Phonetic based sentence level rewriting of questions typed by dyslexic spellers in an information retrieval context

Laurianne Sitbon[1,2], *Patrice Bellot*[1], *Philippe Blache*[2]

[1]Laboratoire d'Informatique d'Avignon, University of Avignon, France
[2]Laboratoire Parole et Langage, University of Provence, France
{laurianne.sitbon, patrice.bellot}@univ-avignon.fr, blache@lpl.univ-aix.fr

10.21437/Interspeech.2007-676

Abstract

This paper introduces a method combining spell checking and phonetic interpretation in order to automatically rewrite questions typed by dyslexic spellers. The method uses a finite state automata framework. Dysorthographics refers to incorrect word segmentation which usually causes classical spelling correctors fail. The specificities of the information retrieval context are that flexion errors have no impact since the sentences are lemmatised and filtered and that several hypothesis can be processed for one query. Our system is evaluated on questions collected with the help of an orthophonist. The word error rate on lemmatised sentences falls from 60% to 22% (falls to 0% on 43% of sentences).

1. Introduction

In an information retrieval (IR) context, taking the user into account involves producing adapted content and conceptualising their needs with a minimal amount of interaction. This means that IR systems might consider the linguistic profile of the user and provide robust processing in case of linguistic impairments.

Dysorthographics refers to dyslexic spellers, whose condition interferes in grapho-phonemic correlation. More specifically, a lack of phonological awareness makes them consider a sentence as a continuum of phonemes instead of a sequence of semantic units [1]. This leads to frequent word segmentation errors in written sentences that implies sentence level processing. This can be mitigate by some speech processing applications.

The first section defines the problem in an information retrieval context and describes the data on which the rewriting system has been designed. The second part of this paper exposes a new approach to sentence rewriting with sentence level processing. This sentence level method bases its approach on two different ways of rewriting data, a phonetic based method inspired by audio transcription systems, and a grapheme based method as used in classical spelling correction. A combination of these methods in a finite state machine (FSM) framework constitutes the final grapho-phonemic method. The performance of the proposed systems are assessed in the third section by computing a word error rate on lemmatised and filtered sentences.

2. Context

2.1. Rewriting for information retrieval systems

Some IR systems include automatic spelling correction or suggestions when unknown or low frequency words are encountered in the query. Many IR systems pre-process the user request to model it in a query. The minimal processing from the typed sentence to the formal query consists in lemmatise the typed sentence, remove stop words, and possibly expand the query. According to this design, the resulting query will contain the same orthographic errors that are present in the user's query. However, thanks to lemmatisation, flexion faults have no impact. This is why the IR process requires a system that can rewrite the question such that it is correct once lemmatised. Depending on the information retrieval model involved, a query can be a boolean expression of these terms or a vector of weighted terms. The query expansion phase can consist in adding or modifying terms. A weighted modelling allows a rewriting system to provide multiple weighted hypotheses.

2.2. Data collection

We first need data typed by dyslexic spellers to classify error types and conceive an adapted system. We focus on the question answering task in order to later evaluate how our question answering system [2] behave with questions typed by dyslexics. The query formulation of this particular task presents the advantage of being a natural language sentence. The typed questions are collected in a semi-spontaneous way. The focus is pre-defined but the formulation is left to the user discretion.

An orthophonist collects the typed questions of 7 children (between 8 and 15 years old) by: 1) Giving to the child the answer to a selected question (*e.g. : The mayor of Bastia is called X*). 2) Asking them what question they would ask to get that answer (*e.g. : What would you ask me if you want me to answer X ?*). 3) Asking the child to type the question. 4) Making the child read and correct the question if needed.

2.3. Data analysis

Whilst the corpus might appear tight, results indicate that it is representative enough to notice a large number of typical linguistic phenomena. The analysis of the questions typed by the children reveals several differences to known features on hand written text. We did not observe any letter confusion (such as between *p* and *b* and *d* and *q*). The number of letters or syllabic inversions is very low (1 word among 37 sentences).

