

# The syntax of concealment: reliable methods for plain text information hiding

Brian Murphy & Carl Vogel\*

Computational Linguistics Lab, Trinity College Dublin, Dublin 2, Ireland

## ABSTRACT

Many plain text information hiding techniques demand deep semantic processing, and so suffer in reliability. In contrast, syntactic processing is a more mature and reliable technology. Assuming a perfect parser, this paper evaluates a set of automated and reversible syntactic transforms that can hide information in plain text without changing the meaning or style of a document. A large representative collection of newspaper text is fed through a prototype system. In contrast to previous work, the output is subjected to human testing to verify that the text has not been significantly compromised by the information hiding procedure, yielding a success rate of 96% and bandwidth of 0.3 bits per sentence.

**Keywords:** plain text, information hiding, steganography, syntactic transform, human judgement

## 1. INTRODUCTION

Information hiding has been suggested for a range of applications including steganography (secret communication) and watermarking (adding attribution data to media files). However, in contrast to rich media files such as audio and images, it has proven difficult to embed hidden data directly in plain text files. Previous work has mainly used formatting information<sup>1-3</sup> and semantic substitutions.<sup>4-6</sup> The first approach is vulnerable to reformatting operations, such as file conversion, while the second depends upon deep semantic processing – word sense disambiguation, semantic role parsing and anaphora resolution – NLP technologies that are not yet robust. In this paper we examine two highly predictable and reasonably common grammatical phenomena in English that can be used in data hiding, the swapping of complementisers (1) and relativisers (2), which rely on a well-established technology: syntactic parsing.

- (1) For example, the New York state counsel for the NAACP said ~~that~~ Mr. Mason is “like a dinosaur.” (2369:348) †
- (2) When I was just a child  
And confronted by my fears,  
The things ~~that~~ which I thought would get me  
Had fangs and pointed ears. (1798:3)

Moreover, this work is to our knowledge unique in that it is rigorously tested for acceptability. The output of a prototype system is evaluated by human judges to measure to what extent the resulting text is well-formed, stylistically consistent and retains the meaning of the original. Previous papers make various claims of hidden information carrying capacity (bandwidth) but give no principled account of the degree to which the text has been compromised. In this paper we address quality *and* quantity. The final bandwidth achieved is lower than some previous work at an average of one bit for every three sentences, but reliability stands at 96% with minimal processing time.

In the following section we review previous work on plain text information hiding and discuss the system design considerations relevant to selecting candidate transforms. In section 3 we examine in detail the two constructions to arrive at algorithms for their correct identification and transformation. The evaluation section describes the main body of the work: implementation of the system and a test-run over a balanced body of newspaper text, extensive evaluation of the output by the investigators, and finally a human judgement experiment with a group of subjects to verify that independent evaluations are both reliable and representative.

---

\* vogel@cs.tcd.ie; brian.murphy@unitn.it

† Source file and line number are given for examples extracted from the Penn Treebank. Removed text is shown struck-through, and added text underlined.

## 2. PLAIN TEXT INFORMATION HIDING

### 2.1. State of the art

It has proven more difficult to hide data in plain text than in noisier data such as sound and images. Su et al.<sup>7</sup> were particularly pessimistic: “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.” Some research has got around this by generating nonsensical text from scratch as a carrier (the mimic method,<sup>8,9</sup> which is vulnerable to human detection), while others have made changes to the orthography and layout in particular textual file formats<sup>1-3</sup> (an approach that can be trivially defeated by file reformatting, or re-capturing text with optical text recognition). The first published account that found ‘perceptual headroom’ in the fabric of text<sup>10</sup> suggested that the substitution of equivalent words (e.g. *large/big*) or constructions (3, 4), could hide bits of data without affecting meaning.

(3) Before the night is over, I will have finished

(4) I will have finished before the night is over

Examples of applications that apply the word substitution method to steganography and traitor tracing<sup>4,5</sup> suffer from indeterminacy in synonym selection. Many words have more than one sense (i.e. meaning usage), and the activation of a particular sense by its context determines what constitutes a synonym. State of the art word sense disambiguation systems<sup>11,12</sup> still struggle to achieve much more than 60% accuracy on general text. For example the verb *fire* might be synonymous with either *sack*, *shoot* or *bake*, depending on whether it is found in an employment, military or manufacturing context. Even when synonyms each have only a single sense, they are often restricted to a particular style or register of text as can be seen in (5) – in formal writing the *fiddle* variant would be unacceptable. While such a system might offer relatively high bandwidth (Winstein claims bandwidth of 0.5% of the original message – or about 6 bits per Penn Treebank sentence), there are three prerequisites for a successful synonym substitution-based algorithm: a comprehensive enumeration of word senses, robust word-sense disambiguation and fine-grained text categorisation. As all remain topics of active research, a reliable system on this model seems, as yet, unattainable.

