# From Extracts to Abstracts: Human Summary Production Operations for Computer-Aided Summarisation

Laura Hasler
Research Group in Computational Linguistics
RIILP/SHLSS, University of Wolverhampton
Stafford St., Wolverhampton, WV1 1SB
L.Hasler@wlv.ac.uk

## Abstract

This paper presents a classification and evaluation of human summary production operations used to transform extracts into more concise, coherent and readable abstracts. Computer-aided summarisation (CAS) allows a user to post-edit an automatically produced extract to improve it. However, unlike other areas of summarisation, no guidance is available to users of CAS systems to help them complete their task. The research reported here addresses this by examining linguistic operations used by a human summariser to transform extracts into abstracts. An evaluation proves that the operations are useful; they do improve coherence when applied to extracts.

## Keywords

Computer-aided summarisation, coherence, post-editing, linguistic transformations, evaluation.

## 1. Introduction

Computer-aided summarisation (CAS) has recently been proposed as a feasible alternative or complement to fully automatic summarisation [20], [21]. CAS allows a human summariser to interact with an automatic system to produce the best possible summary. Users have the opportunity to post-edit the extract produced by the system, thereby integrating aspects of automatic and human summarisation. However, in order to benefit fully from CAS, they need to know exactly how best to improve the automatic output to transform it into a high quality, coherent and readable summary. In the past no guidance has been available for the summariser attempting to post-edit the extract to achieve this. Just as guidelines in automatic and human summarisation (e.g. [1], [11], [2]) are necessary to facilitate consistency and quality, they are also necessary in CAS. To develop such guidelines, an analysis of the operations a human summariser applies is crucial.

This paper presents a corpus-based classification of human *summary production* operations, which was then translated into a set of guidelines (not described further due to their similarity to the classification). It also evaluates the extent to which the operations are useful, by assessing the coherence and readability of abstracts created by applying the operations to extracts. The remainder of the paper is structured as follows: Section 2 discusses the summary production stage of summarisation. Related work on the editing of summaries is detailed in Section 3. Section 4 presents the corpus and an overview of the classification. Section 5 examines the *atomic* summary production operations observed in the corpus, and Section 6 deals with *complex* operations. Section 7 presents an evaluation of the operations. The paper finishes with conclusions.

## 2. Summary production in CAS

CAS utilises Endres-Niggemeyer's model of human summarisation [8], which identifies three different stages in the summarisation process: *document exploration*, *relevance assessment* and *summary production*. Stages one and two correspond to the automatic stage of CAS, where important information in a document is located and assessed for relevance. The third stage of summarising, which, in CAS, is where the user post-edits the output of the system, is most relevant here. In the *summary production* stage, the actual creation of the summary as a discrete unit takes place. This mainly involves *cutting and pasting* material from the original document using sentence patterns typical of the domain. However, because the material is taken from different places in the source, it must be edited to ensure that it contains the most appropriate information in the most appropriate form. Summary production in CAS can be slightly different, in that it can also take relevant sentences together as an extract before operations are applied; however, this extract still needs to be transformed into a more coherent and concise abstract.

Summary production is the least addressed stage of summarisation in terms of human participation in the automatic summarisation (AS) process. In AS, humans are most often asked to annotate sentences which they consider important and worthy of including in a summary of the text they are annotating. However, in CAS there is an element of human interaction in terms of *post-editing*. Whilst there is advice for summarisers regarding content, there is none for improving the readability/coherence of the information selected (the *extract*). This means the summary production stage of CAS can be inconsistent and inefficient.

## 3. Related work

Research on transformations applied during summarisation is relatively sparse and does not attempt to investigate operations for CAS, looking instead at source to summary transformations. Operations observed by some researchers

are given as general categories and although they offer examples, there is no explicit classification describing sub-operations or typical forms which operations usually take (e.g. [7], [13]). The function[1] of the operations or the units to which they are applied is also not addressed in depth. However, when a more detailed description of operations is offered, far fewer general operations are covered (e.g. [8], [5], [6]). The applications of these operations can be so specific that it is difficult to gauge how useful they would be if a set of guidelines was formulated from them. The classification in this paper offers general classes covering more specific individual sub-operations triggered by certain forms, with reasons for application and examples.

In terms of automatic summarisation, various typical problems which affect coherence, such as dangling anaphors and discourse ruptures, have been identified [17], [19], [15]. Ways of dealing with these problems include *surface rejection rules* such as eliminating text in brackets and quotation marks (e.g. [18]), as well as other units such as certain adverbs and prepositional phrases (e.g. [16]). Ignoring sentences containing anaphors or adding sentences containing their antecedents is another option (e.g. [14], [19]). Reference adjustment can also be employed to maximise coherence and minimise repetition (e.g. [16]). However, these operations are very restricted and the classification below shows that human summarisers consider much more than these types of shallow coherence smoothing when producing an abstract.