There is a strong regularity in errors for one child but high variation between children. For example, some children systematically replace one letter by an apostrophe in interrogative pronouns (*Q'el* instead of *Quel*, equivalent in English to *W'at* instead of *What*) while others will leave it out in each sentence they write. This particularity shows that it is impossible to infer a generic graphemic transition model of dysorthographic errors.

The difficulty which is most likely to cause correctors to fail is the inaccurate word segmentation. This means that spaces between words cannot be trusted to distinguish the words of the question. The correction or interpretation necessarily must deal with the whole sentence.

The global phonetic is correct despite the misplaced spaces and completely misspelt words. In most cases the automatic phonetisation does not suffer from syntactical mistakes because in French they are often due to the non-pronunciation of the inflection markers such as plural letters.

2.4. Rewriting for dysorthographics

Commercial spell checkers compute suggestion lists for each out of vocabulary word by computing a distance between the written word and each word in the lexicon. The distance is mainly defined by the Levensthein edition distance, and sometimes includes phonetic features. But these systems work on words separately, and the right correction is rarely in the top position in the suggestion list for impaired spellers such as dysorthographics. [3] highlights issues from identifying "real words" errors to propose a correct assumption.

Error modelling techniques have been proposed. They all consider a correct word segmentation in the sentence. [4] implement for each word an automaton based on confusions learned from a modelling of error causes. This technique supposes isolated and regular errors. [5] also considers mistyping of dyslexic spellers to be worse than the mistyping of regular spellers. He introduces a user specific model which provides results as accurate as commercial spell checkers. This study shows that these systems collapse on such typing. [6] concentrates on the correction and detection of real words errors, using syntactical and semantic context.

A sentence level rewriting system based on phonetics mainly can provide for the detection of both word segmentation and real word errors. Automatic speech recognition systems answer this disambiguation issue with language models.

3. Grapho-phonemic combination

3.1. Phonetic interpretation

The phonological route can be simulated by phonetisation and transcription tools sequencially applied on the initial text. Automatic speech recognition tools are designed to process a continuous stream of phonemes without word segmentation since audio signal provides any. This is a solution to the erroneous word segmentation issue.

The phonetisation is made with LIA_phon [7] tool. This phonetiser combines a phonetic lexicon (for words in vocabulary) and rules (for words out of vocabulary) that are robust for misspelled words. This step transforms a letter sequence in a phoneme sequence (linear graph).

The phonetiser implements academic rules of how the words must be well pronounced, but in fact many people mispronounce some vowels in French, confusing *open* and *close* or *short* and *long* vowels. This pronunciation confusion also reflects on writing confusions (like *living* instead of *leaving*). That is why alternate phonetic hypothesis must be generated on the basis of a confusion matrix. This step finally provides a lattice of phonetic hypothesis.

According to finite state machine transcription work [8] and with the help of AT&T FSM toolkit [9], the phonetic lattice is encoded in a finite state automata (FSA) and composed with a language model automata learned on a journalistic corpus [10] and a phonetic lexicon transducer. The N-best path in this graph are the possible rewritings of the question.

3.2. Spell checkers hypothesis

Some mistyping errors such as omissions or substitution may be done also by dyslexic users, and they cannot be processed by the phonetic interpretation which produces incoherent sentences when they appear. A way of avoiding this low misspellings compromising phonetic lattice is to add graphemic hypothesis on isolated words. These hypothesis can be obtained thanks to a classical spelling corrector which will increase the robustness of the method.

The GNU project Aspell spell checker [1] implements both a phonetic and a Levenshtein edition distance. Its evaluation results, specially on *badspellers* mode, have been shown higher than classical spell checkers. This approach seems the most appropriate for dyslexics as phonetic errors are more frequent than inversion or substitutions. However, it supposes an accurate word segmentation, and can not recover "real word" errors.

3.3. Combination

In order to take advantage on both sentence-level processing of the phonetic interpretation and the graphemic alternatives based on written words, the combination system uses graphemic hypothesis for the construction of an enhanced phonetic lattice. Figure 1 illustrates the whole process.

