

## Co-interpretation Network in English Discourse

Ni Yibin\*

*This paper presents a far more detailed account of co-interpretation network among noun phrases in English discourse than the one laid out in Halliday and Hasan (1976). This system will provide a new foundation for the computational tagging of English texts and assist effective textual processing by the computer.*

### Introduction

There has been little effort made to substantiate Halliday and Hasan's co-reference framework since their work was published 20-odd years ago, excepted for a slight modification in Martin (1992), Fligelstone (1992), Garside (1993), Garside et al. (1997) and Botley & McEnery (eds. forthcoming) have proposed an annotation scheme for tagging texts computationally. However, their theoretical foundation is Halliday & Hasan (1976), which is sketchy and confined to major co-reference relations. In this paper, I will examine thoroughly how English noun phrases are interpreted in discourse. I recognise two roles that a noun phrase plays in discourse. One is that of the introductory expression which may function as the antecedent of a co-referential / co-interpretational network. The other is that of the dependent co-referring / co-interpretational expression, which depends on the introductory expression for the interpretation of its co-referent or descriptive content. I will describe the major types of relation between two co-referring / co-interpretational expressions in a discourse in order to prepare for a comprehensive discursial networking system of English noun phrases.

### Textual co-interpretation / reference

Linguistic interpretation involves the addressee's linguistic and encyclopaedic knowledge, his general reasoning ability, and his quest for relevance in ostensive-inferential communication.<sup>1</sup> When we come across a noun phrase in discourse, we process its linguistically encoded information in context, guided by some pragmatic principles, to recreate a mental representation of what its communicator intended.

**Textual co-interpretation** is a relation between two expressions in the same piece of discourse. One of the expressions is dependent and its interpretation depends on another expression, the antecedent. The dependent noun phrase may or may not co-refer with its antecedent. Both the definite determiners and the third person pronouns may be used to signal a textual dependence for a correct interpretation of what is being expressed with that expression.<sup>2</sup> [1] contains examples of dependent co-interpretational expressions:

- [1a] THE difficulties of being an only child are no more or less than those of any other child - they are simply different. (extract from *Only Child - How to Survive being One* by David Emerson and Jill Pitkeathley, *The Times* 8/27/96)
- [1b] The distinct family environment of the only child totals just three. Within this unit the only social interaction involving the child are with an adult, and the only ones they can observe are between two adults. (ibid.)
- [1c] fox I suppose would take a sitting hen but I mean they're so scarce compared with the abundant red grouse around that they're far more likely to find those (SEU-S-10-08-178)<sup>3</sup>

The intended interpretation of 'they' in [1a] is anaphorically related to 'an only child' and, therefore, 'an only child' is the textual antecedent of 'they'. In [1b], the interpretation of 'they' is related anaphorically to 'the only child',

---

\* Department of English Language and Literature, The National University of Singapore; e-mail: ellniyb@nus.edu.sg

possibly via 'the child', and thus the latter two are the antecedents of 'they'. Similarly, in [1c], the antecedent of the underlined 'they' is the singular 'a sitting hen' and they are not co-referential.

In discourse, the entity or concept represented by an introductory expression may often be talked about again with other appropriate expressions. If the same entity is being referred to, the phenomenon is called textual co-reference. If various realisations of a concept is being talked about, the phenomenon should be termed co-interpretation. Co-reference and co-interpretation give rise to two issues, which relate to certain explicit linguistic features in noun phrases. One is the continuity in a co-referential network between co-referring expressions, indicated by the definiteness of the dependent expressions. The other is the dependence of the interpretation of a pronominal expression on its nominal antecedent. In this section both issues are discussed in relation to the two types of co-reference (interpretation). These are anaphora, relating to a foregoing expression and cataphora, relating to a following expression. Demonstrative pronominal anaphora and cataphora<sup>4</sup> will also be discussed.

In a co-referential network, there are both dependent co-referring expressions and their antecedents (see more in the next section); and the main task of textual reference assignment is to make connections among co-referring expressions, especially with the introductory expression. A dependent co-referring expression may be either a definite common noun phrase or a pronoun,<sup>5</sup> as shown in [2]:

[2] A van driver from Hendon emerged as the most envied white knight in Britain today after he tenaciously took on a gang of teenage girls who robbed *actress Elizabeth Hurley* at knifepoint.

The 32-year-old man, who lives in Aerodrome Road, gave chase when he saw the glamorous girlfriend of actor Hugh Grant being mugged in a street near her Earls Court home last night - even though he did not recognise the actress at the time. (opening of a news bulletin in *Evening Standard*, 24/11/94)

In [2], the underlined definite common noun phrases are used as a continuity device referring to the same representational referent of their antecedent, the italicised noun phrase. Meanwhile, the conceptual content they carry with them contributes to our overall interpretation of the message. On the other hand, if the dependent co-referring expression is pronominal, then its interpretation depends on its antecedent and they carry with them very little conceptual content, as is the case with the underlined pronoun 'her' in [2].

