Research Methods Overview

1. The Problem to be solved

This research is rooted in the real world problem that minority language writers feel or perceive to write their language - especially via digital interfaces. The hypothesis is that the typing experiences can become easier for minority language writers as they encounter the digital medium. However, before we can say 'easier' we must ask "easier than what"? And we must assert that currently there is indeed a challenge to minority language speakers who do write their language and attempt to do this via digital mediums.

1.1. Indicators

In fact we do have several indicators that that the writing challenge is more difficult than it needs to be. Some of these indicators are:

A. Low uptake of minority language use in digital mediums.

i. Accurate assessment of the rate of uptake is compounded by low language up-take in any medium[1] not just digital. - Sociolinguistically known as language shift, even if the language shift is only in the domains of communicative technology.

ii. Additionally some sociolingual-technical issues which affect minority language use in digital devices are as follows:

- Language behavior defaults to the digital device's default language's interface which is often not the minority language.
- Language behavior defaults to use of the language in the socio-linguistic setting in which the device is primarily used which is also often not the minority language.
- Language behavior defaults to use of the language which is the default for the keyboard based on keyboard layout and glyphs printed on the keys.
- Intra-group linguistic formalisms may not exist in the minority language so an external language may serve as the language of formal interactions.

Al-Khatib and Sabbah (2008: 40) discuss theses issues as they pertain to the code switching in computer-mitigated communication in the Arab world:

> As far as the Arab World is concerned, there has also been some interest in studying CS in CMC. In Egypt, Warschauer et al. (2002), for example, examine the use of English and Arabic in online communication by a group of young professionals. The study indicates that English is used predominantly in web use and in formal e-mail communication, whereas a Romanized version of Egyptian Arabic is used extensively in informal e-mail and online chats. They ascribe the anticipated results to four factors: "general dominance of English in the professional milieu, lack of Arabic software standards, computer and internet use learned in English environments and early adopters' fluency in English". Similarly, another study on the ASCII-ized Arabic (AA), (i.e., a form of language in which ASCII "American Standard Code for Information Interchange" symbols are used to represent Arabic in Instant Messaging (IM) and other electronic written communication), was carried out by Palfreyman and al Khalil (2003). The corpus shows that "approximately" 25% of participants use mainly Arabic script in IM, 25% AA, and 50% English. As Palfreyman and al Khalil put it "in the present corpus there was a fair amount of code-

[1] The use of the phrase "any medium" may denote that this language is also becoming less often used in the oral domains as well. However, it does not mean that this is necessarily the case. However, due to the evolutionary nature of language domains it does mean that speakers of a language (which might be an ethnolinguistic-minority) who use a different language to communicate in computer or digital device based communication are undergoing some sort of language shift as these new digitally bounded domains of language use evolve and become a greater part of the language community's daily communication habits.

> switching (changing mid-utterance or mid-sentence from one language to another) and code-mixing (using words or phrases from one language within sentences in the other language)." This mixing of varieties correlates with different functions and topics, with Arabic being used for more formulaic phrases such as greeting, and English for topics such as university courses.

B. *When mother tongue languages are used, writers do not use "correct" characters even if available through indirect means (the use of alt keys or key combinations).*

Sometimes the "correct" characters are not available to language users and intervention is needed to correct this. Bailey (2007: 213) describes the text input method, for users of the Venda [ven][2] language before a sufficient input solution was created.

> … Translate.org.za soon discovered that Venda translators could not physically type the five extra characters needed in the Venda orthography on their computers. Most translators chose to simply ignore these characters or to follow convoluted processes to insert them… Of great concern to Translate.org.za were some of the methods used to circumvent this problem. These included printing and manually adding diacritics, adding characters that looked like the correct ones, and using methods that printed correctly and looked correct but were not actually correct characters.

