Learning the Meaning and Usage of Time Phrases from a Parallel Text-Data Corpus

Ehud Reiter
Department of Computing Science
University of Aberdeen
`ereiter@csd.abdn.ac.uk`

Somayajulu Sripada
Department of Computing Science
University of Aberdeen
`ssripada@csd.abdn.ac.uk`

Abstract

We present an empirical corpus study of the meaning and usage of time phrases in weather forecasts; this is based on a novel corpus analysis technique where we align phrases from the forecast text with data extracted from a numerical weather simulation. Previous papers have summarised this analysis and discussed the substantial variations we discovered among individual writers, which was perhaps our most surprising finding. In this paper we describe our analysis procedure and results in considerably more detail, and also discuss our current work on using parallel text-data corpora to learn the meanings of other types of words.

1 Introduction

NLP systems that interact with the world often need models of what words mean in terms of the non-linguistic world. In this paper, we describe how we have determined the meaning of time phrases in weather forecasts by analysing a parallel corpus of (A) manually-written weather forecast texts and (B) the numerical data (from a weather simulation) that the human forecasters examined when writing the textual forecasts. The analysis procedure first aligns (associates) text fragments with data segments, and then infers the meaning of each time phrase by statistically analysing the time of data segments that are aligned to textual phrases that contain this time phrase. This is broadly similar in concept to the use of parallel multilingual corpora in machine translation (Brown et al., 1990), except that our parallel corpus consists of texts and underlying numeric data, not texts and their translations. In other words, we are trying to learn what words mean in terms of non-linguistic data, not the best translations of words in another language.

Probably the biggest surprise in our analysis was the substantial variation we saw between individuals. For example, *by evening* apparently meant 1800 to some people, but 0000 to others. Although the possibility of such variation in individual idiolects has been acknowledged in the past (for example, (Nunberg, 1978; Parikh, 1994)), it seems to be ignored by most recent work on lexical semantics.

We have published other papers that have summarised our key findings, notably variation between individuals (Reiter and Sripada, 2002a; Reiter and Sripada, 2002b); and also described the corpus itself (Sripada et al., 2003b). The purpose of this paper is to describe our analysis procedure (including alignment) and results in detail, and to also discuss our current work on using parallel text-data corpora to learn the meanings of other types of words.

2 Previous Research

Linguists and lexicographers have used a number of different techniques to determine the meanings of words. These include asking native-speaker informants to judge the acceptability and oddness of test sentences (Cruse, 1986); defining word senses via lexicographic analysis of citations and corpora (Landau, 1984); and asking subjects to respond to 'fill in the blank' questions (Cassidy and Hall, 1996). These techniques have focused purely on texts, and have not analysed how texts and words relate to non-linguistic representations of the meanings of a text, which is our focus.

Psychologists interested in categorisation have done formal experiments to determine which objects human subjects consider to be in a mental category (Rosch, 1978; Malt et al., 1999). If we assume that the meaning of a word is one or more mental categories, then this research has shed considerable light on what words mean. However, like all psychological research, it has examined language usage in an artificial experimental context, not naturally occurring language.

In the NLP community, models of word meanings are typically either entered by a user or developer (for example in Microsoft's English Query natural-language inter-

day	hour	wind dir	wind speed
25-10-00	0	SSW	12
25-10-00	3	SSE	11
25-10-00	6	ESE	18
25-10-00	9	ESE	16
25-10-00	12	E	15
25-10-00	15	ENE	15
25-10-00	18	ENE	18
25-10-00	21	NNE	20
26-10-00	0	NNW	26

Table 1: Wind (at 10m) extract from 24-Oct-00 data file

face) or derived from a hand-built knowledge base (eg, (Reiter, 1991)). There is growing interest in trying to learn word meanings from parallel text-data corpora, for example (Siskind, 2001; Barzilay and Lee, 2002; Roy, 2002). We believe our work is unusual because we are using naturally occurring texts and data. Siskind (2001), in contrast, used data which was explicitly created for his experiments; Barzilay and Lee (2002) used texts which subjects had written for a previous experiment; and Roy (2002) used both data and texts that were created for his experiments.