## Anaphora and cataphora

In textual co-reference, the antecedent may occur either before or after its dependent co-referring expression, i.e. being either anaphoric, as in [1a] and [1b], or cataphoric, as in [1c] and [1d]. In [1a] and [1c], the dependent co-referring expressions are pronominal and, in [1b] and [1d], they are nominal definites.

[1a] In the sea, once upon a time, O my Best Beloved, there was *a Whale*, and he ate fishes. (the opening sentence of R. Kipling's *How the Whale got his Throat*, *Just So Stories*, Penguin Classics)

[1b] A student who had sex with *a college girl* despite her complaints that she was too tired to make love wept yesterday after being cleared of rape. The alleged victim left the court crying after the judge ordered the jury to reach a verdict of not guilty. (*The Times* 11/24/94)

[1c] Asked what he meant by 'Beyond', *Richards* responded, "Beyond anything you can think of." (*I. A. Richards: his life and work* by J. Russo, London: Routledge 1989, p. 609)

[1d] If you are converting your loft, these addresses will be useful: *HighSpaces Ltd. 45 Plover Rise Industrial Estate, Winborough, Gloucestershire, and Morgan & Holmes, 54 High Street, Pensby, Lancashire*

Apart from the sequence of the introductory expression and its co-referring expressions, there is another dimension which is worth studying: the combinations of noun phrases with different nominal content (e.g. common noun phrases, proper names, and pronouns) in a co-referential network. Co-referring expressions may appear in three major patterns with different syntactico-semantic implications:

A. personal pronoun co-reference<sup>4</sup>

(a) anaphoric:

- (1) indefinite - definite (e.g. 'a man ... he ...')
- (2) definite - definite (e.g. 'John / the man in red ... he ...')
  
- (b) cataphoric:
  - (1) definite - definite (e.g. 'he ... the man / John ...')
  
- B. definite determiner co-reference**
  
- (i) same noun phrase repetition: (e.g. 'the wheel ... the wheel ...')
- (a) anaphoric:
  - (1) indefinite - definite (e.g. 'a man ... the man ...')
- (b) cataphoric: (not possible)
  
- (ii) contextually synonymous or hypernymous expression:
  - (a) anaphoric:
    - (1) indefinite - definite (e.g. 'a swallow ... the bird ...')
    - (2) definite - definite (e.g. 'the swallow ... the bird ...'; 'John ... the man')
  - (b) cataphoric:
    - (1) definite - definite (e.g. 'the boy ... Tom ...')
  
- C. repetitive proper name co-reference**
  
- (a) anaphoric:
  - (1) definite - definite (e.g. John ... John ...)
- (b) cataphoric: (not possible)

The following sections contain detailed discussions of these issues.

## Personal pronoun co-reference

In a co-referential network, when the dependent co-referring expressions are pronominal, their interpretation depends wholly on their nominal antecedent. The following three examples are anaphoric, [1a] and [1b] with a definite antecedent; and [1c] and [1d], an indefinite antecedent.

- [1a] *Tom* was in the garden when the telephone rang. He let Mme Annette, his housekeeper, answer it, and went on scraping at the sappy moss that clung to the sides of the stone steps. (opening *Ripley Underground* by P. Highsmith, Penguin Books)
- [1b] *The people of Maili Sai* once considered themselves a favoured people, their shops filled with supplies sent directly from Moscow. For a while, since Kirgizstan cast off from the Soviet Union, they have been feeling less favoured. (the opening of a news bulletin in *The Economist*, Dec. 3, 94)
- [1c] A girl I know has a *glamorous, film-making, motorbike-riding brother*; a tall tearaway who lives high up a Welsh mountain in a house he has built largely with his own hands. (opening of an article in *The Times*, 11/20/94)
- [1d] No concept or term is introduced unless you really need it, ... (*Word Meaning* by Richard Hudson, London: Routledge 1995, p. viii)

Each of the pronominal expressions in [1] depends on its antecedent for its interpretation. At the same time, as a co-referring expression in a co-referential network, each dependant is also used to refer to the shared representational referent, contributing to the network with information attached to it. For example, in [1a] the representational referent of both 'Tom' and 'he' is the fictional character TOM in the novel while the textual antecedent of 'he' and 'his' is the expression 'Tom' in the text. Again, the representational referent of the co-referential network in [1c] is 'the particular brother of a particular friend of the writer's whom she's talking about'.<sup>6</sup> The pronoun 'he' in [1c] not only stands for its italicised antecedent 'a ... brother', but is also used to refer to the shared representational referent of the co-referential network in which it occurs. In the same way, the textual antecedent of the underlined pronouns in

[1b] is the expression 'the people of Maili Sai' and their shared representational referent is, first of all, the construction of a mental representation of what the noun phrase is used to refer to. [1d] is an even more interesting case: though the antecedent of the pronoun 'it' in [1d] is a negative one, 'no concept or term', the pronoun is still used to co-refer with it to potential members of the class denoted by the head of the antecedent.