However even in languages classified as having sufficient digital resources, users are reported to still avoid complex text input operations. In any situation, directly assessing this second indicator is complex because technology users may exhibit any of the following:

i. **Lack of knowledge of how to operate the computer system.**
Compounded not knowing how to access the "correct" characters for the typist's intended orthography.

ii. **Incorrect knowledge about how the computer system functions.**
Compounded by believing that the correct characters are only in a "particular font" rather than an accurate understanding of the relationship between display, and encoding. - This perception is common to many computer users today, not just minority language users.

iii. **Lack of orthography use or written language use competencies.**
Compounded by spelling inhibitions - common in languages where spelling conventions are not clearly set and insufficient technical development has occurred i.e. spell check programs and dictionary development.

Ample examples exist in the literature of text input use cases where users choose to not use "correct" characters. Many of these examples come from resource rich language computing environments. This seems to indicate that the challenges that typists face are not just encountered by minority language[3] users, A few are listed below: instance evidence from German [deu] and Danish [dan] are as follows:

- Native German [deu] speaking colleagues (p.c.) who use English language keyboards report being more prone to type English email to other Native German speakers who also speak German - because they are at an English keyboard. German is a language where all characters which are not in the English alphabet have official spelling options which are available on the English keyboard. This is an indicator that the nature or perceived nature of a keyboard has an influence not just on how a language is typed or not, but actually on language use choice in a domain of language use.
- Portuguese [por] (Jensen 1995)
- French [fra] example (van Compernolle 2011, van Compernolle & Williams 2010, Lally 2000) Supported by (Sturm 2006, 2008, 2010, 2012a, b, 2013)

---

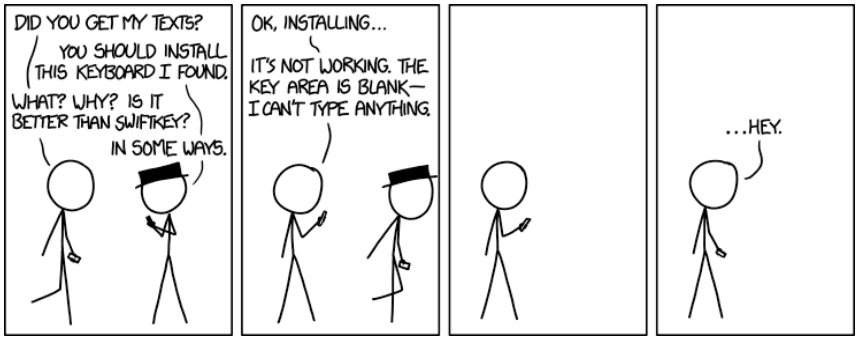
[2] Three letter codes within square brackets are the ISO 639-3 code for the language and are used in this paper to explicitly identify languages.

[3] Languages with fewer than 100,000 speakers.

- Spanish [spa] (Cárdenas-Claros & Isharyanti 2009, Cuetos 1993, Guerrero 2012, Negrón Goldbarg 2009, Ramírez Bustamante & Díaz 2006, Rodríguez 2007)
- Arabic [ara] (Al-Khatib & Sabbah 2008: 57) The results of this study also demonstrated that there are a number of technical elements that might be responsible for the wide use of English or switching between Arabic and English. Among these are: ease and swiftness of writing in English and limited space in Arabic messages.
- Evidence from Danish [dan] text messages show that Danish speakers change their spellings of Danish words when on devices which don't overtly support the text input methods of Danish orthographic characters. The examples below are provided by Eva Skafte Jensen and are typed by a Blackberry user. The Blackberry keyboard does not allow for rapid access to Danish characters therefore some people resort to alternative orthography because of the lack of æ, ø and å on the keyboard. This is the case so often among Danish speakers that the alternate orthography use is a tell-tell sign that the SMS message sender is a Blackberry user. Alternative orthography notation is indicated in bold with English translations below:

  > "Okay - jeg pr**oe**ver at begr**ae**nse fl**oe**de overfloden!"
  >
  > Okay – I try to hold-back-on the-cream extravaganza
  >
  > "Kommer hjem nu! G**aa**r I fakta p**aa** vejen beh**oe**ver vi noget?"
  >
  > Coming home now! Go to fakta (a supermarket) on the-way need we anything?