3   SumTime Project and Corpora

The SUMTIME project is investigating better technology for building software systems that automatically generate textual summaries of time-series data. One of the domains SUMTIME is working in is weather forecasts, and in this domain we acquired a corpus of 1119 weather forecasts (for off-shore oil rigs) written by five professional meteorologists (Sripada et al., 2002; Sripada et al., 2003b). The reports were primarily based on the output of a numerical weather simulation, and our corpus contains this information as well as the forecast texts. Each forecast is roughly 400 words long, giving a total corpus size of about 400,000 words. The forecasts are split into an initial section which gives an overview of the weather, and then additional sections which give detailed forecasts for different periods of time. Figure 1 shows an example extract from a forecast text; this is the detailed description of predicted weather on 25 Oct 2000, from a forecast issued at 3AM on 24 Oct 2000.

Much of our analysis has focused on statements describing predicted wind speed and direction at 10 meters altitude during the first 72 hours after the forecast was issued. In other words, the WIND(10M) field from the detailed weather descriptions up to 3 days after the forecast was issued. One reason for focusing on wind statements is that they are based fairly directly on two fields from the data files, predicted wind direction and speed; the relationship between some of the other statements (such as

weather) and the data files is more complex. The predicted wind (at 10m) speed and direction on 25 Oct 2000, from the 24 Oct 2000 data file, is shown in Table 1. This is the primary information that the meteorologists looked at when writing the wind statement in Figure 1, although they also have access to other information sources, such as satellite weather photographs.

Each forecast contains 3 such wind statements, with an average length of approximately 10 words; hence there are about 30,000 words in our wind-statement subcorpus. This of course is very small compared to many text-only corpora such as the British National Corpus (BNC), but we believe that our weather forecast corpus is one of the largest parallel text-data corpora in existence.

4   Analysis Procedure for Time Phrases

One of SUMTIME's research goals is to learn the meaning of time phrases; in other words, what a forecaster meant when he used a time phrase such as *by evening* or *after midnight*. We also wished to learn which time phrase should be included in a computer-generated weather forecast text to indicate a time; for example, which time phrase should be used to indicate a change in the weather at 1200. Note that it is rare for weather forecasts to explicitly mention numerical times such as 1200, and also that although there are standard terminologies for some meteorological phenomena such as cloud cover and precipitation, we are not aware of any standard terminologies for the use of time phrases in weather forecasts.

We performed this analysis as follows. First we extracted the wind at 10 meters statements for the next 72 hours from all forecasts in our corpus, and parsed these texts with a simple parser tuned to the linguistic structure of these texts. The parser essentially broke sentences up into individual phrases, and then recorded the speed, direction, and time phrase mentioned in each such phrase, along with other information (such as verb) which was not used in the analysis described here. For example the WIND (10M) statement from Figure 1 was broken up by the parser into four wind phrases:

1. SSW 12-16
   (speed:12-16, direction:SSW, timephrase: none)

2. BACKING ESE 16-20 IN THE MORNING,
   (speed:16-20, direction:ESE, timephrase: IN THE MORNING)

3. BACKING NE EARLY AFTERNOON
   (speed:(16-20), direction:NE, timephrase: EARLY AFTERNOON)

4. THEN NNW 24-28 LATE EVENING
   (speed:24-28, direction:NNW, timephrase: LATE EVENING)

```
FORECAST 00-24 GMT, WEDNESDAY, 25-Oct 2000
  WIND(10M):   SSW 12-16 BACKING ESE 16-20 IN THE MORNING, BACKING
               NE EARLY AFTERNOON THEN NNW 24-28 LATE EVENING
      (50M):   SSW 15-20 BACKING ESE 20-25 IN THE MORNING, BACKING
               NE EARLY AFTERNOON THEN NNW 30-35 LATE EVENING
   SIG WAVE:   2.0-2.5 RISING 3.0-3.5 BY AFTERNOON
   MAX WAVE:   3.0-4.0 RISING 5.0-5.5 BY AFTERNOON
    WEATHER:   RAIN SOON, CLEARING TO SHOWERS IN THE EVENING
        VIS:   GOOD BECOMING MODERATE IN RAIN
```

Figure 1: Extract from 5-day forecast issued on 24-Oct-00

If a wind phrase did not specify speed or direction, the parser assumed that this was unchanged from the previous wind phrase; such elision is common in weather forecast texts. Thus, for example, the speed recorded for BACKING NE EARLY AFTERNOON is 16-20, which is the speed from the previous phrase (BACKING ESE 16-20 IN THE MORNING). Our parser successfully parsed 3225 of the 3357 WIND(10M) statements; 132 (4%) of the statements could not be parsed. The parser produced 8198 wind phrases in total.