# 4. Corpus and classification

## 4.1 Corpus of extracts and abstracts

The corpus consists of 43 pairs of extracts and their corresponding abstracts (see Table 1), whose source texts come from *New Scientist*, extracted from the *BNC* [4].

**Table 1. Corpus statistics**

| | Texts | Sentences | Words |
|---|---|---|---|
| **Extracts** | 43 | 538 | 13,275 |
| **Abstracts** | 43 | 328 | 8,652 |
| **Total** | 86 | 866 | 21,927 |

Extracts were created manually by an annotator, simulating the first two stages of summarisation which are performed automatically in CAS. Extracts were manually created because this results in better extracts to be taken as the starting point for transformation. The analysis can then focus solely on summary production stage rather than the previous stages; the post-editor does not have to deal with typical 'mistakes' of automatic summarisation systems in terms of information content (this is considered relevance assessment). The extracts were created using an adapted set of [11]'s annotation guidelines, and are 30% of the source. A human summariser produced abstracts freely by making the extracts more concise, coherent and readable. This step is essential because it simulates the actions of a human summariser using a CAS system to produce abstracts. The only instruction given was that abstracts should be 20% of the source. Human summarisers always shorten a summary when they edit it to make it more concise and coherent; therefore it is reasonable for this to also happen in CAS.

## 4.2 Classification overview

Five general classes of operation were identified, split into two types: *atomic* and *complex* operations. The basic idea is that atomic operations cannot be broken down into further general classes of operation, whereas complex operations can, i.e., complex operations consist of atomic operations. The atomic operations identified in the corpus are *deletion* and *insertion*. As well as being operations which are applied in their own right, they make up the complex operations *replacement*, *reordering* and *merging*.

The operation classes defined here each encapsulate a range of sub-operations, which are determined by certain surface forms acting as *triggers* by which the operation can be identified. The reason for classifying the operations in terms of triggers is that surface forms are easier to recognise than functions of units. Time is of the essence in summarisation, meaning that easily recognisable forms which the summariser can quickly identify are useful. The function of units is vital, as only those fulfilling a certain function are subject to sub-operations. However, it must be remembered that it is impractical to work with functions alone without any means of reliably identifying the units which typically embody them. Triggers are identified, and then the function of the unit is established to assess its suitability for transformation. It should be pointed out here that there are *always* cases where sub-operations are not applied, and that, very often, more than one sub-operation is applied to the same unit, which can complicate matters.

# 5. Atomic operations

## 5.1 Deletion

*Deletion* is defined as *the process of removing a unit*[2] *from a certain place in the extract so that it does not appear in the same place in the abstract*. This is the most basic operation to categorise as it does not involve any reformulation of the extracted text, although it does frequently occur in combination with other operations which do. The reason for using the deletion operation as an atomic operation is simple: to shorten the text. Deletion achieves this 'safely' by removing non-essential units, such as details, repeated information and information which is

[1] The term *function* is used in its most general sense here.

[2] *Unit* covers words, phrases, clauses, sentences and punctuation.

not necessary to disambiguate or clarify. Only information which is either not needed at all or is present explicitly or implicitly elsewhere is deleted. When used to form a complex operation, its function is also to ensure grammaticality and coherence. Deletion can occur at any level of granularity and has ten sub-operations.

The sub-operation **DELETE: SENTENCE** is applied to remove sentences which add unnecessary detail or repeat information present elsewhere. The deleted sentence shown below gives further detail in the form of examples:

- *According to one theory, held by some psychiatrists, patients may find doctors increasingly unsympathetic to their plight.* ~~*They may be neurotics or malingerers or both.*~~

Subordinate clauses are deleted using **DELETE: SUB_CLAUSE** and can be recognised by subordinating conjunctions (e.g. *because*, *while*) and relative pronouns (e.g. *who*, *which*). The fact that particular information is presented in a form signalling subordination means that, as well as being relatively easy to recognise, the summariser can remove these units without having to make further changes to the remaining clause. An example from the corpus, showing information inferable from elsewhere in the text (which is about the complexity of diabetes) is:

- *Three papers published recently in Science move us a little closer to understanding the basis of the disease*~~*, which turns out to be highly complex.*~~

Prepositional phrases (PPs) are deleted using **DELETE: PREP_PHRASE** because they contain information often present explicitly or implicitly elsewhere. In the source they may be necessary for disambiguation, however, this is not always the case in the abstract as it contains fewer topics and sentences referring to the same entity are presented much closer together. A preposition appearing immediately after a noun and being a constituent of a noun phrase is the trigger by which the full PP is recognised, allowing it to be identified as a potential candidate for deletion. PPs are usually deleted when functioning as postmodifiers and adverbials, because this does not necessarily change the meaning of the sentence. E.g.:

- *The four volumes* ~~*of this work*~~ *are soon to be accompanied by a further five…*

Adverb phrase deletion is similar to PP deletion, although in this case the trigger is an adverb. For simplicity and due to the nature of the transformation process, **DELETE: ADVERB** includes adverbs, units introduced by adverbs and phrases which have adverbs as the head. E.g.:

- *Potential and actual savings which have been identified include: shorter lead times,* ~~*presently as long as six months;*~~

**DELETE: REPORT** is used to delete reporting clauses, and is extended to include reported findings in general, rather than just speech, because the structures of reporting in both cases are very similar. In cases of indirect speech or reporting, the conjunction *that* following the reporting verb is also deleted unless it is needed to ensure grammaticality in instances of merging, and is therefore included in this sub-operation. These units are not suitable for deletion when it is important that the reader knows who to attribute the speech or claim to; this is particularly relevant where there are conflicting opinions in the summary. E.g.:

- ~~*Fairbanks found that*~~ *when sea level first began to rise as the ice sheets melted, 17,000 years ago, it did so at a rate of about 4mm per year.*

Modifiers of noun phrases (NPs) such as adjectives and nouns, or even whole noun phrases, are deleted using the sub-operation **DELETE: NP**. Parts of NPs which are no longer needed to disambiguate the head because this has been done earlier in the text can be safely deleted, as can other information which can be inferred from elsewhere in the document. Sometimes, whole NPs appear to further describe another which is present in the text, or are repeated close to another mention of the same entity and therefore are repetitive; these can also be deleted. E.g.:

- *It hit* ~~*the coast at the town of*~~ *Innisfail at 0700* ~~*Eastern Standard Time (AEST) on*~~ *Monday.*

The sub-operation **DELETE: DETERMINER** removes determiners, usually definite articles, and has the main function of shortening the text without impeding readability. An additional function is to transform the first sentence of an abstract into a conventional 'headline' when applied there. It is also used when an NP is mentioned for the first time due to the deletion of other units which previously introduced it. This sub-operation is dependent on each individual text in a different way to the others discussed so far, in that it is used as a 'last resort' to reach the compression rate; some texts are easier to reduce than others without needing to do this. For example:

- ~~*The*~~ *scientists' work confirmed that the undoubted improvement in parachuting ability was a fortuitous result of the development of* ~~*the*~~ *fringes for* ~~*these*~~ *other purposes.*

**DELETE: BE** deletes the verb *be* from constructions such as *are removed* when it is present earlier in the unit. Similar to determiners, the deletion of such small units emphasises the importance of text reduction and depends on the individual text. It also functions to transform the first sentence of a news text into a 'headline'. In fact, the two sub-operations together could be taken as a simple formula by which summarisers of news texts can create a 'headline' for their summary if one is not already included. Other deletion operations may also be applied. E.g.:

- *Britain* ~~*is*~~ *among the front runners as tomorrow's supercomputers take shape.*

**DELETE: FORMAT** deletes text appearing in a certain format, i.e., within brackets, following a colon or a semi-colon, and in between or following dashes. These deletions are relatively few considering the number of instances. Very few cases of specially formatted text were predicted to be in the extracts because of the guidelines used to produce them. The fact that not only was this type of unit present in the extract when the annotator was instructed not to include it, but that it was also retained in the abstract, indicates that some rejection rules used in AS (Section 3),

for the news texts in this corpus at least, are not appropriate. Despite this, specially formatted text is an easily recognisable unit which can be safely deleted in certain circumstances. For example:

*- Over the same distance, people use two and a half times more energy ~~(per kilogram of their bulk)~~ than the average camel.*

**DELETE: PUNCTUATION** removes commas, full stops, semi-colons, colons and quotation marks. This is necessary because punctuation is not always accurate in the source.

## 5.2 Insertion

*Insertion* is defined as *the process of adding a unit which is not present in the extract into the abstract*. As with deletion, it is also relatively easy to recognise and classify as a broad class, as it does not involve any reformulation of the extracted text. Insertion as an atomic operation is concerned with readability/coherence, as no consideration seems to be given overtly to space considerations or compression rate. However, insertion also serves to clarify whilst saving space, by adding smaller elements such as modifiers instead of larger units to perform the same function. Insertion occurs at word, phrase and clause level, but there are no instances of complete sentences being inserted. This strongly suggests that, whilst smaller units are inserted to improve coherence and highlight information, human summarisers would not waste compression to such an extent by inserting a whole sentence from scratch. Four sub-operations were identified.

