S-capade: Spelling Correction Aimed at Particularly Deviant Errors

Emma O'Neill*[1], Robert Young*[2], Elsa Thiaville[2], Muireann MacCarthy[2], Julie Carson-Berndsen[1], and Anthony Ventresque[2]

[1]ADAPT Centre, School of Computer Science, University College Dublin, Ireland
[2]Lero Research Centre, School of Computer Science, University College Dublin, Ireland
{emma.l.oneill,robert.young2,elsa.thiaville,
muireann.maccarthy}@ucdconnect.ie
{julie.berndsen,anthony.ventresque}@ucd.ie

Abstract. S-capade (**s**pelling **c**orrection **a**imed at **pa**rticularly **d**eviant **e**rrors) is a phonemic distance based spellchecking tool[1] intended for the correction of misspellings made by children. Whilst typographic misspellings typically deviate from the target by only one or two characters, children's misspellings tend to be more phonetic. They are influenced both by how the child perceives the pronunciation of a word and by the letters they choose to represent that pronunciation. As such, these misspellings are particularly deviant from the target and can negatively impact the performance of conventional spellcheckers. In this paper we demonstrate that S-capade is capable of correcting a significant portion of misspellings made by children where conventional correction tools fail.

Keywords: Spelling correction · Phonemic distance · Children's spelling

1 Introduction

Spelling errors are often considered one of two types: typographic or cognitive [22]. Typographic errors are the results of motor coordination slips; perhaps substituting a character for an adjacent one on the keyboard. Cognitive errors, on the other hand, stem from a misconception or lack of knowledge regarding the correct spelling of a word. One particular subset of these cognitive errors are known as phonetic errors where the writer produces a misspelling that, whilst not orthographically correct, captures the phonetic sequence of the target word.

These phonetic errors are particularly common in children's spelling, which has long been considered phonetic-based. In an examination of children's early spelling, Read [29] discussed the significant influence of speech sounds and the relationships between them. Thus, whilst some misspellings might appear "bizarre" and deviate heavily from the target word, they tend to reflect the phonetic judgments of the child. Additionally, it is currently common in the classroom

* Both the authors have equal contribution to this paper.

[1] Source code repository may be found in the references section [35].

for children to be taught reading and writing using phonics: an approach which focuses on the relationships between letters and sounds [33]. As such, children are encouraged to use a 'sounding out' method when spelling unfamiliar words - an approach that is relied on heavily by low achieving spellers [7].

Despite the prevalence of phonetic-type errors, conventional spelling correction tools are not fully capable of correcting these types of misspellings and as such exhibit poorer performance on children's spelling. In this paper we present a correction method based on phonemic similarity that is capable of correcting phonetic misspellings of English made by children that deviate heavily from the target word. Kukich [22] grouped work on spelling correction into distinct tasks; detection of errors, isolated error correction, and context dependent error correction. This work focuses on isolated error correction, generating a list of real-word candidate corrections based on the phonetic properties of the misspelling. Similar to Hodge and Austin [17], our goal is to maximise recall through candidate generation as we envision this method as a component of an overall model that will handle candidate selection as a context-dependent task.

2 Related Work

Early spelling correction algorithms typically use character edit distances between misspellings and real-word corrections, relying on the finding that the majority of misspellings differ by a single edit operation (insertion, deletion, substitution, or transposition) [8]. These methods are suited to typographic misspellings. However, phonetic misspellings often deviate more substantially from the real-word target [22]. Improved performance was seen with the use of noisy-channel models which allow for multiple edit operations [4]. In particular, Brill and Moore [3] demonstrated significant performance improvements to the noisy channel model by calculating the probabilities of string-to-string edits and combining these when comparing a misspelling to real-word candidate corrections.