(5) ... the first ~~violin~~ fiddle section of the Berlin Philharmonic ...

Atallah et al.<sup>13</sup> describe a proof-of-concept watermarking implementation based on sentence transforms that they judge to change meaning ‘slightly’, though to what degree is not quantified. Other than the insertion of ‘empty transitional’ phrases such as *basically* and *in general*, the transforms are all syntactic, including passivisation, adjunct movement, clefting and swapping the subject and object in copular sentences, in a similar fashion to Murphy.<sup>14</sup> Their rationale is that any truly meaning-preserving transform could be globally applied to a document to scramble the watermark. A transform 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 a malicious attempt to remove or replace the mark. Later work by the same group<sup>6</sup> discarded the syntactic approach in favour of one that changes the distribution of redundant information throughout a text, resulting in a dramatic jump in bandwidth from 0.5 to 4 bits per sentence. Following the same rationale of subtle changes of meaning, but using semantic transforms, phrases that primarily describe facts circumstantial to each sentence are removed, added or moved.

These last two approaches depend crucially on the quality of syntactic and semantic analysis, necessary to decide in what circumstances a particular transform is appropriate. Semantic role identification remains a difficult problem for NLP – the best scoring systems in the CoNLL-2005 role labeling shared task achieved f-measures of between 70 and 80%,<sup>15</sup> even though the task involved a much more shallow degree of analysis (that of PropBank,<sup>16</sup> which labels arguments numerically for each syntactic frame type) than what Atallah et al.<sup>6,13</sup> attempt (applying semantic labels such as Agent, Recipient and Goal, that are valid across all classes of predicates). Prepositional phrase attachment is also difficult to predict much more than 85% of the time for modern statistical parsers.<sup>17</sup>

While few examples are given, some questionable transforms can be found in the short demo texts on their website.<sup>18</sup> In (6, 7) a simple imperative sentence, instructing the reader to perform some procedure, is replaced by one that might be interpreted as stating that the procedure will be carried out automatically or by some other party (a dangerous prospect, as

the sentences comes from a nuclear submarine accident procedure manual). In example (8) there are several inconsistencies. Presumably Mr. Bush (the younger) is referring to the enemy in its entirety in the first sentence, and not just those *holding the front lines*. It is questionable if *We're making protecting themselves harder* is well-formed, and it is difficult to see how *them* in the original can correspond to *the fundamentalist* in the output.

- (6) Prepare to initiate DECON station procedures.
- (7) Initiation of DECON station procedures shall be prepared.
- (8) “We’re slowly but surely tightening the net on the enemy holding the front lines. We’re making it harder for the enemy to communicate. We’re making ~~it harder for the enemy to protect themselves~~ protecting themselves harder. We’re making it harder for the enemy to hide. And we’re going to get him and ~~them~~ the fundamentalist,” Bush said.

## 2.2. Design considerations

General requirements for the marks are quality and quantity. Broadly the aim is to reliably and robustly make changes to a document, after which it should still convey the same message in the same fashion. The marking structures should also be ubiquitous enough that they are found often in typical texts across a range of styles (e.g. prose, reports, technical documents or journalism). The selection and evaluation of text transforms depends to a degree upon the application. In attribution watermarking applications the information hiding procedure must not perceptively degrade or devalue the document to the end-user. Secondly it should be impossible for an attacker to remove the mark, let alone insert their own to challenge the genuine author’s intellectual ownership. A special case of watermarking is traitor tracing, where confidential documents are individually marked when distributed. In the case of a leak or espionage, the marked document should lead back to the person responsible. Here it is important that the mark cannot be removed or, worse still, replaced to mislead investigators. For steganography the principal requirement is that the marked document does not betray the presence of a hidden message. Once aware that steganography is being used, an eavesdropper may be able to prevent the delivery of such documents, even if they cannot remove the message mark. A second requirement is that the message, even if detected, resists replacement – a third party also might try to substitute messages. And ideally, steganography should operate without the recipient having access to the original unmarked document. Sending documents in marked/unmarked pairs could raise suspicion, while the logistics of both sender and recipient holding large identical banks of long documents might be prohibitive for sustained communication. A typical information hiding scheme could encode 1’s by changing constructions, and 0’s by leaving them as they are. Thus the processed version (10) of the sentence below would encode the sequence of bits ‘101’.