The examples in [2] are cataphoric. While [2a] has co-reference within one orthographic sentence, [2b] and [2c] contain co-reference crossing sentence boundaries.

[2a] Ten weeks before he died, *Mr Mohun Biswas*, a journalist of Sikkim Street, St James, Port of Spain, was sacked. (the opening sentence of V. S. Naipaul's *A House for Mr Biswas*, Penguin Books 1961)

[2b] OMAHA BEACH, France - They do not look like heroes, but they are easy to spot as they stand on the windswept beaches of Normandy or pause by the grave of a fallen buddy, quietly reliving the moments of fear, horror and excitement that marked their unannounced arrival here half a century ago.

Now mostly white-haired, occasionally frail, the youngest of them in their late 60s, the veterans of D-Day keep returning like pilgrims to Normandy, ... (opening of a news bulletin, *International Herald Tribune* 28-9/5/94)

[2c] Sir John Betjeman called it 'easily the most magnificent of all London hotels of its time'. Yet no great Victorian building in the capital has suffered such adversity as *the Midland Grand Hotel at St Pancras*. (the opening of a feature in *The Times*, Dec. 10, 94)

When the reader of [2a] reaches the pronominal expression 'he' - because it is the first mention of the newly-introduced entity - he has to depend on the succeeding part of the text, where its antecedent is, for its interpretation. By the end of the first sentence, the reader may have established a mental representation of the shared fictional referent with 'Mr Mohun Biswas' as its address: 'a journalist of Sikkim Street, St James, Port of Spain, who was sacked two weeks before his death'. In both [2b] and [2c] the writers deliberately delayed supplying the reader with the antecedent to create suspense: who does 'they' refer to? What is this 'it'? Thus, the reader's curiosity is aroused.

When looking for the textual antecedent of a pronominal noun phrase, we generally make use of the principle of proximity and try to match the gender, number, and person, of the pronoun to those of its nominal antecedent. However, ambiguities do arise. For instance, the noun phrases 'shops' and 'supplies' in [1b] above and many plural noun phrases in [2b] are potential candidates of the antecedents of the underlined pronouns, especially from a computer's point of view. When there is such ambiguity, further inference both with our general knowledge and with what is contained in the text is needed, in order to reach the intended antecedent. For example, in [1b] above, we can conclude, based on our general knowledge and the criterion of consistency with the principle of relevance, that 'shops' and 'supplies' are not as likely as 'the people of Maili Sai' to feel 'less favoured'.

In the following example, we use deictic information, the speaker's point of view, to find the textual antecedent of the underlined pronouns: the vocative which one uses to address oneself has to be used to refer to oneself, the speaker. Therefore, 'Joe' can be regarded as the antecedent of the first-person pronouns in [3], which is more suitable than 'I' to serve as the conceptual address for the representation of the fictional character.

[3] I came to Warley on a wet September morning with the sky the grey of Guiseley sandstone. I was alone in the compartment. I remember saying to myself: 'No more zombies, *Joe*, no more zombies'. (the opening paragraph of *Room at the top*, by John Braine, Penguin Books)

When we use different expressions to talk about the same entity, we normally make sure that introductory expressions and their dependent co-referring expressions agree in terms of number, gender and person. However, there are cases in which the matching of person, gender and number between introductory and their dependent expressions is not observed, as shown by examples [1a], [1b] and [1c] in the section on 'Textual Co-interpretation / reference'.

In the following table of personal pronoun co-reference, we can see that one combination is not possible: the combination of an indefinite cataphor with a earlier-appearing pronominal co-referring expression, i.e. the opposite of 'A man came in and he sat down'. This is because the appearance of an indefinite nominal expression may signal the

starting point of a potential co-referential network but it cannot be used to co-refer with any pronominal expressions which precede it.

possibilities of personal pronoun co-reference:

anaphoric:

indefinite nominal

pronoun:

A man came in and *he* sat down.

definite nominal pronoun:

common noun:

The man came in and *he* sat down.

proper name:

John came in and *he* sat down.

cataphoric:

pronoun indefinite nominal: When *it* is cornered, a dog will jump over a wall.

pronoun definite nominal

common noun noun phrase:

When *he* came in, the man walked straight to me.

proper name noun phrase:

When *he* was young, John was very shy.

## Definite Determiner Co-reference

The second group in textual co-reference is called 'definite determiner co-reference' because the dependent co-referring expression in question is a noun phrase with various 'definite determiner + noun' constructions. Though both the dependent expressions and their antecedents in [1a] below are definite, they are definite for different reasons.

