Towards Context-dependent Phonetic Spelling Error Correction in Children's Freely Composed Text for Diagnostic and Pedagogical Purposes

Sebastian Stüker[1], Johanna Fay[2], and Kay Berkling[1]

[1]Research Group 3-01 'Multilingual Speech Recognition'
Karlsruhe Institute of Technology, Karlsruhe, Germany
[2]Institut für deutsche Sprache und Literatur, Pädagogische Hochschule Karlsruhe, Germany
sebastian.stueker@kit.edu, kay@berkling.com, fay@ph-karlsruhe.de

10.21437/Interspeech.2011-480

Abstract

Reading and writing are core competencies of any society. In Germany, international and national comparative studies such as PISA or IGLU have shown that around 25% of German school children do not reach the minimal competence level necessary to function effectively in society by the age of 15. Automized diagnosis and spelling tutoring of children can play an important role in raising their orthographic level of competence. One of several necessary steps in an automatic spelling tutoring system is the automatic correction of achieved text that was freely written by children and contains errors. Based on the common knowledge that children in the first years of school write as they speak, we propose a novel, context-sensitive spelling correction algorithm that uses phonetic similarities, in order to achieve this step. We evaluate our approach on a test set of texts written by children and show that it outperforms Hunspell, a well established isolated error correction program used in text processors.

Index Terms: Spelling Error Correction, Children, Education

1. Introduction

Reading and writing are core competencies for success in any society. In Germany, international and national comparative studies such as PISA or IGLU [1] have shown that around 25% of German school children do not reach the minimal competence level necessary to function effectively in society by the age of 15. In our vision, modern language processing technologies can help to raise the level of competence in spelling for school children, by automatically diagnosing the type of errors committed by the individual student and recommending tailored exercises based on personalized error profiles derived from the diagnosis.

The diagnostic tools that are on the market today offer pricey one-time spelling diagnosis on a fixed test set with high-density error-prone and unnatural text and pre-specified word field analysis. In these tools, usually variants of achieved spellings are predicted based on a-priori known reference words. Potential errors are therefore manually categorized by experts during test set design. Internet-based or paper-based diagnostic tests, such as the 'Diagnostische Rechtschreibtest' (DRT) [2], 'Deutsche Rechtschreibtest' (DERET) [3], and 'Hamburger Schreibprobe' (HSP) [4] work in similar ways to categorize errors.

However, according to recent research by Fay [5], this sort of error analysis deviates, at least in parts, significantly from the error profile derived from a child's spelling skills based on self-chosen and freely written text. The latter therefore presents a more natural picture of the child's competence level. Gathering diagnostic information requires more sophisticated evaluation tools that lay persons can apply frequently and automatically, in order to track progress and maintain effective spelling tutoring as the child's profile changes.

The first step in such a tool has to be the automatic reconstruction of the orthographically correct target text as intended by the child based only on the achieved text that contains the spelling errors. In this paper we present our approach to achieve this. We introduce a context-sensitive spelling error correction system that uses phonetic similarities. While addressing all types of errors, our approach is especially apt at targeting the majority of child-specific spelling errors that are based on the difficulty of representing phonics with graphemes in an ambiguous spelling system such as German. The underlying assumption of this approach then leverages the phonetic similarity between the achieved text and correct text candidates to reconstruct the target text. In our approach the text achieved by the child is first transferred into a phonetic representation. The resulting phoneme string then serves as a feature vector for a speech recognition system that finds the most likely, correctly spelled word sequence given the phonemes as input. For this, it uses an acoustic model that measures similarity of the achieved text and the hypothesized correct text on the phonetic level. The contextual knowledge of this spelling correction system is drawn from the language model of the speech recognition system which consists of an n-gram model.

The steps for automatically profiling the errors committed by a writer and putting together recommendations for him are the topic of a companion paper submitted to the ISCA Special Interest Group on Speech and Language Technology in Education (SLaTE) workshop 2011 [6].

The rest of this paper is organized as follows. Section 2 presents a review of existing approaches for spelling diagnosis on the market as well as the current state-of-the-art in the field of automatic spelling correction. Section 3 then introduces our approach to building a phonetic distance based, context-sensitive spelling correction system. In Section 4 we report the results of our experiments that we conducted to evaluate our approach. Finally, Section 5 gives an outlook on extending the initial system to incorporate more knowledge, in order to improve its reconstruction accuracy, and to recover from more complex types of errors.

