
 (N15:0050.27) Anything that the enemy flew or floated was his target.  
 (N18:0100.48) The arrangement that I had with him was to work four hours a day.

#### *Failed*

- (G05:0910.12) It was to provide a safe and spacious crossing for these caravans, and also to make a  
 pleasure for the city, ~~that~~ Shah Abbas II in about 1657 built, of sun-baked brick, tile, and  
 stone, the present bridge.

### Appendix D3: Relative Clause Reduction Rule

#### **First run: searching for relatives starting with *that* BE followed by a past participle, present participle, adverbial or adjectival clause**

Incidence of candidate constructions: 42/6248 (0.67%)  
 Correctly identified sentences: 37/42 (88%)  
 Correctly transformed sentences: 35/42 (83%)

#### *Successful*

- (A07:0590.18) Corruption is hardly a recent development in the city and state ~~that were~~ widely  
 identified as the locale of Edwin O'Connor's novel, "The Last Hurrah".  
 (A07:1070.18) In the \$40,000,000 budget ~~that has been~~ submitted for Congressional approval,  
 \$26,000,000 would be spent through universities and private voluntary agencies.  
 (A07:1380.09) Indonesia is one of the twenty under-developed countries of Asia, Africa and Latin  
 America ~~that are~~ receiving Soviet aid.  
 (A08:0430.12) Street car tracks run down the center of Pennsylvania, powered with lines ~~that are~~  
 underground.  
 (A09:0210.33) The announcement that the city would sue for recovery on the performance bond  
 was made by City Solicitor David Berger at a press conference following a meeting in the  
 morning with Wagner and other officials of the city and the PTC as well as representatives of  
 an engineering firm ~~that was~~ pulled off the El project before its completion in 1959.  
 (A12:0070.03) Moritz doesn't even have a nose guard or hip pads but he's one of the most valuable  
 members of the Longhorn team ~~that will be~~ heavily favored Saturday over Oklahoma in the  
 Cotton Bowl.  
 (A20:0710.03) Britain began designing the ship in 1956 but got nowhere until the American  
 government decided to end a ban on sharing military secrets with Britain ~~that had been~~  
 imposed after Fuchs blabbed.  
 (G01:0590.03) Theirs is no mere lack of sympathy, but something closer to the passionate hatred  
~~that was~~ directed against Fascism.  
 (G02:0450.12) For lawyers, reflecting perhaps their parochial preferences, there has been a special  
 fascination since then in the role played by the Supreme Court in that transformation — the  
 manner in which its decisions altered in "the switch in time that saved nine", President  
 Roosevelt's ill-starred but in effect victorious "Court-packing plan", the imprimatur of judicial  
 approval ~~that was~~ finally placed upon social legislation.  
 (G03:0460.21) Meanwhile, the experts speak of wars triggered by "false pre-emption",  
 "escalation", "unauthorized behavior" and other terms ~~that will be~~ discussed in this report.  
 (G04:0200.30) It is the gait of the human who must run to live: arms dangling, legs barely swinging  
 over the ground, head hung down and only occasionally swinging up to see the target, a loose  
 motion ~~that is~~ just short of stumbling and yet is wonderfully graceful.  
 (G08:0660.18) Lacking the pioneer spirit necessary to write of a *new* economy, these writers seem  
 to be contenting themselves with an old one ~~that is~~ now as defunct as Confederate money.  
 (G08:0690.06) An example of the changes which have crept over the Southern region may be seen  
 in the Southern Negro's quest for a position in the white-dominated society, a problem ~~that has~~  
~~been~~ reflected in regional fiction especially since 1865.  
 (G08:0840.06) Other examples of *gradual* changes that have affected the Negro have been his  
 moving up, row by row, in the busses; his requesting, and often getting, higher wages, better

- working conditions, better schools — changes ~~that were~~ slowly emerging even before the Supreme Court decision of 1954.
- (G08:1510.54) In the meantime, while the South has been undergoing this phenomenal modernization ~~that is~~ so disappointing to the curious Yankee, Southern writers have certainly done little to reflect and promote their region's progress.
- (G09:1640.03) The only factors ~~that are~~ personally set by the choreographer are the movements themselves, the number of the dancers, and the approximate total duration of the dance.
- (G10:0160.06) The present issue in Atlantica — whether to transform an alliance of sovereign nations into a federal union of sovereign citizens — resembles the American one of 1787-89 rather than the one ~~that was~~ resolved by Civil War.
- (G10:1700.06) Its appeal from ballots to bullets at Fort Sumter ended by costing the Southerners their right to have slaves — a right ~~that was~~ even less compatible with the sovereignty of man.
- (G13:0770.15) The beatnik, seceding from a society ~~that is~~ fatally afflicted with a deathward drive, is concerned with his personal salvation in the living present.
- (G13:1680.39) If he thus achieves a lyrical, dreamlike, drugged intensity, he pays the price for his indulgence by producing work — Allen Ginsberg's "Howl" is a striking example of this tendency — ~~that is~~ disoriented, Dionysian but without depth and without Apollonian control.
- (G18:1500.09) From being a hated tyrant and madman he was now the symbol of all ~~that was~~ noblest and best in the history of Sweden.
- (J03:1140.30) There is another means which should show the direction and relative value of the stresses in viscoelastic fluids ~~that is~~ not mentioned as such in the literature, and that is the shape of the suspended drops of low viscosity fluids in shear fields.
- (J03:1590.27) It is appropriate to call attention to certain thermodynamic properties of an ideal gas ~~that are~~ analogous to rubber-like deformation.
- (J08:1180.39) An enemy would obviously choose an agent ~~that is~~ believed to be highly infectious.
- (J08:1190.42) Agents ~~that are~~ known to cause frequent infections among laboratory workers such as those causing Q fever, tularemia, brucellosis, glanders, coccidioidomycosis, etc., belong in this category.
- (J08:1400.21) Mutants may also be developed with changes in biochemical properties ~~that are~~ of importance in identification.
- (J08:1570.51) This, of course, does not eliminate from consideration for this purpose agents ~~that are~~ associated naturally with epidemic disease.
- (J08:1840.27) This is an enormous industry ~~that is~~ subject to sabotage.
- (N02:0890.27) They passed ranches ~~that were~~ framed dark gray against the black hills.
- (N03:0790.27) You're the one ~~that's~~ taking the big chance.
- (N03:1040.39) They bawled questions ~~that were~~ not answered in the uproar.
- (N05:0200.12) She glanced around the clearing, taking in the wagon and the load of supplies and trappings scattered over the ground, the two kids, the whiteface bull ~~that was~~ chewing its cud just within the far reaches of the firelight.
- (N11:1690.09) The hands and their bosses saw him as a lone knight of the range, waging a dedicated crusade against a lawless new society ~~that was~~ threatening a beloved way of life.
- (N15:0600.24) Greg took the formation wide around three A-26 attack bombers ~~that were~~ headed north over the Gulf.
- (N18:0370.48) There was something about the contour of her face, her smile ~~that was~~ like New Orleans sunshine, the way she held her head, the way she walked — there was scarcely anything she did which did not fascinate me.

#### *Doubtful*

- (A07:0060.24) Some anti-organization Democrats saw in the program an opportunity to end the bitter internal fight within the Democratic party ~~that has been~~ going on for the last three years.
- (N03:1690.03) The bullet had torn through the flesh just above the knee, inflicting an ugly gash ~~that was~~ forming a pool of blood on the floor.

#### *Failed*

- (A05:0760.09) The statutes, similar in both the Bay State and Rhode Island and dating back in some instances to colonial times, severely limit the types of merchandise ~~that may be~~ sold on the Sabbath.

- (G13:1050.33) It is sex that obsesses them, sex ~~that is~~ at the basis of their aesthetic creed.
- (J06:1750.39) The last column shows the rate of exchange ~~that would have been~~ observed at a relative intensity of 4 (14.7 cm. distance) calculated on the assumptions that the incident light intensity is inversely proportional to the square of the distance of the lamp from the cell and that the rate is directly proportional to the incident light intensity.
- (J23:2070.33) Religion has its own supernatural prescriptions ~~that are~~ at the same time codes of behavior for the here and now.
- (N08:1400.18) There was a peculiar density about it, a thick substance ~~that could be~~ sensed but never identified, never actually perceived.

## Second run: same, but with no non-tense modals

Incidence of candidate constructions: 39/6248 (0.62%)  
 Correctly identified sentences: 37/39 (95%)  
 Correctly transformed sentences: 35/39 (90%)

### *Successful*

- (A07:0590.18) Corruption is hardly a recent development in the city and state ~~that were~~ widely identified as the locale of Edwin O'Connor's novel, "The Last Hurrah".
- (A07:1070.18) In the \$40,000,000 budget ~~that has been~~ submitted for Congressional approval, \$26,000,000 would be spent through universities and private voluntary agencies.
- (A07:1380.09) Indonesia is one of the twenty under-developed countries of Asia, Africa and Latin America ~~that are~~ receiving Soviet aid.
- (A08:0430.12) Street car tracks run down the center of Pennsylvania, powered with lines ~~that are~~ underground.
- (A09:0210.33) The announcement that the city would sue for recovery on the performance bond was made by City Solicitor David Berger at a press conference following a meeting in the morning with Wagner and other officials of the city and the PTC as well as representatives of an engineering firm ~~that was~~ pulled off the El project before its completion in 1959.
- (A12:0070.03) Moritz doesn't even have a nose guard or hip pads but he's one of the most valuable members of the Longhorn team ~~that will be~~ heavily favored Saturday over Oklahoma in the Cotton Bowl.
- (A20:0710.03) Britain began designing the ship in 1956 but got nowhere until the American government decided to end a ban on sharing military secrets with Britain ~~that had been~~ imposed after Fuchs blabbed.
- (G01:0590.03) Theirs is no mere lack of sympathy, but something closer to the passionate hatred ~~that was~~ directed against Fascism.
- (G02:0450.12) For lawyers, reflecting perhaps their parochial preferences, there has been a special fascination since then in the role played by the Supreme Court in that transformation — the manner in which its decisions altered in "the switch in time that saved nine", President Roosevelt's ill-starred but in effect victorious "Court-packing plan", the imprimatur of judicial approval ~~that was~~ finally placed upon social legislation.
- (G03:0460.21) Meanwhile, the experts speak of wars triggered by "false pre-emption", "escalation", "unauthorized behavior" and other terms ~~that will be~~ discussed in this report.
- (G04:0200.30) It is the gait of the human who must run to live: arms dangling, legs barely swinging over the ground, head hung down and only occasionally swinging up to see the target, a loose motion ~~that is~~ just short of stumbling and yet is wonderfully graceful.
- (G08:0660.18) Lacking the pioneer spirit necessary to write of a *new* economy, these writers seem to be contenting themselves with an old one ~~that is~~ now as defunct as Confederate money.
- (G08:0690.06) An example of the changes which have crept over the Southern region may be seen in the Southern Negro's quest for a position in the white-dominated society, a problem ~~that has been~~ reflected in regional fiction especially since 1865.
- (G08:0840.06) Other examples of *gradual* changes that have affected the Negro have been his moving up, row by row, in the busses; his requesting, and often getting, higher wages, better working conditions, better schools — changes ~~that were~~ slowly emerging even before the Supreme Court decision of 1954.

- (G08:1510.54) In the meantime, while the South has been undergoing this phenomenal modernization ~~that is~~ so disappointing to the curious Yankee, Southern writers have certainly done little to reflect and promote their region's progress.
- (G09:1640.03) The only factors ~~that are~~ personally set by the choreographer are the movements themselves, the number of the dancers, and the approximate total duration of the dance.
- (G10:0160.06) The present issue in Atlantica — whether to transform an alliance of sovereign nations into a federal union of sovereign citizens — resembles the American one of 1787-89 rather than the one ~~that was~~ resolved by Civil War.
- (G10:1700.06) Its appeal from ballots to bullets at Fort Sumter ended by costing the Southerners their right to have slaves — a right ~~that was~~ even less compatible with the sovereignty of man.
- (G13:0770.15) The beatnik, seceding from a society ~~that is~~ fatally afflicted with a deathward drive, is concerned with his personal salvation in the living present.
- (G13:1680.39) If he thus achieves a lyrical, dreamlike, drugged intensity, he pays the price for his indulgence by producing work — Allen Ginsberg's "Howl" is a striking example of this tendency — ~~that is~~ disoriented, Dionysian but without depth and without Apollonian control.
- (G18:1500.09) From being a hated tyrant and madman he was now the symbol of all ~~that was~~ noblest and best in the history of Sweden.
- (J03:1140.30) There is another means which should show the direction and relative value of the stresses in viscoelastic fluids ~~that is~~ not mentioned as such in the literature, and that is the shape of the suspended drops of low viscosity fluids in shear fields.
- (J03:1590.27) It is appropriate to call attention to certain thermodynamic properties of an ideal gas ~~that are~~ analogous to rubber-like deformation.
- (J08:1180.39) An enemy would obviously choose an agent ~~that is~~ believed to be highly infectious.
- (J08:1190.42) Agents ~~that are~~ known to cause frequent infections among laboratory workers such as those causing Q fever, tularemia, brucellosis, glanders, coccidioidomycosis, etc., belong in this category.
- (J08:1400.21) Mutants may also be developed with changes in biochemical properties ~~that are~~ of importance in identification.
- (J08:1570.51) This, of course, does not eliminate from consideration for this purpose agents ~~that are~~ associated naturally with epidemic disease.
- (J08:1840.27) This is an enormous industry ~~that is~~ subject to sabotage.
- (N02:0890.27) They passed ranches ~~that were~~ framed dark gray against the black hills.
- (N03:0790.27) You're the one ~~that's~~ taking the big chance.
- (N03:1040.39) They bawled questions ~~that were~~ not answered in the uproar.
- (N05:0200.12) She glanced around the clearing, taking in the wagon and the load of supplies and trappings scattered over the ground, the two kids, the whiteface bull ~~that was~~ chewing its cud just within the far reaches of the firelight.
- (N11:1690.09) The hands and their bosses saw him as a lone knight of the range, waging a dedicated crusade against a lawless new society ~~that was~~ threatening a beloved way of life.
- (N15:0600.24) Greg took the formation wide around three A-26 attack bombers ~~that were~~ headed north over the Gulf.
- (N18:0370.48) There was something about the contour of her face, her smile ~~that was~~ like New Orleans sunshine, the way she held her head, the way she walked — there was scarcely anything she did which did not fascinate me.

#### *Doubtful*

- (A07:0060.24) Some anti-organization Democrats saw in the program an opportunity to end the bitter internal fight within the Democratic party ~~that has been~~ going on for the last three years.
- (N03:1690.03) The bullet had torn through the flesh just above the knee, inflicting an ugly gash ~~that was~~ forming a pool of blood on the floor.