- (9) We looked [behind the couches and under the beds]<sub>0</sub>, [in the shed and up in the attic]<sub>0</sub>, and through all the [wardrobes and cupboards]<sub>0</sub>
- (10) We looked [under the beds and behind the couches]<sub>1</sub>, [in the shed and up in the attic]<sub>0</sub>, and through all the [cupboards and wardrobes]<sub>1</sub>

But we can see below that it is possible to recover information without reference to the original as long as the transforms we make are reversible. A modified version of the scheme used in (10) could specify that coordinate constructions with their arguments in alphabetical order encode the bit 0, while those in reverse order encode 1, regardless of their original configuration (11).

- (11) We looked [under the beds and behind the couches]<sub>1</sub>, [in the shed and up in the attic]<sub>0</sub>, and through all the [wardrobes and cupboards]<sub>1</sub>

In the case of non-reversible transforms, such as reducing *be* relatives (12), the original is needed to extract encoded information. In a scheme that represented the bit 0 with full relatives, and the bit 1 with reduced relatives, it is not possible to expand relative clauses reliably. Assuming we wanted to encode the bit 0 in sentence (13) we would need to expand the reduced relative clause *at the end of her tether*, but it is debatable which of the three variants in (14) is the correct one. As reduced relative expansion cannot be automated, a decoding algorithm without access to the original cannot determine which reduced relatives it sees were transformed from full relatives, and which as reduced relatives in the original were skipped during marking.

- (12) Any patients [~~that have been~~ admitted within the last week] and [~~who are~~ feverous] have been isolated
- (13) The woman [at the end of her tether] lashed out
- (14) The woman [who is/was/had been at the end of her tether] lashed out

As we will see later, the transforms proposed in this paper, being largely reversible, are more suited to steganography and traitor tracing, where the existence of a mark should not be known, since removal is trivial once an attacker has access to the transforming algorithm.

### 3. SYNTACTIC TRANSFORMS CONSIDERED

#### 3.1. Relativiser swap

Restrictive relative clauses are incomplete finite sentences that modify a noun phrase in English to help the reader hone in on the particular individual(s) intended by the speaker (15). They can be contrasted with non-restrictive relative clauses, which provide background information on an individual that has already been fully identified in the discourse. Non-restrictive relatives (16) more often modify proper noun or pronoun phrases, and are usually distinguished in text by the insertion of a comma, or in speech by a pause.

- (15) For people [that insist on jumping in now to buy the funds]<sub>rest</sub>, Newgate’s Mr. Foot says: “The only advice I have for these folks is that those [who come to the party late]<sub>rest</sub> had better be ready to leave quickly. (0034:806)
- (16) Rival Boston Herald columnist Howie Carr, [who usually rails at Statehouse “hacks” and nepotism]<sub>non-rest</sub>, argued that the new drawings were designed to hide Mr. Madden’s “rapidly growing forehead” and the facial defects of “chinless” Dan Shaughnessy, a Globe sports columnist. (2010:375)

In the case of restrictive relatives, the word that connects the noun phrase to the clause (the relativiser) varies between *that*, being dropped altogether, and *who* or *which* (depending on whether the noun phrase refers to a person). †

- (17) The theory [that you refer to] is discredited
- (18) The theory [which you refer to] is discredited
- (19) The theory [you refer to] is discredited
- (20) The person [that looks like Mr. T] is an embroidery teacher
- (21) The person [who looks like Mr. T] is an embroidery teacher
- (22) \*The person [looks like Mr. T] is an embroidery teacher §
- (23) John, [who you met yesterday], can’t come
- (24) \*John, [that you met yesterday], can’t come
- (25) \* John, [you met yesterday], can’t come

The first sentence above (17) is object-gapped, in that the clause *you refer to* lacks an object, and allows all three variants. Subject-gapped clauses such as *looks like Mr. T* allow the *that/wh...* alternation, but do not allow the relativiser to be dropped. Non-restrictive relative clauses (23) do not allow any alternation.

Restrictive relative clauses are very common in written English (almost every tenth sentence in the Penn Treebank contains one), and can be easily recognised from syntactic parses:

†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 ...*

§An asterisk \* signifies a badly formed sentence.

- Incomplete finite clauses that are bare or headed by *that* or *who/which*, and that follow noun phrases, are relative clauses.
- Proper nouns or pronouns as the head<sup>¶</sup> of the noun phrase or commas indicate non-restrictive relatives.
- The lack of a noun-phrase before the main verb within the relative clause indicates a subject-gapped clause.

Using this knowledge about the structure of relative clauses in English, an simple algorithm encoded single trits of information in parsed sentences by selectively substituting relativisers.

### 3.2. Complementiser swap

Complement clauses are complete finite sentences that act as arguments to a verb, as well as to some nouns. They often replace noun phrases when we want to talk about events and states of affairs, rather than objects and entities. In the case of verb complements, the complementiser *that* can usually be dropped (26, 27) or inserted (28) freely.