Connectives are inserted into abstracts to make them more coherent by explicitly indicating relationships between units. **INSERT: CONNECTIVE** helps to improve the flow of the text because it helps to avoid ambiguity of interpretation. It includes conjunctions (coordinators and subordinators) and adverbs functioning as conjuncts. Whilst this sub-operation may appear broad because it includes more than one word class, it is suitable for the present analysis because all occurrences function to connect textual units and have a relatively limited set of recognisable triggers. Sentence-initial connectives are easiest to recognise as they are inserted immediately after a full stop. *However* was the most common connective inserted, serving to emphasise a contradiction between a sentence and previously mentioned information. E.g.:

*- He claims that his dummy head does likewise. **However,** it is unclear whether this theory is serious or a misunderstanding.*

**INSERT: FORMULAIC** is particularly interesting because it most obviously illustrates a disregard for the basic concept of summarisation as text reduction. This sub-operation emphasises information and its relationship with other units. The units inserted are larger than the connectives mentioned above and are related to the standard sentence patterns identified in the summary production stage by [8]. This shows that explicit coherence and emphasis is an important issue because of the risk it poses to using up compression. It could be argued that this sub-operation of insertion also fits into the *merging* and *replacement* operations because certain information is inferred from elsewhere in the text and presented in a different way. However, it is dealt with under this heading because it is extremely difficult to match these units with any material in the extract. This type of possible multiple interpretation highlights the difficulty of formulating a clear-cut classification. The following example illustrates an insertion which changes the subject of the sentence to make it the same as other subjects in the abstract, thereby emphasising it and improving coherence:

*- **He sees** the need to raise public awareness and demystify science and technology **as** a key point…* [X sees Y as Z]

In addition to being deleted during the production of abstracts, modifiers are also inserted. The reasons behind specific applications of **INSERT: MODIFIER** are similar to those which explain why some modifiers but not others are deleted. By introducing a modifier of a noun or noun phase, any ambiguity is dealt with in a concise unit rather than requiring a fuller explanation, for example in the form of an additional sentence. The example shows the insertion of a noun phrase necessary because the verb *shown* leaves some ambiguity about what *Men of Science* is:

*- **The TV series** Men of Science is now being shown in a few other areas.*

Punctuation (usually commas, but also hyphens, colons and full stops) is inserted using the sub-operation **INSERT: PUNCTUATION**, to make the text easier to read.

# 6. Complex operations

## 6.1 Replacement

*Replacement* is defined as *the deletion of one unit and the insertion of a different unit in the same place in the text*. This is the 'easiest' of the complex operations to deal with because similar to atomic operations, the summariser works with units which usually (but not always) have recognisable triggers. However, because replacement is often used in combination with *reordering* and *merging*, it is not always easy to identify. The smaller the unit replaced, the easier it is to classify. The main reasons for using the replacement operation include shortening the text, avoiding or minimising repetition, and changing units so that they appear more like those around them to improve readability. Eight sub-operations were identified.

The sub-operation **REPLACE: PRONOMINALISE** involves deleting a noun phrase in an extract and inserting a pronoun in its place in the abstract. It also involves the replacement of determiners with demonstrative and possessive pronouns, or of whole NPs with demonstratives. As well as mnimising repetition, pronominalisation usually (but not always) saves space in the abstract. For example:

*- The Countryside Commission has just announced an imaginative scheme for 12 "community forests" on marginal urban fringe farmland. But ~~the community forests - plus the Commission's other plan for 100,000 acres of new forest in the~~*

~~*Midlands, which will also soak up redundant land - together with new land released for development,*~~ **THIS** *will barely scratch the surface of the problem.*

**REPLACE: LEXEME** replaces one lexical item with another which refers to the same entity, concept or action, or something related to it, by means of synonymy, hyponymy, hypernymy or metonymy. The replaced unit can be more or less precise than its corresponding unit in the extract. REPLACE: LEXEME is applied to NPs, verbs, adjectives and adverbs, but the most obvious cases are those involving NPs. The following example illustrates the replacement of the name of a person working for or representing an organisation by the name of the organisation, which the reader is more likely to recognise. This simplifies the text at the same time as shortening it. The replacement of the verb highlights the effect of style on the summary production process, replacing a relatively neutral word with one typical of news reporting:

- ~~*Zhanat Carr, a radiation scientist with the WHO in Geneva,*~~ **the WHO** ~~*says*~~ **admits** *the 5000 deaths were omitted because the report was a "political communication tool".*

The sub-operation **REPLACE: RESTRUCTURE_NP** restructures noun phrases to shorten units by transforming postmodifiers into premodifiers which usually contain fewer words. The most common case involves PPs as postmodifiers, but there are cases of relative clauses and adverb phrases. Prepositions are deleted from PPs postmodifying nouns and the remaining part of the PP is relocated in front of the original NP head. Similarly, with relative clauses postmodifying nouns, the relative pronoun following the NP head is deleted and the rest of the relative clause is replaced before the head. In both cases, other transformations may be necessary to ensure grammaticality. A MODIFIER+HEAD structure is preferred to HEAD+MODIFIER in abstracts because it makes the text more 'snappy'. Several sub-operations relate to NPs and PPs, and REPLACE: RESTRUCTURE_NP suggests that if these units cannot be deleted, pronominalised or substituted by an appropriate lexical item, they can at least be shortened to pack the necessary information into the smallest possible space. E.g.:

- *Photopia denies that the evident haste is the result of poor* ~~*sales in the US*~~ **US sales***.*

The sub-operation **REPLACE: NOMINALISE** functions to improve readability and shorten the text by expressing the same information in fewer words. It occurs in combination with merging because it can replace a whole clause or sentence but does not form a grammatical sentence on its own. Other operations can also be applied. For example:

- *All this is hardly Culver's fault.* [*The same difficulties are to be found in all other parts of evolutionary ecology.*]→**These general difficulties of evolutionary ecology** *are hardly Culver's fault.*

Pronoun expansion is applied via **REPLACE: REFERRED**, where information (usually an NP) from one extract sentence replaces a pronoun in a different abstract sentence (also present in the extract). In these cases, the first sentence is not important enough to be included in the summary on its own merit, but it includes a small piece of information necessary for the full understanding of a sentence which is included. As an example:

- ~~*You could be forgiven for not having heard of the Treaty of Tlatelolco.*~~ *The non-aligned nations have just dragged* ~~*it*~~ **the Treaty of Tlatelolco** *on the stage…*

**REPLACE: VP** deals with verb changes made necessary by other sub-operations, for example, certain deletions or substitutions, to ensure grammaticality. It also addresses the simplification of units by transforming verb phrases to make the text easier/quicker to read. Lexical substitution also sometimes occurs in combination with the verb change. The general rule gleaned from the application of REPLACE: VP is to prefer the simple present or past tense where possible, but to *always* transform verb phrases in order to ensure grammaticality. The example illustrates REPLACE: VP due to the deletion of a reporting construction:

- ~~*DEFRA told WWT*~~ *samplers* ~~*to moisten*~~ **moistened** *a sterile swab on a stick with saline,* ~~*take*~~ **took** *a faecal sample from the bird, then put the swab back in its dry plastic tube*~~*. The tubes were kept*~~**, kept it** *at refrigerator temperature and* ~~*taken*~~ **took it** *to* ~~*the testing*~~ *laboratories the next day.*

**REPLACE: PASSIVISE** is worth noting because it contradicts advice available to professional abstractors [2], [1], and strategies used by them [8]. This sub-operation depersonalises sentences by passivising them. It removes subjects of sentences which are irrelevant, not mentioned later in the text or cannot easily be resolved (as with *we* in the example below). This shortens the abstract overall.

- *It was a polymer so unlike the polymers known at the time that no one could envisage a use for it. And* ~~*we couldn't make it*~~ **it couldn't be made** *consistently.*

A simple way to utilise space well is to use standard abbreviations, such as acronyms, initialisms, and symbols. Common applications of the sub-operation **REPLACE: ABBREVIATE** involved replacing *kilometres* with *km*, *per cent* with *%* and written numbers with their corresponding figures. Abbreviations referring to organisations or projects were kept rather than replaced with their expanded forms. This is acceptable because they are relatively common and easy to expand in the given context, e.g., *WHO*, *EC*, *UK*. Acronyms and initialisms are only deleted or replaced by their full form if they are not mentioned again in the abstract. These observations suggest that it would also be appropriate to replace full forms of common abbreviations with their short forms as part of the text reduction process.

## 6.2 Reordering

*Reordering* is defined as *the deletion of a unit from one place in the extract and its insertion in a different place in the abstract*. It can also involve aspects of replacement, and may seem to overlap to some extent with merging. However, units can be reordered both within a sentence and without being incorporated into another unit. Merged units are not necessarily reordered, for example, if units in

between them are simply deleted. Unlike the other operations, reordering does not have recognisable triggers because any unit can be repositioned; they are too diverse to classify based on surface forms. Reordering is therefore discussed in terms of the two sub-*functions* identified.

**REORDER: EMPHASISE** repositions information earlier in the abstract than it appears in the extract to emphasise its importance. For example, in a text about face transplants, sentences about tissue-rejection were moved from S7 and S8 position (the last two sentences) in the extract to S3 and S4 in the abstract, which contained 6 sentences. The reason for this is that tissue-rejection was considered more important than scarring, which appeared first in the extract. **REORDER: COHERENCE** improves coherence and makes the abstract easier to read by positioning units about the same thing together. This means that the 'topic' of the abstract does not alternate, making it easier for the reader to process. It also includes repositioning sentences which introduce or summarise the information in the extract to the beginning or end of the abstract. For example, in an abstract about school memories, a sentence referring to teachers was reordered to appear immediately after other 'teacher' sentences. In the extract, these sentences are separated by sentences recalling a comic. As only one of the 'comic' sentences is included in the abstract, it is more coherent to place it after all the 'teacher' sentences.