From these 8198 wind phrases we selected those phrases which (a) included a time phrase, (b) did not use a qualifier such as *mainly* or *occasionally*, (c) did not specify that wind speed or direction was variable, (d) for which we had the corresponding data files, and (e) for which we knew the forecast author. There were 3654 such phrases. The majority (4014 phrases) of the eliminated phrases did not specify a time phrase, such as the first phrase (SSW 12-16) in the above example.

We next associated each wind phrase with an entry in the corresponding data file. In other words, we aligned the textual wind phrases with the numeric data file entries. As in other uses of parallel corpora, good alignment is essential in order for the results to be meaningful (Och and Ney, 2000).

To associate data file entries with wind phrases, we first searched the data file for entries which matched the wind phrase. An entry matched if its speed was within the range defined in the phrase, and if its direction was within 12 degrees of the direction mentioned in the phrase. In 343 cases, no data file entry matched the wind phrase. We believe that such cases were mostly due to (a) forecasters not literally reporting the data file, but instead adjusting what they said based on their meteorological expertise and on information not available to the numerical weather simulation (such as satellite weather images); (b) forecasters reporting a simultaneous change in wind speed and direction, when in fact speed and direction changed at different times (this may be due to forecasters trying to write texts quickly, so that they can use the most up-to-date data (Reiter et al., 2003)); and (c) forecaster errors.

For example, the third phrase in our example, BACKING NE EARLY AFTERNOON, does not match any of the data file entries shown in Table 1. This could be because the forecaster decided that the numerical forecast was underestimating the speed at which the wind was shifting, and hence he believed that the wind would be NE at 12 or 15, even though the data file predicted E and ENE for these times. It could also be that the forecaster made a mistake, and perhaps was intending to write ENE but wrote NE instead because he was writing under time pressure. In any case, wind phrases which did not match any data file entries were dropped from our analysis.

Out of the 3311 matched wind phrases, 1434 (43%) were unambiguous and only matched one data file entry. For example, the fourth wind phrase in our example, THEN NNW 24-28 LATE EVENING, matches only one data file entry, the one for 0000 on 26 Oct 2000.

1877 (57%) of the matched wind phrases were ambiguous and matched more than one data file entry. Typically this happened when the wind was changing slowly and hence two or more adjacent data file entries matched the wind phrase. In such cases we checked if one data file entry had a speed which was was closer than the other data files entries to the middle of the speed range in the textual wind phrase. This heuristic produced a preferred match for 1105 (33%) of the matched wind phrases, and left 772 (23%) phrases as ambiguous and unmatched.

For example, the second wind phrase in our example, BACKING ESE 16-20 IN THE MORNING, matches two data file entries: 0600 (direction ESE, speed 18) and 0900 (direction ESE, speed 16). The midpoint of the 16-20 speed range reported in the forecast is 18, so our speed heuristic matches this wind phrase to the 0600 data file entry, since its speed is closer to the speed range midpoint (indeed, it matches the midpoint).