- (26) When, however, Claudio's sister, Isabella, a novice in a convent, goes to Angelo to plead her brother's case, the obdurate ruler immediately falls in love with her and, in a supreme act of hypocrisy, demands [~~that~~ Isabella yield up her virtue to him]<sub>comp</sub> in exchange for her brother's life. (0819:216)
- (27) The management indicated [~~that~~ they had no idea what they were doing]<sub>comp</sub>
- (28) First, he suggests, [that GPA Group Ltd., the international aircraft leasing company based in Ireland, could lease some of its Boeing jetliners to the Soviet airline]<sub>comp</sub>

Verb complement clauses are extremely common (about one in five sentences in the Penn Treebank contains one) and it is straightforward to extract them from syntactic parses:

- Bare or *that* headed full finite sentences in the position of a verb object are verb complement clauses.

### 3.3. Coding

One issue particular to this approach is how to use the base-three digits that the object-gapped relative transform can carry. A simple solution is to treat pairs of 'trits' as equivalent to triplets of conventional bits of information, since two trits ( $3^2$ ) can store slightly more information than three bits ( $2^3$ ). This would involve separating the bit and trit encoding runs to maximise the possibility of finding sequences of 'trits', and a simple base conversion scheme would be used as such: 000 = 00; 001 = 01; 010 = 02; 011 = 10; etc. When calculating bandwidth, we conservatively treat a trit as having 1.5 times the information carrying capacity of a bit.

## 4. IMPLEMENTATION

The algorithms described were implemented as a prototype system programmed in Perl. We used the Penn Treebank II as input to the system. The Penn Treebank II,<sup>19</sup> published by the University of Pennsylvania is a corpus (i.e. text data collection) of articles selected from the Wall Street Journal daily newspaper. It consists of just over 49 thousand sentences of American English (around 1 million words, 7MB of ASCII) that have been automatically parsed with parts-of-speech and syntactic trees, and then verified by hand. Below is the tree parse of the second relative clause in example (15):

```
(S (NP-SBJ-4 (NP (DT those))
      (SBAR (WHNP-3 (WP who))
        (S (NP-SBJ (-NONE- *T*-3))
          (VP (VBP come)
            (PP-DIR (TO to)
              (NP (DT the)
                (NN party))))
            (ADVP-TMP (RB late))))))
```

<sup>¶</sup>The head of a noun phrase is the word that is syntactically and semantically central to it. E.g. *men* in ... *the three doddery old men that you saw ...*

A verb complement clause, in this case from (1), looks like this:

```
(VP (VBD said)
  (SBAR (IN that)
    (S (NP-SBJ (NNP Mr.)
      (NNP Mason))
      (VP (VBZ is)
        (" ")
        (PP-PRD (IN like)
          (NP (DT a)
            (NN dinosaur)))))))
  (. .)))
```

The objective of using the Penn Treebank as opposed to dynamically parsed free text was two-fold. Firstly, using hand-verified ‘perfect’ parses allows us to measure precisely the reliability of the transforms, independently of the errors of current parsers. Secondly, arbitrarily selected free text may not be representative of text in general. Using a professionally assembled corpus allows us to have confidence that our results are replicable, even if restricted to a particular domain and register (in this case American English journalism).

Every sentence tree in the corpus was read in by the algorithm and represented internally as an  $n$ -ary tree structure. This allowed the algorithm to crawl around the tree searching for particular configurations of nodes that identify candidates for transforms. In this prototype each candidate construction was transformed for evaluation purposes – obviously in an operational system transforms would only be made when required to encode the current bit in the hidden data. The transformed sentences were then output as plain text. The entire corpus of 49 thousand sentences was processed and output in 16 minutes on a PC with 900MB of RAM and a 1GHz CPU running Perl 5.6 on Linux. The processing results are listed in Table 1.

Table 1. transform experiment on the Penn Treebank	
<b>Total number of sentences transformed</b>	12,654
<b>... of which contain one transform</b>	10,892
<b>... of which contain two transforms</b>	1,595
<b>... of which contain three transforms</b>	153
<b>... of which contain four transforms</b>	11
<b>... of which contain five transforms</b>	3
<b>Total number of transforms</b>	14,600
<b>... of which subject-gapped relatives</b>	3,340
<b>... of which object-gapped relatives</b>	1,430
<b>... of which verb complements</b>	9,830

As can be seen from the tree samples just given, the syntactic categories needed by our algorithms are represented in a straightforward manner. For example a finite clause is a verb phrase (VP) whose main verb is not in participle form (e.g. *do/does/did* rather than *done/doing*) and not preceded by the infinitive marker (e.g. *to run*, tagged TO). Many categories, such as subordinate clauses (SBAR) or noun phrases (NP) are represented directly with a single label. Other distinctions, such as those between common nouns and proper nouns or pronouns (to exclude non-restrictive relative clauses) are available from the part-of-speech tags NN, NNP and PP. Identifying the head of a noun phrase is a little more involved, and the guidelines issued by the Penn Treebank makers were used.