#### *Failed*

- (G13:1050.33) It is sex that obsesses them, sex ~~that is~~ at the basis of their aesthetic creed.
- (J23:2070.33) Religion has its own supernatural prescriptions ~~that are~~ at the same time codes of behavior for the here and now.



## Appendix D4: Relative Clause Expansion Rule

### First run: searching for noun phrases directly followed by a past participle, present participle, time/place adverbial or adjectival clause

Incidence of candidate constructions: 143/6248 (2.6%)

Correctly identified sentences: 32/49 (65%)

Correctly transformed sentences: 26/49 (53%)

#### *Successful*

- (A01:1160.18) Before adjournment Monday afternoon, the Senate is expected to approve a study of the number of legislators that are allotted to rural and urban areas to determine what adjustments should be made.
- (A02:0960.24) Under Formby's plan, an appointee would be selected by a board that is composed of the governor, lieutenant governor, speaker of the House, attorney general and chief justice of the Texas Supreme Court.
- (A04:1290.06) Until the Cuban fiasco and the Communist military victories in Laos, almost any observer would have said that President Kennedy had blended a program that respected, generally, the opinions that were voiced both by Mr. Nixon and the professors.
- (A09:0360.27) At Berger's direction, the city also intervened in the Hughes bankruptcy case in U. S. District Court in a move that was preliminary to filing a claim there.
- (A19:1770.12) Although economic and personal circumstances vary widely among those that are now choosing apartments, Leo J. Pantas, vice president of a hardware manufacturing company, pointed out recently that many apartment seekers seem to have one characteristic in common: a desire for greater convenience and freedom from the problems involved in maintaining a house.
- (A20:0240.39) In their suburban cottage the crown charges, the Krogers received secrets from the mystery man, usually on the first Saturday evening of each month, and spent much of the week-end getting the secrets off to Moscow, either on a powerful transmitter that was buried under the kitchen floor or as dots that were posted over period marks in used books.
- (G01:0190.48) This is puzzling to an outsider that is conscious of the classic tradition of liberalism, because it is clear that these Democrats who are left-of-center are at opposite poles from the liberal Jefferson, who held that the best government was the least government.
- (G05:0940.30) From upstream it looks like a long arcaded box that is laid across the river; from downstream, where the water level is much lower, it is a high, elaborately facaded pavilion.
- (G06:0170.30) Seeking an obscure, dark, relatively quiet corner in the airy room that was otherwise suffused with afternoon sunshine, he asked if the soft background music could be turned off.
- (G07:0070.15) Incapable of self-delusion, the Founding Fathers found the crisis of their time to be equally grave, and yet they had confidence that America would surmount it and that a republic of free peoples would prosper and serve as an example to a world that was aching for liberty.
- (G09:0300.12) Thus, the avant-garde choreographers have extended the scope of materials that are available for dance composition.
- (G10:0260.42) To their leaders the Constitution was a compact that was made by the people of sovereign states, who therefore retained the right to secede from it.
- (G11:1610.24) Much of the former extreme uneasiness that is associated with visions and hallucinations and with death has disappeared.
- (G12:0810.15) That is why the form itself becomes a preoccupation, because it exists as a problem that is separate from the material it accommodates.
- (G22:0160.12) Hence the prime issue, as I see it, is whether a democratic or free society can master technology for the benefit of mankind, or whether technology will rule and develop its own society that is compatible with *its* own needs as a force of nature.
- (J03:0210.15) Two of the fluids showed a high-positive normal pressure when undergoing shear, and two showed small negative pressures which were negligible in comparison with the amount of the positive pressures that were generated by the other two.
- (J04:1080.33) Since [FORMULA] is ferromagnetic, we felt that any results that were obtained from the magnetically contaminated [FORMULA] would be suspect.

- (J05:0110.24) In the household market, there are also low-sudsing detergent formulations that are based on nonionic actives with about the same amount of phosphate builder; light-duty synthetic detergents with much less builder; and the dwindling built-soap powders as well as soap flakes and granules, none of which are now nationally advertised.
- (J06:0480.54) Of several methods that were employed for tagging chlorine with radiochlorine, the exchange of inactive chlorine with tagged aluminum chloride at room temperature was found to be the most satisfactory.
- (J08:0450.24) For example, bright sunlight is rapidly destructive for living microorganisms that are suspended in air.
- (J23:1610.09) This is brought out in the common religious ethos that prevails even in the denominationally diverse audiences at many secular semi-public and public occasions in the United States; and it is evidenced in the prayers that are offered, in the frequent religious allusions, and in the confirmation of points on religious grounds.
- (N06:0060.27) Sometimes I was aware of people that were moving about in the darkness.
- (N14:0541.03) Less assured than the tall, wide-shouldered man that was in the lead, Cobb followed alertly, a hand on his gun butt.

#### *Doubtful*

- (G02:0180.24) Beyond the two basic tasks that were mentioned above, no attention was paid by statesman or scholar to an idea of state responsibility, either internally or externally.
- (G17:1830.27) One is that they were established, or gained eminence, under pressure that was provided by these same immigrants, from whom the old families wished to segregate their children.
- (J02:0020.42) Thus, the energy that was transferred from the arc to the anode was partly fed back into the arc.
- (J07:0320.18) According to Whipple (1955), cometary debris is sufficient to replenish the material that is spiraling into the sun, maintaining a fairly steady state.
- (N03:0200.33) No doubt there would be men that are guarding the horses.
- (N04:0510.33) In the brief moment I had to talk to them before I took my post that was on the ring of defenses, I indicated I was sickened by the methods men employed to live and trade on the river.
- (N18:0520.51) True, she was my Aunt, that was married to an Uncle related to me only by marriage, but why she had married a man twice her age, and more, perhaps, I did not know or much care.

#### *Failed*

- (A03:0710.18) After reading his statement that is discharging the 23d ward case, Karns told Wexler that if the seven cases that were scheduled for trial also involved persons who had been subpoenaed, he would dismiss them.
- (A06:0550.03) Earlier, Mitchell said in a statement: "I think that all Americans will resent deeply the statements that are made about President Eisenhower by Richard J. Hughes. His reference to 'discredited carcass' or 'tattered remains' of the president's leadership is an insult to the man who led our forces to victory in the greatest war in all history, to the man who was twice elected overwhelmingly by the American people as president of the United States, and who has been the symbol to the world of the peace-loving intentions of the free nations.
- (A07:0970.30) What could rescue the bill would be some quick progress on a bill that is amending the National Defense Education Act of 1958.
- (A11:1270.27) Unfortunately, Brooks' teammates were not in such festive mood as the Orioles expired before the seven-hit that was pitching of three Kansas City rookie hurlers.
- (A14:1120.15) If such a paragon of perfection as Palmer could commit such a scoring sacrilege, there was hope that was left for all.
- (G03:0200.12) In the command centers there are special clocks that are ready to tick off the minutes that are elapsed since "E hour".
- (G08:0390.18) Undoubtedly even the old Southern stalwart Richmond has felt the new wind: William Styron mentions in his latest novel an avenue that is named for Bankhead McGruder, a Civil War general, now renamed, in typical California fashion, "Buena Vista Terrace".

- (G18:0020.39) Of these there are surely few that would be more rewarding discoveries than Verner von Heidenstam, the Swedish poet and novelist who received the award that was in 1916 and whose centennial was celebrated two years ago.
- (J01:0250.42) For the case of Jupiter, the radio emission spectrum is definitely not like the spectrum of a black-body radiator, and it seems very likely that the radiation that is reaching the earth is a combination of thermal radiation from the atmosphere and non-thermal components.
- (J09:0130.27) In the present work whole sera have been fractionated by chromatography on DEAE-cellulose using single gradients similar to those that were described by Sober and Peterson, and certain chemical and serological properties of the fractions that were containing antibodies of the ABO and Rh systems have been described.
- (J12:0390.51) Some of these were obviously filling from interlobular branches of the bronchial arteries while others were filling from direct hilar branches that were following along the pleural surface.
- (J17:0370.15) Thus cortico-fugal discharges that are induced by topical application of strychnine to a minute area in the neocortex summate with spikes that are present in the hypothalamus and cause increased convulsive discharges.
- (J21:0030.09) These are defined by a simple involutorial transformation of the points that are in which a general line meets a nonsingular quadric surface that is bearing a curve of symbol [FORMULA].
- (J22:0120.39) Nevertheless, with foresight and careful planning, some of the more disruptive and dangerous consequences of social change which have troubled other countries that are passing through this stage can be escaped.
- (J24:0010.03) Overwhelmed with the care of five young children that were and concerned about persistent economic difficulties due to her husband's marginal income, her defense of denial was excessively strong.
- (N01:1260.06) "I don't have many strays that are coming to my front door", he said.
- (N08:0230.33) All of her movements were careful and methodical, partaking of the stealth of a criminal who has plotted his felony that was for months in advance and knows exactly which step to take next in the course of the final execution of his crime.
- (N11:0520.06) "Dead center that was at three hundred yards, that coroner said"! he'd grin.
- (N13:0440.27) Seeing them that were waiting there at the foot of Emigrant Rock was so overwhelming that, for a good minute after they rounded the bend and started down the grade leading toward them, Matilda could not speak at all.

## Appendix D5: Adjunct Movement Rule

### First run: searching for adverbial or prepositional phrase of time, place, manner or contingency, that modifies the sentence at its start or end

Incidence of candidate constructions: 72/6248 (1.2%)

Correctly identified sentences: 21/38 (50%)

Correctly transformed sentences: 17/38 (50%)

#### *Successful*

- (A01:1830.48) ~~During the election campaign,~~ Both candidates, Davis and Bush, reportedly received anonymous telephone calls during the campaign.
- (A02:1630.06) ~~During his college career,~~ Dr. Clark was captain of his basketball team and was a football letterman during his college career.
- (A05:1820.06) ~~At present~~ All offenses must be taken to Sixth District Court for disposition at present.
- (A06:1240.39) ~~Under his supervision,~~ The state fire-fighting agency developed such techniques as plowing of fire lines and established a fleet of tractor plows and tractor units for fire fighting under his supervision.
- (A11:0400.12) ~~In the ninth,~~ Robinson led off with his second double of the night, a blast off the fence 375 feet deep into left, in the ninth.

- (A13:0280.21) ~~Despite the 45-degree weather~~ The game was clicked off in 1:48, thanks to only three bases on balls and some good infield play, despite the 45 degree weather.
- (A14:0360.06) ~~With the announcement of a "special achievement award" to William A. (Bill) Shea,~~ the awards list was completed yesterday for Sunday night's thirty-eighth annual dinner and show of the New York Chapter, Baseball Writers' Association of America, at the Waldorf-Astoria Hotel, with the announcement of a "special achievement award" to William A. (Bill) Shea.
- (G02:0010.06) ~~In the past,~~ The duties of the state, as Sir Henry Maine noted long ago, were only two in number: internal order and external security in the past.
- (G04:0320.30) ~~With a lordly and generous gesture,~~ the discoverer stood up and beckoned to the closest of his fellows with a lordly and generous gesture.
- (G05:1110.12) ~~On the downstream, or "pavilion", side~~ These vaults give out onto terraces twice as wide as the bridge itself on the downstream, or "pavilion", side.
- (G06:0820.24) ~~In 1938,~~ At the insistence of Arturo Toscanini, Steinberg left Germany for the United States, by way of Switzerland in 1938.
- (J03:0120.42) ~~In this test A [FORMULA] tape was pulled between the blocks with a motor and pulley at a rate of [FORMULA] with a clearance of 0.002 on each side of the tape~~ in this test.
- (J06:1020.39) ~~Following reaction~~ The cells were scratched with a file and opened under a 20% aqueous sodium iodide solution following reaction.
- (J09:0130.27) ~~In the present work~~ Whole sera have been fractionated by chromatography on DEAE-cellulose using single gradients similar to those described by Sober and Peterson, and certain chemical and serological properties of the fractions containing antibodies of the ABO and Rh systems have been described in the present work.
- (N01:0260.06) ~~When, in late afternoon on the last day in June, he saw two people top the ridge to the south and walk toward the house,~~ He quit work immediately and strode to his rifle when, in late afternoon on the last day in June, he saw two people top the ridge to the south and walk toward the house.
- (N09:1750.06) ~~When he regained consciousness~~ He was in Lord's house, in the office of Doctor Lord, the deputy's deceased father when he regained consciousness.
- (N10:0410.06) ~~At the moment,~~ The three men were not saying much of anything, at the moment.
- (N11:0750.06) ~~On the morning of September 10, 1895,~~ Powell and Ross rose at dawn and began their day's work on the morning of September 10, 1895.
- (N12:1330.42) ~~By now~~ Curt was seeing clearly again by now.
- (N13:0010.06) ~~Over his shoulder~~ He could see Max's loose grin and the Burnsides' glowering faces over his shoulder.
- (N15:1100.39) ~~In seconds,~~ Greg made his decision in seconds.

#### *Doubtful*

- (A10:0270.42) ~~By leaving our doors open~~ The United States gives other peoples the opportunity to see us and to compare, he said, by leaving our doors open.
- (A19:0140.51) ~~Since the railroad cannot reduce the salary of individual union members under contract,~~ It must accomplish its payroll reduction by placing some of the men on furlough, a B.&O. spokesman said, since the railroad cannot reduce the salary of individual union members under contract.
- (G07:0470.12) ~~ON a military mission for his native Virginia~~ The youthful George Washington touched off the French and Indian War, then guarded his colony's frontier as head of its militia, on a military mission for his native Virginia.
- (J21:0010.15) ~~In~~ We investigate a new series of line involutions in a projective space of three dimensions over the field of complex numbers in 1.
- (N08:0020.24) ~~After they had finished eating,~~ Melissa took Sprite the kitten under her arm — "so that Auntie Grace can teach it about the whistle" — and climbed into the station wagon beside her mother, after they had finished eating.

#### *Failed*

- (A04:1410.27) ~~In Laos,~~ The administration looked at the Eisenhower administration efforts to show determination by sailing a naval fleet into Southeast Asian waters as a useless gesture in Laos.
- (A07:0040.18) ~~At the same time~~ Reaction among anti-organization Democratic leaders and in the Liberal party to the Mayor's reported plan was generally favorable at the same time.

