LACUS FORUM XXXIII

Variation

UNIVERSITY OF
TORONTO

THE VARIABILITY OF REFERRING EXPRESSIONS: AN ALTERNATIVE PERSPECTIVE ON THE NOUN PHRASE IN ENGLISH

Lise Fontaine
Cardiff University

IT COULD BE ARGUED that there are basically two different views concerning language. The first is that language is often changing; there is variation in speakers' language from generation to generation, from region to region, and from context to context. This is seen as a social phenomenon, and often it is something that society has difficulty accepting (for example, lack of acceptance of certain linguistic varieties and the crisis over electronic language). On the other hand, work done in corpus linguistics and psycholinguistics on collocation and formulaic expressions seems to be saying something of the opposite; some parts of language do not change, they stay the same and this is something we can count on (for example, language learning/teaching, text categorisation, and register analysis).

The two different views are not necessarily conflicting. However, if both are true, then the problem is how to resolve them. How much variation is normal? The main interest of this paper is to better understand referring expressions in use. If we accept that individual speakers vary in the way they refer to the objects they want to talk about, then we need to know how much variation is normal or standard for a given speaker. If we also consider that speakers tend to express themselves habitually, with little variation in their use of expressions, then the question to answer would be how much variation, if any, there is among speakers and among text types.

Referring expressions have been studied from a variety of approaches. Philosophers have mostly been concerned with definite descriptions, truth values and existence (cf. Frege 1892). Semanticians look to identifying referents, assigning meaning to expressions. Psycholinguists are interested in how referring expressions are stored, expressed and recognized in the brain and the psychological involvement in the linguistic representation of referring expressions (cf. Cleland & Pickering 2003). Computational linguists have put the theories to the test and have tried to implement systems that are able to generate referring expressions in limited situations (cf. Reiter & Dale 1992). Corpus linguists (cf. Biber *et al.* 1998, Haan 1989, Quirk *et al.* 1985) have begun to look at noun phrase occurrences. Finally, referring expressions have also been studied in a relatively new area of research, formulaic language (cf. Wray 2002).

This paper is concerned with one particular aspect of referring expressions: their realisation in text in terms of their function and form within the framework of Systemic Functional Grammar (SFG). Section 1 discusses the relationship between referring expressions and the noun phrase. Section 2 describes the data under analysis, including a brief overview of the methodology. Then in section 3, the results of this study are presented. Finally section 4 concludes with a summary and some concluding remarks.

1. REFERRING EXPRESSIONS AND THE NOUN PHRASE.[1] Although the term referring expression has its roots in philosophical studies of reference and referring, it is most commonly used in psycholinguistic research and natural language processing. It is a primary concept, related to nearly every type of linguistic analysis from lexicography to discourse. However, linguists tend not to use this term and instead focus on its realization in structural units. Oddly this structural classification is also true for functional linguists, where one might expect to find a more functional approach.

This paper claims that the term needs to be brought to the forefront. We need a term 'for the linguistic units that serve to identify (or refer to) whatever we are talking about when we make a statement about something' (Lyons 1977:23), especially for those of us specifically interested in language from a functional perspective. The term used here is referring expression, to shift perspective from noun phrase to referring expression.

The noun phrase is such a common unit that very few researchers define it. It is a term that is often taken for granted and it is assumed that everyone knows what it is. The following definition is particularly interesting, as it presents the noun phrase as the point of departure, so to say, and suggests that a noun phrase can be used (by a speaker) to refer to an object.

> A noun phrase is a string of words which syntactically is a constituent with an internal structure containing a determiner, a modifier and a head... Semantically, a noun phrase can be used as a referring expression. (Haan, 1989: 8)

Most research on referring expressions would suggest that it is in fact the other way around; that the noun phrase is simply one possible structural realization of a referring expression. The speaker's intention to refer to some object (i.e. to build a referring expression) would precede the building of a noun phrase.