For the semantic distinction between human and non-human nouns (to determine the appropriate usage of *who* or *which*), two lexical resources were used. WordNet<sup>20</sup> is an electronic dictionary produced by Princeton University, which represents word senses as a hierarchical ontology of concepts. It covers 109 thousand nouns in 75 thousand concepts. For more common words it includes counts of each word sense as attested in a small corpus, allowing us to estimate probabilities for each word/sense pair. Once it has been determined in what a word’s likely senses are, it is possible to

**Table 2.** Reliability of transforms over random samples of 100 items

	Successful	Doubtful	Failed
Subject-gapped relatives	98	1	1
Object-gapped relatives	96	3	1
Verb complements	95	4	1
Reliability for single transform in isolation	95.8%		
Three or more combined	87	5	8
Projected reliability for three combined	87.9%		

climb back up the ontology to discover if the concepts in question are descended from the concept *person*. COMLEX is another lexical resource<sup>21</sup> that lists 22 thousand nouns together with various grammatical features, including whether they show human-like syntactic behaviour, such as combining with *who* or being the infixed indirect object of verbs such as *give* or *send* (e.g. compare *She sent Mary a letter* to *She sent Dublin a letter*). In the algorithm an average of the WordNet and COMLEX estimates of ‘humanity’ was used – when this estimate exceeded a threshold of 0.5, a noun was considered to be human.

## 5. RESULTS AND EVALUATION

From the results in Table 1. we can see that verb complements are by far the most ubiquitous of the transforms. The object-gapped relative clause construction, that has a higher information carrying capacity than the other two transforms because of its three-way alternation, is the rarest. On average there is one transform for every three sentences (0.30 transforms per sentence), and the information carrying capacity is 0.31 bits per sentence.

To quantify the reliability of the transforms used, both individually and in combination, random samples of the transformed sentences were extracted. One hundred sentences were extracted in four categories: having one subject-gapped transform; having one object-gapped transform; having one verb complement transform; and having three or more transforms of any kind. First one of the authors marked these sentences for grammaticality (i.e. syntactic well-formedness) and for preservation of meaning and style, using a three way distinction: successful, doubtful and failed (Table 2). Comparing the transforms, we see that all three are highly reliable, and there do not seem to be any significant interference effects. We had thought that multiple transforms, that were acceptable in isolation, might interfere with each other, leading to failure, or at least awkwardness (e.g. 29), but this did not seem to be the case.

- (29) But Mr. Hahn rose swiftly through the ranks, demonstrating a raw intelligence ~~that~~ which he says that he knew that he possessed early on. (0100:524)

The problematic subject gapped sentences were (30), where a non-restrictive relative clause was misdiagnosed as restrictive, and (31), where it is debatable to what extent *company* is human-like. In this case our estimation algorithm gave *company* a probability of 0.503, so raising our probability threshold might eliminate such problems. Three of the four object-gapped sentences that did not succeed were because of ambiguity introduced by the transform, as in (32), and one failed because *all* was erroneously seen to represent people (33). Finally the four doubtful results of the complementiser switch were as a result of the proximity effects. It seems that sentences with long complements, or where an adverb (34) intervenes, resist losing the complementiser. The single failure was due to inconsistent orthography, where the colon must be removed.

- (30) The U.S. officials said ~~that~~ despite the rapid changes under way in Eastern Europe and the Soviet Union, all the Cocom members agreed on “the continuing need for this organization,” ~~which~~ that was founded 40 years ago at the start of the Cold War. (0567:265)
- (31) Prosecutors depict a company ~~that~~ who allegedly sat on damaging evidence of overcharges from 1983 to 1985, despite warnings from an internal auditor. (1184:547)

- (32) Despite the number of fans ~~that~~ who office massage has won, some purists look down on it, arguing ~~that~~-naked, full-body rubs are the only way to go. (2161:1509)
- (33) “All ~~that~~ who this has really established is that MCA and the Bronfmans have agreed on a price at which they can be bought out,” said Jeffery Logsdon, an analyst with Crowell, Weedon in Los Angeles. (0938:318)
- (34) A spokesman for Southern Co. would say only ~~that~~-discussions are continuing between Gulf Power and federal prosecutors. (0619:216)
- (35) After the 1929 crash, Herbert Hoover said: that “The fundamental business of the country... is on a sound and prosperous basis.” (2370:107)