  Instead of æ, the person writes ae; instead of ø, the person writes oe (sometimes just o, not shown in these examples); instead of å, the person writes aa. Jensen notes that unlike German which has officially retained alternate spelling options, Danish has officially given these alternatives up. However, the alternative spelling options being used are actually old conventions for writing the corresponding sounds, but they went out of use a long time ago: aa for å was given up in 1948, and both ae and oe were abandoned in the 15th century.
- Finally one example from English [eng]. It appeals to the commonly held belief that there must be a better method for mobile text input. However, in contrast to the Danish example of text based data this example is brought to us by Randall Munroe (2013) as he illustrates the socially acknowledged elephant in the room - the need for a better text input solution. That is to say, the problem is real enough among English users that they can joke about the need for a better text input or keyboarding option.



1.2.Two classes of typing challenges

When dealing with human interactions, and especially human-computer interactions it is important to realize that users may or may not be aware or able to articulate the factors which are causing “pain points” during the process of accomplishing a task. Difficulties in accomplishing a task may be perceived, or undetected. Other difficulties may be anticipated by the technology user but the due to a lack of knowledge about how the technology functions the task may not be attempted. In this way users may be able articulate, anticipate, or not articulate, difficulties in accomplishing a task. For the remainder of this paper, unless otherwise explicitly stated all three categories of difficulties as perceived by the user will be simply referred to as “difficulties”.

Difficulties with text input are not only evident in languages like German and Danish which use characters not commonly occurring in English (which are also often characters beyond the ASCII range) meaning languages which use a Latin based script but also include characters like < ß, ø, æ, œ, ɗ, ɠ, ə > or also includes characters which include diacritics like < ü, ÿ, à, á, ñ, ō, õ, ł, ż, š, ń, ë, ę, ê, å, ç >, though this is not an exhaustive list. These examples all come from the Latin script, but other languages which are written in other scripts like Arabic, Hebrew or Indic scripts also employ the use of diacritics to various extents. Across the world’s Latin script based orthographies many use both diacritics and single characters outside of the ASCII range. However, based on casual observation, it seems that more orthographies use diacritics than single characters[4]. Also based on casual observation it appears that in those encoded text within orthographies which have both or use diacritics use the diacritics more often than single characters. Therefore the working hypothesis is that diacritics are encountered more often by the world’s writers, than single characters which lie outside of the ASCII range. From a human computer interaction assessment of the typing task this is significant.

2. Orthography research v.s text-input research

It is important to make a clear distinction here between orthography design research and keyboarding/ text-input research. Orthography research is a valid and important research pursuit, however it is a distinct pursuit from keyboard layout design research. This is partially evidenced by the fact that the act of text input or typing is only one of the things that may be done with an orthography. Orthography research classical looks at how an orthography is internally consistent or how the written transcription system matches the linguistic distinctions which enable communication in a language. However, it is equally important to realize that Keyboard layout design research is a distinct line of research. These two separate lines of investigation are intimately connected and ultimately and perhaps also practically very difficulty to separate and focus on independently. This is partially evidenced by the fact that it is impossible to do the text input task (typing) without doing it via the medium of an orthography. Therefore it is important to take a look at some of the relevant claims in recent orthography research.

2.1.Current issues in Orthography Research

In minority language orthography development one of the outstanding questions has been around the use of diacritics in orthographies. Like other languages, minority languages may use diacritics to indicate tone, stress, vowel quality, nasality, or Advanced Tongue Root (ATR) features. For many years the research around orthographies and the use of diacritics in Latin based scripts has centered around the readability of these orthographies for users of latin based scripts (Robberts 2008). Bird (1999a:1) presents the classic problem from an orthography design perspective:

> Tone languages provide some interesting challenges for the designers of new orthographies. One approach is to omit tone marks, just as stress is not marked in English (zero marking). Another approach is to do phonemic tone analysis and then make heavy use of diacritic

[4] Casual observation is based on a scope of the world’s languages and orthographies. For a more statistical analysis of 21 European languages consult Rosenbaum and Fleischmann’s comparison of character frequencies in a multilingual corpus (2002, 2003).