An additional argument against seeing the noun phrase being used as a referring expression is that it is not certain that all languages have noun phrases, yet all language systems need to be able to let speakers refer to things (or objects, conceptual or physical). Field work done on the Iroquois languages, namely Cayuga, Tuscarora, and Mohawk, suggest that these languages do not in fact express referring expressions in the form of noun phrases. In Mohawk 'many entities are identified by means of morphological verbs rather than nouns' (Mithun 1996:636, cited in Rijkhoff 2002:13), as illustrated in the following example (*ibid.*):

(1) rakwá:tihs wahratkáhtho? katéskrahs
ra-kwá:tihs wa-hr-at-káhtho-? ka-téskr-ahs
he-is-young he-looked-at-it it-stinks
'the boy looked at the goat'

Mithun's interpretation (1976:31, cited in Rijkhoff 2002:13) of what she calls descriptive labels is that they are all 'surface morphological verbs... (but) they clearly function in the same way as formal nouns syntactically'.

Text	N Words	N Referent*	N Thing**
GAE (email text –speaker 1)	4,824	1,692	1,510
HES (email text – speaker 2)	4,391	1,169	981
MFN (model forest network)	4,692	552	466
total:	13,907		

Table 1. *Description of data. *Referent = all Participants and Circumstances (Subjects, Complements and Adjuncts). **Thing = all participants and circumstances realized as nominal groups (i.e. referring to some object).*

It is perhaps unreasonable to assume that all languages have nouns and that all referring expressions are realized linguistically as noun phrases. This is clearly an extreme position. It is beyond the scope of this paper to explore this idea any further. It is sufficient evidence however to suggest that the structural realization of a referring expression is not a given. Nevertheless, it does seem clear that the correlation between a referring expression and its realization as a noun phrase is quite strong in English, although this is not always the case.

2. DATA AND METHODOLOGY. The data in this study comprises three small corpora. The first two (called GAE and HES) come from a large archive of personal email messages. The first, GAE, was then broken down over three time periods in order to analyse variation for a single speaker. Texts GAE and HES were selected because they were the only two from the archive to have the following text variables (Fontaine 2006):

- They were composed online using HTTP-based email software. This is important, as it gives less opportunity for editing as compared to POP-based email software.
- Both authors composed linear texts, i.e. their texts did not include any reported or embedded text from previous messages.

The third corpus (MFN) is from an educational brochure on the Model Forest Network. It was included as a control text so that comparisons could be made in several dimensions: texts written by one single speaker; texts written by different speakers within a similar text type (personal email texts); texts written by different speakers in different text types (personal email texts and another type of text).

As can be seen from **Table 1**, the three corpora are comparable with respect to word count, but there is a wide range of variation concerning the number of referents for the word count and especially those expressions realized as nominal groups. **Table 2** (overleaf) presents the relative ratios, and we can clearly see that the ratio between thing and referent is relatively constant across speakers and text types. There is a marked difference in these figures for the MFN text, but this is not surprising, as it was chosen for its differences. If differences are expected, then it will be interesting to see where similar patterns emerge.

This study approaches the nominal group in terms of its function as a Participant or a Circumstance in a situation (where a situation is realized as clause). **Figure 1** (overleaf) offers an example of this, illustrating the analysis of example (2) from the GAE corpus.

Text	Ratio		
	thing:referent	thing:words	referent:words
GAE (email text – speaker 1)	1:1.2	1:3.2	1:2.9
HES (email text – speaker 2)	1:1.9	1:4.5	1:3.8
MFN (model forest network)	1:1.9	1:10.1	1:8.5

Table 2. *Distribution of Referring Expressions in Text*

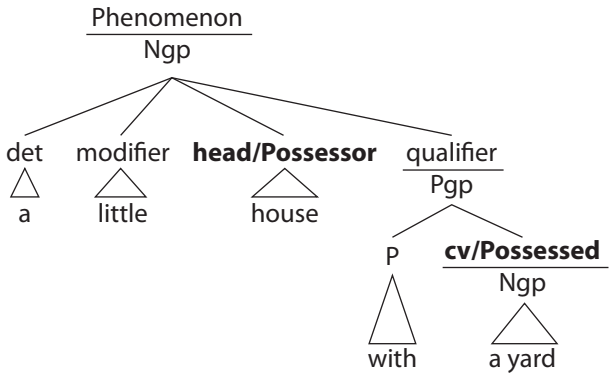

Figure 1. *Partial tree diagram showing analysis of referring expression realized as nominal group.*

(2) I want [a little house [with a yard]].

The referring expression is identified as a Participant or Circumstance, in this case a Participant in the role of Phenomenon (for more detail on the notation used in the tree diagram and the Cardiff grammar, see Fawcett 2000). It is filled by a nominal group with the following components: determiner – modifier – head – qualifier. The tree diagram is not complete; this is only a partial representation. The complex referring expression *a little house with a yard* was further analysed in terms of the functional relationship between the two referring expressions (*with a yard* is treated as an embedded referring expression).