## 6.3 Merging

*Merging* is defined as *taking information from different units in the extract and presenting it as one unit in the abstract*. It can be further described as often including elements of replacement and reordering. Merging can take place within and across sentences, and is believed to be the operation which best captures the essence of abstracting. Merging functions to make the text more concise by fitting as much relevant information as possible into the available space. It allows the most appropriate parts of the most appropriate units to be combined, and as it often occurs with replacement, these can then be expressed in the most appropriate way. Similar to reordering, merging also emphasises information, by combining it with information in other units. Two sub-operations were identified.

**MERGE: RESTRUCTURE** changes the structure of units whilst inserting them into a larger unit. There are numerous variations in triggers, especially as other operations such as replacement and reordering are so often applied to the same unit. This is the reason for grouping them into one large sub-operation. It involves the transformation of units into subordinate clauses introduced by relative pronouns or non-finite verbs in previous sentences. Other realisations are units appearing as prepositional phrases modifying objects of previous sentences, adverb phrases and adjectives. Units are also transformed into coordinate clauses, and declarative sentences containing information from both question and answer sentences are formed from interrogatives. However, there are no apparent patterns as to when this happens. The summariser merged units intuitively to achieve maximum conciseness and coherence depending on the abstract being produced. Because not only existing surface forms are abstracted and merged, but the concepts represented by them, information can take any (appropriate) form. The combination of operations coupled with the difficulty of assigning sub-classes is typical of human summary production operations which embody abstracting, highlighting the complexity of a classification.

The example below is one of the best illustrations of how human summarisers can identify a relevant piece of information, reduce it based on their background or world knowledge, and locate it in the most appropriate place in an abstract. This example shows just how well merging (and replacement) works to reduce the text and make it concise and easier to read, with two sentences (37 words) transformed into one adjective in an earlier sentence:

*- In October 1980 {Zuccarelli} filed a European patent application, covering nine countries including Britain*[*. … The cost of pushing a European patent through in nine countries is around $10000. The cost of application alone is around $2000 and Zuccarelli has already paid an extra $500 for a further stage of official examination*].→*In October 1980, {he} filed an* [{**expensive**}] *European patent application, covering nine countries including Britain.*

**MERGE: PUNCTUATION-CONNECTIVE** deletes/replaces/ adds punctuation and connectives and is necessary to 'join' sentences. The use of commas during merging helps with readability, whilst connectives ensure grammaticality and improve the flow of the text. Commas are most often used, but dashes are also an option. *And* is the preferred connective, with *as*, *so*, *but* and *although* also featuring.

# 7. Evaluation

This section presents the evaluation of the classified operations, applied to a different set of extracts to assess their generality and suitability for the task of post-editing extracts of news texts. The evaluation centres on whether applying the operations results in an abstract which is more coherent than the extract from which it was produced. Existing evaluation methods which focus on coherence and readability [12] are unsuitable for this task for several reasons. Coherence is measured based on features such as dangling anaphors, discourse ruptures, and grammatical accuracy, which should not be issues in the abstracts evaluated here due to their 'creation' by a human summariser. Using intuitive human judgment alone can be subjective. Standard readability measures only consider word/sentence length, which is not enough to account for a coherent and readable abstract. Edit distance cannot be used because many sub-operations are often applied to the same unit, rendering it almost unrecognisable. The human summarisation literature does not address evaluation as

such, focusing instead on checking and editing to ensure that abstracts adhere to an organisation's specifications. Therefore an alternative means of evaluation is necessary.

## 7.1 Centering Theory for evaluation

Centering Theory (CT) [9] is a parametric theory of discourse structure encompassing local coherence and salience. [10] proved its potential suitability for the evaluation of summaries by assessing pairs of extracts and comparing the CT analysis with human judgment. CT deals with coherence by examining repetitions of entities across consecutive utterances, and the relationship between these repetitions based on their position. Entities in utterances are labelled *centers*, and each utterance introduces *Cf*s (forward-looking centers). The highest-ranked *Cf* is the *Cp* (preferred center). The *Cp* is predicted to be the center in the next utterance which provides the link between the two consecutive utterances. The center which does provide the link from one utterance to the previous utterance is the *Cb* (backward-looking center). The *Cb* of any current utterance is the most highly ranked *Cf* of the previous utterance which is realised in the current utterance. The relationship between *Cp*s and *Cb*s results in transitions which reflect the coherence of a discourse (see Table 2). From most to least coherent, these are: CONTINUE, RETAIN, SMOOTH SHIFT, ROUGH SHIFT (e.g. [3], [24]).[3] It is also possible to have NO TRANSITION between utterances. In the present research, NO TRANSITION is split into INDIRECT and NO Cb.