- (A08:0090.12) ~~From June, 1942, Until December, 1945, Pfaff served in the Army Air Corps from June, 1942.~~
- (A12:0480.30) ~~In their first three games,~~ The Longhorns have had the ball 41 times and scored 16 times, or 40 per cent ... their total passing yardage in three games, 447 on 30 completions in 56 attempts, is only 22 yards short of their total passing yardage in 1959, when they made 469 on 37 completions in 86 tries ..., in their first three games.
- (A20:1830.21) ~~At Jenks street,~~ Simms said, the car skidded completely around, just missed two parked cars, and sped east in Jenks, at Jenks street.
- (G08:0290.18) ~~In 1900~~ The South was only 15% urban; in 1950 it had become 47.1% urban in 1900.
- (G18:0010.06) ~~AMONG THE RECIPIENTS of the Nobel Prize for Literature~~ More than half are practically unknown to readers of English among the recipients of the Nobel Prize for Literature.
- (J01:1360.12) ~~In the observations at 4.3 mm (Coates, 1959a),~~ The diameter of the antenna beam, 6'.7, was small enough to allow resolution of some of the larger features of the lunar surface, and contour diagrams have been made of the lunar brightness distribution at three lunar phases, in the observations at 4.3 mm (Coates, 1959a).
- (J12:0700.15) ~~Within certain wide limits~~ Anatomy dictates function and, if one is permitted to speculate, potential pathology should be included in this statement as well within certain wide limits.
- (N07:0020.54) ~~Before they could guess his intention~~ Rankin stepped forward and swung the guard's own gun against the uncovered head, hard, before they could guess his intention-
- (N14:0810.15) ~~Inside the crown, stuffed behind the stained sweatband,~~ Could be seen thin, crumpled wads of currency, inside the crown, stuffed behind the stained sweatband.
- (N18:0780.18) ~~For several weeks~~ We eyed one another almost like sparring partners, and then one day Uncle was slightly indisposed and stayed home; his bride opened the store for several weeks.

**Second run: same, but sensitive to co-ordinate sentences and multiple adjuncts; and excluding subjectless sentences and sentences with adjuncts at both ends of the sentence**

Incidence of candidate constructions: 67/6248 (1.1%)  
 Correctly identified sentences: 33/34 (97%)  
 Correctly transformed sentences: 32/34 (94%)

#### *Successful*

- (A01:1830.48) ~~During the election campaign,~~ Both candidates, Davis and Bush, reportedly received anonymous telephone calls during the campaign.
- (A02:1630.06) ~~During his college career,~~ Dr. Clark was captain of his basketball team and was a football letterman during his college career.
- (A05:1820.06) ~~At present~~ All offenses must be taken to Sixth District Court for disposition at present.
- (A06:1240.39) ~~Under his supervision,~~ The state fire-fighting agency developed such techniques as plowing of fire lines and established a fleet of tractor plows and tractor units for fire fighting under his supervision.
- (A08:0090.12) ~~From June, 1942, until December, 1945,~~ Pfaff served in the Army Air Corps from June, 1942, until December, 1945.
- (A10:0270.42) ~~By leaving our doors open~~ The United States gives other peoples the opportunity to see us and to compare by leaving our doors open, he said.
- (A11:0400.12) ~~In the ninth,~~ Robinson led off with his second double of the night, a blast off the fence 375 feet deep into left, in the ninth.
- (A12:0480.30) ~~In their first three games,~~ The Longhorns have had the ball 41 times and scored 16 times, or 40 per cent, in their first three games ... their total passing yardage in three games, 447 on 30 completions in 56 attempts, is only 22 yards short of their total passing yardage in 1959, when they made 469 on 37 completions in 86 tries ....
- (A13:0280.21) ~~Despite the 45-degree weather~~ The game was clicked off in 1:48, thanks to only three bases on balls and some good infield play, despite the 45 degree weather.

- (A14:0360.06) ~~With the announcement of a “special achievement award” to William A. (Bill) Shea,~~ the awards list was completed yesterday for Sunday night's thirty-eighth annual dinner and show of the New York Chapter, Baseball Writers' Association of America, at the Waldorf-Astoria Hotel, with the announcement of a “special achievement award” to William A. (Bill) Shea.
- (G02:0010.06) ~~In the past,~~ The duties of the state, as Sir Henry Maine noted long ago, were only two in number: internal order and external security in the past.
- (G04:0320.30) ~~With a lordly and generous gesture,~~ the discoverer stood up and beckoned to the closest of his fellows with a lordly and generous gesture.
- (G05:1110.12) ~~On the downstream, or “pavilion”, side~~ These vaults give out onto terraces twice as wide as the bridge itself on the downstream, or “pavilion”, side.
- (G06:0820.24) ~~In 1938,~~ At the insistence of Arturo Toscanini, Steinberg left Germany for the United States, by way of Switzerland in 1938.
- (G07:0470.12) ~~ON a military mission for his native Virginia~~ The youthful George Washington touched off the French and Indian War on a military mission for his native Virginia, then guarded his colony's frontier as head of its militia.
- (G08:0290.18) ~~In 1900~~ The South was only 15% urban in 1900; in 1950 it had become 47.1% urban.
- (J01:1360.12) ~~In the observations at 4.3 mm (Coates, 1959a),~~ The diameter of the antenna beam, 6'.7, was small enough to allow resolution of some of the larger features of the lunar surface in the observations at 4.3 mm (Coates, 1959a), and contour diagrams have been made of the lunar brightness distribution at three lunar phases.
- (J03:0120.42) ~~In this test~~ A [FORMULA] tape was pulled between the blocks with a motor and pulley at a rate of [FORMULA] with a clearance of 0.002 on each side of the tape in this test.
- (J06:1020.39) ~~Following reaction~~ The cells were scratched with a file and opened under a 20% aqueous sodium iodide solution following reaction.
- (J09:0130.27) ~~In the present work~~ Whole sera have been fractionated by chromatography on DEAE-cellulose using single gradients similar to those described by Sober and Peterson, and certain chemical and serological properties of the fractions containing antibodies of the ABO and Rh systems have been described in the present work.
- (J12:0700.15) ~~Within certain wide limits~~ Anatomy dictates function within certain wide limits and, if one is permitted to speculate, potential pathology should be included in this statement as well.
- (N01:0260.06) ~~When, in late afternoon on the last day in June, he saw two people top the ridge to the south and walk toward the house,~~ He quit work immediately and strode to his rifle when, in late afternoon on the last day in June, he saw two people top the ridge to the south and walk toward the house.
- (N07:0020.54) ~~Before they could guess his intention~~ Rankin stepped forward before they could guess his intention and swung the guard's own gun against the uncovered head, hard.
- (N08:0200.33) ~~When the station wagon drew abreast of the dusty dirt road that led up to the porch of the Culver house,~~ Pamela turned the wheel when the station wagon drew abreast of the dusty dirt road that led up to the porch of the Culver house, guiding the car to its familiar parking spot close to the house, and stopped.
- (N09:1750.06) ~~When he regained consciousness~~ He was in Lord's house, in the office of Doctor Lord, the deputy's deceased father when he regained consciousness.
- (N10:0410.06) ~~At the moment,~~ The three men were not saying much of anything, at the moment.
- (N11:0750.06) ~~On the morning of September 10, 1895,~~ Powell and Ross rose at dawn and began their day's work on the morning of September 10, 1895.
- (N12:1330.42) ~~By now~~ Curt was seeing clearly again by now.
- (N13:0010.06) ~~Over his shoulder~~ He could see Max's loose grin and the Burnsides' glowering faces over his shoulder.
- (N14:1020.57) ~~To his faint surprise~~ Russ held up his hand to his faint surprise.
- (N15:1100.39) ~~In seconds,~~ Greg made his decision in seconds.
- (N18:0780.18) ~~For several weeks~~ We eyed one another almost like sparring partners for several weeks, and then one day Uncle was slightly indisposed and stayed home; his bride opened the store.

### Failed

- (A07:0040.18) ~~At the same time~~ Reaction among anti-organization Democratic leaders and in the Liberal party to the Mayor's reported plan was generally favorable at the same time.
- (A04:1410.27) ~~In Laos,~~ The administration looked at the Eisenhower administration efforts to show determination by sailing a naval fleet into Southeast Asian waters as a useless gesture in Laos.

## Appendix D6: Extraposition Rule

**First run: searching for sentences with a subject that is a *that* complement, *for-to* complement or infinitival complement, whether interrogative or declarative, where either the complement starts the sentence, or *it* starts the sentence, the complement ends the sentence and in the case of raising verbs there is a non-clausal object**

Incidence of candidate constructions: 68/6248 (1.1%)

Correctly identified sentences: 37/37 (100%)

Correctly transformed sentences: 30/37 (81%)

### Successful

- (A03:0120.30) ~~It is impossible~~ To get a fair trial when some of the defendants made statements involving themselves and others is impossible.
- (A04:1010.15) ~~It does not take a Gallup poll~~ To find out that most Republicans in Congress feel this understates the situation as Republicans see it does not take a Gallup poll.
- (A10:1630.12) ~~To accomplish this~~ It would necessitate some changes in methods, he said, to accomplish this.
- (A13:1450.33) ~~It How~~ a team will suddenly start hitting after a distressing slump will forever be a baseball mystery how a team will suddenly start hitting after a distressing slump.
- (A20:0620.27) ~~Just how many sub secrets were being handed over when the ring, watched for six months, was broken~~ It remained untold just how many sub secrets were being handed over when the ring, watched for six months, was broken.
- (G01:1300.24) ~~It To see how the situation could be otherwise~~ is hard to see how the situation could be otherwise.
- (G02:0980.21) ~~It That the majority of the delegates to the Congress were from the less developed, former colonial nations is noteworthy that the majority of the delegates to the Congress were from the less developed, former colonial nations.~~
- (G03:0560.03) ~~It To think about the unthinkable~~ is their job to think about the unthinkable.
- (G10:1540.15) ~~It To see that on both sides sovereign Americans had given their lives in the Civil War to maintain the balance between the powers they had delegated to the States and to their Union is much less difficult now than in Lincoln's day to see that on both sides sovereign Americans had given their lives in the Civil War to maintain the balance between the powers they had delegated to the States and to their Union.~~
- (G11:0530.18) ~~It To reconstruct the primeval fears of man~~ is difficult to reconstruct the primeval fears of man.
- (G17:0800.27) ~~It To know how much "integration" there is in the famous, fashionable colleges and prep schools of New England would be interesting to know how much "integration" there is in the famous, fashionable colleges and prep schools of New England.~~
- (J01:0010.15) ~~It To use the radio emission of the moon and planets as a new source of information about these bodies and their atmospheres~~ HAS recently become practical to use the radio emission of the moon and planets as a new source of information about these bodies and their atmospheres.
- (J02:0400.33) ~~It That the cooling requirements for the anode of free burning arcs are large compared with those for the cathode~~ is apparent from the above and from experimental evidence that the cooling requirements for the anode of free burning arcs are large compared with those for the cathode.
- (J03:1590.27) ~~It To call attention to certain thermodynamic properties of an ideal gas that are analogous to rubber-like deformation is appropriate to call attention to certain thermodynamic properties of an ideal gas that are analogous to rubber-like deformation.~~

- (J06:0080.21) It To set a lower limit of about [FORMULA] for the activation energy of the abstraction of a chlorine atom from a carbon tetrachloride molecule by a chlorine atom to form [FORMULA] radical appears possible ~~to set a lower limit of about [FORMULA] for the activation energy of the abstraction of a chlorine atom from a carbon tetrachloride molecule by a chlorine atom to form [FORMULA] radical.~~
- (J07:0350.21) It That some of the dust in the vicinity of the Earth originated from meteoritic impacts upon the moon is also possible ~~that some of the dust in the vicinity of the Earth originated from meteoritic impacts upon the moon.~~
- (J08:1040.30) It That an even greater area was covered, particularly downwind is quite likely ~~that an even greater area was covered, particularly downwind.~~
- (J10:0240.03) It That the stamens of the catkin are always arranged in pairs, and that each individual flower is nothing but one such pair standing on a green, black-tipped little scale is not difficult to see ~~that the stamens of the catkin are always arranged in pairs, and that each individual flower is nothing but one such pair standing on a green, black-tipped little scale.~~
- (J12:0570.12) It That many marked and striking differences exist between lungs when an inter-species comparison is made is evident ~~that many marked and striking differences exist between lungs when an inter-species comparison is made.~~
- (J21:0410.42) It That a 1:1 correspondence can be established between the lines of two such pencils, so that in a sense a unique image can actually be assigned to each tangent is interesting ~~that a 1:1 correspondence can be established between the lines of two such pencils, so that in a sense a unique image can actually be assigned to each tangent.~~
- (J24:1300.24) That we are experiencing an upsurge of interest in the many formulations and preventive adaptations of brief treatment in social casework It is evident from even a small sampling of current literature ~~that we are experiencing an upsurge of interest in the many formulations and preventive adaptations of brief treatment in social casework.~~
- (N04:1580.30) It To see the thin ranks of warriors, old and young, wheeling and twisting their ponies frantically from side to side only to be tumbled bleeding from their saddles by the relentless slam, slam of the cruelly efficient Hawkinses was pitiful ~~to see the thin ranks of warriors, old and young, wheeling and twisting their ponies frantically from side to side only to be tumbled bleeding from their saddles by the relentless slam, slam of the cruelly efficient Hawkinses.~~
- (N05:0120.51) It To compose himself took him a long time ~~to compose himself.~~
- (N06:0210.09) It That he wished himself different from the sort of person he thought he was was obvious ~~that he wished himself different from the sort of person he thought he was.~~
- (N08:0430.12) It That she would have to enact her renunciation, beg forgiveness was there ~~that she would have to enact her renunciation, beg forgiveness.~~
- (N10:0370.39) It That she probably would never know bothered her ~~that she probably would never know.~~
- (N11:0420.12) It To locate Horn took some time ~~to locate Horn.~~
- (N15:1540.39) It That they hadn't been spotted did not seem possible ~~that they hadn't been spotted.~~
- (N18:0500.06) That she impressed me instantly It was obvious ~~that she impressed me instantly; I was aware that when our eyes met we both quickly averted them.~~

### *Doubtful*

- (G07:0050.06) It That in this hour both of "national peril" and of "national opportunity" we can take counsel with the men who made the nation is well then ~~that in this hour both of "national peril" and of "national opportunity" we can take counsel with the men who made the nation.~~
- (G08:0180.09) It That despite this strong upsurge in Southern writing, almost none of the writers has forsaken the firmly entrenched concept of the white-suited big-daddy colonel sipping a mint julep as he silently recounts the revenue from the season's cotton and tobacco crops; of the stereotyped Negro servants chanting hymns as they plow the fields; of these and a host of other antiquated legends that deny the South its progressive leaps of the past century is interesting, however, ~~that despite this strong upsurge in Southern writing, almost none of the writers has forsaken the firmly entrenched concept of the white-suited big-daddy colonel sipping a mint julep as he silently recounts the revenue from the season's cotton and tobacco crops; of the stereotyped Negro servants chanting hymns as they plow the fields; of these and a host of other antiquated legends that deny the South its progressive leaps of the past century.~~



- (G12:1060.15) It ~~That our situation becomes inexplicable~~ is because there is not only darkness but also light ~~that our situation becomes inexplicable~~.
- (J05:1830.33) It ~~That the requirements imposed by these effects upon any one detergent constituent acting alone are severe~~ is evident ~~that the requirements imposed by these effects upon any one detergent constituent acting alone are severe~~.