The incorporation of phonetic information into these methods proves advantageous to the correction of cognitive misspellings. Veronis [36] used a weighted edit-distance algorithm where the costs of edit operations were based on the phonetic similarity between graphemes. It is also common to convert words from their orthographic form to one which captures their phonetic features. For example, Soundex, described by Kukich [22] and patented by Russel and Ordell [30], maps words to a fixed length alpha-numeric code based on its characters. Numeric values are assigned to groups of letters that are phonetically similar. Thus words which are pronounced similarly will have the same encoding (e.g. 'sure' and 'shore' both have encoding S600). Edit-distance algorithms can be applied to these encodings to find real-word candidate corrections that are phonetically similar to a misspelling. However, Soundex has been criticised as being too general given its limited permutations [17, 23]. Thus, phonetic transformation rules, determined by linguistic knowledge of the target language, are often used before encoding [17, 28]. Alternatively, phonemic forms can be used directly by transforming a misspelling to its corresponding phoneme sequence using letter-to-sound-rules [9, 21, 23, 34]. Other approaches to spelling correction include tackling the problem as one of Machine Translation [2, 31] or as a synthesis/recognition task [32].

The method described in this paper combines elements from a number of these approaches. Misspellings are converted to their corresponding phonemic sequences using a machine learned grapheme-to-phoneme tool [5] instead of explicit letter-to-sound rules. Weighted edit distances are calculated between misspellings and real-word candidate corrections using a phoneme-to-phoneme distance matrix based on both the acoustic and distributional properties of the phonemes. Section 3 describes this method in detail, whilst Section 4 details the experimental setup for comparing this method with other spelling correction tools on various datasets. The results of this are presented in Section 5 where we demonstrate that S-capade is capable of correcting a significant proportion of children's misspellings beyond those corrected by other tools.

3 S-capade Method

When a child uses a 'sounding out' approach to spelling they are approximating the sounds they perceive in the target word with letters they believe represent those sounds. As such, deviations from the correct spelling occur both as a result of incorrect phonemes being perceived, e.g. phoneme /V/[2] being perceived as /F/ resulting in the misspelling 'gif' (give), and of incorrect letters being chosen, e.g. representing the /AY/ phoneme with an 'i' in the misspelling 'ciber' (cyber). The majority of misspellings resulting from the latter case are handled by converting the graphemic misspelling to its phonemic form. In these instances the phonemic form typically matches that of the correct spelling. However, misspellings of the former variety map to phoneme sequences that are similar to that of the correct spelling but not necessarily identical. In these cases we require some measure of similarity at a phoneme level so that, for example, we can predict that 'gif' is more likely to be 'give' than 'gig' due to the /F/ phoneme being more similar to /V/ than /G/.

In this work, similarity is modelled using two features which have been shown to influence a native speaker's perception of phonemic similarity; namely the acoustic and distributional properties of the phonemes. Phonemic similarity is considered to be a function of confusability [13] - two phonemes can be thought of as similar if one is often mistakenly identified as the other. Previous work by Kane and Carson-Berndsen [19] investigated phoneme confusability using an underspecified recognition system. A target phoneme was removed during training so that at test time the system was forced to select an alternative phoneme - one which was acoustically similar to the target. The frequency with which one phoneme was identified as another was used as a measure of their acoustic similarity. The potential influence of a phoneme's distributional properties on perceived similarity was demonstrated in previous work by O'Neill and Carson-Berndsen [27]. Here, phonemes that often occurred in the same environment

[2] Throughout this paper we use the ARPAbet notation when referring to phonemes.

(having the same preceding and following phonemes) were shown to be perceived as more similar. A word2vec model, trained on the Brown Corpus [11], was applied at the phonemic level and used to generate phoneme embeddings. The distances between these embeddings, or vector representations, represented the distributional similarity between the corresponding phonemes. Both the acoustic and distributional properties were combined to form a phoneme-to-phoneme distance matrix where smaller distance values represented more similar phonemes. Significantly the distance matrix is not symmetric i.e. the distance between a target phoneme X being perceived as Y is not necessarily the same as the target phoneme Y being perceived as X. For example, it is likely that the /NG/, as in 'walking', will be pronounced as /N/; resulting in the misspelling 'walkin'. However it is much less likely that the /N/ phoneme, as in 'happen', will be pronounced as /NG/; resulting in the misspelling 'happeng'. The distance matrix employed in this work is able to make this distinction.

