1 What This Document is About

This document provides bibliographic details for the various materials cited in the summer school course on Automated Writing Assistance: Grammar Checking and Beyond, held in Tarragona, Spain as part of the International Summer School in Language and Speech Technologies (SSLST 2011) on 31st August and 1st September 2011. The class material was broken down into five topics; the references here are organized under that same structure.

2 The Nature of the Problem

We set the scene using the conduit metaphor as introduced by Reddy (1979), and adopted a definition of technical writing due to Britton (1965). The example of a stage model of the writing process was due to Rohman (1965); the more sophisticated cognitive process model is from Flower and Hayes (1981). We also examined a taxonomy of revision processes proposed by Faigley and Witte (1981).

We discussed an extensive analysis of student writing errors carried out by Connors and Lunsford (1988), and briefly reviewed a number of taxonomies of error (Douglas and Dale, 1991; Becker et al., 2003; Busta et al., 2009).

3 Spell Checking

The now somewhat mundane concerns of early spelling checkers are discussed by Peterson (1980). We discussed a range of approaches to correcting non-word spelling errors:

- Angell, Freund, and Willett (1983) on trigram analysis;
- Yannakoudakis and Fawthrop (1983) on error patterns;
- van Berkel and de Smedt (1988) on triphone analysis;
- Kernighan, Church, and Gale (1990) on the noisy channel model;
- Agirre et al. (1998) on the use of context;
- Brill and Moore (2000) on integrating a model of string-to-string edits into the noisy channel model; and
- Toutanova and Moore (2002) on integrating pronunciation modeling into the noisy channel model.

We also discussed Mays et al.’s (1991) approach to using trigrams to detect and correct real-word errors, and Hirst and Budanitsky’s (2005) attempt to use semantics for the same problem. Finally, we described the work of Whitelaw et al. (2009) on using the web as a source of information for spelling correction.

We also alluded to, but did not discuss, a number of approaches to real-word error correction based on the idea of confusion sets: (Golding, 1995; Golding and Schabes, 1996; Mangu and Brill, 1997).
4 Grammar Checking

The range of different grammatical errors we discussed at the outset were drawn from (Douglas and Dale, 1991). The Unix Writer’s Workbench is described in (Macdonald, 1983); Atwell’s approach to Constituent-Likelihood Error Detection is described in (Atwell, 1987).

The description of EPISTLE.CRITIQUE was drawn from (Heidorn et al., 1982; Jensen et al., 1983). A number of relevant papers are collected together in (Jensen, Heidorn, and Richardson, 1993). The Microsoft Word grammar checker is described in some detail by Heidorn (2000).

In terms of specific techniques for grammar checking, an early example of relaxation is presented by Weischedel and Black (1980). Douglas and Dale (1992) provide a description of relaxation in the PATR grammatical framework; Schneider and McCoy (1998) provide a description of the mal-rules approach.

We also mentioned the following other techniques:

• fitted parsing (Jensen et al., 1983);
• mixed bottom-up and top-down parsing (Mellish, 1989); and
• minimum edit distance parsing (Lee et al., 1995).

Kohut and Gorman (1995) provide an evaluation of five commercial grammar-checking packages that were available in the mid-1990s.

5 Handling ESL Errors

Bolt (1992) tested seven grammar-checking programs of the time against 35 sentences containing ESL errors. Donahue (2001) provides an analysis of ESL errors that is contrasted with the findings of Connors and Lunsford (1988) for native speakers. The counts of errors in the Cambridge Learners Corpus were taken from Leacock et al. (2010), as were a number of other tabulations of data used in this lecture. The ‘Helping Our Own’ (HOO) task is described in (Dale and Kilgarriff, 2010).

The three approaches to article errors we discussed are (Knight and Chander, 1994; Han, Chodorow, and Leacock, 2006; De Felice and Pulman, 2008). Pelletier’s universal grinder and universal packager are introduced in (Pelletier, 1975).

The papers cited in the tabular summary of work on preposition errors are as follows:

• Papers which address preposition selection in well-formed text: (Lee and Seneff, 2008; Chodorow, Tetreault, and Han, 2007; De Felice and Pulman, 2007; De Felice and Pulman, 2008; Tetreault and Chodorow, 2008a; Gamon et al., 2008; Bergsma, Lin, and Goebel, 2009);

• Papers which address preposition error detection on learner data: (Eeg-Olofsson and Knuttson, 2003; Tetreault and Chodorow, 2008b; De Felice and Pulman, 2009; Hermet, Dsilets, and Szpakowicz, 2008; Tetreault and Chodorow, 2009; Gamon, 2010; Han et al., 2010).

We also briefly mentioned Lee and Seneff’s (2008) work on verb form errors. Other supporting tools that were cited in passing were Collin’s parser (Collins, 1999) and Lin’s work on automatic thesaurus construction (Lin, 1999).

6 Beyond the Sentence

We revisited Flower and Hayes’ (1981) cognitive process model of writing and Faigley and Witte’s (1981) taxonomy of revision operations to motivate looking at other aspects of the writing process where assistance might be provided.

Our discussion of architectures for natural language generation was drawn from (Reiter and Dale, 2000); the primary reference on Rhetorical Structure Theory is (Mann and Thompson, 1988).

References