We evaluated our alignment process by applying it to the subset of wind phrases which used a time phrase which we thought had a clear and unambiguous interpretation, namely an absolute time (such as *by 0600*), *by*

time	F1	F2	F3	F4	F5	total
0000	2	**9**	**80**	5	**14**	**110**
0300					1	1
0600					1	1
0900						0
1200		1				1
1500	2	1	1		2	6
1800	**30**	5	2	**27**	13	77
2100	13	6	8	2	11	40
total	37	22	91	34	42	236

Significance of differences: p < 0.001

Table 2: Usage of *by evening*, by forecaster (mode in **bold**)

time	F1	F2	F3	F4	F5	total
0000			**215**	9	15	**239**
0300					1	1
0600						0
0900						0
1200			1			1
1500						0
1800						0
2100			3	3	2	8
total	0	0	219	12	18	249

Significance of differences p < 0.001 (ANOVA: p = 0.06)

Table 4: Usage of *by late evening*, by forecaster (mode in **bold**)

time	F1	F2	F3	F4	F5	total
0000			2	1		3
0300				1		1
0600				1		1
0900	3		1	7	2	13
1200	**23**		**71**	**86**	**11**	**191**
1500	7	**1**	9	5	2	24
1800			2	2	1	5
2100	1					1
total	34	1	85	103	16	239

Significance of differences: p > 0.1

Table 3: Usage of *by midday*, by forecaster (mode in **bold**)

midday (which means 1200), *by midnight* (which means 0000), and *by end of period* (which means either 0000 or 0600, depending on the forecast section). The matching process was fairly accurate; in 86% of cases it produced the expected meaning (such as 0000 for *by midnight*).

Clearly we would benefit from better matching and alignment techniques, and we wonder if perhaps some of the alignment techniques used for parallel multi-lingual corpora (Och and Ney, 2000) could be adapted to help align our text-data corpora. This is a topic we plan to investigate in future research.

This matching/alignment procedure is different in detail from the preliminary analysis procedure reported in (Reiter and Sripada, 2002b). The procedure used in our earlier paper aligned fewer phrases (1382 vs. 2539) and had a higher error rate (22% vs. 14%), so it was inferior.

5 Results

We examined the association between time phrase and time in the 2539 aligned (wind phrase, data file entry) pairs. In this analysis, we regarded time phrases as different if they involved different head nouns (for example,

by evening and *by afternoon*), different prepositions (for example, *midday* and *by midday*) and/or different adjectives (for example, *by afternoon* and *by late afternoon*). However, we ignored determiners (for example, *by this evening* was regarded as the same phrase as *by evening*).

Tables 2, 3, and 4 gives details of the usage of the three most common non-contextual time phrases: *by evening*, *by midday*, and *by late evening*. This tables also shows the statistical significance of differences between forecasters, calculated with a a chi-square test (which treats time as a categorical variable). As some colleagues have expressed an interest in a one-way ANOVA analysis (which compares mean time), we show this as well where it gives a substantially different value from the chi-square analysis. This data suggests that

- *by evening* means different things to different people; for example, F1 and F4 primarily use this phrase to mean 1800, while F3 primarily uses this phrase to mean 0000.

- *by midday* was used in a very similar way by all forecasters (ignoring F2, who only used the term once).

- *by late evening* was used by all forecasters (who used this term) primarily to mean 0000. However, the usages of the different forecasters was still significantly different. This reflects a difference in the distribution of usage; in particular, F3 almost always (98% of cases) used this phrase to mean 0000, while F4 and F5 used this phrase to mean 0000 in about 80% of cases.

These patterns are replicated across the corpus: some phrases (such as *by midday* and *by morning*) are used in the same way by all forecasters; some phrases (such as *by evening* and *by late morning*) are used in very different ways by the forecasters; and some phrases (such as *by late evening* and *by midnight*) have the same core meaning (eg, 0000) but different distributions around the core.

We have, incidentally, looked for seasonal variations in meaning (for example, *by evening* meaning one thing in the winter and another in the summer), but we have found no evidence of such variation.

Roy (2002) has also noted variation in the meanings that individuals assign to words, in his parallel text-data study of object descriptions. For example, one object might be described as having the colour *pink* by one subject, but other subjects might have problems identifying the object when it was described as *pink*, because they did not consider it to have this colour.

Table 5 presents the most common time-phrase used by each forecaster for each time, including context-dependent phrases such as *later*. This highlights major 'stylistic' differences between forecasters in terms of which time phrases they prefer to use. For example, F1 and F2 make heavy use of contextual time phrases such as *later* and *soon*, while F5 (and to a lesser extent F4) seem to prefer to avoid such terms. It is also interesting that contextual time phrases are especially commonly used to refer to the time 0300. We wonder if this could reflect a 'lexical gap' in English; there are no commonly used time phrases in English for times around 0300, and perhaps this encourages the forecasters to use contextual time phrases to refer to 0300.