> symbols to distinguish the 'tonemes' (exhaustive marking). While orthographies based on either system have been successful, this may be thanks to our ability to manage inadequate orthographies rather than to any intrinsic advantage which is afforded by one or the other approach. In many cases, practical experience with both kinds of orthography in sub-Saharan Africa has shown that people have not been able to attain the level of reading and writing fluency that we know to be possible for the orthographies of non-tonal languages. In some cases this can be attributed to a sociolinguistic setting which does not favour vernacular literacy. In other cases, the orthography itself might be to blame. If the orthography of a tone language is difficult to user or to learn, then a good part of the reason, I believe, is that the designer either has not paid enough attention to the function of tone in the language, or has not ensured that the information encoded in the orthography is accessible to the ordinary (non-linguist) user of the language. If the writing of tone is not going to continue to be a stumbling block to literacy efforts, then a fresh approach to tone orthography is required, one which assigns high priority to these two factors.

Therefore one of the questions has been: "How much tone is too little tone to mark?" Which is rapidly followed by the second side of the same question: "How much tone is too much tone to mark?" Various authors ____() and _____() have made the appeal towards functional load. Saying that functional load should determine the upper bounds of the quantity of diacritics used. __(Define functional load and other loads)__ Introduce the problem of the designer bias to the orthography equation.

__Issue of transparent orthography and orthographic depth__

__Bottom up or top down strategy in reading affecting the orthography decisions__

__ where do I address the issue of:

Functional load - measuring the linguistic distinctions

Perceptual load - measuring the difficulty in visual distinctions

Cognitive load - measuring how much "brain power" is used during the reading or writing process.

END of Load issues__

Inherent to any design process is the perspective and bias that the designer brings to the process. This is no less true with the process of orthography design. In the case of orthography design one of the biases which designers bring, perhaps unintentionally, is their familiarity with a typographical tradition. A second bias is how that typographical tradition interacts in the cultural context in which that designer is most familiar. In language development among minority languages we can see the effects of cultural norms and typographical traditions of the native culture of the orthography designers on the designs of end products. __Smalley pg. 71__

For instance languages in Ethiopia and Nigeria where German missionaries or linguists/ anthropologists have engaged in *language development* often contain "German" like characters in the orthography such as the use of trigraphs like <sch> and <tch> or the use of <h> after a vowel to indicate a long vowel or the use of umlauts to disambiguate vowel quality. Another example is that French speaking people often do not see a problem with diacritics while English speaking people often despise diacritics and see them as an unnecessary burden in the reading process.

__insert examples of german like orthography__

For instance there is currently the TORP Research group headed by Dave Roberts of CNRS has been looking at how tone is marked in various orthographies(__ADD 2015 Project proposal__) (Roberts 2009, 2011a, b, Roberts & Walter 2012, Roberts, Walter & Snider 2014) and the resulting effects on literacy in languages which use those orthographies. The TORP group follows on the foundational research by Bird (1999a, b, 2001) which looked at the effects of diacritic marking on the fluency of

reading in the Dschang [ybb] language of Cameroon. The TORP project seeks to expand the kinds of tonal languages[5] and the kinds of representations their orthographies make.

Fundamentally there are two aspects to orthography assessment and the assessment of the actual load of diacritics on the usability of an orthography. The first is the assessment of the 'perception of encoded content' or reading. The second is the assessment of 'ease of acquisition' and 'ease of use in production tasks'. While research to this point (Bird) has mostly focused on the first task of reading efficiency, I hope to point my research in the second area of interest - especially where it intersects with digital text production.