Systemic Functional Linguistics considers language as a social activity whereby speakers make choices as they produce language. These choices are a reflection of the options available to the speaker. The options are semantic options, reflecting the function that the speaker wants his or her language to have. Then the meaning must be turned into form and realized through the grammar of the language. 'The grammar of any language can be represented as a very large network of systems, an arrangement of options in simultaneous and hierarchical relationship' (Halliday 1969:3).

Unfortunately space does not permit us to view the full system network for referent-thing as it is far too large. **Figure 2** offers a partial view, modelling the initial semantic options when entering the system. For a fuller discussion of this network, see Fawcett 1980.

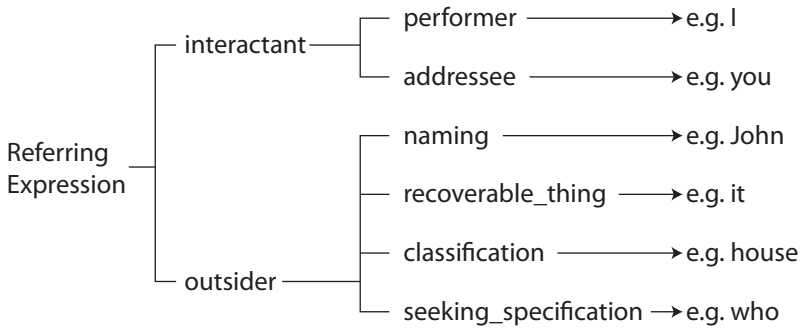

Figure 2. *Partial view of the system network for referent thing (Fawcett 1980).*

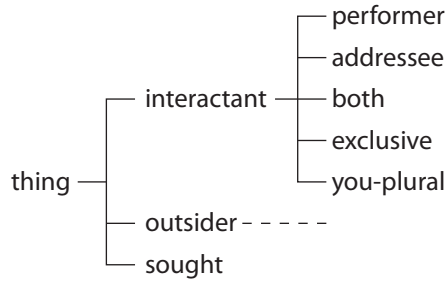

Figure 3. *Partial view of the features coded for thing.*

Converting the system network into an analytical methodology was challenging. Although direct modelling was impossible, every attempt was made to maintain a balance among the theoretical framework of analysis, the existing network for referring expressions, and the constraints imposed by the software. The analysis was completed using a theory-neutral computer assisted manual analysis, called the Systemic Coder (O'Donnell 1995, available online: *http://www.wagsoft.com/software.html*).

The input to the software is the text to be analysed (as unformatted text). The user determines the boundaries for the units under analysis by inserting markers into the text. In this study, boundaries were inserted around all Participants and Circumstances (for a full description of these terms and SFG, see Halliday 1994). The software then interprets these markers as boundaries and in Coding mode, the user is prompted to code the text with user-defined features.

The Coder enables the user to define the features to be coded by developing a scheme which represents the relationships among features through the use of system notation. The coding scheme developed is also far too large to represent here. As an illustration, **Figure 3** shows (using network notation) some of the features analyzed in the texts. The full set of coded features is not given here, again due to space limitations.

Text	'thing' Ngp	'event' Clause	'description' Qlgp
GAE	89%	3%	8%
HES	84%	5%	11%
MFN	84%	12%	4%

Table 3. *Frequency of structural class realizing Referring Expression.*

3. RESULTS. As stated above, the means of identifying a referring expression was to rely on the transitivity of the clause and to identify all Participants and Circumstances. Then these were categorized structurally (i.e. nominal groups, clauses and quality groups). In this section we consider the results from the three corpora in the following four areas: the structural realization of referring expressions, Interactant and Outsider reference, referring expressions in Subject role, and complex referring expressions.

3.1. THE FORM OF REFERRING EXPRESSIONS. The distribution frequency of the structural realization of referring expressions is presented in **Table 3**. It is clear that there is a very strong correlation between referring expressions and nominal groups in English, regardless of speaker or text type. There is no similar consistency with respect to the occurrence of clauses or quality groups, for which we find variation by speaker and text type. This variation is a reflection of speaker choice in referring and it is also a variable of Field of Discourse and Ideational meaning, i.e. the speaker's semantic choices in representing experience.