The definiteness of dependent co-referring expressions indicates the continuity in the co-referential network whilst the definiteness of the introductory expression carries clues for the interpretation of its representational referent. When the antecedent is indefinite, as is the case in [1b] and [1c] below, there is an overt shift from an indefinite noun phrase to a definite noun phrase in the co-referential network, which signals the continuity between the two but does not affect the nature of their shared referent. It is the definite determiner in the dependent expression that indicates the existence of the antecedent and thus retains the continuity of the co-referential network. Once the addressee has decided that the expression is co-referentially used, what he then has to do is to infer its antecedent, which will be nearby in the text: a particular expression (or groups of expressions) which the communicator has used (anaphoric) or is about to use (cataphoric). The head noun of the co-referring expression can be a repetition of that of the antecedent, as shown in the examples in [1].

[1a] A huge mudslide last spring nearly pushed a waste site into *the nearby Maili Sai river*<sup>i</sup>. The river<sup>i</sup> flows through the middle of the town and on to *the fertile Fergana valley*<sup>ii</sup>, one of the most densely populated areas of the former Soviet Union.

Most of the valley<sup>ii</sup> lies in the neighbouring republic of Uzbekistan. (*The Economist*, Dec. 3, 94)

[1b] The Republican Convention in Houston last August has become a byword for bigotry. Yet the worst examples of intolerance were to be found not in the Astrodome, but outside *a low cement building a few streets away*. The building was an abortion clinic; ... (*The Economist*, 12-18 June, 93)

[1c] *American scientists*<sup>s</sup> have reported a *new strain of antibiotic-resistant bacterium*<sup>ii</sup> and warned that its appearance could mean antibiotics may soon become useless.

The bacterium<sup>ii</sup> would mean patients may not be able to avoid picking up infections in hospitals, the scientists<sup>s</sup> said. (the opening of a news bulletin in *International Herald Tribune*, 11/11/94)

[1d] He had not told her, and could not tell her, about his secret theft. Later that night, he looked at the letters again, in the bathroom. . . . But these dead letters troubled him, physically even, because they were only beginnings. (*Possession*, A. S. Byatt, London: Vintage, pp. 20f.)

The interpretation of a dependent expression in a co-referential (or co-interpretational) relation involves inference and, if the head noun in the co-referring expressions is a repetition of that in the antecedent, then less inference is needed. In [1b], the expression 'the building' not only is related to its textual antecedent 'a low building a few streets away' but is also used to refer to the shared representational referent of the co-referential network in which it occurs. In other words, textually, its referent is 'the building which the author has been talking about' and, representationally, it is 'the particular building that lay a few streets away from the Astrodome and existed in August 1992 in Houston, USA'.

If there is not a head noun repetition between the antecedent and its dependent co-referring expression (even repetition is not necessarily sufficient to establish such a relation), then we have to rely more on inference with our general knowledge, contextual information and the criterion of consistency with the principle of relevance.

The example in [1a] above is counted as anaphoric because there is no reason to call the later-appearing co-referring expression the antecedent when the co-referring expressions are both definite and have the same head noun.

On the other hand, if we believe that a proper name given to an entity is intuitively more legitimate as the antecedent than a description of the entity, then in the textual relations between two noun phrases such as 'John Major . . . the Prime Minister . . .' or 'The Prime Minister . . . John Major . . .' we can always call the proper name the antecedent and the description of the entity which has that name its co-referring expression. Given all this, we can find cataphoric examples for this category, as in [2]:

[2a] Here was this fat man up on a piano, dancing in a way that suggests that, during rehearsals, he had miscalculated a pervy groin grind and ended up on the floor, screaming for his manager to put a detachable, rubber, non-slip surface on the old Joanna ...

'Aye, aye,' my dad said, looking up from the '\$400 and Under' second-hand cars' section of the local paper, 'it's *old Tommy Jones*. I know him.' (opening of a feature in *The Times*, Dec. 10, 94)

[2b] The boy with fair hair lowered himself down the last few feet of rock and began to pick his way towards the lagoon. ... The fair boy stopped and jerked his stockings with an automatic gesture that made the jungle seem for a moment like the Home Counties . . . *Ralph* . . . (the opening of *Lord of the Flies*, by W. Golding, Penguin Modern Classics 1954)

In [2b], the reader has to wait until he reaches the second page to find out the proper name antecedent of description 'the fair boy' is 'Ralph'. Within [2b], there are also some pronouns using the two definite descriptions as their antecedents.