### *Failed*

- (J22:0230.30) It ~~That success can be purchased much less dearly by fishing in the murky waters of international politics than by facing up to the intractable tasks at home~~ may seem to some of them ~~that success can be purchased much less dearly by fishing in the murky waters of international politics than by facing up to the intractable tasks at home~~.
- (N07:0470.30) It ~~That Barton had sent Carl Dill on Dill's release from the prison~~ was to him ~~that Barton had sent Carl Dill on Dill's release from the prison~~.
- (N14:1010.06) It ~~For you to return this money to Colcord and tell him about it'll be a pleasure for you to return this money to Colcord and tell him about it~~.

**Second run: same, but in the case of copula and raising verbs there must be a non-clausal direct object, and the main verb must not be contracted**

Incidence of candidate constructions: 63/6248 (1.0%)  
 Correctly identified sentences: 35/35 (100%)  
 Correctly transformed sentences: 31/35 (88%)

### *Successful*

- (A03:0120.30) It ~~is impossible~~ To get a fair trial when some of the defendants made statements involving themselves and others is impossible.
- (A04:1010.15) It ~~does not take a Gallup poll~~ To find out that most Republicans in Congress feel this understates the situation as Republicans see it does not take a Gallup poll.
- (A10:1630.12) To ~~accomplish this~~ It would necessitate some changes in methods, he said, to accomplish this.
- (A13:1450.33) It ~~How a team will suddenly start hitting after a distressing slump~~ will forever be a baseball mystery how a team will suddenly start hitting after a distressing slump.
- (A20:0620.27) Just ~~how many sub secrets were being handed over when the ring, watched for six months, was broken~~ It remained untold just how many sub secrets were being handed over when the ring, watched for six months, was broken.
- (G01:1300.24) It ~~To see how the situation could be otherwise~~ is hard ~~to see how the situation could be otherwise~~.
- (G02:0980.21) It ~~That the majority of the delegates to the Congress were from the less developed, former colonial nations~~ is noteworthy ~~that the majority of the delegates to the Congress were from the less developed, former colonial nations~~.
- (G03:0560.03) It ~~To think about the unthinkable~~ is their job ~~to think about the unthinkable~~.
- (G10:1540.15) It ~~To see that on both sides sovereign Americans had given their lives in the Civil War to maintain the balance between the powers they had delegated to the States and to their Union~~ is much less difficult now than in Lincoln's day ~~to see that on both sides sovereign Americans had given their lives in the Civil War to maintain the balance between the powers they had delegated to the States and to their Union~~.
- (G11:0530.18) It ~~To reconstruct the primeval fears of man~~ is difficult ~~to reconstruct the primeval fears of man~~.
- (G17:0800.27) It ~~To know how much "integration" there is in the famous, fashionable colleges and prep schools of New England~~ would be interesting ~~to know how much "integration" there is in the famous, fashionable colleges and prep schools of New England~~.
- (J01:0010.15) If ~~To use the radio emission of the moon and planets as a new source of information about these bodies and their atmospheres~~ HAS recently become practical ~~to use the radio emission of the moon and planets as a new source of information about these bodies and their atmospheres~~.
- (J02:0400.33) It ~~That the cooling requirements for the anode of free burning arcs are large compared with those for the cathode~~ is apparent from the above and from experimental

- evidence that the cooling requirements for the anode of free burning arcs are large compared with those for the cathode.
- (J03:1590.27) It To call attention to certain thermodynamic properties of an ideal gas that are analogous to rubber-like deformation is appropriate to ~~call attention to certain thermodynamic properties of an ideal gas that are analogous to rubber-like deformation.~~
- (J06:0080.21) It To set a lower limit of about [FORMULA] for the activation energy of the abstraction of a chlorine atom from a carbon tetrachloride molecule by a chlorine atom to form [FORMULA] radical appears possible to ~~set a lower limit of about [FORMULA] for the activation energy of the abstraction of a chlorine atom from a carbon tetrachloride molecule by a chlorine atom to form [FORMULA] radical.~~
- (J07:0350.21) It That some of the dust in the vicinity of the Earth originated from meteoritic impacts upon the moon is also possible that ~~some of the dust in the vicinity of the Earth originated from meteoritic impacts upon the moon.~~
- (J08:1040.30) It That an even greater area was covered, particularly downwind is quite likely that ~~an even greater area was covered, particularly downwind.~~
- (J10:0240.03) It That the stamens of the catkin are always arranged in pairs, and that each individual flower is nothing but one such pair standing on a green, black-tipped little scale is not difficult to see that ~~the stamens of the catkin are always arranged in pairs, and that each individual flower is nothing but one such pair standing on a green, black-tipped little scale.~~
- (J12:0570.12) It That many marked and striking differences exist between lungs when an inter-species comparison is made is evident that ~~many marked and striking differences exist between lungs when an inter-species comparison is made.~~
- (J21:0410.42) It That a 1:1 correspondence can be established between the lines of two such pencils, so that in a sense a unique image can actually be assigned to each tangent is interesting that ~~a 1:1 correspondence can be established between the lines of two such pencils, so that in a sense a unique image can actually be assigned to each tangent.~~
- (J22:1000.51) It Not only to encourage effective land reform programs but also to underline the relation of such reforms to the economic growth and modernization of the society should be American policy ~~not only to encourage effective land reform programs but also to underline the relation of such reforms to the economic growth and modernization of the society.~~
- (J24:1300.24) That we are experiencing an upsurge of interest in the many formulations and preventive adaptations of brief treatment in social casework It is evident from even a small sampling of current literature ~~that we are experiencing an upsurge of interest in the many formulations and preventive adaptations of brief treatment in social casework.~~
- (N04:1580.30) It To see the thin ranks of warriors, old and young, wheeling and twisting their ponies frantically from side to side only to be tumbled bleeding from their saddles by the relentless slam, slam of the cruelly efficient Hawkinses was pitiful to ~~see the thin ranks of warriors, old and young, wheeling and twisting their ponies frantically from side to side only to be tumbled bleeding from their saddles by the relentless slam, slam of the cruelly efficient Hawkinses.~~
- (N05:0120.51) It To compose himself took him a long time to ~~compose himself.~~
- (N06:0210.09) It That he wished himself different from the sort of person he thought he was was obvious that ~~he wished himself different from the sort of person he thought he was.~~
- (N07:1740.24) It to fix them cost us a hundred thousand dollars and thirty days lost time to ~~fix them.~~
- (N08:0430.12) It That she would have to enact her renunciation, beg forgiveness was there that ~~she would have to enact her renunciation, beg forgiveness.~~
- (N10:0370.39) It That she probably would never know bothered her that ~~she probably would never know.~~
- (N11:0420.12) It To locate Horn took some time to ~~locate Horn.~~
- (N15:1540.39) It That they hadn't been spotted did not seem possible that ~~they hadn't been spotted.~~
- (N18:0500.06) That she impressed me instantly It was obvious ~~that she impressed me instantly~~; I was aware that when our eyes met we both quickly averted them.

#### *Doubtful*

- (G07:0050.06) It That in this hour both of "national peril" and of "national opportunity" we can take counsel with the men who made the nation is well then that ~~in this hour both of "national peril" and of "national opportunity" we can take counsel with the men who made the nation.~~

- (G08:0180.09) ~~It~~ That despite this strong upsurge in Southern writing, almost none of the writers has forsaken the firmly entrenched concept of the white-suited big-daddy colonel sipping a mint julep as he silently recounts the revenue from the season's cotton and tobacco crops; of the stereotyped Negro servants chanting hymns as they plow the fields; of these and a host of other antiquated legends that deny the South its progressive leaps of the past century is interesting, however, that despite this strong upsurge in Southern writing, almost none of the writers has forsaken the firmly entrenched concept of the white-suited big-daddy colonel sipping a mint julep as he silently recounts the revenue from the season's cotton and tobacco crops; of the stereotyped Negro servants chanting hymns as they plow the fields; of these and a host of other antiquated legends that deny the South its progressive leaps of the past century.
- (G12:1060.15) ~~It~~ That our situation becomes inexplicable is because there is not only darkness but also light that our situation becomes inexplicable.
- (J05:1830.33) ~~It~~ That the requirements imposed by these effects upon any one detergent constituent acting alone are severe is evident that the requirements imposed by these effects upon any one detergent constituent acting alone are severe.

## Appendix D7: Subject Raising Rule

**First run: searching for sentences that have *seem, appear, be certain, be likely* as their main verb, that are followed by a logical subject *that* complement and preceded by a surface subject *it*, or that are followed by an infinitival complement**

Incidence of candidate constructions: 74/6248 (1.2%)  
 Correctly identified sentences: 32/40 (80%)  
 Correctly transformed sentences: 25/40 (62%)

### *Successful*

- (A04:1590.30) ~~The pro-Western government, which the United States had helped in a revolt against the Souvanna Phouma "neutralist" government; It never did appear that the pro-Western government, which the United States had helped in a revolt against the Souvanna Phouma "neutralist" government sparked to spark much fighting spirit in the Royal Lao Army.~~
- (A07:0920.39) ~~The education bill It appears that the education bill is to be temporarily stalled in the Rules Committee, where two Northern Democratic members who usually vote with the Administration are balking because of the religious controversy.~~
- (A08:1180.24) ~~The Gov. Jimmie H. Davis administration It appears that the Gov. Jimmie H. Davis administration faces to face a difficult year in 1961, with the governor's theme of peace and harmony subjected to severe stresses.~~
- (A10:0620.48) ~~Dr. Barnes said that there it seemed that there was to be feeling that evacuation plans, even for a high school where there were lots of cars "might not be realistic and would not work".~~
- (A19:1770.12) ~~Although economic and personal circumstances vary widely among those now choosing apartments, Leo J. Pantas, vice president of a hardware manufacturing company, pointed out recently that many apartment-seekers it seems that many apartment seekers have to have one characteristic in common: a desire for greater convenience and freedom from the problems involved in maintaining a house.~~
- (G01:0520.39) ~~His assumption It seems that his assumption is to be that any such friends, being tolerable humans, must be more liberal than most Southerners and therefore at least partly in sympathy with his views.~~
- (G06:1580.21) ~~Since attack serves to stimulate interest in broadcasts, I added to my opening statement a sentence in which I claimed that German youth it seemed that German youth lacked to lack the enthusiasm which is a necessary ingredient of anger, and might be classified as uninterested and bored rather than angry.~~
- (G11:0370.18) ~~The most effective political inventions It seems that the most effective political inventions make to make maximum use of natural harbors and are aware that restraining breakwaters can play only a minor part in the whole scheme.~~
- (G17:0340.03) ~~They It didn't seem that they were to be able to think of any.~~

- (J01:0250.42) For the case of Jupiter, the radio emission spectrum is definitely not like the spectrum of a black-body radiator, and it the radiation reaching the earth seems very likely to be that the radiation reaching the earth is a combination of thermal radiation from the atmosphere and non-thermal components.
- (J03:1200.09) ~~These droplets~~ It appears that these droplets are to be ellipsoids, and it is mathematically convenient to assume that they are.
- (J04:0240.12) [FORMULA] It appeared that [FORMULA] was to be well suited for the study of these matters, since it is a normal paramagnet, with three unpaired electrons on the chromium, its crystal structure is very simple, and the unknown position of the hydrogen in the strong [FORMULA] bond provides structural interest.
- (J05:1350.39) For anionics, ~~these micelles~~ it appears that these micelles are to be roughly spherical assemblages in which the hydrocarbon tails come together so that the polar groups (the ionized ends) face outward towards the aqueous continuous phase.
- (J06:0110.21) ~~The rate of the gas phase exchange reaction~~ It appears that the rate of the gas phase exchange reaction is to be proportional to the first power of the absorbed light intensity indicating that the radical intermediates are removed at the walls or by reaction with an impurity rather than by bimolecular radical combination reactions.
- (J09:1620.24) ~~There~~ It appeared that there was to be no difference in the distribution of anti-A and anti-B activity in group O serum, though in two group O donors (J. F. and E. M.) only one type of agglutinin was found in the regions of low anionic binding capacity (Figs. 1 and 2).
- (J10:0480.18) ~~Most species~~ It seems that most species prefer to prefer a ready-made hollow such as a deserted mouse nest, a bird house, or the hole made by a woodpecker; some show a definite liking for making their nest in moss.
- (N03:1140.24) ~~The orderly room~~ It seemed that the orderly room was to be deserted.
- (N04:0050.15) ~~The men~~ It seems that the men think to think so.
- (N06:0570.03) ~~The presence of the two exhaust fans~~ It seemed that the presence of the two exhaust fans indicated to indicate that the hall could become crowded for air.
- (N08:1250.39) ~~She~~ It seemed that she had to have come such a long distance — too far for her destination which had wilfully been swallowed up in the greedy gloom of the trees.
- (N09:1490.27) ~~He~~ It seemed that he was to be fighting not one man but a dozen.
- (N10:1460.09) His face took on a sudden pallor, became beaded with sweat, and ~~he~~ it seemed that he had to have trouble with his breathing.
- (N11:0010.06) ~~The author of the anonymous notes~~ It seemed that the author of the anonymous notes was to be all-knowing.
- (N12:0480.18) ~~It~~ Jess was certain now that Jess was to be in the house, but also, presumably, was Stacey Black.
- (N13:1180.03) ~~Even strange names~~ It seemed that even strange names made to make them feel closer to some kind of civilization when stumbled across out here in this wilderness.
- (N15:1190.15) ~~They~~ It appeared that they were to be the enemy.
- (N18:0430.06) You must forgive me if ~~I~~ It seems that I dwell to dwell too much on her physical aspects but I am an artist, accustomed to studying the physical body.

#### *Doubtful*

- (G09:0160.27) He began with masks to make the dancer identify himself with the creature ~~he~~ it appeared that he was to be.
- (N07:0120.12) ~~It~~ The green eyes seemed to Barton that the green eyes mocked to mock him, the thin-lipped smile held insolence, but he had no time to waste now.

#### *Failed*

- (A11:1460.03) Richard M. Forbes's Paget, which had ~~what~~ it seemed that what was to be a substantial lead in the early stages, tired rapidly nearing the wire and was able to save place money only a head in front of Glen T. Hallowell's Milties Miss.
- (A12:0420.12) I've tried to teach some of the other boys to kick and ~~some of them~~ can't it seem that some of them get to get the feel.
- (A13:1010.09) JUST WHEN IF baseball SEEMS baseball might to might be losing its grip on the masses up pops heroics to start millions of tongues to wagging.
- (G04:1350.51) Everything was burnished with sweat and grease so that ~~all of the objects~~ it seemed that all of the objects had to have been carved from the same material and to be ageless.