Candidate corrections are the possible real-word targets of a misspelling. Both the real-word candidate correction and the misspelling were first converted to their corresponding phonemic forms; the former using the CMU Pronouncing Dictionary [38] and the latter with a grapheme-to-phoneme tool trained on this dictionary [5]. To determine the degree of similarity between the two phoneme sequences, a distance score was calculated between the misspelling and real-word candidate using a weighted edit distance algorithm akin to that of Wagner and Fischer [37]. The cost for performing a substitution operation was defined as the distance between the two phonemes per the distance matrix. Deletion and insertion operations were treated as substitutions of a phoneme with the empty string and vice versa. Distance values for these operations were chosen heuristically based on existing literature regarding which phonemes typically undergo insertion (epenthsesis) and deletion (ellision) in speech [6, 10, 15, 18, 42]. A comparison of the character-level edit-distance and S-capades' phoneme-level weighted edit-distance used in this paper is given in Table 1. The misspelling 'sichweshan' and its real-word target 'situation' have a high character edit-distance. As such, non-phonetic spelling correction approaches are unable to correct this error. However, their phonetic similarity results in a very small edit distance using S-capade, thus making it more easily correctable.

Table 1. Character-level edit-distance vs S-capade's phonemic edit-distance

	Character-level										S-capade							
Misspelling	s	i	c	h	w	e	s	h	e	n	S	IH	CH	W	EH	SH	AH	N
Real-Word Target	s	i	t		u	a	t	i	o	n	S	IH	CH	UW	EY	SH	AH	N
Edit-Distance	7										1.1							

4 Experimental Setup

In this section we describe the experimental setup designed to test the ability of S-capade to correct particularly deviant errors. As stated previously, the S-capade method is envisioned as a component of a larger system. It is not intended for the correction of typos but rather is specifically aimed at misspellings which lie beyond the scope of conventional spelling correction tools. As such, S-capade is not expected to outperform other tools but instead to uniquely target a greater proportion of errors on datasets likely to contain these particularly deviant errors i.e. those consisting of misspellings made by children.

4.1 Datasets

The baseline spelling correction methods, see Section 4.2, and the phonemic distance method discussed in Section 3 were evaluated and compared using a collection of five different misspelling datasets. Four of these datasets are publicly available, and were obtained in a pre-processed format from the Birkbeck University of London [25]. The fifth was acquired through a collaboration with an Irish Educational company, Zeeko [43]. Details of all datasets may be seen in Table 2. For each dataset, only misspellings where the target correction is one word are included. A single word target correction would be the misspelling 'hapen' corrected to 'happen'. An example of a target correction which is two words would be the misspelling 'alot' corrected to 'a lot'.

Table 2. Misspelling datasets

Dataset	Misspellings	Misspellings Used	Target Words	Publicly Available
Birkbeck	36,133	33,887	6,068	Yes
Holbrook	1,791	1,562	1,177	Yes
Wikipedia	2,455	2,230	1,909	Yes
Aspell	531	515	437	Yes
Zeeko	232	232	163	No

- **Birkbeck** - native-speaker errors (British or American) [25]. Majority of errors from schoolchildren, university students or adult literacy students [24].
- **Holbrook** - extracts of writings from British secondary school students in their penultimate year of school [25].
- **Wikipedia** - common misspellings made by editors (British or American) on Wikipedia [25]. A common misspelling is one that occurs at least once a year on the site [39].
- **Aspell** - GNU spell checker dataset (British forms). Comprised of common misspellings [1].
- **Zeeko** - comprised of spelling mistakes from Irish primary school students [43]. The age range of respondents is 8-14 years old.

Across these five datasets there is a broad spectrum of literacy demographics; namely primary school students, secondary school students, university students, and Wikipedia article editors. We hypothesised that due to children's first efforts in spelling being based on a 'sounding out' approach, as discussed in Section 1, S-capade will perform better on datasets containing children's phonetic spelling attempts, which tend to have larger character edit distances. For example, in the Holbrook dataset 53% of misspellings have a character edit distance of 1, 31% have a character edit distance of 2 and 16% have an edit distance greater than 2. Conversely, in the Wikipedia dataset 69% of misspellings have an edit distance of 1, 28% of 2 and only 3% have an edit distance greater than 2. In the Zeeko dataset 91% of the misspellings are of two character edit distance or less.