2. Related Work

For the purpose of this work two distinct fields of study are of relevance. First, there is the area of educational science, especially the area that deals with how children acquire the skill of writing. German education theorists have looked at the problem of spelling errors and their development in children for decades. Standardized tests have been developed, applied and normed according to grade levels. This progress has to be taken into account as well as the technological aspect, in our case automatic spelling error correction. Automatic spell checkers that have been developed in the past are usually interactive and in common use. We will extend the current-state-of-the art in this paper by presenting an approach that is at the same time context-sensitive and incorporates phonetic knowledge.

2.1. Standardized Spelling Diagnosis

The individual support for children according to their personal strengths and weaknesses is key to the process of learning to write and spell correctly. This, in turn, requires the ability to regularly diagnose the child's performance accurately and frequently, as well as to determine the exact type of errors committed. Over the last 20 years, German linguists and educators have been developing such diagnostic tools. As a result a number of pencil and paper tests have been developed as standardized tests with large data collections to form statistically accurate diagnoses, normed for specific grade levels such as the aforementioned grade-level specific DRT, DERET, and HSP. They are expensive to administer and cover word level and sentence level spelling errors where both words and sentences are predetermined and either dictated to the child or elicited via pictures. Administration of these tests have been facilitated by providing on-line 'fill-in-the-blank' approaches to the tests. But as mentioned above, research by Fay [7] indicates, that the analysis of freely written text is superior, but tools with this capability have not yet been developed.

2.2. Automatic Spell Checkers

In literature two tasks related to dealing with spelling correction are usually considered. The first task deals with *error detection*, i.e. to mark the place at which spelling errors have occurred, while the second tasks addresses the problem of *error correction*, that is correcting a misspelled string [8]. The latter task is the one of interest for our purposes.

Kukich [8] further distinguishes between *isolated word error correction* and *context-dependent word error correction*. While the former concentrates only on single, misspelled words and their correction in isolation from the surrounding text, the latter takes the textual context of a word into account for potential correction. Thus, context dependent word error correction is potentially more powerful than the isolated form.

With respect to the type of spelling errors encountered, Kukich [8] distinguishes between *typographic*, *cognitive*, and *phonetic* errors. Typographic errors are assumed to result from wrong input on the keyboard that are caused by motor slips, while the typer knows the correct spelling. These types of errors are currently not of interest in this work, as for our initial experiments we assume that we have an electronic representation of the children's writing that correctly reflects the hand-written letter sequence. Cognitive errors are of interest, however, as they are assumed to be caused by a misconception or lack of knowledge on behalf of the writer. Especially interesting for us are the class of phonetic errors, in which the child writes a phonetically correct, but orthographically incorrect letter sequence, as we work under the hypothesis that a major part of the errors committed by the children are of this type. With regard to the nature of the spelling errors found in text, the findings from [9],[10], and [11] which are cited in [8] do not directly apply to our work, as we are a) dealing with children still learning how to write, and b) with handwritten text. It will be part of our future research to collect a sufficient amount of data to allow similar analysis on children's freely hand-written texts.

Regarding phonetic errors, Kukich [8] cites scenarios, where people have to spell names or other unknown words, e.g., when querying telephone directories [12], encyclopedia [13], or law enforcement databases [14]. We believe, that our application presents a similar scenario in which all words are subject to errors of phonetic nature committed by the learning writers.

Most isolated word error correction systems surveyed in [8] work with a background dictionary. Once a misspelled word is detected, the correct word or a ranked list of candidate words is selected based on a distance measure, such as the minimum edit distance, a simple n-gram vector distance, or a singular value decomposition based n-gram vector distance. Alternatively, probabilistic models can be used to select the most likely correct word(s), or an artificial neural network classifier can be trained to directly select the correct word [15].

In order to be able to detect real-word errors, that is errors that result in another valid word, the use of context-dependent word error correction systems is required. Kukich [8] divides these systems into those utilizing traditional natural language processing techniques, and those that use statistical language models, such as tri-grams. Verbene [16] similarly describes a tri-gram based spell checking and correction system, but suffers, as also described in Kukich [8], from a high false-alarm rate.