This type of cataphoric reference with a long delay is a literary technique, *in medias res*,<sup>8</sup> and appears in discourse when the writer wants to plunge 'the reader into the apparent middle of a situation' (Wales 1989:240). It may be used to create suspense, or to make the delayed, more informative expression a focus of attention.<sup>7</sup> The literary effect may be achieved because the norm or default sequence is that proper names presented as homophors<sup>9/10</sup> precede their co-referring definite descriptions, as shown in [3] below ('Friedrich Nietzsche' appears before the noun phrase in italics) and [2] in the section on 'Textual co-interpretation / reference' above:

[3] Rocken, Germany - The change in Friedrich Nietzsche's status in Eastern Germany can be seen on the welcome sign of this tiny hamlet in the state of Saxon-Anhalt where *the philosopher* was born 150 years ago and lies buried. (the opening of a feature in *International Herald Tribune* 11/11/94)

The underlined noun phrase in the following example may be regarded as an *esphora*<sup>11</sup> because of the modification in the noun phrase. At the same time, it may be perceived as cataphoric because its textual referent are those indefinite expressions occurring afterwards.

[4] I looked at the ornaments on the desk. Every one standard and all copper. A copper lamp, a pen set and a pencil tray, a glass and copper ashtray with a copper elephant on the rim, a copper letter opener, a copper thermo bottle on copper tray copper corners on the bottle holder. (*The Raymond Chandler Omnibus*, London: Hamish Hamilton 1953)

There is a cognitive constraint which says that in lexical anaphora one should go from a more specific expression to a more general one.<sup>12</sup> [5a] is odd and a better version would be either [5b] or [5c], where the more specific description is introduced as new information:

- [5a] Peter shot a bird. *The swallow* fell down.
- [5b] Peter shot a swallow. *The bird* fell down.
- [5b] Peter shot a bird. It was *a swallow*.

However, writers may exploit the violation of this cognitive constraint to create suspense in order to achieve the effect of a close-up. In [6], the underlined expression, headed by a general term 'object' was used to introduce a new entity while the italicised expression with more specific information was used as a subsequent co-referring expression.

- [6] Closer in time to us by 2,000 years, another object of even greater rarity than the Assyrian relief provides a measure of the current sense of urgency when an opportunity to buy something extraordinary arises. *The carved narwhal horn of the 12th century*, which was sold at Christie's on Tuesday for \$441,500, is a deeply mysterious object . . . The Christie's horn, and a closely related piece in the Victoria & Albert Museum are the only known horns to display such intricate, highly sophisticated decoration. (*Room at the Top for the Truly Rare*, *International Herald Tribune*, July 9-10, 94)

When the writer uses a definite co-referring expression with a different head noun, not only does it serve as a stylistic variation but it may also give the reader new information, as is the case in [7].

- [7a] The giraffe is an unmistakable inhabitant of Africa's savannah country. *The tallest land mammal of all* can browse on vegetation which other animals simply cannot reach. (adapted from *Wildlife Fact-file: Giraffe*)
- [7b] I have been testing crab and lobster recipes and have come to the conclusion that to get the freshest, sweetest product, you should cook *the crustacean* yourself, rather than buying it ready-cooked. (*Tips on the most humane way to dispatch lobsters and crabs*, *The Times*, July 3, 93)

In [7a] the head noun of the antecedent is the standard subordinate (hyponym) of the head noun of the co-referring expression. In addition, because the italicised co-referring expression is restrictively modified, it is intensionally as well as extensionally equivalent to its antecedent<sup>13</sup>. In [7b] the taxonomic name of the class which contains lobsters and crabs is used to co-refer with them. If the reader does not know the word but managed to infer the co-referential relation between 'the crustacean' and 'crab and lobster', then he can also get the new information conveyed by the hyponymic co-reference.

[8] is an example illustrating the same point but the co-referring relation is between a proper name and a definite common noun phrase; such a relation consequently requires more inference from the addressee's general knowledge than a standard hyponymic one discussed above:

- [8] For a man who once commanded a \$4 billion lemons-to-electronics empire, Asil Nadir appears surprisingly content to be exiled in northern Cyprus, a self-styled republic recognised only by Turkey.  
A year and a half after he jumped \$3.5 million bail, *the former head of Polly Peck International* faces having many of the company's remaining riches removed from his control. (opening of a feature in *The Times*, Dec. 10, 94)

The following is the scheme of definite determiner co-reference in English:

definite determiner co-reference

same noun phrase repetition:

definite noun phrase - definite noun phrase:

The wheel was invented tens of thousands of years ago and the wheel is still in use today.

same noun phrase head repetition:

anaphoric:

indefinite noun phrase - definite noun phrase:

A man came in and the man sat down.

other co-referring expressions:

semantically synonymous or hyponymous expression:

anaphoric:

indefinite - definite:

A book was sold and the volume was old.

A swallow fell and the bird was dying.

definite - definite:

The book in red was sold and the volume was old.

The swallow I shot fell and the bird was dying.

contextually co-referring expression:

anaphoric:

definite noun phrase - definite noun phrase:

Tom turned up again and the boy was always

restless.

non-nominal - definite noun phrase:

Tom passed all the exams. The wonder made his mother very happy.

cataphoric:

definite noun phrase - definite noun phrase:

The star died. Olivier was a real genius.

definite noun phrase - non-nominal:

The wonder made his mother very happy: Tom passed all the exams.