- (G05:0250.21) It seems that for Persia, and especially for this city, there are only two times: the glorious past and the corrupt, depressing, sterile present.
- (G08:0080.39) Indeed, it Tobacco Road seems ~~that~~ only in today's Southern fiction does ~~Tobacco Road~~, with all the traditional trimmings of sowbelly and cornbread and mint juleps, ~~continue to continue~~ to live — but only as a weary, overexploited phantom.
- (G12:0330.30) He seems, by some unconscious division of labor, to have given them that one function and no other, leaving communication to the rest of the face.
- (J08:1040.30) It An even greater area is quite likely ~~that an even greater area was~~ to be covered, particularly downwind.
- (J12:0210.39) In this instance, because of incomplete septation, the secondary lobule does not constitute in itself ~~what it appears that what is~~ to be a small individual lung as in type I.
- (J17:0130.12) It seems to follow that by and large an antagonism exists between the paleo- and the neocortex as far as emotional reactivity is concerned, and that the balance between the two systems determines the emotional responsiveness of the organism.
- (J22:0230.30) It may seem to some of them that success can be purchased much less dearly by fishing in the murky waters of international politics than by facing up to the intractable tasks at home.
- (N05:0800.21) “He works for my father”, the girl said, and then seemed to change her mind.
- (N14:0160.30) Suddenly ~~it they~~ seemed to him insane ~~that they might to might~~ hope to locate Gyp Carmer so casually, even were he to prove the thief.

**Second run: same but excluding sentences with modals other than HAVE, WILL and SHALL, and sentences with *wh...* subjects and sentences where the predicate has other non-PP arguments, and sentences with co-ordinate structures**

Incidence of candidate constructions: 45/6248 (0.72%)

Correctly identified sentences: 30/30 (100%)

Correctly transformed sentences: 29/30 (97%)

#### *Successful*

- (A04:1590.30) ~~The pro-Western government, which the United States had helped in a revolt against the Souvanna Phouma “neutralist” government,~~ It never did appear that the pro-Western government, which the United States had helped in a revolt against the Souvanna Phouma “neutralist” government sparked to spark much fighting spirit in the Royal Lao Army.
- (A07:0920.39) ~~The education bill~~ It appears that the education bill is ~~to be~~ temporarily stalled in the Rules Committee, where two Northern Democratic members who usually vote with the Administration are balking because of the religious controversy.
- (A08:1180.24) ~~The Gov. Jimmie H. Davis administration~~ It appears that the Gov. Jimmie H. Davis administration faces to face a difficult year in 1961, with the governor's theme of peace and harmony subjected to severe stresses.
- (A10:0620.48) Dr. Barnes said that ~~there it~~ seemed that there was ~~to be~~ feeling that evacuation plans, even for a high school where there were lots of cars “might not be realistic and would not work”.
- (A19:1770.12) Although economic and personal circumstances vary widely among those now choosing apartments, Leo J. Pantas, vice president of a hardware manufacturing company, pointed out recently that ~~many apartment seekers~~ it seems that many apartment seekers have to have one characteristic in common: a desire for greater convenience and freedom from the problems involved in maintaining a house.
- (G01:0520.39) ~~His assumption~~ It seems that his assumption is ~~to be~~ that any such friends, being tolerable humans, must be more liberal than most Southerners and therefore at least partly in sympathy with his views.
- (G06:1580.21) Since attack serves to stimulate interest in broadcasts, I added to my opening statement a sentence in which I claimed that ~~German youth it~~ seemed that German youth lacked to lack the enthusiasm which is a necessary ingredient of anger, and might be classified as uninterested and bored rather than angry.
- (G08:0660.18) Lacking the pioneer spirit necessary to write of a *new* economy, ~~these writers it~~ seems that these writers are ~~to be~~ contenting themselves with an old one that is now as defunct as Confederate money.

- (G11:0370.18) ~~The most effective political inventions~~ It seems that the most effective political inventions make to make maximum use of natural harbors and are aware that restraining breakwaters can play only a minor part in the whole scheme.
- (G12:0330.30) ~~He~~ It seems that, by some unconscious division of labor, ~~he has to have~~ given them that one function and no other, leaving communication to the rest of the face.
- (G17:0340.03) ~~They~~ It didn't seem that they were to be able to think of any.
- (J01:0350.03) The low intensity of the radiation from Saturn has limited observations, but again ~~the measured radiation it seems that the measured radiation is to be~~ consistent with a thermal origin.
- (J04:0240.12) [FORMULA] It appeared that [FORMULA] was to be well suited for the study of these matters, since it is a normal paramagnet, with three unpaired electrons on the chromium, its crystal structure is very simple, and the unknown position of the hydrogen in the strong [FORMULA] bond provides structural interest.
- (J05:1350.39) For anionics, ~~these micelles~~ it appears that these micelles are to be roughly spherical assemblages in which the hydrocarbon tails come together so that the polar groups (the ionized ends) face outward towards the aqueous continuous phase.
- (J06:0110.21) ~~The rate of the gas phase exchange reaction~~ It appears that the rate of the gas phase exchange reaction is to be proportional to the first power of the absorbed light intensity indicating that the radical intermediates are removed at the walls or by reaction with an impurity rather than by bimolecular radical combination reactions.
- (J09:1620.24) ~~There~~ It appeared that there was to be no difference in the distribution of anti-A and anti-B activity in group O serum, though in two group O donors (J. F. and E. M.) only one type of agglutinin was found in the regions of low anionic binding capacity (Figs. 1 and 2).
- (J10:0480.18) ~~Most species~~ It seems that most species prefer to prefer a ready-made hollow such as a deserted mouse nest, a bird house, or the hole made by a woodpecker; some show a definite liking for making their nest in moss.
- (J12:0810.30) Anatomically, ~~the horse lung~~ it appears that the horse lung is to be remarkably like that of man, insofar as this can be ascertained from comparison of our findings in the horse with those of others (Birnbaum,'54) in the human being.
- (J17:1000.42) The Mecholyl and noradrenalin tests applied with certain precautions are reliable indicators of this central autonomic balance, but for the sake of correlating autonomic and clinical states, and of studying the effect of certain therapeutic procedures on central autonomic reactions, ~~additional tests~~ it seems that additional tests are to be desirable.
- (J22:0440.51) ~~The society~~ It is likely that the society is to be characterized by having a fairly modernized urban sector and a relatively untouched rural sector, with very poor communications between the two.
- (N03:1140.24) ~~The orderly room~~ It seemed that the orderly room was to be deserted.
- (N04:0050.15) ~~The men~~ It seems that the men think to think so.
- (N06:0570.03) ~~The presence of the two exhaust fans~~ It seemed that the presence of the two exhaust fans indicated to indicate that the hall could become crowded for air.
- (N08:1250.39) ~~She~~ It seemed that she had to have come such a long distance — too far for her destination which had wilfully been swallowed up in the greedy gloom of the trees.
- (N09:1490.27) ~~He~~ It seemed that he was to be fighting not one man but a dozen.
- (N10:1460.09) His face took on a sudden pallor, became beaded with sweat, and ~~he~~ it seemed that he had to have trouble with his breathing.
- (N11:0010.06) ~~The author of the anonymous notes~~ It seemed that the author of the anonymous notes was to be all-knowing.
- (N13:1180.03) ~~Even strange names~~ It seemed that even strange names made to make them feel closer to some kind of civilization when stumbled across out here in this wilderness.
- (N15:1190.15) ~~They~~ It appeared that they were to be the enemy.

### *Doubtful*

- (G09:0160.27) He began with masks to make the dancer identify himself with the creature ~~he~~ it appeared that he was to be.

## Appendix D8: Object Raising Rule

**First run: searching for sentences that have *be easy, difficult, tough, impossible* or *hard* as their main verb, with *it* as surface subject and are followed by an infinitival complement containing an bare or PP embedded object phrase, or followed by an object gapped infinitival complement and preceded by the gapped logical object in surface subject position**

Incidence of candidate constructions: 8/6248 (0.13%)

Correctly identified sentences: 8/8 (100%)

Correctly transformed sentences: 7/8 (87%)

### *Successful*

- (A03:0120.30) It ~~A fair trial~~ is impossible to get ~~a fair trial~~ when some of the defendants made statements involving themselves and others.
- (G08:1080.51) Writers openly admit that ~~the Negro~~ it is easier to write ~~the Negro~~ than the white man; but they obviously mean by this, not a Negro personality, but a Negro type.
- (G11:0530.18) ~~It is~~ The primeval fears of man are difficult to reconstruct ~~the primeval fears of man~~.
- (G12:0740.21) Plays ~~more highly formalized than "Waiting for Godot", "Endgame", and "Krapp's Last Tape"~~ It would be hard to find plays more highly formalized than "Waiting for Godot", "Endgame", and "Krapp's Last Tape".
- (J04:0820.09) Hence ~~it a packing of the atoms in this material in which the oxygen atoms are far from geometrical equivalence~~ is difficult to conceive of ~~a packing of the atoms in this material in which the oxygen atoms are far from geometrical equivalence~~.
- (J05:1890.09) In the over-all process, it a "pure" role is difficult to assign a "pure" role to each constituent of a built-detergent formulation; and, indeed, there is no more reason to separate the interrelated roles of the active, builder, antiredeposition agent, etc. than there is to assign individual actions to each of the numerous isomers making up a given commercial organic active.
- (N09:0380.42) She studied him hopefully, yearningly; against the limitless background of sky and wasteland ~~it her analysis~~ was easy to confirm ~~her analysis~~.

### *Failed*

- (J12:0780.24) In addition, ~~it chronic generalized emphysema occurring in a cow~~ would be difficult to imagine ~~chronic generalized emphysema occurring in a cow~~, considering its marked lobular development but, conversely, not difficult to imagine this occurring in the horse or the dog.

## Appendix D9: *There* Insertion and Deletion Rule

**First run: searching for a BE sentence, with existential *there* in the subject position, and followed by its semantic subject, and a past participle or present participle complement, OR searching for a *be* sentence whose subject is a common noun phrase carrying either *an* or no determiner and whose complement is a present participle or past participle clause**

Incidence of candidate constructions: 32/6248 (0.51%)

Correctly identified sentences: 32/32 (100%)

Correctly transformed sentences: 24/32 (75%)

### *Successful*

- (A01:1420.48) Pelham said Sunday night ~~there~~ research was ~~research~~ being done on whether the "quickie" vote on the increase can be repealed outright or whether notice would have to first be given that reconsideration of the action would be sought.

- (A06:0180.21) Sunday he had added, “We can love Eisenhower the man, even if we considered him a mediocre president ... but ~~there nothing is nothing~~ left of the Republican Party without his leadership”.
- (A06:1710.15) He said “Morris County is rapidly changing and unless steps are taken to preserve the green areas, ~~there no land~~ will be ~~no land~~ left to preserve”.
- (A06:1870.12) ~~A special presentation~~ There was a special presentation made to Mrs. Geraldine Thompson of Red Bank, who is stepping down after 35 years on the committee.
- (A07:0490.24) ~~A wave of public resentment against corruption in government~~ There is a wave of public resentment against corruption in government rising in Massachusetts.
- (A14:0520.12) ~~A crowd of 1,400~~ There is a crowd of 1,400 expected for the ceremonies, which will be followed by the show in which the writers will lampoon baseball personalities in skit, dance and song.
- (G04:0380.12) They are huge areas which have been swept by winds for so many centuries that ~~there no soil is no soil~~ left, but only deep bare ridges fifty or sixty yards apart with ravines between them thirty or forty feet deep and the only thing that moves is a scuttling layer of sand.
- (G04:0860.15) It took me a moment to realize what was odd about that panel: ~~there a gimbaled compass~~ was a gimbaled compass welded to it, which rocked gently back and forth as the Land Rover bounced about.
- (G08:0310.27) ~~There A New South~~ is a New South emerging, a South losing the folksy traditions of an agrarian society with the rapidity of an avalanche — especially within recent decades.
- (G08:0970.15) ~~A new order~~ There is a new order thrusting itself into being.
- (G08:0980.03) ~~A new South~~ There is a new South emerging after the post-bellum years of hesitation, uncertainty, and lack of action from the Negro in defining his new role in the amorphously defined socio-political organizations of the white man.
- (G08:1060.39) While ~~there many Faulknerian Lucas Beauchamps~~ may still be many Faulknerian Lucas Beauchamps scattered through the rural South, such men appear to be a vanishing breed.
- (G11:1110.33) If ~~an automobile~~ there were an automobile approaching him, he would know what was required of him, even though he might not be able to act quickly enough.
- (G22:1310.30) ~~Leadership~~ There is leadership lacking in our society because it has no legitimate place to develop.
- (J01:0290.33) Of the remaining planets, only Mars and Saturn have been observed as radio sources, and ~~not very much information~~ there is not very much information available.
- (J03:0190.33) Although ~~there only four fluids~~ were only four fluids tested, it was apparent that there were two distinct types.
- (J03:1470.18) ~~There No data~~ are no data published in the literature on the shape of low viscosity drops to confirm the above formulas. [tense]
- (J03:1700.12) This conformational entropy is, in this case, equal to the usual entropy, for ~~there no other changes or other energies~~ are no other changes or other energies involved.
- (J12:0390.51) ~~Some of these~~ There were obviously some of these filling from interlobular branches of the bronchial arteries while ~~others there were others~~ filling from direct hilar branches following along the pleural surface.
- (N02:0300.39) Now ~~there nothin's nothin~~ left of me.
- (N03:1100.30) ~~A man~~ There was a man standing in the open door of the lighted orderly room a few yards to Mike's left, but he, too, suddenly made up his mind and went racing to join the confused activity at the east end of the stockade.
- (N03:1380.21) ~~Guerrillas~~ There were guerrillas racing toward him.
- (N04:1430.27) ~~Other Indians~~ There were other Indians running at the ponies, shrilling and waving blankets.
- (N04:1510.30) Above me ~~a dark rider~~ there was a dark rider whipping his pony with a quirt in an attempt to hurdle the bales.

#### *Doubtful*

- (A06:1800.15) ~~Fifteen members of the Republican State Committee who are retiring — this year~~ There were fifteen members of the Republican State Committee who are retiring — voluntarily — this year honored yesterday by their colleagues.



### *Failed*

- (A04:1620.42) There Not any more energy certainly was ~~not any more energy~~ displayed after it was clear the United States would not back the pro-Western government to the hilt.
- (A13:1010.09) JUST WHEN IT SEEMS baseball ~~there~~ might be baseball losing its grip on the masses up pops heroics to start millions of tongues to wagging.
- (J04:0210.33) It is important to consider and experimentally verify this influence since ~~quantitative nuclear resonance~~ there is quantitative nuclear resonance becoming increasingly used in investigations of structure.
- (J12:1680.03) ~~Lobularity~~ There is extremely well lobularity developed in type I; absent in type II; imperfectly developed in type III.
- (J22:0410.18) The men in power are committed in principle to modernization, but ~~economic and social changes~~ there are economic and social changes proceeding only erratically.
- (N06:0180.12) ~~Baldness~~ There was baldness attacking his pate.
- (N11:1350.39) ~~Jury, judge and executioner~~ There were jury, judge and executioner riding the range in the form of a single unknown figure that could materialize anywhere, at any time, to dispense an ancient brand of justice the men of the new West had believed long outdated.