4.2 Conventional Spelling Correction Comparison Tools

Three different conventional spelling correction tools are used for comparison in this paper - PySpellChecker, SymSpell and Aspell. All three tools are based on a character edit distance limit of 2, use British English dictionaries and generate a suggested spelling correction and a list of candidate corrections. S-capade is limited to a distance of two insertions and deletions of phonemes in a misspelling sequence for edit candidate generation (adapted from SymSpellPy [14, 40]) and lookup. Any target words in the datasets that were not present in the tools' default dictionaries were manually added to ensure fairness of results.

- **PySpellChecker** - word permutations were created via insertions, deletions, replacements and transpositions [26] which were then compared to known words in a frequency dictionary [12].
- **SymSpell** - generates word permutations for comparison via the misspelling and valid words in the dictionary using deletes only [14]. Selection based on the smallest edit distance and highest frequency word [16] [41].
- **Aspell** - performed word comparisons in a given dictionary and uses phonetic comparisons with other words [20]. This was done via table driven phonetic code allowing 'sounds like' word comparison and suggestions. This makes it the most relevant tool to compare to this paper's S-capade method.

4.3 Metrics

In Section 5, we compare the accuracy and recall of S-capade across the five datasets against the three conventional spelling correction tools, Pyspell, Symspell, and Aspell. For each misspelling, real-word candidates are ranked in order of distance and subsequently frequency. We define accuracy as whether or not the closest candidate matches the real-word target and recall as whether the real-word target is found in the top 10 closest candidates. Word correction overlap graphs, based on accuracy, are presented for each of the datasets comparing S-capade with the Aspell spelling corrector. In these graphs, the common corrections between each method and the misspelling corrections made only by one or other method are shown. Aspell was chosen as the comparison for S-capade given its use of the 'sounds like' word correction approach, see Section 4.2.

5 Results and Discussion

The Birkbeck dataset results are presented in Table 3. With respect to accuracy, S-capade is comparable to PySpell and SymSpell, and outperforms both in terms of recall. Aspell outperforms S-capade in accuracy and recall. Of the 17,029 misspellings corrected between both methods, ∼48% were word misspelling corrections common to both, ∼32% were corrections made only by Aspell, and ∼20% were corrections made only by S-capade. The coverage of misspelling corrections for the dataset between the two methods can be seen in Figure 1.

Method	Accuracy	Recall
PySpell	35.3%	42.6%
SymSpell	34.74%	43.04%
Aspell	39.89%	66.03%
S-capade	34.43%	51.49%

Table 3. Birkbeck correction scores

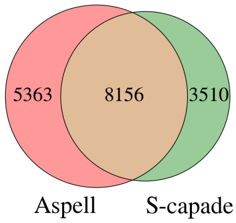

Fig. 1. Birkbeck: Aspell vs S-capade

Table 4 displays the results for the Holbrook dataset. Compared to the Birkbeck dataset results, the overall scores are similar. The most interesting result from this dataset can be seen in Figure 2, which compares the misspelling correction coverage of the two methods for the Birkbeck dataset. Of the 626 misspellings corrected between both methods, ∼35.3% were word misspelling corrections common to both, ∼32.3% were corrections made only by Aspell, and ∼32.4% were corrections made only by S-capade. It can be seen that S-capade has a slightly greater correction coverage of misspellings when compared with Aspell. The Holbrook dataset is more likely to be comprised of phonetic spelling mistakes given its demographic, discussed in Section 4.1, and as such shows how our approach corrects misspelling errors different to those corrected by Aspell.

Method	Accuracy	Recall
PySpell	29.32%	42.06%
SymSpell	27.46%	42.51%
Aspell	27.08%	67.93%
S-capade	27.14%	52.82%

Table 4. Holbrook correction scores

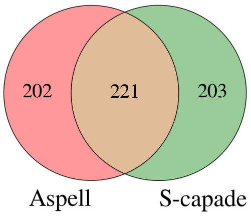

Fig. 2. Holbrook: Aspell vs S-capade

Of the five datasets under analysis, S-capade performed the worst on the Wikipedia dataset, relative to the other spelling correction methods. This is visible in Table 5, where it obtained the lowest score for accuracy and recall. Compared with the overlap scores for the other datsets in Figures 1, 2, 4, and 5, S-capade also had the smallest proportion of misspelling corrections. Of the 1,984 misspellings corrected between both methods, ∼61% were word misspelling corrections common to both, ∼29% were corrections made only by Aspell, and ∼10% were corrections only made by S-capade. As discussed in Section 4.1, the Wikipedia dataset is made up of Wikipedia editors' common spelling mistakes. These are typically typographic misspellings, and as expected our phonetic approach does not produce competitive results for this dataset.