While all the systems above only use the written form of the word, Berkel [13] and Toutanova [17] proposed to take the pronunciation of the misspelled words into consideration by using text-to-phoneme conversion technology as it is for example used in speech synthesis systems. While Berkel uses the information value of tri-phones within a word, Toutanova works with a Bayesian, probabilistic classification frame work. However, both papers only propose an isolated word error correction system, that combines phonetic and letter similarities for correcting single words, but does not incorporate the context of the misspelled words.

In contrast, the work presented in this paper proposes a context-dependent word error correction system that uses phonetic information in combination with context information in the form of an n-gram language model. To do so, we make use of algorithms and tools from automatic speech recognition research that also uses a Bayesian classification frame work.

3. Context-dependent Phonetic Spelling Error Correction

In order to build a phonetic, context-dependent spelling error correction system we make use of algorithms from *large vocabulary continuous speech recognition* (LVCSR). Similar to the noisy channel approach in isolated word error correction in [18], LVCSR searches for the most likely word sequence $\hat{W}$ given an audio recording X. By applying Bayes' theorem we obtain the fundamental equation of speech recognition:

$$\begin{aligned} \hat{W} &= \underset{W}{\operatorname{argmax}} P(W|X) = \underset{W}{\operatorname{argmax}} \frac{P(X|W)P(W)}{P(X)} \\ &= \underset{W}{\operatorname{argmax}} P(X|W)P(W) \end{aligned} \tag{1}$$

$P(X|W)$ is called the acoustic model and describes the relation between the observed sound wave, produced by the act of speaking, and the hypothesized word sequence. $P(W)$ is called the language model and describes the a-priori probability of observing a specific word sequence without considering the recorded audio. State-of-the-art systems realize acoustic models with *Hidden Markov Models* (HMMs) [19] and language models with n-gram models [20]. The $argmax$ operator in the the equation is usually referred to as as search, as it finds the most likely word sequence out of the search space of all possible word sequences, and is often realized as a Viterbi beam search [21]. We conducted our recognition experiments with the Janus Recognition Toolkit that features the Ibis single pass decoder [22].

3.1. Pre-processing and Acoustic Model

In order to formulate the spelling correction problem in terms of speech recognition, we take the achieved word string potentially containing errors and transform it into a feature vector X. The feature vector consists of the phoneme sequence of the achieved string, as we want to operate with phonetic similarities for finding the correct word sequence.

The acoustic model for computing the probability $P(W|X)$, that a (correctly spelled) word sequence W has produced the feature vector X, uses HMMs. A correct word sequence is first transformed into a sequence of phonemes using a pronunciation dictionary. The phonemes of this sequence are then modeled with context-independent HMMs, where each phoneme is modeled with one state that either allows a self-loop transition or a transition into the next phoneme state in sequence. In order to model the case that the child has achieved a text that translates into fewer phonemes than the correct text has, we duplicate each phoneme in the feature vector three times. The duplicated states can be easily absorbed by the self-loops in the HMM topology, and the achieved string now may have as little as a third of the phonemes of the correct string—a reasonable lower limit. In order to translate the achieved string into a phoneme sequence, as well as building the pronunciation dictionary of the recognition system, we used the speech synthesis tool MARY [23] developed at DFKI. Since MARY also has a rule based component for deriving the pronunciation of unknown words, it can be easily applied to misspelled text.

The emission probability distributions of the HMM states are for now not learned on training data, but are pre-defined by us, based on the phonetic similarities of the phoneme that a state represents and the phoneme that it supposedly emitted. A confusion matrix based on linguistic features [24] is used for defining these probability distributions. First a score for every phoneme state is calculated for the emission of every other by summing up points over all shared phonetic categories between the state phoneme and the phoneme to be emitted. Points approximately reflect the degree of relatedness between two phonemes containing this feature. Thus, if we were to make a tree of all phoneme features, then the number reflects the depth of the tree at which a particular feature is located. For example, phonemes can be either vowels or consonants (1 point), vowels can be short or long (1.5 points), short vowels can be either back or front (2 points). From this basic method, ambiguities are resolved with linguistic knowledge and points are altered by looking at the relative similarity of phonemes at different depths in the tree. The resulting scores are normalized and used as emission probabilities for the HMM states.