However, grammaticality judgements can be unreliable<sup>22</sup> and could be further biased since they were being made by one of the investigators. Accordingly we took a much smaller sample of the sentences tested, and had them verified by a separate group of subjects. We chose three sentences randomly from each of the four categories above (subject and object-gapped relatives, verb complements and triple-transform sentences), and padded them out with three random untransformed filler sentences, and five sentences that we judged to be faulty. The subjects were colleagues of the authors, and were unaware of the object of the experiment – three were Irish, two French and one American. They were first asked to rate the 20 sentences for grammaticality on a +3 (‘perfect’) to –3 (‘unacceptable’) scale. Then they were asked to compare the original and transformed versions of the same sentences (this time 17, since the fillers had been removed) and decide to what extent meaning and style had been preserved. The questionnaire is given below in an appendix that follows the references.

**Table 3.** Comparison of grammaticality judgements

Question	Category	Well-Formed	Subjects' mean	Standard deviation
8	Triple transform	✓	3	0
13	No transform	✓	3	0
15	Subject-gapped relative	✓	3	0
17	Verb complement	✓	2.83	0.41
3	No transform	✓	2.67	0.52
16	Verb complement	✓	2.67	0.52
4	Object-gapped relative	✓	2.5	0.55
18	Failed transform	?	2.33	1.21
12	Subject	✓	2.17	1.60
19	No transform	✓	2.17	1.17
2	Object-gapped relative	✓	2	1.55
9	Object-gapped relative	✓	2	1.55
7	Verb complement	✓	1.83	1.33
14	Subject-gapped relative	✓	1.5	1.97
10	Triple transform	✓	0.5	2.81
6	Triple transform	?	–0.33	2.25
11	Failed transform	*	–0.5	2.16
1	Failed transform	*	–0.67	1.86
5	Failed transform	*	–0.67	2.88
20	Failed transform	*	–1.67	1.97

Table 3. collates and compares the grammaticality results. Here we can see that other than in question 18, our judgements coincide with those of the other subjects. All that we judge to be successful have an average positive evaluation, and failed or doubtful sentences get an average negative evaluation. It is also interesting to note, that the degree of disagreement amongst subjects (represented by the standard deviation) is highest for the most problematic sentences. In Table 4 we show judgements of meaning and style. Here the results are less clear cut, but the sentences we judged as flawed still cluster towards the bottom the subjects' evaluation. Finally we carried out a Chi-squared test on the subjects' evaluation of the



**Table 4.** Comparison of meaning and style judgements

Question	Category	Meaning	Style	Subjects' mean	Standard deviation
15	Verb complement	✓	✓	3	0
2	Object-gapped relative	✓	✓	3	0
6	Verb complement	✓	✓	2.83	0.41
7	Triple transform	✓	✓	2.67	0.82
14	Verb complement	✓	✓	2.67	0.82
11	Subject (SC)	✓	✓	2.67	0.82
12	Subject-gapped relative	✓	✓	2.67	0.82
8	Object-gapped relative	✓	✓	2.6	0.55
3	Object-gapped relative	✓	✓	2.33	0.52
16	Failed transform	?	✓	2.33	1.63
13	Subject-gapped relative	✓	✓	1.83	2.4
10	Failed transform	*	✓	1.5	2.35
9	Triple transform	✓	✓	0.5	1.76
1	Failed transform	*	✓	0.17	2.86
5	Triple transform	✓	✓	-0.167	2.86
17	Failed transform	*	✓	-0.33	2.5
4	Failed transform	*	*	-2	2.45

three transforms, to detect any significant distributional difference. The result was not significant ( $p \leq 0.20$ ), due in part to the small number of sentences analysed, but is consistent with our own judgements that indicate the three transforms to be of roughly equal reliability.

Overall we see persuasive evidence that the transforms chosen are highly reliable, commonly found in text and can be quickly executed (0.02 seconds per sentence). State-of-the-art parsers<sup>23</sup> typically achieve over 90% recall and precision for syntactic labels, and some more recent parsers<sup>24</sup> reliably identify functional categories such as subject and object more than 95% of the time. Thus we can reasonably expect reliability of around 90% using free text in an operational system.

### Acknowledgements

We gratefully acknowledge the support of Science Foundation Ireland funding in the Research Frontiers Programme project 05/RF/CMS002.