Method	Accuracy	Recall
PySpell	78.39%	88.48%
SymSpell	80.99%	92.11%
Aspell	79.96%	97.04%
S-capade	63.14%	77.80%

Table 5. Wikipedia correction scores

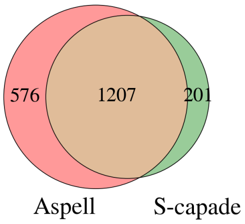

Fig. 3. Wikipedia: Aspell vs S-capade

The scores for the Aspell dataset can be seen in Table 6, where S-capade had similar performance to PySpell and SymSpell. Of the 351 misspellings corrected between both methods, ∼50% were word misspelling corrections common to both, ∼32% were corrections made only by Aspell, and ∼18% were corrections made only by S-capade, as may be seen in Figure 4. The Aspell dataset focuses on particularly bad spelling attempts; those which deviate from the real-word target by multiple edit operations. However, these are not necessarily phonetic misspellings and, as such, S-capade performs satisfactorily on this dataset.

Method	Accuracy	Recall
PySpell	49.32%	62.33%
SymSpell	53.20%	67.18%
Aspell	55.73%	85.6%
S-capade	46.41%	65.24%

Table 6. Aspell correction scores

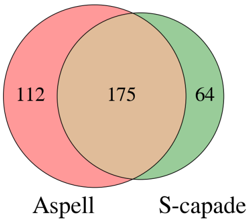

Fig. 4. Aspell: Aspell vs S-capade

The scores for the Zeeko dataset may be seen in Table 7. After Holbrook, the Zeeko dataset resulted in the second best performance for S-capade with respect to recall relative to the comparison methods. Figure 5 shows that, of the 164 misspellings corrected by Aspell and S-capade, ∼47.5% were common corrections, ∼30% were corrections only made by Aspell, and ∼22.5% were corrections only made by S-capade. As discussed in Section 4.1, 91% of the Zeeko dataset misspellings are of 2 character edit distance or less. We believe this shows that despite the misspellings edit distance falling within the boundary of conventional tools, phonetic spelling errors require a different correction approach.

Method	Accuracy	Recall
PySpell	56.90%	72.41%
SymSpell	54.74%	70.69%
Aspell	54.74%	86.64%
S-capade	49.57%	76.29%

Table 7. Zeeko correction scores

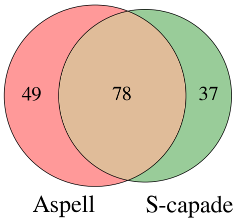

Fig. 5. Zeeko: Aspell vs S-capade

The CMU pronouncing dictionary was used by the S-capade method for phoneme-sequence-to-word lookups when correcting word misspellings. Figure 6 displays the breakdown in corrections made by S-capade, showing the split between either exact match lookup in the CMU dictionary using a phoneme sequence, in which case the edit distance is equal to zero, or using S-capade to calculate the phonemic distance between the misspelling and the real-word target. It can be seen that for four out of the five datasets, S-capade's distance method accounted for over 50% of the real-word candidate corrections made.

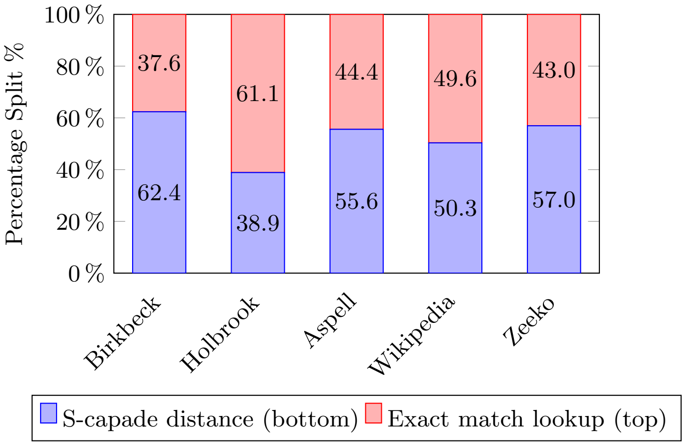

Fig. 6. S-capade distance calculation vs Exact dictionary match

The approach taken by S-capade resulted in interesting word corrections of phonetically spelt words with large character edit distances that conventional spelling correction tools were unable to correct. Table 8 shows some selected misspellings from the datasets that S-capade corrected, and the traditional character edit distance versus the phonemic edit distance generated by our approach.