## Repeated proper name co-reference

If the same proper name is used as the subsequent co-referring expression in a co-referential network, then the first is naturally the introductory expression, as those in [1].

- [1a] It was dusk when the inhabitants of the leproserie heard the bell of the long-overdue boat; the sound came to Colin and Query where they sat over the first drink of the evening on the doctor's verandah, "At last," Colin said, finishing his whisky, "if only they have brought the new X-ray ..." (*A Burnt-Out Case* by Graham Green, London: Heinemann, p. 118)
- [1b] Kissing, tickling and being bored may seem as familiar and instinctive as life itself, but it took the British writer Adam Phillips to make them part of the complex language of psychoanalysis. ...  
Though *Phillips* chooses to explore tantalizingly simple subjects, his work can at times be difficult to penetrate. (the opening of a feature in *International Herald Tribune*, Nov. 8, 94)
- [1c] You may not think of David Mamet, the prolific author of angriified and angriifying plays and films, as an insecure fellow.

...

It's a funny story for *Mamet* to tell on himself, a twinkly-eyed acknowledgement of his reputation as difficult, thorny and impatient. (the opening of a feature in *International Herald Tribune*, 11/21/94)

In such cases, the subsequent use of the same name (or part of it) assumes the introductory expression as the textual antecedent and shares with it its representational referent. Examples [1b] and [1c] also show two common ways of introducing into discourse a proper name which the communicator assumes is unfamiliar to the addressee: description + proper name or proper name + description. The information provided in the appositive noun phrase gives the addressee some clue to establish a representation of its referent, a similar process to esphora (see Note 10).



In the examples in [2], the writers expected that the reader would be able to retrieve the referents of the proper names homophorically (see Note 9).

- [2a] John Major is facing a new move to oust him as Tory leader by his right-wing critics, . . .  
The move will fuel the debate inside the Tory party over whether *Major* should lead it into the next general election. (the opening of a news bulletin in *The Sunday Times*, 13/11/94)
- [2b] Since it was first mentioned in chronicles dating from 1147, Moscow has played a vital role in Russian history; indeed, the history of the city and of the state are closely interlinked. For more than 600 years *Moscow* has also been the spiritual centre of the Orthodox Church of Russia. (the opening of 'Moscow', *Encyclopedia Britannica* 15th edn.)

In the case of [2b], if we have 'Moscow' in our mind solely as the name of the capital of Russia, then the notion of the textual antecedent is not useful any more because the subsequently mentioned 'Moscow' is not textually restricted. The interpretation of the subsequent co-referring expressions is not dependent on the earlier-appearing one - each of them can be independently homophoric and each of them is used to refer directly to the same representational referent. Of course, the introductory and subsequently-mentioned expressions of 'Moscow' still share the same referent in the reader's general knowledge, which may be enriched when he finds new information in the article.

Sometimes, co-referring proper names are non-restrictively modified, mostly in newspaper writing, so that more information can be packed in the limited space, as shown by the italicised expressions in [3].

- [3] The brave motorist who helped police arrest the four-girl gang that attacked actress Liz Hurley has been praised for his swift action.  
... He stopped his van immediately, phoned the police on his mobile and told *devastated sexbomb Liz* to stay near his vehicle.  
... *Screen lovely Liz* returned to live in the Big Orange after two years in crime-ridden Los Angeles .  
.. (London *Tonight* 11/24/94)

In this section, I have clarified various ways proper names are introduced into discourse and their possible co-referential relations in a text.

## Demonstrative pronominal anaphora and cataphora

All of the demonstratives *this*, *that*, *these* and *those* can be used in textual reference as determiners as well as heads of noun phrases, as shown in [1], [2] and [3] and all of the four can be used anaphorically, as shown by examples in [1] and [2].

- [1a] The alarm system does contain two panic alarm devices. *These* are located in reception ( our street entrance ) and the cashier 's office ( 4th floor ). (W1B-028-103f)<sup>14</sup> [anaphoric; noun phrase head; nominal antecedent]
- [1b] Where have the time and the distance gone? I have defeated them, these 2 old enemies of lovers. I don't want them to stand between us as they are very strong and subtle. (W1B-007-128ff) [anaphoric; determiner; nominal antecedent]
- [2a] I regret that I am unable to change the date that the Personal Loan transfer goes through, so you will need to bear *this* in mind when you calculate the funds you have available to you. (W1B-016-22) [anaphoric; noun phrase head; discursal antecedent]
- [2b] (Carrying a cup of tea, for example, is an absolute disaster area.) You have to concentrate more - and *that* makes you more tired.' (W2B-001-84) [anaphoric; noun phrase head; discursal antecedent]
- [2c] I found it easy to cast myself as the man with a legion of evil spirits as they suggested - *Those* being , Anger , fear etc. but I couldn't really accept that J. C. could get them all out of me and into the pigs and drop them into the sea . (W1B-003-80f) [anaphoric; noun phrase head; nominal antecedent]

*This* can be used cataphorically,<sup>15</sup> as shown in [3].