### **Second run: same but with an indefinite determiner**

Incidence of candidate constructions: 24/6248 (0.38%)

Correctly identified sentences: 24/24 (100%)

Correctly transformed sentences: 24/24 (100%)

### *Successful*

- (A01:1420.48) Pelham said Sunday night ~~there~~ research was ~~research~~ being done on whether the “quickie” vote on the increase can be repealed outright or whether notice would have to first be given that reconsideration of the action would be sought.
- (A04:1620.42) ~~There~~ Certainly not any more energy was ~~not any more energy~~ displayed after it was clear the United States would not back the pro-Western government to the hilt.
- (A06:0180.21) Sunday he had added, “We can love Eisenhower the man, even if we considered him a mediocre president ... but ~~there~~ nothing is ~~nothing~~ left of the Republican Party without his leadership”.
- (A06:1710.15) He said “Morris County is rapidly changing and unless steps are taken to preserve the green areas, ~~there no land~~ will be ~~no land~~ left to preserve”.
- (A06:1870.12) ~~A special presentation~~ There was a special presentation made to Mrs. Geraldine Thompson of Red Bank, who is stepping down after 35 years on the committee.
- (A07:0490.24) ~~A wave of public resentment against corruption in government~~ There is a wave of public resentment against corruption in government rising in Massachusetts.
- (A08:0440.30) ~~Many spectators~~ There will be many spectators occupying seats and vantage points bordering Lafayette Square, opposite the White House.
- (A14:0520.12) ~~A crowd of 1,400~~ There is a crowd of 1,400 expected for the ceremonies, which will be followed by the show in which the writers will lampoon baseball personalities in skit, dance and song.
- (G04:0380.12) They are huge areas which have been swept by winds for so many centuries that ~~there no soil~~ is ~~no soil~~ left, but only deep bare ridges fifty or sixty yards apart with ravines between them thirty or forty feet deep and the only thing that moves is a scuttling layer of sand.
- (G04:0860.15) It took me a moment to realize what was odd about that panel: ~~there a gimbaled compass~~ was a gimbaled compass welded to it, which rocked gently back and forth as the Land Rover bounced about.
- (G08:0310.27) ~~There~~ A New South is a New South emerging, a South losing the folksy traditions of an agrarian society with the rapidity of an avalanche — especially within recent decades.
- (G08:0970.15) ~~A new order~~ There is a new order thrusting itself into being.
- (G08:0980.03) ~~A new South~~ There is a new South emerging after the post-bellum years of hesitation, uncertainty, and lack of action from the Negro in defining his new role in the amorphously defined socio-political organizations of the white man.

- (G08:1060.39) While ~~there~~ many Faulknerian Lucas Beauchamps may still be ~~many~~ Faulknerian Lucas Beauchamps scattered through the rural South, such men appear to be a vanishing breed.
- (G11:1110.33) If ~~an automobile~~ there were ~~an automobile~~ approaching him, he would know what was required of him, even though he might not be able to act quickly enough.
- (J01:0290.33) Of the remaining planets, only Mars and Saturn have been observed as radio sources, and ~~not very much information~~ there is not very much information available.
- (J03:0190.33) Although ~~there~~ only four fluids were ~~only four fluids~~ tested, it was apparent that there were two distinct types.
- (J03:1470.18) ~~There~~ No data are ~~no data~~ published in the literature on the shape of low viscosity drops to confirm the above formulas. [tense]
- (J03:1700.12) This conformational entropy is, in this case, equal to the usual entropy, for ~~there~~ no other changes or other energies are ~~no other changes or other energies~~ involved.
- (J12:0390.51) ~~Some of these~~ There were obviously ~~some of these~~ filling from interlobular branches of the bronchial arteries while others were filling from direct hilar branches following along the pleural surface.
- (N02:0300.39) Now ~~there~~ nothin's ~~nothin~~ left of me.
- (N03:1100.30) ~~A man~~ There was ~~a man~~ standing in the open door of the lighted orderly room a few yards to Mike's left, but he, too, suddenly made up his mind and went racing to join the confused activity at the east end of the stockade.
- (N04:1510.30) Above me ~~a dark rider~~ there was ~~a dark rider~~ whipping his pony with a quirt in an attempt to hurdle the bales.
- (N11:0590.30) ~~A good many beef-hungry settlers~~ There were ~~a good many beef-hungry settlers~~ accepting the death of William Lewis as proof that the warning notes were not idle threats.

## Appendix D10: Conjunction Movement Rule

**First run: searching for *and/or/nor* co-ordinate structures with pairs of balanced noun, prepositional, adjectival or adverbial phrases, that do not contain any verbal, temporal or numerical constituents**

Incidence of candidate constructions: 498/6248 (8.0%)

Correctly identified sentences: 61/61 (100%)

Correctly transformed sentences: 49/61 (80%)

### *Successful*

- (A01:0030.21) The jury further said in term-end presentments that the City Executive Committee, which had over-all charge of the election, “deserves the thanks and praise ~~and thanks~~ of the City of Atlanta” for the manner in which the election was conducted.
- (A04:0120.03) The nightmare of a clash between those in trouble in Africa, exacerbated by the difficulties, changes, and tragedies facing them, and other allies who emotioinally and intellectually ~~and emotioinally~~ disapprove of the circumstances that have brought these troubles about, has been conspicuous by its absence.
- (A05:0780.51) The Central Falls City Council expressed concern especially that more foods be placed on the eligible list and that neighborhood variety and grocery ~~and variety~~ stores be allowed to do business on Sunday.
- (A06:0320.24) The audience last night did not respond with either boos or applause ~~or boos~~ to mention of Hughes' remark.
- (A07:0140.03) Some Democratic county and district ~~and county~~ leaders are reported trying to induce State Controller Arthur Levitt of Brooklyn to oppose Mr. Wagner for the Mayoral nomination in the Sept. 7 Democratic primary.
- (A08:0260.30) Within view of the avenue are some of the United States government's tremendous buildings, plus monuments and shrines ~~and monuments~~.
- (A09:0150.30) The signature of Harold V. Varani, former director of engineering and architecture ~~and engineering~~ in the Department of Public Property, appeared on payment vouchers certifying work on the project.

- (A10:0170.06) "We must persuade them to enjoy a way of life which, if not identical, is congenial with ours", he said but adding that if they do not develop the kind of society they themselves want it will lack loyalty and rituality and ~~loyalty~~.
- (A11:1180.21) Baltimore's bulky spring-training contingent now gradually will be reduced as Manager Paul Richards and his coaches seek to trim it down to a more workable and streamlined and ~~workable~~ unit.
- (A13:0250.18) The Bears added their last run in the sixth on Alusik's outfield and double and ~~outfield-flies~~ by Porter and Wert.
- (A14:0330.06) Skorich and Shaw and ~~Skorich~~ headed a group of players, coaches and team officials who received an engrossed copy of an official city citation and a pair of silver cufflinks shaped like a football.
- (A20:1210.39) "When they became members of the city police narcotics unit", Sokol said, "they were told they would have to get to know certain areas of Chicago in which narcotics were sold and they would have to get to know people in the narcotics racket. They, on occasion, posed as peddlers and addicts and ~~peddlers~~".
- (G02:0030.03) By prevailing over other claimants for the loyalties of men, the nation-state maintained an adequate measure of order and certainty and ~~order~~ within its territorial borders.
- (G03:0520.24) One arm represented our missiles and bombers and ~~missiles~~, the other arm "theirs".
- (G04:0180.42) It has nothing of the proud stride of the trained runner about it, it is not a lope, it is not done with verve or style or ~~verve~~.
- (G05:0020.30) The architects, the carpet and tile and ~~carpet~~-makers, the potters, painters, calligraphers, and metalsmiths worked through Abbas's reign and those of his successors to enrich the city.
- (G06:0010.03) *Die Frist ist um, und wiederum verstrichen sind sieben Jahr*, the Maestro quoted *The Flying Dutchman*, as he told of his wanderings and career and ~~wanderings~~, explaining that the number seven had significantly recurred in his life several times.
- (G07:0210.15) Their writings assume more than patriotic or dramatic or ~~patriotic~~ interest because of their conviction that the struggle in which they were involved was neither selfish nor parochial but, rather, as Washington in his last wartime circular reminded his fellow countrymen, that "with our fate will the destiny of unborn millions be involved".
- (G08:0180.09) It is interesting, however, that despite this strong upsurge in Southern writing, almost none of the writers has forsaken the firmly entrenched concept of the white-suited big-daddy colonel sipping a mint julep as he silently recounts the revenue from the season's tobacco and cotton and ~~tobacco~~-crops; of the stereotyped Negro servants chanting hymns as they plow the fields; of these and a host of other antiquated legends that deny the South its progressive leaps of the past century.
- (G10:0030.03) The truth in their conflicting concepts was expounded by statesmen of the calibre of Calhoun and Webster and ~~Calhoun~~, and defended in the end by leaders of the nobility of Lee and Lincoln and ~~Lee~~.
- (G11:0620.27) They are in general those fears that once seemed to have been amenable to ritual and prayer or ~~ritual~~.
- (G12:0290.30) One might say it combines the man's own humility and pride and ~~humility~~.
- (G13:0160.27) He has elevated sex — not libido or Eros or ~~libido~~ but pure, spontaneous, uninhibited sex — to the rank of the godhead; it is Astarte, Ishtar, Venus, Yahwe, Dionysus, Christ, the mysterious and divine orgone energy flowing through the body of the universe.
- (G17:0520.39) To guard against the tyranny of a numerical majority, Calhoun developed his theory of "concurrent majority", which, he said, "by giving to each portion of the community which may be unequally affected by the action of government, a negative on the others, prevents all local or partial or ~~local~~ legislation".
- (G18:0020.39) Of these there are surely few that would be more rewarding discoveries than Verner von Heidenstam, the Swedish novelist and poet and ~~novelist~~ who received the award in 1916 and whose centennial was celebrated two years ago.
- (G22:0160.12) Hence the prime issue, as I see it, is whether a free or democratic or ~~free~~ society can master technology for the benefit of mankind, or whether technology will rule and develop its own society compatible with *its* own needs as a force of nature.
- (J01:0010.15) IT HAS recently become practical to use the radio emission of the planets and moon and ~~planets~~ as a new source of information about these bodies and their atmospheres.

- (J02:0760.15) The anode in figure 2 was mounted by means of the anode holder which was attached to a steel disk and plug ~~and disk~~.
- (J03:0020.30) There is a small well in the top in which the paste or fluid ~~or paste~~ to be tested is placed.
- (J05:0160.30) A well-publicized entrant which has achieved success only recently is the built liquid detergent, with which the major problem today is incorporation of active and builder ~~and active~~ into a small volume using a sufficiently high builder/active ratio.
- (J06:0190.12) It would appear that it should be possible to determine unique mechanisms for the photochemical and thermal ~~and photochemical~~ reactions in both the liquid and gas phases and to determine values for activation energies of some of the intermediate reactions of atoms and free radicals, as well as information on the heat of dissociation of the carbon-halogen bond.
- (J10:0010.06) Apart from the honeybee, practically all bumblebees and bees ~~and bumblebees~~ hibernate in a state of torpor.
- (J12:0300.27) The bronchial artery in its distribution and course ~~and distribution~~ differs somewhat from that found in other mammals.
- (J17:0210.33) MacLean stressed correctly the importance of the visceral brain for preservation of the species and the individual ~~and the species~~, as evidenced by the influence of the limbic brain (including the hypothalamus) on emotions related to fight and flight and also on sexual functions.
- (J22:0010.39) As the historic processes of modernization gradually gain momentum, their cohesion will be threatened by divisive forces, the gaps between subjects and rulers ~~and subjects~~, country and town ~~and country~~, will widen; new aspirants for power will emerge whose ambitions far exceed their competence; old rulers may lose their nerve and their sense of direction.
- (J23:0080.15) Nevertheless, for most of the population of heterogeneous advanced societies, though less for the less religious portion, religion does perform certain modal social and individual ~~and social~~ functions.
- (N01:0160.03) Because the summer was unusually hot and dry ~~and hot~~, the spring produced a smaller stream than in ordinary years.
- (N02:0690.18) It was nice then, so quiet and peaceful ~~and quiet~~.
- (N04:0610.27) Montero had set up a strong position, using every box and bale ~~and box~~ we had in addition to barricades of logs and brush.
- (N05:0200.12) She glanced around the clearing, taking in the wagon and the load of trappings and supplies ~~and trappings~~ scattered over the ground, the two kids, the whiteface bull that was chewing its cud just within the far reaches of the firelight.
- (N06:0350.09) Noticing my disappointment he attempted to salvage what shreds and scraps ~~and shreds~~ of authority he felt might still be clinging to his person.
- (N07:0380.03) He had been worried that with Rankin and Miller ~~and Rankin~~ added to the escape party they would be short.
- (N08:0230.33) All of her movements were methodical and careful ~~and methodical~~, partaking of the stealth of a criminal who has plotted his felony for months in advance and knows exactly which step to take next in the course of the final execution of his crime.
- (N09:0340.15) Never, he'd once told Joyce, had he encountered any situation or man ~~or situation~~ that called for a gun.
- (N11:0120.12) There had been no sign of a rifleman and no trace or track ~~or trace~~ to show that anyone had been near.
- (N13:0860.24) Rod shifted his eager eyes from the milling group out in the circle long enough to reply, "I ain't much of a hand for Farmer-in-the-Dell and Dare-Base ~~and Farmer-in-the-Dell~~, but I'd sure like to get in on the handhold and wrestles".
- (N14:0450.15) While five minutes ago the place had presented a scene of easy revelry, with Gyp Carmer a prominent figure, it was now as dull and somnolent ~~and dull~~ as the day before payday.
- (N15:0290.15) The expression was his trade-mark, his open sesame to good luck, and his prayer that plane and pilot ~~and plane~~ would always return.
- (N18:0110.51) The rest of the time I devoted to painting or to those other activities a healthy and young ~~and healthy~~ man just out of college finds interesting.