Table 8. S-capade interesting word corrections

Target	Misspelling	Character Distance	S-capade Distance
necessarily	nessecarryally	8	1.62
philosophy	folocify	7	0.54
situation	sichweshen	7	1.10
ecstasy	extersee	6	0.46
sufficient	servishant	6	0.93
procedure	prosiegeur	5	0.59
whistled	wisheld	4	1.14
council	cousall	3	1.00
actually	achuly	3	1.28

6 Conclusions

In this paper, we have presented a phonemic-distance based approach to spelling correction that is capable of handling phonetic misspellings which conventional tools are unable to correct. The creativeness of children's spelling attempts has been shown to produce phonetic misspellings that heavily deviate from the real-word target. As such we see poorer performance of conventional spelling correction tools on datasets specifically consisting of children's spelling errors. The method described in this paper is shown to correct a significant portion of misspellings within these datasets that one of the top performing English spellcheckers, Aspell, cannot.

The phonemic-distance based approach is envisioned as a component in a fully context-dependent spelling correction system. Future work will look to incorporate this method into a spellchecker capable of handling both typographic and phonetic misspellings and of choosing the correct real-word target from a list of candidates based on the context of the misspelling. Further plans for improvement include investigating the effects of accent on the phonetic misspellings produced and the potential benefits of an accent-specific system on spelling correction accuracy.

Acknowledgements This work was supported with the financial support of the Science Foundation Ireland grant 13/RC/2094 to Lero - the SFI Research Centre for Software (`www.lero.ie`). The ADAPT Centre for Digital Content

Technology (www.adaptcentre.ie) is funded under the SFI Research Centres Programme (Grant 13/RC/2106). The authors would like to thank the team at Zeeko (https://zeeko.ie/) for supporting their research.

References

1. Atkinson, K.: Aspell spell checker test data (2002), http://aspell.net/test/cur-all/batch0.tab, last accessed 19 May 2020
2. Aw, A., Zhang, M., Xiao, J., Su, J.: A phrase-based statistical model for sms text normalization. In: COLING/ACL. pp. 33–40 (2006)
3. Brill, E., Moore, R.C.: An improved error model for noisy channel spelling correction. In: ACL. pp. 286–293 (2000)
4. Church, K.W., Gale, W.A.: Probability scoring for spelling correction. Statistics and Computing 1(2), 93–103 (1991)
5. CMUSphinx: Grapheme-to-phoneme tool based on sequence-to-sequence learning (2016), https://github.com/cmusphinx/g2p-seq2seq
6. Collins, B., Mees, I.M.: Practical phonetics and phonology: A resource book for students. Routledge (2013)
7. Daffern, T., Critten, S.: Student and teacher perspectives on spelling. Australian Journal of Language and Literacy 42(1), 40–57 (2019)
8. Damerau, F.J.: A technique for computer detection and correction of spelling errors. Commun. ACM 7(3), 171–176 (Mar 1964)
9. Fisher, W.M.: A statistical text-to-phone function using ngrams and rules. In: ICASSP. vol. 2, pp. 649–652 (1999)
10. Fourakis, M., Port, R.: Stop epenthesis in english. Journal of Phonetics 14(2), 197–221 (1986)
11. Francis, W.N., Kucera, H.: Brown corpus manual (1979)
12. FrequencyWords: Frequency word list generator (2020), https://github.com/hermitdave/FrequencyWords, last accessed 21 May 2020
13. Gallagher, G., Graff, P.: The role of similarity in phonology. Lingua 2(122), 107–111 (2012)
14. Garbe, W.: Symspell (2020), https://github.com/wolfgarbe/symspell, last accessed 21 May 2020
15. Gimson, A.C., Ramsaran, S.: An introduction to the pronunciation of English, vol. 4. Edward Arnold London (1970)
16. Google: Google books ngram viewer (2012), http://storage.googleapis.com/books/ngrams/books/datasetsv2.html, last accessed 21 May 2020
17. Hodge, V.J., Austin, J.: An evaluation of phonetic spell checkers (2001)
18. Itô, J.: A prosodic theory of epenthesis. Natural Language & Linguistic Theory 7(2), 217–259 (1989)
19. Kane, M., Carson-Berndsen, J.: Enhancing data-driven phone confusions using restricted recognition. In: INTERSPEECH. pp. 3693–3697 (2016)
20. Kevin Atkinson, G.A.: How aspell works) (2004), http://aspell.net/0.50-doc/man-html/8_How.html, last accessed 21 May 2020
21. Khoury, R.: Microtext normalization using probably-phonetically-similar word discovery. In: WiMob. pp. 384–391 (2015)
22. Kukich, K.: Techniques for automatically correcting words in text. Acm Computing Surveys (CSUR) 24(4), 377–439 (1992)