2.2. Text-input Research Questions

If we consider the text input task with a keyboard[6] there are two fundamental components of this task. The manual key strike to produce the characters and the visual feedback loop where the typists looks at the input to determine that they made the correct key strike. An analysis of behavior after incorrect key strikes is also desirable because this constitutes part of the overall 'typing' experience.

__(Introduce Fitts' law and Shannon's theorem, Zipf's law and Information theory analysis of channel depth)__

2.2.1. Error analysis

Dealing with "errors" is not always strait forward. There are basically two classes of errors, manual performance errors and spelling errors. Manual performance errors are often categorized as action slips and further divided by cause. Norman (1981) provides a categorization of action slips encountered by typists. Chen, Yesilada and Harper (2010) provide an ad hoc taxonomy of errors and compare error rates between small device users (mobile) and standard laptop/desktop users. In part Chen et. al's error typology is based on work done by MacKenzie and Soukoreff (2002). However as pointed out by Paterson (2014: 57-8), the minority language context which can in various contexts from time to time lead to typing mistakes based on visually perceived glyph/character similarity. This is not accounted for in previous taxonomies of typing errors. Regarding the second class of errors loosely held as 'spelling errors'; In some cases 'spelling errors' are not known by the typists and therefore go undetected, thereby creating "errors" but "errors where each key struck was intentional". This means that no performance error was performed because each intended stroke was struck successfully. These kinds of errors point to user challenges with the orthography rather than the text input system itself. However, some errors (both spelling and performance) are detected after the key strike and are corrected by the typist (but causing more key strikes than was originally necessary if we assume 100% accuracy), based on the typist's understanding of orthographical conventions. A third kind of error is where the correct spelling of a word is know and recognized but due to typing speed or habits an invalid sequence of characters is input this is a performance error in the manual part of the task. In a similar vein as the third kind of error a fourth kind of error might also be considered and that is of the high variety and the low variety. That is if "proper orthography compliance" is of a high speech variety and speech in low varieties need not comply with "proper orthography conventions" then some aspects of the orthography may be dropped to optimize the communicative channel. It needs to be

IS this in the correct place for readers to process?

[5] By kinds of tonal languages I mean both tonal languages in other language families, and other languages with other orthographic traditions for marking tone. This also implies that other languages will have different patterns of tonal melodies following Snider (1999), and will have other patterns of tonal sandhi, and grammar distinctions made with grammatical tone in addition to lexical tone.

[6] Keyboarding is commonly called typing, however in English, especially in the sciences, typing can refer to the process of assigning examples to a taxonomy or the activity of sorting. In text based searches the term 'typing' therefore provides quite a bit of noise or confusion.

determined if this variation is 1) intentional or 2) an effect on the capacity of the communication channel due to the technology involved. Look at this: (Tavosanis 2007)

2.2.2. What to analyze

Analysis of the text input process needs to cover several dynamics. First the manual stance that the typist is using. e.g. finger pecking v.s. full hand typing. Video and motion detection can capture this as typing rate alone can not detect this.

Second analyzing typing is often fundamentally considered as a time over distance problem. That is, the time it takes to strike keys over the distances that fingers must travel to strike those keys. To analyze the key strike time, a keystroke recorder is needed. To work cross-linguistically this recorder needs to not only record the actual key used by the typists but also the intended character that the stroke (or stroke set) produced, targeting both key-down and key-up events on a time based log. Additionally it would be good to mark the kinds of characters the user is seeing on the keys they are striking.

The third set of important data to bring to bear is EEG data of the typists to look at some of the psychological “pain points” and to create an indexing of responses to errors and long distance characters.