For our initial experiments in this paper we restrict the search space of word sequences that can be hypothesized to those that contain exactly the same number of words as the initial word sequence achieved by the writer. We do so, by introducing a special phoneme *WB* that indicates word boundaries in the feature vector, and by demanding that every word in the search space ends with this special phoneme. The probability that the HMM state that represents WB produces any other phoneme is set to 0.

3.2. Language Model and Vocabulary

As language model for our experiments we used a German 4-gram language model that was developed for the speech recognition task within the *Quaero* project [25]. The language model is an interpolation of language models trained on a wide variety of corpora, such as newspaper and news wire texts, transcriptions of web video and pod casts, and data collected from the World Wide Web (WWW).

The vocabulary that we used is the vocabulary of the Hunspell dictionary [26]. Since in our preliminary experiments we did not want to deal with the problem of out-of-vocabulary (OOV) words, we also added missing words from the reference texts to our vocabulary. In total, the dictionary contains approx. 67,000 words.

4. Experiments

4.1. Database

Very little data is available concerning children's free writing. For the purpose of evaluating the proposed system, two existing databases have been merged. The first is taken from the dissertation of Fay [7]. The data from that work has been transcribed yielding an achieved and target text with word level correspondence so that it is known which words are misspelled. The Fay database was collected after the teacher read a story to the children about a king's battle with soldiers. The texts were written by children based on the task of telling their own story about that topic. The data includes 10 texts from three different classes at each of four grade levels spanning the entire German primary school—120 texts in total.

The second set of texts has been taken from the dissertation of Thelen [27]. It is based on children who write a text corresponding to a picture story about a man and a dog. The resulting texts are mostly from second grade children. Text from three classes have been transcribed at the word level resulting in 15 (2nd grade) + 19 (2nd grade) + 26 (4th grade) texts. In total, both databases contain 558 sentences and almost 54,000 words.

4.2. Results

Using our proposed system described in Section 3 we automatically corrected the texts in the database described above. In order to compare our approach against another spelling correction approach, we also performed the error correction of the children's texts with the program Hunspell [26]. Since we are not interested in interactive error correction, we automatically selects the first correction proposal by Hunspell. For all corrections we used the word error rate (WER) of the corrected text vs. the orthographical reference as measure of quality.

For our context-sensitive error correction procedure we need to determine the optimal language model weight in the overall score computation. Since we are lacking a development set, we optimized the language model weight directly on our test set. The results of the experiments are summarized in Table 1.

The texts written by the children show a WER of 20.1%

compared to the orthographically corrected text. Using Hunspell to correct the children's text reduces the WER to 15.6%. Using our approach, but without a language model only improves the word error rate to 17.9%, worse than with Hunspell. But when we incorporate the language model knowledge, the WER drops down to 9.7% a clear improvement over the result from Hunspell. Thereby the language model used has a perplexity of 448 on the test set.

Table 1: *Results*

Approach	WER in %	PPL
original text achieved by the children	20.1	—
Hunspell	15.6	—
no LM	17.9	—
gen. German LM	9.7	448

5. Conclusion and Future Work

In this paper we have introduced a novel approach to automatic spelling correction which uses phonetic similarities in a context-sensitive frame work. We have evaluated the performance of the approach in the context of diagnosing the spelling errors of German children learnering to write. The results on real children's texts prove the feasibility of our approach and that it outperforms a well established isolated word error correction program.

Future research will be directed at extending the approach. By conducting a large scale data collection campaign, we plan to actually train the emission probabilities of the HMM states, instead of manually defining them. Some superficial analysis of the results show, that the current definition of the probabilities still has a clear room for improvement. We further plan to lift the restrictions imposed by the word boundaries in order to detect, e.g., compounding errors. The results and the comparison to Hunspell also show, that not only phonetic similarities but also graphemic similarities are important. We therefore aim to enhance the performance of the correction system by also taking into account the distance between graphemes in addition to the phonemes.

6. Acknowledgements

We would like to thank Kevin Kilgour for providing the language model from the 2010 Quaero evaluation. 'Research Group 3-01' received financial support by the 'Concept for the Future' of Karlsruhe Institute of Technology within the framework of the German Excellence Initiative.