The following are examples of referring expressions labelled as 'thing', 'event' and 'description'[2] respectively:

(3) A model forest is a place where the best sustainable forest management practices are developed, tested and shared across the country.
(4) Surveying stream crossings and their condition determines which structures require upgrading or repair.
(5) Aboriginal involvement is significant.

What should be noted in **Table 3** is that although there is statistically significant variation both between text types and speakers of the same text type for the distributions as events and descriptions, there is no significant variation for a single speaker.

As Biber *et al.* note, 'noun phrases are the major grammatical device used to refer to people, objects, or other entities in texts'. This tendency seems to hold. However, their work shows that 'texts from different registers often differ dramatically in the use of these referring expressions' (1998:108). One main area in which they consider these expressions to differ based on register is in their structural realization (or the lexical density of the noun phrase). From the ratios presented in **Table 2**, this seems to be true. Quirk *et al.* also make claims about the noun phrase serving as an index: '(these) comparisons... make clear how sensitive the noun phrase is as an index of style and how responsive it can be to the basic purpose and subject matter in varying types of discourse' (1985:1352). It is difficult to make any comparisons with these studies, or indeed others, as each study either counts

Text	Interactant	Outsider
GAE-1	28.4%	71.4%
GAE-2	29.9%	70.1%
GAE-3	18.3%	81.0%

Table 4. *Frequency distribution for Interactant and Outsider Referent for the GAE text.*

Text	Interactant	Outsider
GAE	25.7%	74.0%
HES	17.9%	81.7%
MFN	0.0%	100.0%

Table 5. *Frequency distribution for Interactant and Outsider Referent for texts GAE, HES and MFN .*

the components in a different way or groups them in a different way (e.g. what counts as a modifier, defining a simple noun phrase, etc.).

3.2. INTERACTANT AND OUTSIDER REFERENCE. Another main area in which texts of different registers are said to vary significantly is with respect to endophoric and exophoric reference. This relates most closely to the distinction made in this study between the semantic options of Interactant and Outsider referents. The results for their frequency distribution are presented in **Tables 4** and **5**. **Table 4** presents the results of this distribution for the GAE texts only. The amount of variation we find is statistically significant, indicating that there is a range, for this speaker at least, of 10% variation between Interactant reference and Outsider reference. This needs to be explored further, as the variation is only statistically significant between times 2 and 3; there was no variation between times 1 and 2. These results show that endophoric and exophoric reference varies not only between registers but also for a given speaker (or at least for this particular speaker).

When we consider the results across speakers for the same text type and across different text types, we find that the variation in this distribution is statistically significant both for text type and individual speakers. As we can see in **Table 5**, there is indeed substantial variation between different speakers and between different text types. The most striking difference is of course the relative absence of Interactant reference in the MFN text. This is not at all surprising, as it is a more formally written informative text.

3.3. THE ROLE OF SUBJECT. There were in fact many areas where variation occurred, but what is surprising is finding features that do not exhibit any variation. One such feature seems to be that of Subject. When we look at what we find in the role of Subject, the referring expression is nearly always 'thing' realized as nominal group. This holds for all time periods for the GAE text, therefore we can conclude that this is a feature that does not vary for an individual speaker. Further there is no variation across speakers of the same text type or across different text types. However, this lack of variation does not hold for Complement and Adjunct roles.

Type of REx	GAE	HES	MFN
Thing	99.6%	99.8%	99.1%
Event	0.1%	0.2%	0.9%
Description	0.0%	0.0%	0.0%

Table 6. *Frequency Distribution of type of Referring Expression in Subject Role for all texts.*

Text	Subject	N	Non-Subject (Complement or Adjunct)
GAE-1	44.5%	292	55.5%
GAE-2	47.8%	235	52.2%
GAE-3	44.5%	234	55.5%

Table 7. *Frequency of Referring Expressions in Subject role for GAE.*

Text	Subject	N	Non-Subject (Complement or Adjunct)
GAE-1	44.5%	761	55.5%
HES	40.8%	477	59.2%
MFN	38.4%	212	61.6%

Table 8. *Frequency of Referring Expressions in Subject role for all texts*

As an alternate view on the relationship between Subject role and referring expression type, we can look at the distribution of all referring expressions in Subject role. **Table 7** gives the results for this distribution for the GAE text and **Table 8** gives the results for all texts. The GAE text does not display any variation in this area and consequently, it seems that this single speaker does not vary in the frequency distribution for the role of Subject. However, there is no statistically significant variation between text HES and MFN where one might expect there to be (see **Table 8**). The variation that we do find is that GAE varies significantly from both HES and MFN. Although not reported here, results for Complement position need to be explored further, since it is not an obligatory element and its occurrence depends largely on the process type involved in the utterance. The results from GAE as compared to HES seem to indicate a peculiarity in the speaker of the GAE texts. This is an area that requires further investigation.