The fourth and final kind of desired data is eye-tracking data of the subject typist. The reason for this is to determine where the typist is looking at the moment they are typing a given keystroke. From an English orthography and typing perspective we intrinsically expect the typists to be looking either at the output or the keyboard. However, in languages with lots of diacritics it remains an open question as to where the typist is looking, if they are looking for the diacritics or if they are looking at the base character or if they are looking at the keyboard and looking for instructions on how to type a modification to a base character.

Latin Scripts are not the only scripts which use diacritics or small marks above or below the main “reading” line of the script. The Hebrew script, the Arabic Script and the Hindi family of scripts also use diacritics.

Nadine Chahine recently (2012) completed a Ph.D where she used eye-tracking software.

2.2.3. Classification of participants

There are three groups of participants

Classes of participants	*Green Group*	*Yellow Group*	*Red Group*
Start typing with original layout for L1 language	No typing experience in L1 No typing experience in L2	No typing experience in L1 Some typing experience in L2	Some typing experience in L1 Some typing experience in L2
Start typing with new layout for L1 language	No typing experience in L1 No typing experience in L2	No typing experience in L1 Some typing experience in L2	Some typing experience in L1 Some typing experience in L2

The challenge is that the Red Group will be a really small group of people for most language groups. This means that the to successfully compare them we might have to lump several different L1’s together.

Specific Tools to keep in Mind

Logging of keystrokes

http://www.nada.kth.se/iplab/trace-it/

http://www.writingpro.eu/logging_programs.php

http://www.inputlog.net/download_software.html

http://www.translog.dk/

http://bridge.cbs.dk/platform/?q=Translog-II

Tobii Studio http://www.tobii.com/en/eye-tracking-research/global/products/software/tobii-studio-analysis-software/news-in-tobii-studio-3-1/

E-Prime Works with Video events of Tobii Studio: https://www.youtube.com/watch?v=TpBMB9BSJI8

StimTracker http://cedrus.com/stimtracker/ EEG with Tobii Stuido - Instantly compatible with ERP/EEG devices from ADInstruments, Biopac, BioSemi, and Brain Products.

http://www.tobii.com/en/eye-tracking-research/global/products/hardware-accessories/stimtracker-for-tobii-tx300-eye-tracker/

Pit those tools against http://imotionsglobal.com/software/add-on-modules/mobile-eye-tracking-module-glasses/

EEG: http://www.emotiv.com/store/compare/

3. Bibliography

Bailey, Dwayne. 2007. Creating a single South African keyboard layout to promote language. *Lexikos* 17.1: 212-25. http://www.ajol.info/index.php/lex/article/view/51533

Bird, Steven. 1999a. Strategies for Representing Tone in African Writing Systems. *Written Language & Literacy* 2.1: 1-44. http://dx.doi.org/10.1075/wll.2.1.02bir

Bird, Steven. 1999b. When Marking Tone Reduces Fluency: An Orthography Experiment in Cameroon. *Language and Speech* 42.1: 83-115. http://las.sagepub.com/content/42/1/83.abstract

Bird, Steven. 2001. Orthography and identity in Cameroon. *Written Language & Literacy* 4.2: 131-62. http://dx.doi.org/10.1075/wll.4.2.02bir

Cárdenas-Claros, Mónica Stella & Neny Isharyanti. 2009. Code switching and code mixing in Internet chatting: between

'yes', 'ya', and 'si' a case study. *jaltcalljournal* 5.3: 67–78.

Chahine, Nadine. 2012. *Reading Arabic: legibility studies for the Arabic script.* Ph.D dissertation, Leiden University, Leiden, Netherlands.