**Table 2. Centering Theory transitions**

| | $Cb(U_{n+1}) = Cb(U_n)$ or $Cb(U_n)$ undefined | $Cb(U_{n+1}) \neq Cb(U_n)$ |
|---|---|---|
| $Cb(U_{n+1}) = Cp(U_{n+1})$ | CONTINUE | SMOOTH SHIFT |
| $Cb(U_{n+1}) \neq Cp(U_{n+1})$ | RETAIN | ROUGH SHIFT |

## 7.2 The evaluation procedure

A metric was developed to reflect both the traditional preference order for transitions and the effect of the transitions in summaries. This was necessary to reward transitions positively affecting coherence and penalise those with a negative effect. The differences between the weights rather than the values of the weights themselves are most important, as they reflect the comparative effects of transitions. The metric is as follows: CONTINUE: +3; RETAIN: +2; NO TRANSITION (INDIRECT): +1; SMOOTH SHIFT: -1; ROUGH SHIFT: -2; NO TRANSITION (NO Cb): -5.

To evaluate the summary production operations, a new set of (extract, abstract) pairs was created. 25 texts were taken from the *Reuters* corpus [23] and *New Scientist*, covering the domains of science, business, politics and sport. 30% extracts of these texts were produced manually and automatically (using term weighting in the CAST system), and were then transformed into 20% abstracts by a human summariser using summary production guidelines derived from the classification above. This resulted in 50 pairs of extracts and abstracts, which were analysed using CT with the following parameter instantiations. Utterance: sentence, realisation: direct + possessive pronouns, *Cf* list ranking: grammatical (see [22] for an overview of alternative instantiations). The average transition score per summary was calculated, which is necessary for an accurate comparison of each pair because abstracts virtually always contained fewer sentences (and therefore transitions) than their corresponding extracts. The summary with the higher transition score is the more coherent.

## 7.3 Results and discussion

CT evaluated 78% of abstracts as more coherent than the extracts from which they were produced. 2% (one pair) were considered equal, leaving 20% of extracts evaluated as more coherent than the post-edited abstracts. Extracts were sometimes evaluated as better than abstracts due to the entity-based nature of coherence in CT, coupled with the tendency of automatic summarisation methods to select repetitive sentences. CT evaluates two sentences with the same entities in the same position as very coherent, whereas in a summary it is not desirable to have such repetitions. When the summariser created abstracts, they dealt with this issue, but their deletions and reordering applied to improve readability affected the CT assessment.

Human judgment was also obtained to ensure that other aspects of coherence/readability were not overlooked (e.g. connectives, replacement, restructuring). The judge was asked to select the more readable/coherent summary out of an (extract, abstract) pair, but was not told which text was which. 82% of abstracts were judged as better than the extracts from which they were produced. 2% (one pair) could not be evaluated, leaving 16% of extracts evaluated as more coherent than their corresponding abstracts. It was predicted that a judge would always prefer abstracts to extracts. However, automatic extracts do not always focus on the main topic of the source, or the same aspects of it, and in some cases the judge preferred summaries which gave information about various topics or aspects rather than focusing on the main topic. Where operations were applied to delete units about different topics or aspects, the summaries contained different information. The judge found it impossible to assess the summaries on readability/ coherence alone, and commented that the information in a summary was an important element in the evaluation.

The agreement between the CT and human evaluations was 70%, disagreement was 26%. Two pairs (4%) were not compared because they could not be evaluated (see above). Chi-square was applied, and revealed that the disagreement between the two evaluations on (extract, abstract) pairs was

[3] The smooth/rough shift distinction is not made by [9].

not statistically significant, at a level of $p \leq 0.001$. The differences are due to each evaluation considering different aspects of summary coherence in their assessment.

# 8. Concluding remarks

Whilst CAS has been proposed as a feasible alternative to automatic summarisation, to date no guidance has been available to users of CAS systems to help them in their summary production task. Guidelines are used in other areas of summarisation to ensure consistency and quality, and so it is only fair to allow people involved in the CAS process access to such resources. This paper presented a corpus-based classification of linguistic operations used by a human summariser to transform extracts into more concise, coherent and readable abstracts. These operations were translated into a set of guidelines and were applied to a different set of extracts to assess their suitability for the task of post-editing extracts during the summary production stage of CAS. The evaluation, using a metric developed for CT and a human judge, proved that the operations observed in the corpus are useful and that they do improve the coherence and readability of summaries of news texts. They can therefore be issued to users of CAS systems to ensure that the summaries they produce with the help of such systems are consistently and reliably coherent.