### REFERENCES

1. J. Brassil, S. H. Low, N. F. Maxemchuk, and L. O’Gorman, “Electronic marking and identification techniques to discourage document copying,” *IEEE Journal on Selected Areas in Communications* **13**(8), pp. 1495–1504, 1995.
2. W. Bailer and L. Rathner, “Linguistic information hiding.” <http://www.wbailer.com/wbstego>, 2001.
3. Compris GmbH, “Texthide [application].” <http://www.compris.com/TextHide/en/>, 2001.
4. K. Winstein, “Lexical steganography through adaptive modulation of the word choice hash.” <http://www.imsa.edu/keithw/tlex/>, 1998.
5. C. Jensen, “Fingerprinting text in logical markup languages,” *Lecture Notes in Computer Science* (2200), pp. 433–445, 2001.
6. M. Atallah, V. Raskin, C. Hempelmann, M. Karahan, R. Sion, and K. Triezenberg, “Natural language watermarking and tamperproofing,” *Lecture Notes in Computer Science and Proceedings of the 5th International Information Hiding Workshop*, 2002.
7. J. K. Su, F. Hartung, and B. Girod, “Digital watermarking of text and image and and video documents,” *Computers and Graphics* **22**, pp. 687–695, 1971.
8. P. Wayner, *Disappearing Cryptography - Being and Nothing on the Net*, Academic Press, Massachusetts, 1996.

9. M. Chapman and G. Davida, "Hiding the hidden: A software system for concealing ciphertext as innocuous text," in *Proceedings of the International Conference on Information and Communications Security*, (Beijing), 1997.
10. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal* **35**(3-4), pp. 313–336, 1996.
11. R. Mihalcea and A. Csomai, "Senselearner: Word sense disambiguation for all words in unrestricted text," in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, (Ann Arbor), June 2005.
12. T. Pedersen, S. Banerjee, and S. Patwardhan, "Maximizing semantic relatedness to perform word sense disambiguation," tech. rep., University of Minnesota Supercomputing Institute, Minnesota, March 2005.
13. M. Atallah, V. Raskin, M. Crogan, C. Hempelmann, and K. Kerschbaum, "Natural language watermarking: Design and analysis and a proof-of-concept implementation," in *Proceedings of Forth International Workshop on Information Hiding*, 2001.
14. B. Murphy, "Syntactic information hiding in plain text," Master's thesis, Trinity College Dublin, [https://www.cs.tcd.ie/research\\_groups/clg/brian/murphy01hidingMasters.pdf](https://www.cs.tcd.ie/research_groups/clg/brian/murphy01hidingMasters.pdf), September 2001.
15. S. Pradhan, K. Hacioglu, W. Ward, J. H. Martin, and D. Jurafsky, "Semantic role chunking combining complementary syntactic views," in *Proceedings of the Association for Computational Linguistics 43rd annual meeting*, (Ann Arbor), June 2005.
16. P. Kingsbury and M. Palmer, "From treebank to propbank," in *Proceedings of LREC 2002*, (Las Palmas), 2002.
17. M. Collins, *Head-Drive Statistical Models for Natural Language Parsing*. PhD thesis, University of Pennsylvania, 1999.
18. M. Atallah and V. Raskin, "Demo texts for watermarking of natural language." <http://www.cerias.purdue.edu/homes/wmnlnt/>, 2002.
19. M. Marcus, G. Kim, M. A. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger, "The Penn treebank: Annotating predicate argument structure," in *Proceedings of the ARPA Human Language Technology Workshop*, 1994.
20. G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, , and K. Miller, "Introduction to WordNet: an on-line lexical database," *International Journal of Lexicography* **3**(4), pp. 235–244, 1990.
21. R. Grisham, C. Macleod, and A. Meyers, "Complex syntax: Building a computational lexicon," in *Proceedings of the 15th International Conference on Computational Linguistics*, pp. pp.268–272, 1994.
22. C. T. Schütze, *The Empirical Basis of Linguistics*, University of Chicago Press, Chicago, 1996.
23. E. Charniak, "A maximum-entropy-inspired parser," in *Proceedings of NAACL-2000*, 2000.
24. M. Burke, A. Cahill, M. McCarthy, R. O'Donovan, J. van Genabith, and A. Way, "Evaluating automatic LFG F-structure annotation for the Penn-II treebank," *Journal of Research on Language and Computation; Special Issue on Treebanks and Linguistic Theories* , 2004.

### Appendix: questionnaires

The first questionnaire is given in full. The second used identical materials, with the omission of the three filler sentences, items number 3, 13 and 19. Only the first item is given for illustration.

#### Questionnaire 1: Grammaticality

Please read through the sentences below and decide whether they are grammatical or not. Don't worry about style too much (that's for the next test). Mark them on a scale of +3 (perfect) to -3 (unacceptable, in need of correction). If you like you can underline any parts of the text that you think are suspect. An example of a perfect sentence is "The cat sat on the mat". An unacceptable sentence might be "The cat sat on mat".