### *Doubtful*

- (A03:0200.24) President Kennedy today pushed aside other White House business to devote all his attention and time ~~and-attention~~ to working on the Berlin crisis address he will deliver tomorrow night to the American people over nationwide radio and television ~~and-radio~~.
- (G01:0120.03) It is these other differences between South and North ~~and-South~~— other, that is, than those which concern discrimination or social welfare — which I chiefly discuss herein.
- (J04:0950.18) A randomization of “downs” and “ups” ~~and-“downs”~~ is more likely than “downs” and “ups” ~~and-“downs”~~ in position (3) since the hydrogen atoms are well separated and so the position of one could hardly affect the position of another, and also since ordered “down” and “up” ~~and-“down”~~ implies a larger unit cell, for which no evidence exists.
- (J08:0030.27) It is the exploitation of the inherent potential of infectious disease agents by scientific development and research ~~and-development~~, resulting in the production of BW weapons systems.
- (J09:0010.06) A variety of techniques have been directed toward the study and isolation ~~and-study~~ of blood group antibodies.
- (J24:0290.15) In a joint interview Mrs. and Mr. ~~and-Mrs.~~ B. were helped to understand the meaning of a younger son's wandering away from home in terms of his feelings of displacement in reaction to the arrival of the twins.
- (N10:0250.39) He'd come alone, without his child and wife ~~and-child~~.

### *Failed*

- (A02:0030.21) Daniel personally led the fight for the measure, which he had watered down considerably since its rejection by two previous Legislatures, in a public hearing before the House Committee on Taxation and Revenue ~~and-Taxation~~.
- (A19:0010.06) The Ohio and Baltimore ~~and-Ohio~~-Railroad announced yesterday it would reduce the total amount of its payroll by 10 per cent through salary cuts and lay-offs effective at 12.01 A.M. next Saturday.
- (G09:0060.48) He has designed a matching costumes and backdrop ~~and-costumes~~ of points of color on white for Mr. Cunningham's *Summerspace* , so that background and dancers ~~and-background~~ merge into a shimmering unity.
- (N03:0020.03) He had been carrying an Enfield rifle and a holstered navy ball-and-cap ~~cap-and-ball~~ pistol.
- (N12:0700.51) I and you ~~and-I~~ have a little talking to do.

### **Second run: same, BUT excluding arguments that include proper nouns, personal pronouns and capitalised letters**

Incidence of candidate constructions: 423/6248 (6.8%)  
Correctly identified sentences: 60/60 (100%)  
Correctly transformed sentences: 52/60 (87%)

### *Successful*

- (A01:0030.21) The jury further said in term-end presentments that the City Executive Committee, which had over-all charge of the election, “deserves the thanks and praise ~~and-thanks~~ of the City of Atlanta” for the manner in which the election was conducted.
- (A04:0120.03) The nightmare of a clash between those in trouble in Africa, exacerbated by the difficulties, changes, and tragedies facing them, and other allies who emotioinally and intellectually ~~and-emotioinally~~ disapprove of the circumstances that have brought these troubles about, has been conspicuous by its absence.
- (A05:0780.51) The Central Falls City Council expressed concern especially that more foods be placed on the eligible list and that neighborhood variety and grocery ~~and-variety~~-stores be allowed to do business on Sunday.
- (A06:0320.24) The audience last night did not respond with either boos or applause ~~or-boos~~ to mention of Hughes' remark.
- (A07:0140.03) Some Democratic county and district ~~and-county~~-leaders are reported trying to induce State Controller Arthur Levitt of Brooklyn to oppose Mr. Wagner for the Mayoral nomination in the Sept. 7 Democratic primary.

- (A08:0260.30) Within view of the avenue are some of the United States government's tremendous buildings, plus monuments and shrines ~~and monuments~~.
- (A09:0150.30) The signature of Harold V. Varani, former director of engineering and architecture ~~and engineering~~ in the Department of Public Property, appeared on payment vouchers certifying work on the project.
- (A10:0170.06) "We must persuade them to enjoy a way of life which, if not identical, is congenial with ours", he said but adding that if they do not develop the kind of society they themselves want it will lack loyalty and ritual ~~and loyalty~~.
- (A11:1180.21) Baltimore's bulky spring-training contingent now gradually will be reduced as Manager Paul Richards and his coaches seek to trim it down to a more workable and streamlined ~~and workable~~ unit.
- (A13:0250.18) The Bears added their last run in the sixth on Alusik's outfield and double ~~and outfield~~ flies by Porter and Wert.
- (A14:0360.06) With the announcement of a "special achievement award" to William A. (Bill) Shea, the awards list was completed yesterday for Sunday night's thirty-eighth annual show and dinner ~~and show~~ of the New York Chapter, Baseball Writers' Association of America, at the Waldorf-Astoria Hotel.
- (A19:0090.54) There are about 3,325 employess and officers ~~and employees~~ in this class.
- (A20:1210.39) "When they became members of the city police narcotics unit", Sokol said, "they were told they would have to get to know certain areas of Chicago in which narcotics were sold and they would have to get to know people in the narcotics racket. They, on occasion, posed as peddlers and addicts ~~and peddlers~~".
- (G01:0380.30) As for states' rights, they have never counted in the thinking of my liberal friends except as irritations of a immoral and minor ~~and immoral~~ nature which exist now only as anachronisms.
- (G02:0030.03) By prevailing over other claimants for the loyalties of men, the nation-state maintained an adequate measure of order and certainty ~~and order~~ within its territorial borders.
- (G03:0520.24) One arm represented our missles and bombers ~~and missiles~~, the other arm "theirs".
- (G04:0180.42) It has nothing of the proud stride of the trained runner about it, it is not a lope, it is not done with verve or style ~~or verve~~.
- (G05:0020.30) The architects, the carpet and tile ~~and carpet~~ makers, the potters, painters, calligraphers, and metalsmiths worked through Abbas's reign and those of his successors to enrich the city.
- (G06:0010.03) *Die Frist ist um, und wiederum verstrichen sind sieben Jahr*, the Maestro quoted *The Flying Dutchman*, as he told of his wanderings and career ~~and wanderings~~, explaining that the number seven had significantly recurred in his life several times.
- (G07:0210.15) Their writings assume more than patriotic or dramatic ~~or patriotic~~ interest because of their conviction that the struggle in which they were involved was neither selfish nor parochial but, rather, as Washington in his last wartime circular reminded his fellow countrymen, that "with our fate will the destiny of unborn millions be involved".
- (G08:0180.09) It is interesting, however, that despite this strong upsurge in Southern writing, almost none of the writers has forsaken the firmly entrenched concept of the white-suited big-daddy colonel sipping a mint julep as he silently recounts the revenue from the season's tobacco and cotton ~~and tobacco~~ crops; of the stereotyped Negro servants chanting hymns as they plow the fields; of these and a host of other antiquated legends that deny the South its progressive leaps of the past century.
- (G10:0890.21) "We the people of the Confederate States, each state acting in its independent and sovereign ~~and independent~~ character, in order to form a permanent federal government, establish justice, insure domestic tranquility, and secure the blessings of liberty to ourselves and our posterity — invoking the favor and the guidance of Almighty God — do ordain and establish this Constitution for the Confederate States of America".
- (G11:0620.27) They are in general those fears that once seemed to have been amenable to ritual and prayer ~~or ritual~~.
- (G12:0290.30) One might say it combines the man's own humility and pride ~~and humility~~.
- (G13:0160.27) He has elevated sex — not libido or Eros ~~or libido~~ but pure, spontaneous, uninhibited sex — to the rank of the godhead; it is Astarte, Ishtar, Venus, Yahwe, Dionysus, Christ, the mysterious and divine orgone energy flowing through the body of the universe.

- (G17:0520.39) To guard against the tyranny of a numerical majority, Calhoun developed his theory of "concurrent majority", which, he said, "by giving to each portion of the community which may be unequally affected by the action of government, a negative on the others, prevents all local or partial or local legislation".
- (G18:0020.39) Of these there are surely few that would be more rewarding discoveries than Verner von Heidenstam, the Swedish novelist and poet ~~and novelist~~ who received the award in 1916 and whose centennial was celebrated two years ago.
- (G22:0160.12) Hence the prime issue, as I see it, is whether a free or democratic ~~or free~~ society can master technology for the benefit of mankind, or whether technology will rule and develop its own society compatible with *its* own needs as a force of nature.
- (J01:0010.15) IT HAS recently become practical to use the radio emission of the planets and moon ~~and planets~~ as a new source of information about these bodies and their atmospheres.
- (J02:0760.15) The anode in figure 2 was mounted by means of the anode holder which was attached to a steel disk and plug ~~and disk~~.
- (J03:0020.30) There is a small well in the top in which the paste or fluid ~~or paste~~ to be tested is placed.
- (J05:0160.30) A well-publicized entrant which has achieved success only recently is the built liquid detergent, with which the major problem today is incorporation of active and builder ~~and active~~ into a small volume using a sufficiently high builder/active ratio.
- (J06:0190.12) It would appear that it should be possible to determine unique mechanisms for the photochemical and thermal ~~and photochemical~~ reactions in both the liquid and gas phases and to determine values for activation energies of some of the intermediate reactions of atoms and free radicals, as well as information on the heat of dissociation of the carbon-halogen bond.
- (J10:0010.06) Apart from the honeybee, practically all bumblebees and bees ~~and bumblebees~~ hibernate in a state of torpor.
- (J12:0300.27) The bronchial artery in its distribution and course ~~and distribution~~ differs somewhat from that found in other mammals.
- (J17:0210.33) MacLean stressed correctly the importance of the visceral brain for preservation of the species and the individual ~~and the species~~, as evidenced by the influence of the limbic brain (including the hypothalamus) on emotions related to fight and flight and also on sexual functions.
- (J22:0010.39) As the historic processes of modernization gradually gain momentum, their cohesion will be threatened by divisive forces, the gaps between subjects and rulers ~~and subjects~~, country and town ~~and country~~, will widen; new aspirants for power will emerge whose ambitions far exceed their competence; old rulers may lose their nerve and their sense of direction.
- (J23:0080.15) Nevertheless, for most of the population of heterogeneous advanced societies, though less for the less religious portion, religion does perform certain modal social and individual ~~and social~~ functions.
- (J24:0580.21) To establish an emotionally meaningful relationship the worker must demonstrate potential or actual ~~or potential~~ helpfulness immediately, preferably within the first interview, by meeting the client's specific needs.
- (N01:0160.03) Because the summer was unusually hot and dry ~~and hot~~, the spring produced a smaller stream than in ordinary years.
- (N02:0690.18) It was nice then, so quiet and peaceful ~~and quiet~~.
- (N04:0610.27) Montero had set up a strong position, using every box and bale ~~and box~~ we had in addition to barricades of logs and brush.
- (N05:0200.12) She glanced around the clearing, taking in the wagon and the load of trappings and supplies ~~and trappings~~ scattered over the ground, the two kids, the whiteface bull that was chewing its cud just within the far reaches of the firelight.
- (N06:0350.09) Noticing my disappointment he attempted to salvage what shreds and scraps ~~and shreds~~ of authority he felt might still be clinging to his person.
- (N07:1400.09) It was partially cemented by pressure and ages ~~and pressure~~, yet it crumpled before the onslaught of the powerful streams, the force of a thousand fire hoses, and with the gold it held washed down through the long sluices.
- (N08:0230.33) All of her movements were methodical and careful ~~and methodical~~, partaking of the stealth of a criminal who has plotted his felony for months in advance and knows exactly which step to take next in the course of the final execution of his crime.

- (N09:0340.15) Never, he'd once told Joyce, had he encountered any situation or man or situation that called for a gun.
- (N11:0120.12) There had been no sign of a rifleman and no trace or track or trace to show that anyone had been near.
- (N12:1190.33) With a roar of fury and pain and fury Jess made his attack.
- (N13:0860.24) Rod shifted his eager eyes from the milling group out in the circle long enough to reply, "I ain't much of a hand for Farmer-in-the-Dell and Dare-Base and Farmer-in-the-Dell, but I'd sure like to get in on the handhold and wrestles".
- (N14:0450.15) While five minutes ago the place had presented a scene of easy revelry, with Gyp Carmer a prominent figure, it was now as dull and somnolent ~~and dull~~ as the day before payday.
- (N15:0290.15) The expression was his trade-mark, his open sesame to good luck, and his prayer that plane and pilot and plane would always return.
- (N18:0110.51) The rest of the time I devoted to painting or to those other activities a healthy and young ~~and healthy~~ man just out of college finds interesting.

#### *Doubtful*

- (A02:0860.33) THE OTHER BILL , by Sen. A. M. Aikin Jr. of Paris, would relieve real estate brokers, who pay their own annual licensing fee, from the \$12 annual occupation license on brokers in such as bonds and stocks ~~and bonds~~.
- (A03:0200.24) President Kennedy today pushed aside other White House business to devote all his attention and time and attention to working on the Berlin crisis address he will deliver tomorrow night to the American people over nationwide radio and television and radio.
- (J04:0950.18) A randomization of "downs" and "ups" and "~~downs~~" is more likely than "downs" and "ups" and "~~downs~~" in position (3) since the hydrogen atoms are well separated and so the position of one could hardly affect the position of another, and also since ordered "down" and "up" and "~~down~~" implies a larger unit cell, for which no evidence exists.
- (J08:0030.27) It is the exploitation of the inherent potential of infectious disease agents by scientific development and research ~~and development~~, resulting in the production of BW weapons systems.
- (J09:0010.06) A variety of techniques have been directed toward the study and isolation ~~and study~~ of blood group antibodies.
- (N10:0250.39) He'd come alone, without his child and wife ~~and child~~.

#### *Failed*

- (G09:0060.48) He has designed a matching costumes and backdrop ~~and costumes~~ of points of color on white for Mr. Cunningham's *Summerspace* , so that background and dancers ~~and background~~ merge into a shimmering unity.
- (N03:0020.03) He had been carrying an Enfield rifle and a holstered navy ball-and-cap cap and ball pistol.