7. References

[1] W. Bos, "IGLU: Einige Länger der BRD im nationalen und internationalen Vergleich," Münster, 2004.

[2] R. Müller, "Diagnostischer Rechtschreibtest für 3. Klassen. DRT 3: Manual," Göttingen, 2004.

[3] C. Stock and W. Schneider, "Deutscher Rechtschreibtest für das 1.-4. Schuljahr," Göttingen, 2008.

[4] P. May, U. Vieluf, and V. Malitzky, "Diagnose orthographischer Kompetenz: Hamburger Schreibprobe: Handbuch, Manual," Hamburg, 2007.

[5] J. Fay, "Kompetenzfacetten in der Rechtschreibdiagnostik. Rechtschreibleistung im Test und im freien Text," in *Didaktik Deutsch: Symposium Deutschdidaktik*, A. Bermerich-Vos, Ed. Schneider Verlag, 2010, vol. 29, pp. 15–36.

[6] K. Berkling, J. Fay, and S. Stüker, "Speech technology based framework for quantitative analysis of german spelling errors in freely composed children's texts," in *SLaTE*, 2011.

[7] J. Fay, *Die Entwicklung der Rechtschreibkompetenz beim Textschreiben. Eine empirische Untersuchung in Klassen 1 bis 4*. Frankfurt: Peter Lang, 2010.

[8] K. Kukich, "Techniques for automatically correcting words in text," *ACM Computing Surveys*, vol. 24, December 1992.

[9] J. J. Pollock and A. Zamora, "Collection and characterization of spelling errors in scientific and scholarly text," *Journal of the American Society for Information Science*, vol. 34, pp. 51–58, January 1983.

[10] R. Mitton, "Spelling checkers, spelling correctors and the misspellings of poor spellers," *Information Processing & Management*, vol. 23, pp. 495–505, 1987.

[11] F. J. Damerau, "A technique for computer detection and correction of spelling errors," *Communications of the ACM*, vol. 7, pp. 171–176, March 1964.

[12] J. Veronis, "Morphosyntactic correction in natural language interfaces," in *COLING*. Budapest, Hungary: ACL, August 1988, pp. 708–713.

[13] B. van Berkel and K. D. Smedt, "Triphone analysis: a combined method for the correction of orthographical and typographical errors," in *ANLC*. ACL, 1988.

[14] B. T. Oshika, B. Evans, F. Machi, and J. Tom, "Computational techniques for improved name search," in *ANLC*. ACL, 1988.

[15] V. J. Hodge and J. Austin, "A comparison of a novel neural spell checker and standard spell checking algorithms," *Pattern Recognition*, vol. 35, pp. 2571–2580, 2002.

[16] S. Verberne, "Context-sensitive spell checking based on word trigram probabilities," Master thesis Taal, Spraak & Informatica, University of Nijmegen, 2002.

[17] K. Toutanova and R. C. Moore, "Pronunciation modeling for improved spelling correction," in *ACL*. Philadelphia, PA, USA: ACL, July 2002.

[18] E. Brill and R. C. Moore, "An improved error model for noisy channel spelling correction," in *ACL*, 2000.

[19] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257–286, February 1989.

[20] F. Jelinek, *Statistical Methods for Speech Recognition*. Cambridge, Massachusetts: The MIT Press, 1997.

[21] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Upper Saddle River, New Jersey: Prentice Hall, 2001.

[22] H. Soltau, F. Metze, C. Fügen, and A. Waibel, "A one pass-decoder based on polymorphic linguistic context assignment," in *ASRU*, Madonna di Campiglio Trento, Italy, December 2001, pp. 214–217.

[23] Marc Schröder and Jürgen Trouvain, "The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching," in *International Journal of Speech Technology*, 2003, pp. 365–377.

[24] K. Berkling, "SCoPE, syllable core and periphery evaluation: automatic syllabification and foreign accent identification," *Speech Commun.*, vol. 35, pp. 125–138, 2001.

[25] S. Stüker, K. Kilgour, and J. Niehues, "Quaero speech-to-text and text translation evaluation systems," in *Results and Review Workshop of the HLRS*, 2010.

[26] "Hunspell," http://sourceforge.net/projects/hunspell.

[27] T. Thelen, *Automatische Analyse orthographischer Leistungen von Schreibanfängern. Dissertation*, 2010.