In future, the evaluation should be extended to include more summarisers and judges, and to compare the coherence of abstracts produced both with and without the summary production guidelines. In addition, the integration of informativeness and coherence in the evaluation is desirable. It would also be interesting to investigate operations applied in other genres, such as research papers, which are used in AS and summarised by professionals.

# 9. References


[1] American National Standards Institute. *Guidelines for Abstracts*. ANSI/NISO Z39.14-1997, National Information Standards Organization (NISO) Press, Bethesda, 1997.

[2] H. Borko and C. Bernier. *Abstracting Concepts and Methods*. Academic Press, San Diego, 1975.

[3] S.E. Brennan, M.A. Friedman and C.J. Pollard. A Centering Approach to Pronouns. In *Proceedings of ACL'87*, 155-162. Stanford, 6-9 July 1987.

[4] L. Burnard. *Users Reference Guide: British National Corpus Version 1.0*. Oxford University Computing Services, 1995.

[5] C.-K. Chuah. Just What May be Deleted or Compressed in Abstracting? In *Proceedings of TALN2001*, 339-344. Tours, 2-5 July 2001.

[6] C.-K. Chuah. Aggregation by Conflation of Quasi-Synonymous Units in Author Abstracting. In *Proceedings of TALN2001*, 143-152. Tours, 2-5 July, 2001.

[7] E.T. Cremmins. *The Art of Abstracting*. Information Resources Press, Arlington, 1996.

[8] B. Endres-Niggemeyer. *Summarizing Information*. Springer, Berlin, 1998.

[9] B.J. Grosz, A.K. Joshi and S. Weinstein. Centering: a Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics* 21(2): 203-225, 1995.

[10] L. Hasler. An Investigation into the Use of Centering Transitions for Summarisation. In *Proceedings of CLUK'04*, 100-107. Birmingham, 6-7 January 2004.

[11] L. Hasler, C. Orasan and R. Mitkov. Building Better Corpora for Summarisation. In *Proceedings of Corpus Linguistics 2003*, 309-319. Lancaster, 28-31 March 2003.

[12] L. Hirschman and I. Mani. Evaluation. In R. Mitkov (ed.). *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford, 414 – 429, 2003.

[13] H. Jing and K. McKeown. The Decomposition of Human-Written Summary Sentences. In *Proceedings of SIGIR-99*, 129-136. Berkeley, 15-19 August 1999.

[14] F.C. Johnson, C.D. Paice, W.J. Black and A.P. Neal. The Application of Linguistic Processing to Automatic Abstract Generation. *Journal of Document and Text Management* 1(3): 215-239, 1993.

[15] I. Mani. *Automatic Summarization*. John Benjamins, Amsterdam/Philadelphia, 2001.

[16] I. Mani, B. Gates and E. Bloedorn. Improving Summaries by Revising Them. In *Proceedings of ACL'99*, 558-565. College Park, 20-26 June 1999.

[17] J.-L. Minel, S. Nugier and G. Piat. How to Appreciate the Quality of Automatic Text Summarization? Examples of FAN and MLUCE Protocols and their Results on SERAPHIN. In *Proceedings of the ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization*, 25-31. Madrid, 11 July 1997.

[18] R. Mitkov, D. Le Roux and J.P. Descles. Knowledge-based automatic abstracting: Experiments in the sublanguage of elementary geometry. In C. Martin Vide (ed.). *Current Issues in Mathematical Linguistics*. North Holland, Amsterdam, 415-421, 1994.

[19] H. Nanba and M. Okumura. Producing More Readable Extracts by Revising Them. In *Proceedings of COLING-2000*, 1071-1075. Saarbruecken, 31 July-4 August 2000.

[20] C. Orasan and L. Hasler. Computer-aided summarisation: How much does it really help? In *Proceedings of RANLP'07*. Borovets, 27-29 September 2007, forthcoming.

[21] C. Orasan, R. Mitkov and L. Hasler. CAST: a Computer-Aided Summarisation Tool. In *Proceedings of EACL'03*, 135-138. Budapest, 12-17 April 2003.

[22] M. Poesio, R. Stevenson, B. di Eugenio and J. Hitzeman. *Centering: A Parametric Theory and its Instantiations*. CS Technical Report CSM-369, University of Essex, 2004.

[23] T.G. Rose, M. Stevenson and M. Whitehead. The Reuters Corpus Volume 1 - from Yesterday's News to Tomorrow's Language Resources. In *Proceedings of LREC2002*, 827-833. Las Palmas de Gran Canaria, 29-31 May 2002.

[24] M.A. Walker, A.K. Joshi and E.F. Prince. Centering in Naturally Occurring Discourse: An Overview. In M.A. Walker, A.K. Joshi and E.F. Prince (eds.). *Centering Theory in Discourse*. Oxford University Press, Oxford, 1-28, 2004.