Chen, Tianyi, Yeliz Yesilada & Simon Harper. 2010. What input errors do you experience? Typing and pointing errors of mobile Web users. *International Journal of Human-Computer Studies* 68.3: 138-57. http://www.sciencedirect.com/science/article/pii/S1071581909001578

van Compernolle, Rémi A. 2011. Use and variation of French diacritics on an Internet dating site. *Journal of French Language Studies* 21.02: 131-48. <Accessed: 27. October 2014>. http://dx.doi.org/10.1017/S0959269510000293

van Compernolle, Rémi A. & Lawrence Williams. 2010. Orthographic Variation in Electronic French: The Case of l'accent aigu. *The French review* 83.4: http://www.refdoc.fr/Detailnotice?idarticle=29333835

Cuetos, Fernando. 1993. Writing processes in a shallow orthography. *Reading and Writing* 5.1: 17-28. http://dx.doi.org/10.1007/BF01026916

Guerrero, Fabio G. 2012. On the Entropy of Written Spanish. *Revista Colombiana de Estadistica* 35.3: 425-42.

Jensen, John B. 1995. Writing Portuguese Electronically: Spontaneous Spelling Reform. *Hispania* 78.4: 837-45. http://www.jstor.org/stable/345160

Al-Khatib, Mahmoud A. & Enaq H. Sabbah. 2008. Language Choice in Mobile Text Messages among Jordanian University Students. *SKY Journal of Linguistics* 21. 37-65. <Accessed: 27. October 2014>. http://www.linguistics.fi/julkaisut/SKY2008/Al-Khatib%20et%20al%20NETTIVERSIO.pdf

Lally, Carolyn Gascoigne. 2000. The Effect of Keyboarding on the Acquisition of Diacritical Marks in the Foreign Language Classroom. *The French Review* 73.5: 899-907. http://www.jstor.org/stable/398298

MacKenzie, I. Scott & R. William Soukoreff. 2002. A Character-level Error Analysis Technique for Evaluating Text Entry Methods. *A character-level error analysis technique for evaluating text entry methods. Proceedings of the Second Nordic Conference on Human-Computer Interaction -- NordiCHI 2002*, 241-4. New York: ACM.

Munroe, Randall. 2013. Improved Keyboard. xkcd.com. 30 October 2013. <Accessed: 10. May 2014>. http://xkcd.com/1284/

Negrón Goldbarg, Rosalyn. 2009. Spanish-English Codeswitching in Email Communication. *Language@Internet* 6.3: http://nbn-resolving.de/urn:nbn:de:0009-7-21398

Norman, Donald A. 1981. Categorization of action slips. *Psychological Review* 88.1: 1-15. http://psycnet.apa.org/journals/rev/88/1/1/

Palfreyman, David & Muhamed al Khalil. 2003. "A Funky Language for Teenzz to Use": Representing Gulf Arabic in Instant Messaging. *Journal of Computer-Mediated Communication* 9.1: http://onlinelibrary.wiley.com/enhanced/doi/10.1111/j.1083-6101.2003.tb00355.x/

Paterson, Hugh J., III. 2014. Keyboard layouts: Lessons from the Me'phaa and Sochiapam Chinantec designs. In Mari C. Jones (ed.), *Endangered Languages and New Technologies*, 49-66. Cambridge, UK: Cambridge University Press.

Ramírez Bustamante, Flora & Enrique López Díaz. 2006. Spelling Error Patterns in Spanish for Word Processing Applications. Paper presented at Fifth International Conference on Language Resources and Evaluation, LREC 2006, Magazzini del Cotone, Genoa (22-28 May 2006). <Accessed: 3 January 2013>. http://www.cs.brandeis.edu//~marc/misc/proceedings/lrec-2006/pdf/119_pdf.pdf

Robberts, David. 2008. Thirty years of tone orthography testing in West African languages (1977 - 2007). *Journal of West African Languages* 35.1-2: 199-242.

Roberts, David. 2009. Visual Crowding and the tone orthography of African languages. *Written Language & Literacy* 12.1: 140-55. http://dx.doi.org/10.1075/wll.12.1.07rob

Roberts, David. 2011a. Autosegmental and pedagogical considerations in preparation for a tone orthography experiment in Kabiye. *Journal of West African Languages* 38.2: 87-106.