- [3a] . . . You won't get this letter till Monday by which time I'll no doubt have spoken to you, but never mind. This is the record of how I'm feeling tonight. Thank you my love for all your help and encouragement during my revision which enabled me to get my work to a standard where I could get a 3. (W1B-007-74ff) [cataphoric; noun phrase head; discorsal antecedent]
- [3b] Every male teacher lives with this fear. But when Chris, a 41-year-old drama teacher, was called to the head's office on Wednesday morning in February last year, it was the last thing on his mind.  
He was met at the door by his union representative, who told him he had been the subject of a serious complaint. A parent had written a letter claiming that Chris had indecently assaulted her daughter. (the opening paragraphs of 'Every teacher's nightmare', *The Independent* 6/28/94) [cataphoric; determiner; discorsal antecedent]

The case with *that* is controversial.<sup>16</sup> However, the following two examples can be formally regarded as *that* being used cataphorically.<sup>17</sup>

- [4a] . . . It was like being married to a novelist, he never knew when something he said or did would be reborn in one of her speeches. That's what he'd expected: that she'd come back and they'd have a drink or two and she'd shout at him ( he worried about the Goldmans hearing, the walls were like cardboard) then it'd be over - until the next time. (W2F-016-119ff) [cataphoric; noun phrase head; discorsal antecedent]
- [4b] . . . But Cossiga has agreed to say his piece before a Parliamentary Commission or before a secret service committee with evidence in camera.  
Casson says: 'That is all I wanted - that new light should be shed on these years of lies and mysteries: that Italy should for once know the truth.' (W2C-010-53f) [cataphoric; noun phrase head; discorsal antecedent]

Though *these* can be used cataphorically,<sup>18</sup> such use is very rare. Of the 1512 occurrences of *these* in the ICE-GB,<sup>19</sup> none has been used cataphorically. Nor has any of the 1043 occurrences of *those*.

Whether the antecedent is a well-defined nominal constituent or not is very important from the point of view of processing a text. When the uses of the four demonstratives (when being used to indicate textual reference) in the ICE-GB written corpus is compared, it has been found that *this* and *that* are often used with discorsal antecedent<sup>20</sup> ([2a], [3a], [3b] and [5a] for *this* and [2b], [4a] and [4b] for *that*). By contrast, *these* and *those* are used most frequently with a nominal antecedent ([1a] and [1b] for *these* and [2c] and [5b] for *those*).

- [5a] You're lying down, probably beneath a line of dripping wet clothing, the radio on (BBC of course), and underneath and around you lies . . . God only knows what! I don't mean to sound religious, God forbid no!  
No, no more of this conjecture, you're well I know. (W1B-001-12ff) [anaphoric; determiner; discorsal antecedent]
- [5b] In case I was tempted to gloss over certain thoughts, I decided to share the diary which had helped me admit to myself what I was now admitting to the listeners. Here are some of those thoughts. (W2B-001-116) [anaphoric; determiner; nominal antecedent]

Discorsal antecedents are much more difficult to identify in textual reference assignment because they can be clusters of linguistic expression of various sizes.

The following is the scheme of demonstrative pronoun co-reference:

- I. anaphoric  
i. noun phrase head  
a. nominal antecedent  
this:  
that:  
these:  
those:

- b. discursal antecedent
              - this:
              - that:
              - these:
              - those: examples have not been found
  - ii Det.
    - a. nominal antecedent
      - this:
      - that:
      - these:
      - those:
    - b. discursal antecedent
      - this:
      - that:
      - these:
      - those:
- II. cataphoric
- i. noun phrase head
    - a. nominal antecedent
      - this:
      - that:
      - these:
      - those:
    - b. discursal antecedent
      - this:
      - that:
      - these:
      - those:
  - ii. Det.
    - a. nominal antecedent
      - this: examples have not been found
      - that:
      - these:
      - those:
    - b. discursal antecedent
      - this:
      - that:
      - these:
      - those:

## Conclusion

In view of the lack of detail in Halliday & Hasan's (1976) framework of discursal relations in English texts, this paper has present a comprehensive picture of all such major relations in English. The detailed system described in this paper will be of practical use in attempts to tag discursal relations in English texts by the computer.

## Notes:

1. See Ni (1996), Chapter 2.
2. See Brown & Yule 1983:6.3; Quirk et al. 1985:267. In co-reference, third person pronouns are our main concern; normally, first- and second-person pronouns do not directly co-refer. But see example [3] in the section on 'personal pronoun co-reference'.