Table 6 presents the most common (mode) meaning of non-contextual time phrases, for each forecaster. Perhaps not surprisingly, the greatest variability occurs when a time phrase denoting a time period (*morning*, *afternoon*, or *evening*) occurs without being modified by an adjective (*early*, *mid*, or *late*). The data also suggests that the forecasters may disagree about the meaning of *morning*, with F4 in particular considering *morning* to be the period 0300-0900, while F5 considers *morning* to be the period 0600-1200.

6 Current and Future Work

6.1 Verb Choice

We would like to use our corpus to learn choice rules for verbs which are near-synonyms (Edmonds and Hirst, 2002). We are currently attempting to learn rules which predict which of three possible verbs – *decreasing*, *easing*, and *falling* – are used when the wind speed decreases.

We have conducted two experiments. The first was a semantic analysis, where we attempted to learn a choice rule based on features extracted from the numerical data. To do this, we used our aligned corpus to extract semantic features which we thought could be relevant to this decision (such as the amount by which the wind speed has decreased), and then analysed this with Ripper (Cohen, 1995). This gave the rules shown in Figure 2; these again show substantial variation between individual fore-

verb	F1	F2	F3	F4	F5	total
decreasing	0	0	3	**2**	0	5
easing	1	**19**	0	0	0	20
falling	**4**	0	**61**	0	0	**65**

Table 7: Choice of wind decrease verb when subsequent word is *variable*, by forecaster (mode in **bold**)

casters. These rules are mildly effective; 10-fold cross validation error is 25%, compared to a baseline error rate of 33% from always choosing the most common verb (*easing*). These rules suggest that at least for some forecasters, *decreasing* suggests a larger change in the wind speed than *easing*; this is the sort of near-synonym connotational difference that we expected to find. More surprisingly (at least to us), the presence of forecast date in some of the rules suggests that forecasters change how they write over time. Perhaps in retrospect this should not have been a surprise, because we have also observed changes over time in how people write in a previous project (Reiter et al., 2000).

We also analysed collocation effects, that is whether we could predict verb choice based purely on the words immediately preceding and following the verb (and hence ignoring the numerical prediction data). This was done on the complete corpus (not just verbs that were part of successfully aligned phrases). It is difficult to directly compare the collocation analysis with the semantic one due to differences in the corpora texts used, but in general terms the reduction in baseline error rate seems comparable to the semantic analysis. Some of the collocation effects were both strong and forecaster-dependent. For example, Table 7 shows the choice of wind decrease verb when the word following the verb was *variable* (indicating wind direction was variable). In this context, forecasters F1 and F3 usually used *falling*; F2 always used *easing*; and F4 always used *decreasing* (F5 never used *variable* in his forecasts). Similar individual differences were observed in other collocations. For example, when the word preceding the verb was *gradually*, F3, F4, and F5 preferred *decreasing*, but F2 always used *easing* (F1 never used *gradually* in his forecasts).

In summary, it seems that the choice between the near synonyms *decreasing*, *easing*, and *falling* depends on

- *semantics:* how much the actual wind speed has changed;

- *collocation:* immediately preceding and following words in the sentence;

- *author:* which forecaster wrote this particular text;

- *date:* when the text was written.

time	F1	F2	F3	F4	F5	all
0000	**later**	**later**	**by late evening**	by midnight	in evening	later
0300	later	soon	**soon**	**soon**	tonight	**soon**
0600	**later**	**overnight**	**soon**	by morning	later in period	later
0900	**soon**	**soon**	soon	by midday	in morning	**soon**
1200	by midday	soon	**by midday**	**by midday**	**in morning**	**by midday**
1500	**by afternoon**	soon	**by mid afternoon**	by mid afternoon	early afternoon	by mid afternoon
1800	**by evening**	by evening	**by late afternoon**	**by evening**	**by evening**	**by evening**
2100	later	later	by evening	later	in evening/ by evening	later/ by evening

bold font means this phrases was at least twice as common as the second-most common term.