1. (0317:812) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Mr. Hahn said that Georgia-Pacific has accounted in its strategy for a "significant downturn" in the pulp and paper industry, an event which he said that would temporarily dilute earnings.
2. (0411:41) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
We did experience some difficulties with the initial LBO terms and, as your article made clear, successfully restructured our debt earlier this year, something that those other retailers have yet to accomplish.

3. (0449:821) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
“Our decision had nothing to do with any pressure Mr. Welch received,” Mr. Carpenter said.
4. (0569:850) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
It’s all in this handout that you don’t want to look at.
5. (0578:204) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Mr. Peters claims that he reminded Mr. Ross Robert Daly and Terry Semel, the top executives of the Warner Brothers studio, had “repeatedly agreed we had every right to accept” an offer such as Sony’s.
6. (0944:1144) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
The quotes could not be found on the tapes, and the two judges that decided the case for Ms. Malcolm and her publishers conceded , for the purpose of their decision, “we assume that the quotations were deliberately altered.”
7. (1013:663) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
The thrift emphasized it has a large portfolio of equity securities issued in connection with corporate restructurings and leveraged buy-outs, which has a book value of \$ 90 million.
8. (1101:216) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Separately, the New York Times reported the Israeli government had provided its correspondent in Jerusalem with different documents which Israel said prove that the PLO has been conducting terrorism from the occupied Arab territories.
9. (1145:378) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
The company has filed an internal reorganization plan that it valued at \$ 2.2 billion that would require 5.5 % rate increases.
10. (1154:990) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Common sense suggests people that play for a company who charges about half what those houses do for a ticket are not in the same market.
11. (1294:240) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
The evidence does not clearly support the view that the downward bias in output growth has become greater during the 1948-89 period, but all which I am claiming is that the growth trend is understated.
12. (1351:25) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Amtech, which makes an automated process system which improves the yields of semiconductor manufacturers, said that profit for the year ended Sept. 30 rose to more than \$ 800,000 from \$ 446,000 last year.
13. (1442:902) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Many Arizona real-estate experts think the worst may be yet to come.
14. (1587:284) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
The company has begun offering shots of gamma globulin, which will diminish the flu-like symptoms of hepatitis A, in anyone that has contracted the disease, Mr. Mattausch said.
15. (1691:1053) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Vista Chemical Co., with three chemical plants in and near Lake Charles, La., “prepares for every hurricane which enters the Gulf of Mexico,” says Keith L. Fogg, a company safety director.
16. (1700:153) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
The usual problem with including communists in “interim” coalition governments is that their ideology and methods require that they squeeze out everyone else.
17. (1826:133) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Pittsburgh figures that it will be dethroned but plans to accept its ouster graciously.

18. (1857:1140) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Orin Kramer, an insurance consultant in New York, estimates the 1906 San Francisco destruction, on an inflation-adjusted basis, included insured losses of \$ 5.8 billion.
19. (2151:55) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
First, he suggests, GPA Group Ltd., the international aircraft leasing company based in Ireland, could lease some of its Boeing jetliners to the Soviet airline.
20. (2329:2) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Ideal Basic Industries Inc. said that its directors reached an agreement in principle calling that HOFI North America Inc. to combine its North American cement holdings with Ideal in a transaction which will leave Ideal's minority shareholders with 12.8 % of the combined company.

## Questionnaire 2: Preservation of meaning and style

The sentences you have just read had been altered. Below most of the sentences are repeated, showing what changes were made to them – what was inserted (underlined) and deleted (struck-through) as such: “some text ~~that~~ which has been altered”.

Please read through the sentences and decide whether they retain the meaning and style of the original (this time don't worry about grammaticality). Mark them on a scale of +3 (identical in meaning and style) to -3 (unacceptably different, in need of correction). A perfect transformation might be “John gave Mary a flower ~~to Mary~~”, while an unacceptable one might be “John gave Mary to a flower ~~to Mary~~”.

Note that you only have to compare the original and final versions of each sentence. So for “~~a~~ the sentence with two ~~changes~~ alterations” you need only compare the original “a sentence with two changes” with the final “the sentence with two alterations” (i.e don't worry about “a sentence with two alterations” and “the sentence with two changes”).

1. (0317:812) [perfect ( +3 +2 +1 0 -1 -2 -3 ) unacceptable]  
Mr. Hahn said that Georgia-Pacific has accounted in its strategy for a “significant downturn” in the pulp and paper industry, an event ~~that~~ which he said that would temporarily dilute earnings.