### Appendix D11: Verb Particle Movement Rule

**First run: searching for verbs directly followed by a particle and then a direct object, or for verbs followed by a direct object and then a particle**

Incidence of candidate constructions: 97/6248 (1.6%)  
 Correctly identified sentences: 35/41 (85%)  
 Correctly transformed sentences: 33/41 (80%)

#### *Successful*

- (A01:1140.06) The Georgia Legislature will wind up its 1961 session up Monday and head for home — where some of the highway bond money it approved will follow shortly.
- (A04:0120.03) The nightmare of a clash between those in trouble in Africa, exacerbated by the difficulties, changes, and tragedies facing them, and other allies who intellectually and



- emotionally disapprove of the circumstances that have brought about these troubles ~~about~~, has been conspicuous by its absence.
- (A05:0590.30) While the city council suggested that the Legislative Council might perform the review, Mr. Notte said that instead he will take ~~up~~ the matter up with Atty. Gen. J. Joseph Nugent to get "the benefit of his views".
- (A07:0210.21) In an apparent effort to head ~~off~~ such a rival primary slate off, Mr. Wagner talked by telephone yesterday with Representative Charles A. Buckley, the Bronx Democratic leader, and with Joseph T. Sharkey, the Brooklyn Democratic leader.
- (A08:0750.12) Some predict the administration will settle down during 1961 and iron ~~out~~ the rough edges out which it has had thus far.
- (A09:0290.18) On Monday, the Hughes concern was formally declared bankrupt after its directors indicated they could not draw ~~up~~ a plan for reorganization up.
- (A10:1600.57) To step ~~up~~ the denomination's program up, the Rev. Mr. Brandt suggested the vision of 8,000 new Assemblies of God churches in the next 10 years.
- (A11:0360.06) Keegan, a 6-foot-3-inch 158-pounder, gave ~~up~~ the Orioles' last two safeties up over the final three frames, escaping a load of trouble in the ninth when the Birds threatened but failed to tally.
- (A12:0690.54) I think the big thing about the game was that our kids for the third straight week stayed in there pitching and kept on the pressure ~~on~~.
- (A13:1120.54) The modest and happy Spahn waved ~~off~~ his new laurels off as one of those good days.
- (A14:0750.18) Bill Jr., 20; Kathy, 15, and Patricia, 9, round ~~out~~ the Shea family out.
- (G01:0660.24) In fact it has caused us to give serious thought to moving our residence south, because it is not easy for the most objective Southerner to sit calmly by when his host is telling a roomful of people that the only way to deal with Southerners who oppose integration is to send in troops and shoot down the bastards ~~down~~.
- (G03:1010.42) Even the President cannot pick ~~up~~ his telephone up and give a "go" order.
- (G04:1650.12) The stink is all the same to me, but I really think they can make out one another ~~out~~ blindfolded.
- (G07:0470.12) ON a military mission for his native Virginia the youthful George Washington touched ~~off~~ the French and Indian War off, then guarded his colony's frontier as head of its militia.
- (G09:1270.15) It may establish the relation of the figure of the dancer to light and color, in which case changes in the light or color will set ~~off~~ a kaleidoscope of visual designs off.
- (G11:1060.03) We have staved ~~off~~ a war off and, since our behavior has involved all these elements, we can only keep adding to our ritual without daring to abandon any part of it, since we have not the slightest notion which parts are effective.
- (G17:0060.03) In town after town my companion pointed ~~out~~ the Negro school and the White school out, and in every instance the former made a better appearance (it was newer, for one thing).
- (G18:1170.15) To carry ~~out~~ this exalted conception out the author has combined the vivid realism and imaginative power we have noticed in his early poetry and carried them out on a grand scale.
- (G22:0710.03) No amount of ballyhoo will cover ~~up~~ the sordid facts up.
- (J07:0100.03) Further, the corpuscular radiation, *i.e.*, the solar-wind protons, must sputter ~~away~~ the surface atoms of the dust away and cause a slow diminution in size, with a resultant increase in both the Poynting-Robertson effect and the ratio of the repulsive force to the gravitational force.
- (J10:0020.24) Occasionally, you may come across one or two bumblebees in the cold season, when you are turning ~~over~~ sods over in your garden, but you have to be a really keen observer to see them at all.
- (J22:1110.09) By holding ~~out~~ prospects out for external capital assistance, the United States can provide strong incentives to prepare for the concerted economic drive necessary to achieve self-sustaining growth.
- (N03:0910.39) That touched ~~off~~ a total stampede off.
- (N04:0610.27) Montero had set ~~up~~ a strong position up, using every bale and box we had in addition to barricades of logs and brush.
- (N05:0030.51) You could burn ~~down~~ this whole mountainside down with a fire that size.

- (N06:0370.39) When he saw me coming he turned off his radio ~~off~~.
- (N09:0980.12) That was the day that he had practically mopped ~~up~~ the main street of Big Sands up with Aaron McBride, field boss for the Highlands Oil & Gas Company.
- (N10:0170.21) The Maguire family was setting ~~up~~ a separate camp up nearby.
- (N11:0700.15) "Fred was mighty crude about the way he took ~~in~~ cattle in" his own hired man, Andy Ross, mentioned later.
- (N12:0470.51) Curt snuffed ~~out~~ the match out.
- (N15:0890.21) Only one of the flight scored a direct hit and the rest blew ~~up~~ jungle up.
- (N18:0820.03) Even as she was telling me about it I became aware of a give-away flush that suffused her neck and moved upwards to her cheeks, and subconsciously I realized that when she entered the store she did not switch ~~on~~ the lights on.

#### *Doubtful*

- (A02:0150.21) It permits the state to take ~~over~~ bank accounts, stocks and other personal property of persons missing for seven years or more over.
- (G02:1680.21) Jean Bodin, writing in the sixteenth century, may have been the seminal thinker, but it was the vastly influential John Austin who set ~~out~~ the main lines of the concept as now understood out.
- (N07:0880.33) Twenty years before a group of Easterners had bought ~~out~~ the Haskell claims in the rocky hills south of Grass Valley out.
- (N13:0130.06) Dan could hear Clayton Burnside and Eben Jackson summing ~~up~~ their final reckoning for rental on the oxen up.

#### *Failed*

- (A20:0240.39) In their suburban cottage the crown charges, the Krogers received secrets from the mystery man, usually on the first Saturday evening of each month, and spent much of the week-end getting ~~off~~ the secrets ~~off~~ to Moscow, either on a powerful transmitter buried under the kitchen floor or as dots posted over period marks in used books.
- (J05:1890.09) In the over-all process, it is difficult to assign a "pure" role to each constituent of a built-detergent formulation; and, indeed, there is no more reason to separate the interrelated roles of the active, builder, antiredeposition agent, etc. than there is to assign individual actions to each of the numerous isomers making ~~up~~ a given commercial organic active up.
- (N08:0810.21) The forest took ~~on~~ an impersonal aspect on.
- (N14:0510.21) Looking over the setup ~~over~~, Stevens started coolly for the rear of the place.

#### *Excluded*

- (A13:0180.27) The Bears added two more in the fifth when McAuliffe dropped a double into the leftfield corner, Paschal doubled down the rightfield line and Cooke singled ~~off~~ Phil Shartzter's glove off.
- (A11:0220.06) Rookie southpaw George Stepanovich relieved Hyde at the start of the ninth and gave ~~up~~ the A's fifth tally up on a walk to second baseman Dick Howser, a wild pitch, and Frank Cipriani's single under Shortstop Jerry Adair's glove into center.
- (J24:1550.33) While there are many different possibilities for the timing of casework intervention, the experiments recently reported from a variety of traditional settings all point ~~up~~ the importance of an *immediate* response to the client's initial need for help up.
- (N03:0240.51) "Dean and myself will try to cut ~~out~~ horses to ride out", he said.
- (G08:1140.15) Just as the Negro situation points ~~up~~ the gradual and abrupt changes affecting Southern life up, it also points ~~up~~ the non-representation of urbanism in Southern literature up.

## References

- Abney, S. 1992: "Prosodic Structure, Performance Structure and Phrase Structure", in Proceedings of the 5th DARPA Workshop on Speech and Natural Language (Harriman, NY). Morgan Kaufmann.
- Abney, S. 1995: "Chunks and dependencies: Bringing processing evidence to bear on syntax", in Cole, Green & Morgan, (eds): *Computational Linguistics and the Foundations of Linguistic Theory*, pp.145--164. CSLI, Stanford.
- Anderson, R. J. & Petitcolas, F. A. P., 1998: "On the Limits of Steganography", in *IEEE Journal of Selected Areas in Communications*: v 16 no 4, 1998, pp.474-481.
- Anderson, R. J. & Petitcolas, F. A. P., 2001: *Information Hiding An Annotated Bibliography*. Unpublished.
- Atallah, M. J., McDonough, C. J., Raskin, V. & Nirenburg, S., 2000: " Natural Language Processing for Information Assurance and Security: An Overview and Implementations", to appear in Schaeffer, M. (ed), 2000: *NSPW '00*. ACM Publications.
- Atallah, M. J., Raskin, V., Crogan, M., Hempelmann, C. F. & Kerschbaum, F., 2001: " Natural Language Watermarking: Design, Analysis and a Proof-of-Concept Implementation", to appear in *Information Hiding: Forth International Workshop*. Springer Verlag.
- Bailer, W., 2001: wbStego Steganography Tool [web page], <http://www.wbailer.com/wbstego>.
- Bailer, W. & Rathner, L. 2001: *Linguistic Information Hiding*. Unpublished.
- Baker, C. L., 1989: *English syntax*. Cambridge, MA: MIT Press.
- BBC, 2001: "Tackling terror with technology" [web page], *BBC News Online*: 21<sup>st</sup> September 2001. [http://news.bbc.co.uk/hi/english/sci/tech/newsid\\_1555000/1555981.stm](http://news.bbc.co.uk/hi/english/sci/tech/newsid_1555000/1555981.stm).
- Bell, A. M., 1951: *The Compounding and Hyphenation of English Words*. New York: Funk & Wagnalls Press.
- Bender, W., Gruhl, D., Morimoto, N. & Lu, A., 1996: "Techniques for Data Hiding", in *IBM Systems Journal*: v 35 no 3-4, 1996, pp.313-336.
- Boneh, D. & Shaw, J., 1995: "Collusion-Secure Fingerprinting for Digital Data" in *Crypto'95*: LNCS 963, 1995, pp.452-465. Berlin: Springer-Verlag.
- Borsley, R., 1991: *Syntactic Theory: A Unified Approach*. London: Arnold Press.
- Brassil, J. T., Low, S., Maxemchuk, N. F. & O'Gorman, L., 1994: "Electronic Marking and Identification Techniques to Discourage Document Copying", in *Proceedings of IEEE INFOCOM '94*, (Toronto, Canada), 1994, pp.1278-1287.
- Chapman, M. & Davida, G., 1997: *Hiding the Hidden: A software System for Concealing Ciphertext as Innocuous Text*. International Conference on Information and Communications Security, November 11-13, 1997, Beijing, P. R. China.
- Chapman, M. & Davida, G., 2001: *NICETEXT* [web application], <http://www.ctgi.net/NICETEXT>.
- COBUILD, 1996: *Grammar Patterns 1: Verbs*. London: HarperCollins.

- Compris [company], 2001: TextHide [web page], <http://www.compris.com/TextHide/en/>.
- Copestake, A., Flickinger, D., Malouf, R., Riehemann, S. & Sag, I., 1995: "Translation using Minimal Recursion Semantics", in *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95)*, Leuven, Belgium.
- Copestake, A., 1999. *The (new) LKB System*. Stanford: CSLI.
- Culicover, P. W., 1982: *Syntax*. New York: Academic Press.
- Greenbaum, S., 1969: *Studies in English Adverbial Usage*. London: Longman.
- Grisham, R., Macleod, C. & Meyers, A., 1994: "COMLEX Syntax: Building a Computational Lexicon" in *Proceedings of the 15th International Conference on Computational Linguistics*, 1994, pp.268-272.
- Hewings, M., 1999: *Advanced grammar in use : a self-study reference and practice book for advanced learners of English*. Cambridge: Cambridge University Press.
- Katzenbeisser, S., 2000: "Principles of Steganography" in Katzenbeisser, S. & Petitcolas, F.A.P. (eds), 2000: *Information hiding techniques for steganography and digital watermarking*. Boston: Artech House.
- Kutter, M. & Hartung, F., 2000: "Introduction to Watermarking Techniques" in Katzenbeisser, S. & Petitcolas, F.A.P. (eds), 2000: *Information hiding techniques for steganography and digital watermarking*. Boston: Artech House.
- Langendoen, D. T., 1970: *Essentials of English Grammar*. New York: Holt, Rinehart and Winston.
- Levin, B., 1993: *English verb classes and alternations : a preliminary investigation*. Chicago: University of Chicago Press.
- Liles, B. L., 1979: *A Basic Grammar of Modern English*. New Jersey: Prentice-Hall.
- Mathews, P. H., 1981: *Syntax*. Cambridge: Cambridge University Press.
- McArthur, T., 1972: *Using Compound Words*. London: Collins.
- McCawley, J. D., 1988: *The Syntactic Phenomena of English*. Chicago: University of Chicago Press.
- McEnery, T. & Wilson, A., 1996: *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J., 1990: "Introduction to WordNet: an on-line lexical database", in *International Journal of Lexicography*, 1990, v 3 no 4, pp.235-244.
- Oakes, M. P., 1998: *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Perez-Gonzalez, F. & Hernandez, J. R., 1999: *A Tutorial in Digital Watermarking*. Unpublished.
- Petitcolas, F. A. P., Anderson, R. J. & Kuhn, M. G., 1999: "Information Hiding – A Survey", in *Proceedings of the IEEE*: v 87 no 7, 1999, pp.1062-1078.
- Pustejovsky, J. & Boguraev, B., 1996: *Lexical Semantics: The Problem of Polysemy*. Oxford: Clarendon Press.

- Quirk, R. & Crystal, D., 1985: *A Comprehensive Grammar of the English Language*. London: Longman.
- Quirk, R. & Greenbaum, S., 1973: *A University Grammar of English*. London: Longman.
- Raskin, V. and Nirenburg, S., 1995: "Lexical Semantics of Adjectives: A Microtheory of Adjectival Meaning" in *Memoranda in Computer and Cognitive Science*: MCCS-95-288. Las Cruces, N.M.: New Mexico State University.
- Ryder, M. E., 1994: *Ordered Chaos: The Interpretation of English Noun-Noun Compounds*. Berkely: University of California Press.
- Sag, I. A. & Wasow, T., 1999: *Syntactic theory: a formal introduction*. Stanford: CSLI.
- Sampson, G., 1995: *English for the computer: the SUSANNE corpus and analytic scheme*. Oxford: Clarendon.
- Sampson, G., 2001: *SUSANNE corpus and analytic scheme* [web page], <http://www.cogs.susx.ac.uk/users/geoffs/RSue.html>.
- Schütze, C. T., 1996: *The Empirical Basis of Linguistics*. Chicago: University of Chicago Press.
- Shaw, S., 1999: *Overview of Watermarks, Fingerprints, and Digital Signatures* [web page], <http://www.jisc.ac.uk/jtap/htm/jtap-034.html>.
- Sorace, A., Gurman Bard, E. & Robertson, D., 1996: "Magnitude Estimation of Linguistic Acceptability", in *Language*: Vol. 72, No. 1, pp.32-68.
- Su, J. K., Hartung F. & Girod B., 1998: "Digital watermarking of text, image, and video documents", in *Computers & Graphics*: Vol. 22, No. 6, pp.687-695.
- Trask, R. L., 1993: *A Dictionary of Grammatical Terms in Linguistics*. London: Routledge.
- Utts, J., 1996: *Seeing Through Statistics*. London: Duxbury Press.
- Wayner, P., 1996: *Disappearing Cryptography □ Being and Nothing on the Net*. Massachusetts: Academic Press.
- Winstein, K., 1998: *Lexical Steganography Through Adaptive Modulation of the Word Choice Hash*. Unpublished. <http://www.imsa.edu/~keithw/tlex/>