23. de Mendonça Almeida, G.A., Avanço, L., Duran, M.S., Fonseca, E.R., Nunes, M.d.G.V., Aluísio, S.M.: Evaluating phonetic spellers for user-generated content in brazilian portuguese. In: International conference on computational processing of the Portuguese language. pp. 361–373 (2016)
24. Mitton, R.: Birkbeck spelling error corpus (1980), `https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/0643`, last accessed 19 May 2020
25. Mitton, R.: Corpora of misspellings for download (2007), `https://www.dcs.bbk.ac.uk/~ROGER/corpora.html`, last accessed 19 May 2020
26. Norvig, P.: Pyspellchecker (2020), `https://pypi.org/project/pyspellchecker/`, last accessed 21 May 2020
27. O'Neill, E., Carson-Berndsen, J.: The effect of phoneme distribution on perceptual similarity in English. INTERSPEECH pp. 1941–1945 (2019)
28. Philips, L.: The double metaphone search algorithm. C/C++ users journal 18(6), 38–43 (2000)
29. Read, C.: Children's creative spelling. Routledge (2018)
30. Russell, R., Odell, M.: Soundex. US patent 1,261,167 (1918)
31. Silfverberg, M., Kauppinen, P., Lindén, K.: Data-driven spelling correction using weighted finite-state methods. In: SIGFSM Workshop on Statistical NLP and Weighted Automata. pp. 51–59 (2016)
32. Stüker, S., Fay, J., Berkling, K.: Towards context-dependent phonetic spelling error correction in children's freely composed text for diagnostic and pedagogical purposes. In: INTERSPEECH (2011)
33. Torgerson, C., Brooks, G., Hall, J.: A systematic review of the research literature on the use of phonics in the teaching of reading and spelling. DfES Publications Nottingham (2006)
34. Toutanova, K., Moore, R.C.: Pronunciation modeling for improved spelling correction. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. pp. 144–151 (2002)
35. University College Dublin: S-capade github repository (2020), `https://github.com/ucd-csl/Scapade`, last accessed 15 July 2020
36. Veronis, J.: Computerized correction of phonographic errors. Computers and the Humanities 22(1), 43–56 (1988)
37. Wagner, R.A., Fischer, M.J.: The string-to-string correction problem. JACM 21(1), 168–173 (1974)
38. Weide, R.L.: The CMU pronouncing dictionary (1998), `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`
39. Wikipedia: Wikipedia:lists of common misspellings (2020), `https://en.wikipedia.org/wiki/Wikipedia:Lists_of_common_misspellings`, last accessed 19 May 2020
40. Wolf Garbe, S.: Symspellpy (2020), `https://github.com/mammothb/symspellpy`, last accessed 21 May 2020
41. Wordlist, A.: Scowl (spell checker oriented word lists) (2019), `http://wordlist.aspell.net`, last accessed 21 May 2020
42. Yip, M.: English vowel epenthesis. Natural Language & Linguistic Theory pp. 463–484 (1987)
43. Zeeko: Zeeko: Free text survey responses (2020), `https://zeeko.ie`, last accessed 19 May 2020