X/Y means X and Y were equally common

Table 5: Most common time-phrases for each time, by forecaster

Forecaster	rule
F1	never
F2	never
F3	speed change >= 10 knots AND time interval >= 15 hours
F4	speed change >= 9 knots OR forecast date >= 2-November-2001
F5	forecast date >= 30 March 2001

Choose *decreasing* if

Otherwise choose *easing*

Figure 2: Verb choice rule based on data features

All of these factors are important, and in particular the kind of semantic differences investigated by (Edmonds and Hirst, 2002) are only one factor among many, and do not dominate the choice decision. We plan to continue working on this and other analyses of near-synonyms, and obtain a better idea of how these factors interact.

6.2 Other corpora

In addition to the weather corpus, the SUMTIME project also has access to a parallel text-data corpus of doctors describing signal data from a neonatal intensive care unit (Alberdi et al., 2001). We would like to analyse this corpus to determine the meanings of words such as *steady* and *oscillations*. However, a preliminary analysis has suggested to us that we cannot conduct such an analysis until we remove non-physiological events from the data (Sripada et al., 2003a). For example, a doctor may describe a signal as *steady* even when it contains a large spike, if the doctor believes the spike is due to a non-physiological event (such as a sensor falling off the baby and then being replaced by a nurse). Hence non-physiological events (known in this domain as 'artifacts') must be removed from the data before it is possible to analyse word meaning. We are currently working on artifact removal, and once this is complete we will start our analysis of word meanings.

SUMTIME is also working on generating textual summaries of gas turbine sensor data (Yu et al., 2003). Unfortunately in this domain, as in many other NLG applications (Reiter et al., 2003), there is no existing corpus of manually written texts describing the input data. We have explicitly asked two experts to write descriptions of 38 signal fragments. This very small corpus showed that once again there were major differences between individuals (Reiter and Sripada, 2002a), but the corpus is too small to allow meaningful statistical analysis of word meanings.

7 Conclusion

To conclude, we believe that parallel text-data corpora are a valuable resource for investigating lexical semantic and pragmatic issues, and can help shed valuable light on the fundamental question of how words relate to the non-linguistic world. We have described in detail how we have used such a corpus to analyse the meaning and usage of time phrases in weather forecasts, and also sketched our current work on other analyses of text-data corpora. We hope that other researchers interested in semantics and pragmatics will find our techniques interesting, and consider whether they might be useful in exploring other semantic and pragmatic questions about word meaning usage.

phrase	F1	F2	F3	F4	F5	combined
after midnight					0600	0600
afternoon				*	1630	1630
around midday	*			*	*	1200
by 0600					0600	0600
by afternoon	1500	1200		1200	1200	1200
by early afternoon	*	*		*	1330	1330
by early evening				1800	*	1800
by early morning				0300		0300
by evening	1800	0000	0000	1800	0000	0000
by late afternoon			1800		*	1800
by late evening			0000	0000	0000	0000
by late morning	*			0900	1200	1030
by mid afternoon			1500	1500	*	1500
by mid evening			*	2100	*	2100
by mid morning			*	*	*	0900
by midday	1200	*	1200	1200	1200	1200
by midnight	0000			0000	0000	0000
by morning		0600		0600	*	0600
during afternoon		*			*	1500
during evening		0000		0000	0000	0000
during morning	*	1030		*	0900	0900
early afternoon	*	*		*	1500	1500
early evening				*	1800	1800
evening	1612			2100	0000	0000
from midday				1200		1200
in afternoon	*			*	1800	1800
in evening			0000		0000	0000
in morning					1200	1200
late evening	*		0000		0000	0000
later in evening					0000	0000
later in night					0600	0600
mid morning			*	*	*	0900
overnight	0600	0600		*		0600
tonight	*				0000	0000

* means phrase was used fewer than five times by this forecaster.

Phrases used less than 5 times by all forecasters combined are omitted.

Contextual phrases (such as *later*) are also omitted.

If 2 or more times are equally common, their average is shown.