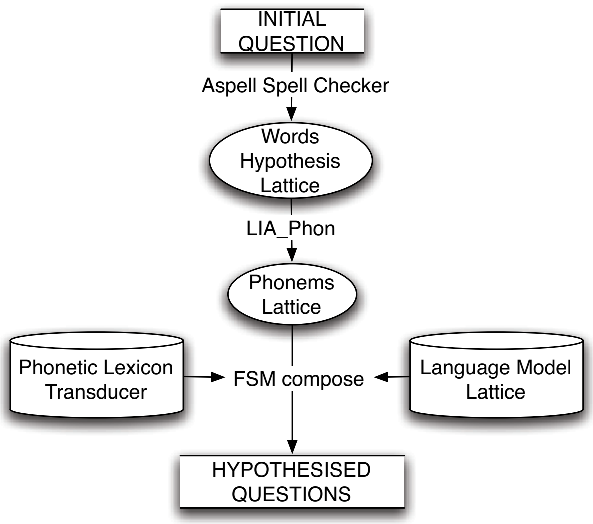

Figure 1: Grapho-phonemic processing of a question.

All alternative hypothesis, graphemic or phonemic, must be associated to a path cost in order to privilege initial path. Indeed, in order to avoid the production of new errors (on proper nouns essentially), the system must primarily suppose that the user writes well and keep the initial production as an hypothesis. Hypothesised words H provided by spell checker might be assign a graphic weight :

$$Wg(H) = f(d(H, I)) \quad (1)$$

where f is a normalisation function of distance $d(H, I)$ between the hypothesised and the initially written word. The distance can be a score provided by the spell checker. Phonetic alternatives H obtained from a confusion matrix might also be

[1] `http://aspell.sourceforge.net/`

assigned transition costs on their alternative path :

$$Wp(H) = g(m(H, I)) \tag{2}$$

where g is a normalisation function of confusion score $m(H, I)$ between the alternate and the initial phoneme (obtained from the confusion matrix).

comment s' appelle le maire de bail à	52.2848701
comment s' appelle le maire de bahia	54.6559029
comment s' appelle le maire de bastia	54.8422737

Figure 4: Top 3 rewritings of typed sentence : *koman sapel le mer de batia*

Consider for example the typed sentence *koman sapel le mer de batia ?* (approximatively equivalent to *O izkold the my or of batia?*) instead of *Comment s'appelle le maire de Bastia ?*, which means *What is the name of the mayor of Bastia ?*. Figure 2 illustrates the graphemic hypothesis on words produced by Aspell encoded in a finite state automata where transition are the words and their associated cost. The word *mer* has no alternate because it exists in the lexicon (*mayor* and *sea* are homonymic words in French) . Figure 3 is the phonetic lattice resulting from the phonetisation of all sentences acceptable by the preceding automata. Figure 4 contains the top 3 results of understanding obtained when composing preceding lattice with a phonemes-to-words transducer and a language model. The two first hypothesis are partially correct while the third one is the correct sentence, on both sense and syntactic levels.

In the following experiments the cost normalisation functions for alternative hypothesis are empirically set to :

$$f(d(H, I)) = \begin{cases} 0\, if\ H = I \\ 0.1\, if\ H \neq I \end{cases} \tag{3}$$

$$g(m(H, I)) = \begin{cases} 0\, if\ H = I \\ 0.1\, if\ H \neq I \end{cases} \tag{4}$$

The confusion phonemes matrix is restricted to confusion between $open$ and $closed$ vowels, and the graphemic alternatives are the three first hypothesis provided by Aspell in $badspellers$ mode.

4. Evaluation

In order to evaluate the accuracy of our method, we perform a word error rate with a dynamic programming comparison between tested sentences and correct sentences. Correct sentences are produced by an agreement of three human correctors. The NIST furnishes SCLITE toolkit [2] implements the dynamic programming comparison and counts substitution, deletions and insertions of words in tested sentences.

Evaluated sets are : the original typed sentences, the first and top three hypothesis from Aspell (Asp 1 and Asp 3), and the first and top three hypothesis from our combination tool (FSM 1 and FSM 3). Each set is evaluated according to two measures. The first one is the mean word error rate (WER) per sentence which weights substitutions, insertions and deletion together. The second one is the percentage of fully correct sentences (FC).