3.4. COMPLEX REFERRING EXPRESSIONS. Finally, we now consider the frequency occurrence and distribution of modification, which is also an area where texts of different registers are said to differ (Quirk *et al.* 1985). The area we focus on here is complex referring expressions; those referring expressions having an embedded referring expression (i.e. post-modification, in structural terms), as in examples (6) and (7), both from GAE.

(6) I bought [a game [called Junior Trivia]]

GAE (set)	HES	MFN
25.9%	21.9%	52.2%

Table 9. *Comparative Frequency of Complex Referring Expressions.*

Text:	GAE (set)		HES		MFN	
Frequency:	**%**	**N**	**%**	**N**	**%**	**N**
Subject	12.5	11	21.3	17	28.7	52
Complement	79.5	70	52.5	42	54.1	98
Adjunct	8.0	7	26.3	21	17.1	31

Table 10. *Comparative Frequency of Complex Referring Expressions by role in clause.*

Text:	GAE (set)		HES		MFN	
Frequency:	**%**	**N**	**%**	**N**	**%**	**N**
Subject	6.3	6	13.2	15	53.4	79
Complement	70.8	68	46.5	53	27.7	41
Adjunct	22.9	22	40.4	46	18.9	28

Table 11. *Comparative Frequency of Simple Referring Expressions by role in clause.*

(7) I want [a little house [with a yard]]

The occurrence of complex referring expressions is relatively low, as we can see in **Table 9**. This is, however, an area of the grammar of the noun phrase that still poses challenges from a functional perspective. For the purposes of this study, we will simply consider their relative frequency in the texts. Academic and scientific texts are known to have a higher frequency of these expressions than other texts. This is suggested above in the ratios given in **Table 2** and also confirmed by various studies (for example, Biber *et al.* 1999, Quirk *et al.* 1985). In the frequency of complex referring expressions within the texts for a single speaker (GAE) there is no statistically significant variation; the rate is constant. Further, there is no statistically significant difference between speakers of the same text type (GAE and HES). However, the MFN text contains far more qualifiers and in fact, a referring expression is more likely to be complex than simple for the MFN text, see **Table 9**.

Table 10 presents the results for the frequency of complex referring expressions in terms of distribution within the clause (Interpersonal role). The figures here represent the distribution of complex referring expressions only. There is very clearly a clustering effect in the role of Complement. For GAE and HES, Subject is the least 'preferred' role for complex referring expressions. The opposite is true for MFN, where Circumstance is the least favoured role for these expressions. Simple referring expressions seem to have a different distribution in the clause than complex referring expressions, as we can see in **Table 11**. We must take care in interpreting these results, because there is obviously an intervening factor; any given clause will have only one Subject but may have two Complements and two

or more Circumstances, for example. Nevertheless the most balanced distribution of roles is found in the MFN text, where there is a very distinct difference between the distribution of simple and complex referring expressions (see actual number of occurrences, N). Simple referring expressions seem to prefer Subject role while complex referring expressions find themselves most frequently in Complement role.

The patterns found in GAE and HES seem to suggest key differences between the two speakers, despite consistency in overall frequency. GAE maintains a clustering effect for Complement role while for HES, the preference for Complement role in the distribution of complex referring expressions does not hold in the distribution of simple referring expressions, where the frequency of occurrence seems to balance out between Adjunct and Complement.

4. SUMMARY AND CONCLUDING REMARKS. This paper set out to consider variation in the use and distribution of referring expressions and to compare results of a single speaker, two speakers of the same text type and then to compare these to a very different text type. Regardless of speaker and text type, there seems to be a strong indication that the relationship between 'thing' referent and nominal group is very strong in English. Further investigation into Participants and Circumstances filled by clauses and quality groups should yield interesting results as this was identified as having significant variation.