Table 6: Most common (mode) usage of time phrases, by forecaster

References

Eugenio Alberdi, Julie-Clare Becher, Ken Gilhooly, Jim Hunter, Robert Logie, Andy Lyon, Neil McIntosh, and Jan Reiss. 2001. Expertise and the interpretation of computerized physiological data: implications for the design of computerized monitoring in neonatal intensive care. *International Journal of Human-Computer Studies*, 55:191–216.

Regina Barzilay and Lillian Lee. 2002. Bootstrapping lexical choice via multiple sequence alignment. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP-2002)*.

Peter F. Brown, John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A statistical approach to machine translation. *Computational Linguistics*, 16(2):79–85.

Frederick Cassidy and Joan Hall, editors. 1996. *Dictionary of American Regional English*. Belknap.

William Cohen. 1995. Fast effective rule induction. In *Proc. 12th International Conference on Machine Learning*, pages 115–123. Morgan Kaufmann.

D. Cruse. 1986. *Lexical Semantics*. Cambridge University Press.

Philip Edmonds and Graeme Hirst. 2002. Near-synonymy and lexical choice. *Computational Linguistics*, pages 105–144.

Sidney Landau. 1984. *Dictionaries: the art and craft of lexicography*. Scribner.

Barbara Malt, Steven Sloman, Silvia Gennari, Meiyi Shi, and Yuan Wang. 1999. Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, 40:230–262.

Geoffrey Nunberg. 1978. *The Pragmatics of Reference*. University of Indiana Linguistics Club, Bloomington, Indiana.

Franz Och and Herman Ney. 2000. A comparison of alignment models for statistical machine translation. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, pages 1086–1090.

Rohit Parikh. 1994. Vagueness and utility: The semantics of common nouns. *Linguistics and Philosophy*, 17:521–535.

Ehud Reiter and Somayajulu Sripada. 2002a. Human variation and lexical choice. *Computational Linguistics*, 28:545–553.

Ehud Reiter and Somayajulu Sripada. 2002b. Should corpora texts be gold standards for NLG? In *Proceedings of the Second International Conference on Natural Language Generation*, pages 97–104.

Ehud Reiter, Roma Robertson, and Liesl Osman. 2000. Knowledge acquisition for natural language generation. In *Proceedings of the First International Conference on Natural Language Generation*, pages 217–215.

Ehud Reiter, Somayajulu Sripada, and Roma Robertson. 2003. Acquiring correct knowledge for natural language generation. *Journal of Artificial Intelligence Research*. Forthcoming.

Ehud Reiter. 1991. A new model of lexical choice for nouns. *Computational Intelligence*, 7(4):240–251.

Eleanor Rosch. 1978. Principles of categorization. In E. Rosch and B. Lloyd, editors, *Cognition and Categorization*, pages 27–48. Lawrence Erlbaum, Hillsdale, NJ.

Deb Roy. 2002. Learning visually grounded words and syntax for a scene description task. *Computer Speech and Language*, 16:353–385.

Jeffrey Siskind. 2001. Grounding the lexical semantics of verbs in visual perspective using force dynamics and event logic. *Journal of Artificial Intelligence Research*, 15:31–90.

Somayajulu Sripada, Ehud Reiter, Jim Hunter, and Jin Yu. 2002. SUMTIME-METEO: Parallel corpus of naturally occurring forecast texts and weather data. Technical Report AUCS/TR0201, Computing Science Dept, Univ of Aberdeen, Aberdeen AB24 3UE, UK.

Somayajulu Sripada, Ehud Reiter, Jim Hunter, and Jin Yu. 2003a. Exploiting a parallel text-data corpus. In *Proceedings of Corpus Linguistics 2003*. UCREL, Lancaster University, UK.

Somayajulu Sripada, Ehud Reiter, Jim Hunter, and Jin Yu. 2003b. Summarising neonatal time-series data. In *Proceedings of EACL-2003*. Forthcoming.

Jin Yu, Ehud Reiter, Jim Hunter, and Somayajulu Sripada. 2003. Sumtime-turbine: A knowledge-based system to communicate gas turbine time-series data. In *Proceedings of IEA/AIE-2003*.