The first evaluation concerns whole sentences, and is spell checking oriented. The results of this evaluation presented in Table 1 show a significant improvement of sentences quality by FSM combination, when evaluating the top three hypothesis. The 40% fully correct sentences is the most significant indicator while comparing with Aspell accuracy.

Measure	Initial	Asp 1	Asp 3	FSM 1	FSM 3
WER	45.1	32.4	30.2	28.6	**22**
FC	**2.7**	10.8	10.8	18.9	**40.5**

Table 1: Accuracy of different systems on whole sentences

The second evaluation automatic processing oriented. Both reference and hypothesised sentences are preliminary lemmatised and filtered. This reflects the computational acceptability of alternative hypothesis. This acceptability is not experimented on QAS directly because the formulation of the questions may also infer on the system ability to answer correctly.

Measure	Initial	Asp 1	Asp 3	FSM 1	FSM 3
WER	**51**	35.7	30.8	23.0	**19.9**
FC	5.4	13.5	18.9	**43.2**	45.9

Table 2: Accuracy of different systems on lemmatised and filtered sentences

The results of this second evaluation are reported in Table 2. The word error rate of initial sentences suggests that in previous evaluation on whole sentences, stop words were mostly correctly typed. This raised global accuracy of Aspell system. Considering three hypothesis from the FSM combination system leads word error rate decreasing 38%, and raises multiply by nine the percentage of fully correct. It is also interesting to notice that the fully correct sentences percentage is already excellent if considering the first hypothesis only. This suggests that provided hypothesis are mainly morpho-syntactical variation of the same sentence. In regard of this, the gap between first and top 3 hypothesis on whole sentences evaluation means that around 50% of first hypothesis contains grammatical errors only. In both evaluations, already correct sentences are not degraded by our system. They would be if missing proper nouns, but the high score of such misinterpretation would discriminate the approach.

P	IWER	WER	FC
1	36	9	75
2	67	39	40
3	49	18	40
4	50	30	20
5	73	12	60
6	30	27	40
7	46	21	40
8	60	30	50

T	IWER	WER	FC
1	58	47	0
2	52	42	0
3	40	0	100
4	43	17	62
5	65	11	57

Table 3: Accuracy distribution for lemmas interpreted by first hypothesis of FSM combination, by person (P) or by topic (T) of question, according to Initial typed sentence Word Error Rate (IWER), Word Error Rate of hypothesis (WER), and percentage of Fully Correctly interpreted sentence (FC)

Average accuracy measures attest the global efficiency of FSM combination system, but it is also interesting to watch closely how the results are distributed depending on the topic or on the person who typed the question. Table 3 provides such information. The per person distribution or word error rate on initially typed sentences indicates that there is no excellent speller,

[2] `http://www.nist.gov/speech/tools/index.htm`

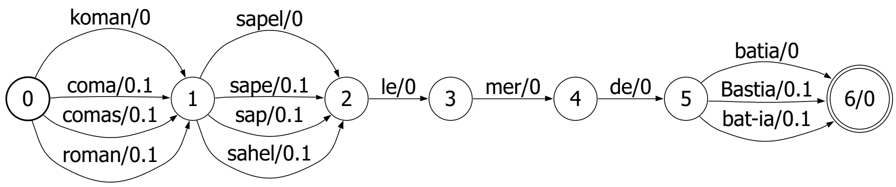

Figure 2: Lattice hypothesised word corrections for *koman sapel le mer de batia*