With respect to expressions speakers use to refer to themselves or the addressee (Interactant reference), it was not surprising to find that an educational text like MFN had no instances of Interactant referent. However, considerable variation was found between speakers in the same register, and, more surprisingly, for the same speaker within the same register. Further work should consider factors, especially discourse factors, that could explain this latter type of variation.

Despite these differences, there is a relative lack of variation with respect to referring expressions in Subject role. This remains unexplained at this point, although it is clear that Subject is a special role where many functions are simultaneously mapped onto each other. It is a focus point with key information in English: identifying roles in Ideational meaning, identifying mood in Interpersonal meaning, and identifying Theme in Textual meaning. There is no other element of the clause with so much potential. It is perhaps unsurprising that there is a kind of coming together pattern associated with it and as a result it is a very low point for variation, preferring instead grammatical consistency.

The results concerning complex referring expressions demonstrate a lack of variation in frequency when frequency is considered as a percentage of the corpus. However, when considered in terms of the distribution across the clause, it is less clear what the patterns are, if any. The results suggest that simple referring expressions have a different distribution pattern for different speakers and for different text types or registers. This is another area requiring detailed research.

The work presented here reflects a first step towards a better understanding of variation in the production of referring expressions. In order to truly begin to understand variation in this area, research will have to go beyond frequencies and attempt to identify the contexts and loci of speaker choice.

[1] The term most commonly used in Systemic Functional Grammar (SFG) is *nominal group* (see Halliday & Hasan 1976:39). In this paper, *noun phrase* and *nominal group* are used interchangeably, with a tendency to use *nominal group* when discussing it from a Systemic Functional perspective.

[2] Fawcett's term (1980:93) is 'referent regarded as quality'. The terms used in this paper, (i.e. three types of referring expression: thing, event and description) correlate to his three classifications of referent: 'referent regarded as thing', 'referent regarded as situation' and 'referent regarded as quality'. 'Description' is the term given here to those expressions which refer to some 'thing' but which do so with respect to its description (e.g. That man is kind). Further, for the purposes of this study, all description-referring expressions have a participant role in the transitivity of the clause (namely Attribute). However, their analysis was not considered in any detail other than frequency of occurrence. In fact, both event- and description-referring expressions were not included in the detailed analysis, since the focus was on 'thing' and the noun phrase.

REFERENCES

Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Biber Douglas, S. Johansson, G. Leech, S. Conrad & E. Finegan. 1999. *Longman grammar of spoken and written English*. New York: Longman.

Cleland, A.A. & M. J. Pickering. 2003. The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of memory and language* 49:214–30.

Fawcett, Robin. 1980. *Cognitive linguistics and social interaction*. Heidelberg: Groos.

———. 2000. *A Theory of syntax for systemic functional linguistics.* Amsterdam: Benjamins.

Fontaine, Lise. 2006. Where do 'we' fit in? Linguistic inclusion and exclusion in a virtual community. In *Beyond misunderstanding, the linguistic reconstruction of intercultural communication*, ed. by Kristin Bührig and Jan D. ten Thije, 319–56. Amsterdam: Benjamins.

Frege, Gottlob. 1892 [1960]. On sense and reference. Reprinted in *Translations from the philosophical writings of Gottlob Frege*, ed. by Peter Geach & Max Black, Oxford: Blackwell.

de Haan, Pieter. 1989. *Postmodifying clauses in the English noun phrase: A corpus-based study.* Amsterdam: Rodopi.

Halliday, M. A. K. 1969 (1976). Types of process. In *Halliday: System and function in language*, ed. by Gunther Kress, 159–73. London: Oxford University Press.

———. 1994. *An introduction to functional grammar*. London: Edward Arnold.

——— & Ruqaiya. Hasan. 1976. *Cohesion in English*. London: Longman.

Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.

O'Donnell, Michael. 1995. From corpus to codings: Semi-automating the acquisition of linguistic features. *Proceedings of the AAAI spring symposium on empirical methods in*

discourse interpretation and generation, Stanford University, California, March 27–29. Available online at *http://www.wagsoft.com/Coder* [accessed October 25 2006].

Quirk, R., S. Greenbaum, S. Leech, & J. Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Reiter, E. & R Dale. 1992. A FAST algorithm for the generation of referring expressions. In *Proceedings of COLING-92*, Nantes, France, July 1992.

Rijkhoff, J. 2002. *The noun phrase*. Oxford: Oxford University Press.

Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.