3. 'SEU' stands for the Survey Corpus, a corpus developed at the Survey of English Usage, University College London. 'S' stands for the spoken part of the Survey Corpus.
4. Quirk et al. (1985:375f) discusses this phenomenon. Martin (1992:123) uses the term 'cohesive cataphora'.
5. Halliday & Hasan (1976:19) use the term 'explicitly anaphoric' and 'explicitly cataphoric' to refer to the cases where dependent co-referring expressions are pronouns since the interpretation of pronouns in non-deictic context has to rely explicitly on their antecedents.
6. Sells & Wasow 1994:116ff gives a brief account of pronominal anaphora.
7. See Halliday & Hasan 1976:309 for a different approach.
8. See Wales 1989:59 and Leech & Short 1981:108f.
9. A homophor is a definite expression used to refer to an entity known to both the speaker and hearer.
10. That is, in cases when the communicator believes that the addressee has a mental slot for the particular individual. One exception to this sequence is when the person referred to is well-known because of his present position, as in the sequence of 'the President' and 'Bill Clinton' in news bulletins during Clinton's presidency
11. An esphor is a restrictively modified definite noun phrase.
12. See Sanford & Garrod 1981
13. Two expression with different heads and without modification can be only extensionally equivalent. For example, in the context of [5b] the expression 'a swallow' is extensionally equivalent to the expression 'the bird' because they are used to refer to the same entity on this occasion. But intensionally these two expressions are not equivalent.
14. This is a file name for data from the British component of the International Corpus of English (ICE-GB) collected and analysed at the Survey of English Usage, University College London. The corpus consists of 200 written texts, marked by 'W', and 300 spoken texts, marked by 'S'. Each text, consisting of approximately 2,000 words, has a label such as 'W1B-001' or 'S1A-001' encoding the text type. The extension after the label, like '-10' in 'W1B-001-10', is the number of a text unit, the basic unit within a text.
15. Martin (1992:123) calls this phenomenon 'cohesive cataphor'.
16. Quirk et al. (1985:375) claimed that 'the "distant" demonstrative *that / those* can have only anaphoric reference'.
17. In [4a] the colon after the word 'expected' is used to 'precede an explanation or particularisation' (*Plain Words* by E. Gowers and revised by Greenbaum & Whitcut 1986, London: HMSO, p. 155) and in [4b] the dash after the word 'wanted' is 'the colon's weaker relative' (*ibid.*)
18. See Quirk et al. 1985:375 and COBUILD English Dictionary 1995.
19. 'ICE-GB' stands for 'International Corpus of English-Great Britain'.
20. In Quirk et al. (1985:375) this type of antecedent is referred to as 'sentential antecedent'. I call it discorsal antecedent because often the pronoun is used to refer to a piece of discourse longer than a sentence (clause).

## References:

1. Asher, R. E. ed. (1994) *The encyclopedia of language and linguistics*, Oxford: Pergamon Press
2. Botley, S. P. & A. M. McEnery (forthcoming) eds. *Corpus-based and Computational Approaches to discourse anaphora*, London: UCL Press
3. Brown, Gillian & G. Yule (1983) *Discourse analysis*, Cambridge: Cambridge University Press
4. COBUILD on CD-ROM (1994) HarperCollins Publishers Ltd.
5. Fligelstone, Steve (1992) 'Developing a scheme for annotating text to show anaphoric relations', in Leitner ed. 1992, pp. 153 - 70
6. Garside, R. (1993) 'The marking of cohesive relationships: tools for the construction of a large bank of anaphoric data', *ICANE Journal* 17, pp. 5 - 27
7. Garside, R., Steve Fligelstone & Simon Botley (1997) *Discourse annotation: anaphoric relations in corpora*, in Garside et al. eds. 1997, pp. 66 - 84
8. Garside, R., G. N. Leech, & A. McEnery eds. (1997) *Corpus Annotation: Linguistic Information From Computer Text Corpora*, London: Longman
9. Halliday, M. A. K. & R. Hasan (1976) *Cohesion in English*, London: Longman
10. Leech, Geoffrey N. & M. H. Short (1981) *Style in fiction*, London: Longman

11. Leitner, G. ed. (1992) *New directions in English Language corpora*, Berlin: Mouton de Gruyter
12. Martin, James R. (1992) *English text: system and structure*, Amsterdam: Benjamins
13. Ni, Yibin (1996) *The Interpretation of English Noun Phrases with Particular Regard to Generic Reference*, Ph.D. dissertation, University College London
14. Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985) *A comprehensive grammar of the English language*, London: Longman
15. Sanford, A. J. & S. C. Garrod (1981) *Understanding written language*, Chichester: John Wiley & Sons
16. Sells, P. & T. Wasow (1994) 'Anaphora', in Asher ed. 1994, 116–122
17. Wales, Katie (1987) *Dictionary of Stylistics*, London: Longman