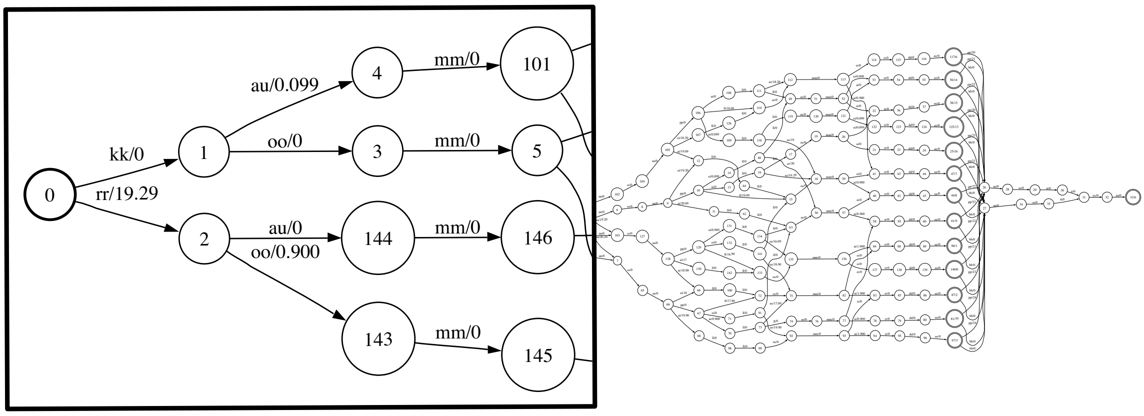

Figure 3: Lattice of hypothesised phonemes for *koman sapel le mer de batia*

and there is no correlation between the amount of errors and the ability of the system to solve them. The per topic distribution reveals a topic dependent accuracy. As each topic is related to a proper noun, the systematic errors for topic 1 and 2 are related to the absence of this noun in the lexicon.

5. Conclusion and future work

An analysis of questions written by dyslexic children highlights the need to process sentences as a whole instead of word-by-word. The combined system based on an FSM framework is efficient in terms of spell checking and very efficient in terms of automatic system interpretation needs. This system shows a good accuracy by decreasing the lems error rate from 60% to 22% and allowing a correct automatic interpretation for 43% of sentences on the first hypothesis.

In most cases, remaining errors come from missing proper nouns language model and lexicon, which help them to be recognised. An adaptive language model could avoid this. In a question answering system the adaptation could be based on the target corpus. Under this condition, the risk of integrating new errors apply only on questions that the system cannot answer anyway. Since the initial typing is kept as an hypothesis, the introduced method could be used for any user.

Moreover, it should be language independent since the phonological awareness deficience is. The phoneme confusion matrix might be enhanced with a larger corpus of corrected texts typed by dyslexic spellers. Under an assumption that the human phonetic production system is related to the computer phonetic interpreter, confusion matrices used by speech transcription systems should be efficient.

6. References

[1] G. T. Gillon, *Phonological Awareness- From Research to Practice*. Guilford Press, 2004.

[2] L. Gillard, L. Sitbon, E. Blaudez, P. Bellot, and M. El-Bèze, "The lia at qa@clef2006," in *Cross Language Evaluation Forum (CLEF) 2006*, Alicante, Espagne, 20-22 septembre 2006 2006.

[3] A. James and E. Draffan, "The accuracy of electronic spell checkers for dyslexic learners," *PATOSS bulletin*, August 2004.

[4] S. Deorowicz and M. G. Ciura, "Correcting spelling errors by modelling their causes," *International journal of applied mathematics and computer science*, vol. 15, no. 2, pp. 275–285, 2005.

[5] R. Spooner, "A spelling checker for dyslexic users : user modelling for error recovery," Ph.D. dissertation, Human Computer Interaction Group, Department of Computer Science, University of York, Heslington, York,, September 1998.

[6] J. Pedler, "The detection and correction of real-word spelling errors in dyslexic text," in *Proceedings of the 4th Annual CLUK Colloquium*, 2001.

[7] F. Bechet, "Lia_phon - un systeme complet de phonetisation de textes," *Traitement Automatique des Langues (T.A.L.)*, vol. 42, no. 1, 2001.

[8] M. Mohri, F. C. N. Pereira, and M. Riley, "Weighted finite-state transducers in speech recognition," *Computer Speech and Language*, vol. 16, no. 1, pp. 69–88, 2002.

[9] M. Mohri, F. C. N. Pereira, and M. D. Riley. (1997) At&t fsm librarytm – finite-state machine library. [Online]. Available: http://www.research.att.com/~fsmtools/fsm/

[10] C. Allauzen and M. Mohri, "The design principles and algorithms of a weighted grammar library," *International Journal of Foundations of Computer Science*, vol. 16, no. 3, pp. 403–421, 2005.