Roberts, David. 2011b. A tone orthography typology. *Written Language & Literacy* 14.1: 82-108. <Accessed: 15 May 2012>. http://dx.doi.org/10.1075/wll.14.1.05rob

Roberts, David & Stephen L. Walter. 2012. Writing grammar rather than tone: An orthography experiment in Togo. *Written Language & Literacy* 15.2: 226-53. <Accessed: 02. May 2014>. http://dx.doi.org/10.1075/wll.15.2.06rob

Roberts, David, Stephen L. Walter & Keith Snider. 2014. Testing the optimal depth of a diacritic tone orthography for Kabiye (Togo): a classroom experiment.

Rodríguez, Néstor J. and Diaz, Maria I. . 2007. Word Processing in Spanish Using an English Keyboard: A Study of Spelling Errors. . In Nuray M. Aykin (ed.), *UI-HCII 2007 - Usability and Internationalization. Global and Local User Interfaces - Part II July 22-27, 2007, Beijing, China* (Lecture Notes in Computer Science 4560), 219-27. Berlin and Heidelberg: Springer.

Rosenbaum, Richard & Michael Fleischmann. 2002. Character Frequency in Multilingual Corpus 1--Part 1. *Journal of Quantitative Linguistics* 9.3: 233. http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=9897590&site=ehost-live&scope=site

Rosenbaum, Richard & Michael Fleischmann. 2003. Character Frequency in Multilingual Corpus 1 – Part 2. *Journal of Quantitative Linguistics* 10.1: 1-39. <Accessed: 2014/10/26>. http://www.tandfonline.com/doi/abs/10.1076/jqul.10.1.1.16550

Snider, Keith L. 1999. *The geometry and features of tone*. Dallas: Summer Institute of Linguistics.

Sturm, Jessica Lynn. 2006. The Effect of Keyboarding and Presentation Format on the Recall of Accent Marks in L2 Learners of French. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics* 6.2: Article 4 (15 pages). <Accessed: 22. July 2014>. http://journals.tc-library.org/index.php/tesol/article/viewFile/197/192

Sturm, Jessica Lynn. 2008. *Accent marks in L2 French: Keyboarding, presentation format, working economy, and pronunciation ability*. 3337934 dissertation, University of Illinois at Urbana-Champaign, Ann Arbor.

Sturm, Jessica Lynn. 2010. The effect of keyboarding and presentation format on accent marks in L2 French. In F. Neveu et al. (eds.), *Proceedings of the 2nd Congres Mondial de Linguistique Français (Congrès Mondial de Linguistique Française - CMLF 2010)*, 1591-606. Paris: Institut de Linguistique Française.

Sturm, Jessica Lynn. 2012a. Attention, awareness, and accents in L2 French. *Language Awareness* 22.2: 146-60. <Accessed: 22. July 2014>. http://dx.doi.org/10.1080/09658416.2012.670243

Sturm, Jessica Lynn. 2012b. Meaning and orthography in L2 French. *Writing Systems Research* 4.1: 47-60. <Accessed: 2014/10/26>. http://dx.doi.org/10.1080/17586801.2011.635950

Sturm, Jessica Lynn. 2013. Liaison in L2 French: The Effects of Instruction. In J. Levis & LeVelle. K. (eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, 157-66. Ames, IA: Iowa State University.

Tavosanis, Mirko. 2007. A Causal Classification of Orthography Errors in Web Texts. In Craig Knoblock, Daniel Lopresti, Shourya Roy & L. Venkata Subramaniam (eds.), *Proceedings of IJCAI-07 Workshop on Analytics for Noisy Unstructured Text Data (AND-07) January 8, 2007 Hyderabad, India*, 99-106. Hyderabad, India:

Warschauer, Mark, Ghada R. El Said & Ayman G. Zohry. 2002. Language Choice Online: Globalization and Identity in Egypt. *Journal of Computer-Mediated Communication* 7.4: 0-. http://dx.doi.org/10.1111/j.1083-6101.2002.tb00157.x